计 算 机 科 学 丛 书

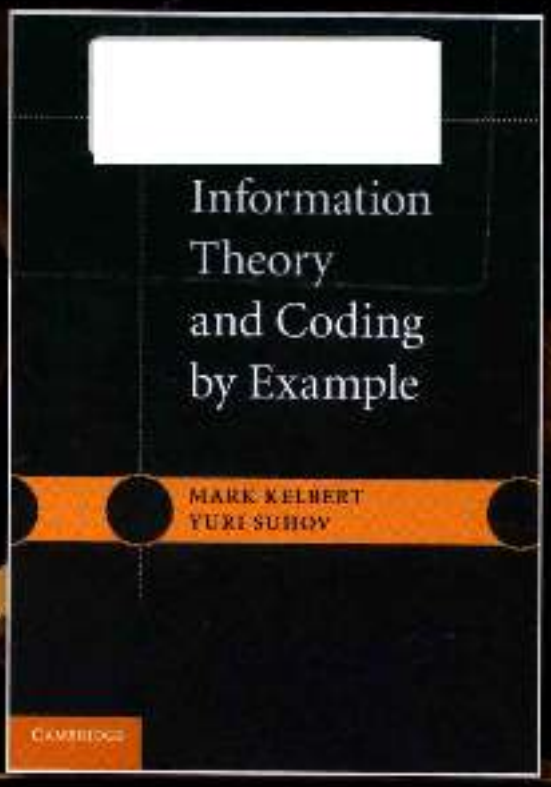
CAMBRIDGE

信息论与编码理论

剑桥大学真题精解

[英] 马克·凯尔伯特（Mark Kelbert）
[俄] 尤里·苏霍夫（Yuri Suhov） 著
高晖 吕铁军 译

Information Theory and Coding by Example



机械工业出版社
China Machine Press

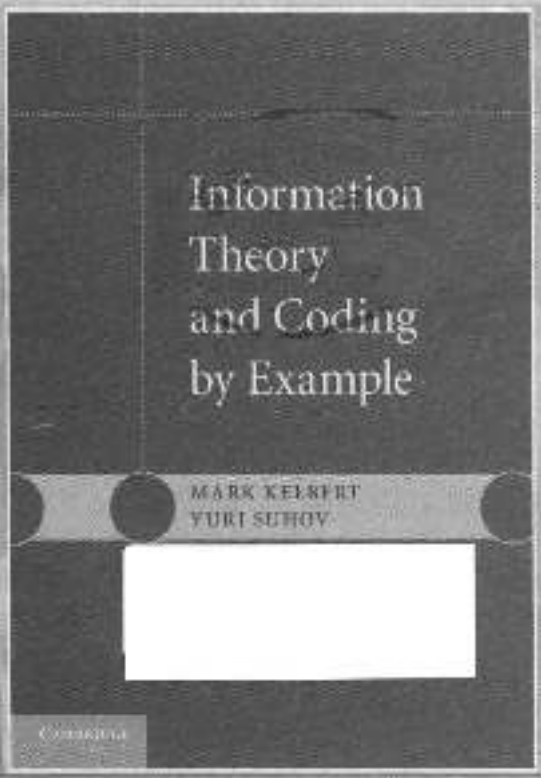
计 算 机 科 学 丛 书

信息论与编码理论

剑桥大学真题精解

[英] 马克·凯尔伯特（Mark Kelbert）
[俄] 尤里·苏霍夫（Yuri Suhov） 著
高晖 吕铁军 译

Information Theory and Coding by Example



机械工业出版社
China Machine Press

图书在版编目（CIP）数据

信息论与编码理论：剑桥大学真题精解 /（英）马克·凯尔伯特（Mark Kelbert）等著；高岳等译．—北京：机械工业出版社，2016.12
（计算机科学丛书）
书名原文：Information Theory and Coding by Example

ISBN 978-7-111-55352-6

I. 信… II. ①马… ②高… III. ①信息论－高等学校－教材 ②信源编码－高等学校－教材 IV. TN911.2-44

中国版本图书馆 CIP 数据核字（2016）第 274897 号

本书版权登记号：图字：01-2016-3783

This is a Chinese simplified edition of the following title published by Cambridge University Press: Mark Kelbert, Yuri Suhov, Information Theory and Coding by Example, ISBN 978-0-521-13988-5.

© Cambridge University Press 2013.

This Chinese simplified edition for the People's Republic of China (excluding Hong Kong, Macau and Taiwan) is published by arrangement with the Press Syndicate of the University of Cambridge, Cambridge, United Kingdom.

© Cambridge University Press and China Machine Press in 2017.

This Chinese simplified edition is authorized for sale in the People's Republic of China (excluding Hong Kong, Macau and Taiwan) only. Unauthorized export of this simplified Chinese is a violation of the Copyright Act. No part of this publication may be reproduced or distributed by any means, or stored in a database or retrieval system, without the prior written permission of Cambridge University Press and China Machine Press.

本书原版由剑桥大学出版社出版。

本书简体字中文版由剑桥大学出版社与机械工业出版社合作出版。未经出版者预先书面许可，不得以任何方式复制或抄袭本书的任何部分。

此版本仅限在中华人民共和国境内（不包括香港、澳门特别行政区及台湾地区）销售。

本书讲解信息论与编码理论，涵盖原本两个版块。书中素材来自剑桥大学本科生理想“信息论”“编码与密码学”以及几门数学方向的研究生课程。全书最大的特色是极富干货，并将 Shannon 等科学家的学术历程贯穿其中，在透彻讲解理论知识的同时介绍读者感兴趣主题。

欢迎高校学生、研究者和工程师阅读此书，你不仅可以将以往出现在计算机、电子工程等分散学科中的信息论知识融会贯通，还能够通过自测判断自己已达到或未达到的学习程度。

出版发行：机械工业出版社（北京市百万庄大街 22 号 邮政编码：100037）
责任编辑：曲 熙　　责任校对：殷 虹
印　　刷：中国电影出版社印刷厂　　版　　次：2017 年 1 月第 1 版第 1 次印刷
开　　本：185mm × 260mm　1/16　　印　　张：22
书　　号：ISBN 978-7-111-55352-6　　定　　价：89.00 元

凡购本书，如有缺页、倒页、脱页，由本社发行部调换
客服热线：（010）88379426　88361066　　购书热线：（010）68326294　88379649　68995259
投稿热线：（010）88379604　　读者信箱：hzjsj@hzbook.com

版权所有·侵权必究
封底无防伪标均为盗版
本书法律顾问：北京大成律师事务所　韩光 / 邹晓东

文艺复兴以来，源远流长的科学精神和逐步形成的学术规范，使西方国家在自然科学的各个领域取得了垄断性的优势；也正是这样的优势，使美国在信息技术发展的六十多年间名家辈出、独领风骚。在商业化的进程中，美国的产业界与教育界越来越紧密地结合，计算机学科中的许多泰山北斗同时身处科研和教学的最前线，由此而产生的经典科学著作，不仅擘划了研究的范畴，还揭示了学术的演变，既遵循学术规范，又自有学者个性，其价值并不会因年月的流逝而减退。

近年，在全球信息化大潮的推动下，我国的计算机产业发展迅猛，对专业人才的需求日益迫切，这对计算机教育界和出版界都既是机遇，也是挑战；而专业教材的建设在教育战略上显得举足轻重。在我国信息技术发展时间较短的现状下，美国等发达国家在其计算机科学发展的几十年间积淀和发展的经典教材仍有许多值得借鉴之处。因此，引进一批国外优秀计算机教材将对我国计算机教育事业的发展起到积极的推动作用，也是与世界接轨、建设真正的世界一流大学的必由之路。

机械工业出版社华章公司较早意识到“出版要为教育服务”。自1998年开始，我们就将工作重点放在了遴选、移译国外优秀教材上。经过多年的不懈努力，我们与Pearson，McGraw-Hill，Elsevier，MIT，John Wiley & Sons，Cengage等世界著名出版公司建立了良好的合作关系，从他们现有的数百种教材中甄选出Andrew S. Tanenbaum，Bjarne Stroustrup，Brian W. Kernighan，Dennis Ritchie，Jim Gray，Afred V. Aho，John E. Hopcroft，Jeffrey D. Ullman，Abraham Silberschatz，William Stallings，Donald E. Knuth，John L. Hennessy，Larry L. Peterson等大师名家的一批经典作品，以“计算机科学丛书”为总称出版，供读者学习、研究及珍藏。大理石纹理的封面，也正体现了这套丛书的品位和格调。

“计算机科学丛书”的出版工作得到了国内外学者的鼎力相助，国内的专家不仅提供了中肯的选题指导，还不辞劳苦地担任了翻译和审校的工作；而原书的作者也相当关注其作品在中国的传播，有的还专门为其书的中译本作序。迄今，“计算机科学丛书”已经出版了近两百个品种，这些书籍在读者中树立了良好的口碑，并被许多高校采用为正式教材和参考书籍。其影印版“经典原版书库”作为姊妹篇也被越来越多实施双语教学的学校所采用。

权威的作者、经典的教材、一流的译者、严格的审校、精细的编辑，这些因素使我们的图书有了质量的保证。随着计算机科学与技术专业学科建设的不断完善和教材改革的逐渐深化，教育界对国外计算机教材的需求和应用都将步入一个新的阶段，我们的目标是尽善尽美，而反馈的意见正是我们达到这一终极目标的重要帮助。华章公司欢迎老师和读者对我们的工作提出建议或给予指正，我们的联系方法如下：

华章网站：www.hzbook.com
电子邮件：hzjsj@hzbook.com
联系电话：(010)88379604
联系地址：北京市西城区百万庄南街1号
邮政编码：100037

华章科技图书出版中心

Mark Kelbert 与 Yuri Suhov 的这本书可谓信息论研究学习中的经典好书。本书涉及信息论和编码理论相关的多个领域，当我们接到翻译此书的任务时，多少有些惶恐，担心不能将书中的精髓充分呈现给读者。之前国内关于信息论与编码领域的书籍大多集中在理论研究方面，而本书提供了丰富的例题，可以弥补国内教材在实例应用上的欠缺。我们欣然接受了此项翻译任务，并且力争不辱使命。

本书涵盖信息论与编码理论的方方面面，信息量大，内容丰富，既详尽地讲解了基础内容，比如熵、信源、信道以及编译码规则，又讨论了大量相关领域中的进阶话题，计算机科学、密码学、电子工程以及概率与统计等学科的教学内容在本书中都有体现。

本书包含信息论与编码中的概率和代数两个方向，为了保持其在不同领域的特色，同时使风格尽可能一致，我们在翻译的过程中反复斟酌，力求完美。还虚心向相关领域的专业人员请教，在此对他们表示感谢。最后，我们还要对机械工业出版社的编辑们表示感谢，他们的尽职尽责以及热情合作给予了我们莫大的帮助。

译者

2016 年 11 月

本书的素材取自剑桥大学数学荣誉学位考试的几门相关课程：本科二年级的"信息论"（该课程已历经40余年的教学与发展，期间仅仅在课程名称上略有调整），"编码与密码学"（一门新开设的简明课程，省去了繁杂的技术细节），以及一些更为前沿的第三部分课程（相当于数学硕士研究生课程）。本书的内容安排围绕以下核心概念：概率分布的熵——一种不确定性的度量（也包括随机过程的熵率——样本轨迹变化率的度量），编码——一种度量及利用随机过程中冗余信息的方法。

因此，本书的内容大致涵盖了当前全球范围内与信息论相关的典型数学素材，这些教学内容通常安排在计算机科学、电子工程以及概率与统计等学科中。然而，本书与其他著作的首要不同在于丰富的例题（其模式遵循了我们在剑桥大学出版社推出的本系列图书第一本——《Probability and Statistics by Example》）。书中绝大部分例题来源于剑桥大学数学荣誉学位考试。因此，读者可以通过本书判断自己所达到或者期望达到的学习程度。

本书与其他信息论和编码相关著作的第二个不同之处在于，它包含了两个可能的方向：概率和代数。通常而言，这两个方向往往出现在不同的专著、教材或者课程中，所涉及的人员也来自不同的领域。本书成形得益于两段经历。我们曾经在位于莫斯科的俄罗斯科学院下属的信息传输问题研究所工作。俄罗斯科学院一直具有跨学科研究科学问题的优良传统，特别值得一提的是，Roland Dobrushin，Raphail Khasminsky，Mark Pinsker，Vladimir Blinovsky，Vyacheslav Prelov，Boris Tsybakov，Kamil Zigangirov（从事概率和统计研究），Valentin Afanasiev，Leonid Bassalygo，Serguei Gelfand，Valery Goppa，Inna Grushko，Grigorii Kabatyansky，Grigorii Margulis，Yuri Segalovich，Alexei Skorobogatov，Mikhail Tsfasman，Victor Zinov'yev，Victor Zyablov（从事代数、组合数学、几何和数论研究）等学者都曾经工作或依然工作于俄罗斯科学院（曾经有一段时期，这些学者都在莫斯科中心一幢改建楼同一层的五个房间中工作）。我们也具有在剑桥大学的工作经历，这段经历同样十分重要。剑桥大学教授信息论和编码理论相关课程时，具有与俄罗斯科学院相似的跨学科精神。这种风格主要起始于 Peter Whittle（从事概率和最优化研究）及其后的 Charles Goldie（从事概率研究）、Richard Pinch（从事代数和几何研究）、Tom Körner 和 Keith Carne（从事分析研究），还有 Tom Fisher（从事数论研究）。

需要补充的是，作为训练有素的数学家（并且骨子里也是数学基因），尽管我们也有很强的应用背景，但在完成本书的过程中依然经历着这样一些折磨：表述模糊不清，不精确，真假可疑（这包含了个人因素），当然还有将完美的数学思想付诸实践所需要的代价。然而，我们依然坚定地认为数学思维依然是在当今充满竞争的世界上生存并自我完善的主要途径。因此，数学需要被认真地对待并加以学习（或许不需要理由）。

作为面向随机过程的信息论方法基础，上述两个概念（熵和编码）已由 Shannon 在20世纪40年代发表的代表性论文[1,2]中完整地引入。当然，熵的概念早在一个世纪前就已被 Boltzmann 和 Gibbs 在热力学中使用，而编码已被（高效地）应用在实际生活当中很久了。但是，Shannon 是第一个充分意识到这些概念在信息领域的作用并用现代数学框架加以阐述的开创者，尽管 Shannon 从未经历成为数学家的训练，也并不总能完整地给出关于

自己的理论的一些证明(或许他并不觉得有任何不妥)。在本书的相关章节中，我们会点评一些 Shannon 与数学界的关系发展中非常引人注目的场景。幸运的是，这些纷杂并没有给 Shannon 造成困扰(Shannon 和 Boltzmann 不同，后者对外界的评论十分敏感且十分在意)。Shannon 一定知道他所发现的理论背后的巨大价值：在我们的眼中，他的地位与伟大的数学家 Wiener 和 von Neumann 相当。

客观地说，Shannon 的名字依然主导着当前信息与编码理论中概率和代数的方向，这样强大的影响力是非同寻常的。特别是当我们意识到 Shannon 的学术活跃期已过去 40 多年时，(虽然在一些先进的话题方面，Shannon 或许会沿用 Einstein 的话："数学家们已经涌入通信理论，现在连我自己都搞不清楚这理论了。")

在 Shannon 的创建及发明之后，数学、电子工程、计算机科学等学科都经历了巨大的变化。谁又能预见在 20 世纪 40～50 年代，原本相互对立的 Shannon 信息论与 Wiener 控制论能够融合？事实上，后者包含造福全人类的宏伟(甚至是不切实际的)愿景，而前者仅仅设定了一个谦卑的目标以将信息传输中的误差控制在某些极限当中。Wiener 的著作[37]塑造了 20 世纪 50～60 年代思想家们所开展智力活动的几乎所有维度，特别地，控制论在苏联及其卫星国成为严肃的政治议题：最初它被认为是"一个资产阶级的反科学理论"，然后又被过度狂热地追捧。(1953 年发表在苏联主要意识形态期刊《哲学问题》上的关于控制论的评价是："帝国主义者没有办法消除摧毁资本主义社会的根本矛盾，他们不能阻止即将发生的经济危机。所以，他们尝试从狂热的军备竞赛和意识形态战争中寻找答案，在深层的绝望中，他们寻求伪科学带来的一线希望以苟延残喘。"在 1954 年版的苏联《简明哲学词典》中有成百上千条关于控制论的定义："反动的伪科学，首先出现在二战后的美国，后广泛传播于资本主义国家，是一种现代的机械论。"然而，受困于参与苏联核试验且掌握实权的一些顶尖物理学家，之前反对控制论的《哲学问题》期刊在 1955 年发表了鼓吹控制论积极面的文章，该文章的作者包括 Alexei Lyapunov 和 Sergei Sobolev 等苏联卓越的数学家。)

奇怪的是，最近关于 Wiener 的自传[38]显示，曾经存在"秘密的(美国)文档指出 FBI 和 CIA 如何在冷战期间追踪 Wiener 以阻挠他的社会激进主义并压制控制论在国内外的巨大影响"。文献[35]中也提到了这种有趣的对比。

然而，历史总是以自己的脚步前进。如 Freeman Dyson 在对文献[35]的评述[39]中指出："(Shannon 的理论)在数学方面是优雅和清晰的，它能够应对通信所涉及的许多实际问题，它比控制论更易于使用。它奠定了一门崭新的学科——信息论……(在当代)电子工程师将学习 Shannon 创建的信息论作为基本训练，而控制论逐渐被遗忘。"

事实上控制论并未被遗忘，在苏联依然有至少七个研究院或机构以控制论命名，其中俄罗斯的莫斯科和白俄罗斯的明斯克分别有两所，爱沙尼亚的塔林、乌兹别克斯坦的塔什干和乌克兰的基辅(苏联计算机科学的中心)也分别坐落着一所。在美国，至少有四所大学设置了控制论相关的院系，分别是波尔顿大学、布拉德福德大学、赫尔大学和瑞丁大学，这项统计事实上不包括其他相关的学术组织和学会。在全球范围内来看，控制论相关的学会看起来非常繁荣，具有长短不一、各式各样的名字，比如瑞士的方法研究所、意大利的控制论学会、阿根廷布宜诺斯艾利斯的普适系统理论和控制论学会。我们也十分欣喜地发现剑桥控制论协会坐落于美国加州的贝尔蒙，与控制论情形不同，以信息论命名的研究机构屈指可数。显然，关于 Shannon 和 Wiener 的经典争论还会继续。

无论如何，Wiener 在数学领域的个人声誉依然坚实，我们能够说出好几个能理论中

的珍宝，比如 Paley-Wiener 定理（在 Wiener 无数次到访剑桥的过程中创造）和 Wiener-Hopf 方法，当然还有 Wiener 过程——代表他在科学研究及应用方面的重要地位。然而，当前针对这位科学巨擘的一些回忆录展示出他复杂而困惑的人格。（从关于 Wiener 的传记[31]题名不难发现这种特点，但是这些观点仍然有争议，比如文献[107]的评论，而在本书中，我们尝试采用文献[75]中第 386～391 页关于 Wiener 的温和口吻加以阐述。）另一方面，关于 Shannon 的生平记录（这些论述来自其他信息和编码理论创始人，如 Richard Hamming）则给出了一致的描绘——他是一位安静、睿智和幽默的人。我们希望现有这些说法不要成为人们描写 Shannon 传记的障碍，也希望未来能有更多关于 Shannon 的书，正如现在关于 Wiener 的书那样。

如前所述，本书的目的是双重的：一方面通过丰富的例题和例子对信息论中概率与几何方面的知识做系统的介绍，另一方面讨论一些很少在其他主流教材中涉及的有益话题。本书第 1～3 章介绍信息论和编码理论的基础知识并对一些相关前沿话题展开讨论。内容组织安排方面，我们主要关注具有代表性的问题和例题（其中很多源自剑桥大学的课程），而不对背后的理论做过于细致的阐述。第 4 章对信息论相关的一系列深层主题进行介绍，其表述风格十分简洁，因此一些重要的结论并未给出证明。

本书的很大一部分内容源自课堂讲义和对课堂习题或考试题的解答，所以某种程度上的内容重复难以避免，并且有可能出现符号的多重定义或者非规范的语言表述。对此，我们顺其自然，我们觉得这些不完美恰好营造了教学和考试过程中的真实氛围。

本书行文安排深受两部优秀著作[15,16]的影响。我们与 Charles Goldie 长久的友谊以及同 Tom Cover 和陆的交往均对本书产生了有益的帮助。我们同样受益于对文献[18]、[110]、[130]和[98]的阅读及借鉴。此外，感谢剑桥大学牛顿研究院 2002～2010 年的一系列课程，特别是通信科学中的随机过程（2010 年 1～7 月）。本书中的诸多内容都经过与来自不同研究机构的同行的交流和讨论，其中最为重要的就是位于贝斯科的信息传输问题研究所和数学地理及地震预测研究所（我们曾经是其中忠诚的一员）。我们还要感谢来自剑桥大学 Statslab 的 James Lawrence 为本书提供了图片。

本书中 PSE I 和 PSE II 分别代表本书作者所著由剑桥大学出版社出版的《Probability and Statistics by Example》第 1 卷和第 2 卷。我们采用 PSE II 的风格，呈现了许多带有答案的例题。这些例题都以问题的形式出现（其中很多源自于剑桥数学荣誉学位的考试试卷，其形式和风格均得以保留）。

目 录

Information Theory and Coding by Example

中版序言
译者序
前言

第1章 信息论基础 …… 1

1.1 基本概念，Kraft不等式，Huffman编码 …… 1
1.2 熵：简介 …… 11
1.3 Shannon第一编码定理，Markov信源的熵率 …… 26
1.4 信道，解码规则，Shannon第二编码定理 …… 38
1.5 微分熵及其性质 …… 54
1.6 本章附加问题 …… 60

第2章 编码理论简介 …… 93

2.1 Hamming距离，码字的几何特征，码本规模的基本界 …… 93
2.2 Shannon第二编码定理的几何证明，码本规模的粘组界 …… 104
2.3 线性码：基本构造 …… 119
2.4 Hamming码，Golay码，Reed-Muller码 …… 129
2.5 循环码和代数多项式，BCH码简介 …… 139
2.6 本章附加问题 …… 158

第3章 编码理论的深层主题 …… 176

3.1 有限域入门 …… 176
3.2 Reed-Solomon编码，再论BCH编码 …… 191
3.3 再论循环码，BCH解码 …… 197
3.4 MacWilliams标识和线性规划界 …… 206
3.5 渐近好码 …… 216
3.6 本章附加问题 …… 224

第4章 信息论的深层主题 …… 242

4.1 Gauss信道 …… 242
4.2 连续时间集的渐近均分性 …… 262
4.3 Nyquist-Shannon公式 …… 270
4.4 空间点过程和网络信息论 …… 287
4.5 密码学选例与问题 …… 298
4.6 本章附加问题 …… 316

参考文献 …… 330

索引 …… 337

信息论基础

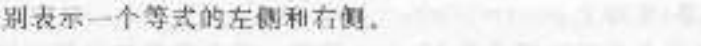
全书当中，$\mathbb{P}$表示各类概率分布。特别地，在第 1 章当中，$\mathbb{P}$表示信源输出随机变量序列的概率。作为约定，我们假设这些序列是由独立同分布的随机变量构成的，或者来源于离散时间 Markov 链，即 $\mathbb{P}(U_1=u_1, \cdots, U_n=u_n)$是随机变量 $U_1, \cdots, U_n$ 分别取值 $u_1, \cdots, u_n$ 的联合分布，$\mathbb{P}(V=v \mid U=u, W=w)$表示在给定随机变量 U 取值 u、随机变量 W 取值 w 条件下，随机变量 V 取值 v 的条件概率。同样地，$\mathbb{E}$表示对分布 $\mathbb{P}$ 求期望。

符号 p 和符号 P 用来表示各种概率(或者与概率相关的对象)。符号 $\#A$ 表示有限集合 A 的势。符号 $\mathbb{1}$ 表示指示函数。本书也采用以下对数符号和准则：$\ln=\log_e$，$\log=\log_2$，对任意 $b>1$，$0\cdot\log_b 0=0\cdot\log_b\infty$。对于给定的 $x>0$，$\lfloor x\rfloor$和$\lceil x\rceil$分别表示对 x 的下取整和上取整，因此有$\lfloor x\rfloor\leqslant x\leqslant\lceil x\rceil$，当 x 取正整数时等式成立($\lfloor x\rfloor$也可被认为是 x 的整数部分)。

LHS 和 RHS 分别表示一个等式的左侧和右侧。

1.1 基本概念，Kraft 不等式，Huffman 编码

信息传输过程中一个典型的方案如下图所示：



例子 1.1.1 (a) 信源：剑桥大学学院咱诗班。

(b) 编码器：一个 BBC 的录音单元，它将声音转换为二进制序列然后写入 CD 音轨，随后 CD 发行上市。

(c) 信道：一位消费者购买了一张 CD 并从英国邮递到澳大利亚。这个信道受到“噪声”的影响，CD 在传输(或者运输)过程中可能会受到损坏(器械的、电子的、化学的等)。

(d) 译码器：在澳大利亚的 CD 播放机。

(e) 信宿：在澳大利亚的一位听众。

(f) 目的：确保有损情况下的高质量音频。

事实上，即便用针头在 CD 上戳一个小洞或者滴一滴酸(当然，这样的实验并不值得提倡)，它依然能够经受考验而保持音频质量。用术语来说，信息传输的目标主要包括：

(i) 对信息的快速编码。

(ii) 经过编码后的消息需易于传输。

(iii) 有效利用信道(即最大化单位时间内信息的传输量)。

(iv) 快速译码。

(v) (尽可能多地)纠正由噪声引起的错误。

这些目标往往是相互冲突的，因此人们必须寻找优化方案。这就是本章即将讨论的。然而，人们不能奢求完美的解决方案，接下来要介绍的理论旨在提供与基本原理相关的知识。关于方案的最终决定取决于负责人(或团体)。

本节的很大部分(也包括整个第 1 章)都将讨论编码问题，而编码的意义在于：

(i) 压缩数据以减少消息中的冗余信息。

(ii) 防止非法用户获取消息。

(iii) 使纠错成为可能。

接下来我们从信源和编码器开始展开研究。信源发出一串字母(或者符号)，

$$u_1 u_2 \cdots u_n \cdots \tag{1.1.1}$$

其中 $u_j \in I$，$I(=I_m)$是一个具有 m 个元素的集合，通常表示为$\{1, \cdots, m\}$(一个信源字母表)。以文学英语为例，$m=26+7$，这包含了 26 个字母和 7 个标点符号 . , : ; ' ( )(有时候人们也会增加如?! ''和 "")。电报英语则对应 $m=27$。

一种常用的方法是将式(1.1.1)视为一个随机源的采样点，即一个随机变量序列

$$U_1, U_2, \cdots, U_n, \cdots \tag{1.1.2}$$

然后尝试建立一种方法以将这样的序列做合理的分类。

2 **例子 1.1.2** (a) 最简单的随机信源是一串独立同分布(IID)的随机变量

$$\mathbb{P}(U_1 = u_1, U_2 = u_2, \cdots, U_n = u_n) = \prod_{j=1}^{n} p(u_j) \tag{1.1.3a}$$

其中 $p(u)=\mathbb{P}(U_j=u)$，$u \in I$ 是一个随机变量的边缘分布。一个具有 IID 符号的随机信源通常被称作 Bernoulli 信源。

关于 $p(u)$的一个特殊例子是等概率的 Bernoulli 信源，其中概率分布 $p(u)$与具体事件 $u \in U$ 没有关系(实际上 $p(u)=1/m$)。

(b) 一个更为普遍的例子是 Markov 信源，其中信源输出符号构成了一个离散时间 Markov 链(DTMC)

$$\mathbb{P}(U_1 = u_1, U_2 = u_2, \cdots, U_k = u_k) = \lambda(u_1) \prod_{j=1}^{k-1} P(u_j, u_{j+1}) \tag{1.1.3b}$$

其中 $\lambda(u)=\mathbb{P}(U_1=u)$，$u \in I$ 是初始概率，$p(u, u')=\mathbb{P}(U_{j+1}=u' \mid U_j=u')$，$u, u' \in I$ 是转移概率。当 $\mathbb{P}(U_j=u)=\lambda(u)$，$j \geq 1$ 时，Markov 信源被认为是平稳的，也就是说，$\lambda=\{\lambda(u), u=1, \cdots, m\}$相对矩阵 $P=\{P(u, v)\}$是旋转不变的行向量，满足等式约束 $\sum_{u \in I} \lambda(u) P(u, v)=\lambda(v)$，$v \in I$，或简记为 $\lambda P=\lambda$。

(c) 一个退化的 Markov 信源的例子是信源发出多个重复的符号。这里，

$$\begin{gathered} \mathbb{P}(U_1 = U_2 = \cdots = U_n = u) = p(u), u \in I \\ \mathbb{P}(U_k \neq U_k') = 0, 1 \leq k < k' \end{gathered} \tag{1.1.3c}$$

其中 $0 \leq p(u) \leq 1$ 且 $\sum_{u \in I} p(u) = 1$。

序列(1.1.1)中的起始块

$$\boldsymbol{u}^{(n)} = (u_1, u_2, \cdots, u_n) \text{ 或简写为 } \boldsymbol{u}^{(n)} = u_1 u_2 \cdots u_n$$

被称作(源)样本 n-字符串，或 n-字，其字母表是 I。$\boldsymbol{u}^{(n)}$通常被视为消息，相应地，可以引入随机的 n-字符串(随机消息)

$$\boldsymbol{U}^{(n)} = (U_1, U_2, \cdots, U_n) \text{ 或简写为 } \boldsymbol{U}^{(n)} = U_1 U_2 \cdots U_n$$

编码器使用符号集 $J(=J_q)$，通常写为$\{0, 1, \cdots, q-1\}$，而编码符号的数目通常满
3 足$q \leq m$(有时甚至是 $q \ll m$)。在很多案例中，通常采用二元编码，即 $q=2$，$J=\{0, 1\}$。一个码(有时也称为编码)是一个映射 f，它将一个符号 $u \in I$ 转换为一个有限字符串，$f(u)=x_1 \cdots x_s$，其字母来自于 J。换句话说，f 将 I 映射到所有可能的集合 J^* 上，即

$$f: I \to J^* = \bigcup_{s \geq 1} (J \times \cdots (s\text{次}) \times J)$$

字符串 $f(u)$被称为码字(在编码 f 中)，它是在 f 映射下的符号 $u \in I$ 的像。如果上式

中 z 的取值对所有码字都一致等于 N，那么这个编码具有(定)长 N。消息 $u^{(k)}=u_1u_2\cdots u_k$ 可以表示为码字的级联

$$f(u^{(k)})=f(u_1)f(u_2)\cdots f(u_k)$$

它实际上也是 J^* 中的字符串。

定义 1.1.3 如果 $u\neq u'$ 使得 $f(u)\neq f(u')$(即映射 $f: I\to J^*$ 是一对一的)，那么编码是无损的。如果任何集合 J^* 中的任何字符串是至多一个信息的象，那么这种编码就是可译的。如果 $y=xz$，那么字符串 x 是另一个字符串的前置。例如，y 可以用相互联系的 x，z 来表示。如果没有一个码字是其他码字的前缀，那么这个码字就是无前缀的(例如定长码就是无前缀的)。

无前缀编码是可译码，但是可译码不一定是无前缀编码。

例子 1.1.4 具有三个源字母 1，2，3 和二元编码字母表 $J=\{0,1\}$ 的编码被表示为

$$f(1)=0,f(2)=01,f(3)=011$$

此码是可译的，但不是无前缀的。

定理 1.1.5 (Kraft 不等式)*给定正整数 $s_1,\cdots,s_m$，存在一个可译码 $f: I\to J^*$，码字长度是 $s_1,\cdots,s_m$，当且仅当*

$$\sum_{i=1}^{m}q^{-s_i}\leqslant 1 \tag{1.1.4}$$

而且，在式(1.1.4)下，存在一个码字长度为 $s_1,\cdots,s_m$ 的无前缀编码。

证明 (i) 充分性。令式(1.1.4)成立。我们的目的是用码长 $s_1,\cdots,s_m$ 的码字建立无前缀编码。改写式(1.1.4)

$$\sum_{l=1}^{s}n_lq^{-l}\leqslant 1 \tag{1.1.5}$$

4

或者

$$n_sq^{-s}\leqslant 1-\sum_{l=1}^{s-1}n_lq^{-l}$$

其中 n_l 是长度为 l 的码字的数量，并且 $s=\max s_i$。上式可等效地表述为

$$n_s\leqslant q^s-n_1q^{s-1}-\cdots-n_{s-1}q \tag{1.1.6a}$$

因为 $n_{s-1}\geqslant 0$，可推得

$$n_{s-1}q\leqslant q^s-n_1q^{s-1}-\cdots-n_{s-2}q^2$$

或者

$$n_{s-2}\leqslant q^{s-2}-n_1q^{s-3}-\cdots-n_{s-3}q \tag{1.1.6b}$$

重复这个过程可得

$$\begin{aligned} n_{s-2} &\leqslant q^{s-2}-n_1q^{s-3}-\cdots-n_{s-3}q \\ &\vdots \\ n_2 &\leqslant q^2-n_1q \end{aligned} \tag{1.1.6.s-1}$$

$$n_1\leqslant q \tag{1.1.6.s}$$

可以看到实际上对于所有 $i=1,\cdots,s-1$，如果满足 $n_{i+1}=0$ 或 n_i 小于不等式的 RHS(根据定义，$n_i\geqslant 1$，所以对于 $i=s-1$，第二种可能性发生)，那么我们能完成下面的构造。首先选择码长为 1 的 n_1 个码字，抽取 J 中不同的符号，对于式(1.1.6.s)来说这是可能的，剩下了$(q-n_1)$个未用的符号；通过附加一个符号，我们可以组成$(q-n_1)q$ 个长度为 2 的码字；利用式(1.1.6.s-1)，我们可以选择 n_2 个码字，这样仍然还有 $q^2-n_1q-n_2$ 个

未用的码字来组成 n_s 个码字……从这个构造的过程来看，没有任何新的码字把前面的码字作为前缀。所以，构造出来的码是无前缀码。

(ii) 必要性。假设存在一个码字长度为 s_1，…，s_m 关于 J^* 的可译码。令 $s=\max s_i$，对任意正整数 r，可以得到

$$(q^{-s_1}+\cdots+q^{-s_m})^r=\sum_{l=1}^{rs} b_l q^{-l}$$

5 其中 b_l 是 r 码字放在一起组成一个长度为 l 的字符串的组合数目。

由于可译性，这些字符串必须是不同的。因此，$b_l\leqslant q^l$，又因为 q^l 是 l 字符串的总数，那么

$$(q^{-s_1}+\cdots+q^{-s_m})^r\leqslant rs$$

和

$$q^{-s_1}+\cdots+q^{-s_m}\leqslant r^{1/r}s^{1/r}=\exp\left(\frac{1}{r}(\log r+\log s)\right)$$

对于任意 r，上式都成立。当 $r\to\infty$，RHS 趋近于 1。 □

备注 1.1.6 一个满足式(1.1.4)的编码不一定是可译的。

Leon G. Kraft 于 1949 年在他的麻省理工博士论文中介绍了不等式(1.1.4)。

信息论的一个重要的目标是找到“最好”(即最短的)可译(无前缀)码。我们现在采用概率论的观点，假设 $u\in I$ 中的符号是由一个概率为 $p(u)$ 的信源发送的：

$$\mathbb{P}(U_k=u)=p(u)$$

(在这里，没有必要说明多个发送符号的联合概率。)

同顾一下，给定一个编码 f：$I\to J^*$，通过一个指定码字长度为 $s(i)$ 的码字 $f(i)=x_1\cdots x_{s(i)}$，我们对字母 $i\in I$ 进行编码。对于一个随机符号，所产生的码字变为一个来自 J^* 的随机字符串。当 f 是无损的，对于一个符号，所产生的字符串作为一个码字的概率刚好等于 $p(i)$，前提是这个字符串刚好和 $f(i)$ 吻合；如果不存在具有这样性质的字母 $i\in I$，这个概率就是 0。如果 f 不是一一映射，字符串的概率等于对应码字 $f(i)$，即这个字符串的所有 $p(i)$ 的和。那么码字的长度变为一个随机变量 S，概率分布是

$$\mathbb{P}(S=s)=\sum_{1\leqslant i\leqslant m}\mathbf{1}(s(i)=s)p(i) \tag{1.1.7}$$

我们寻找一个可译码来最小化码字长度的期望：

$$\mathbb{E}S=\sum_{s\geqslant 1}s\mathbb{P}(S=s)=\sum_{i=1}^{m}s(i)p(i)$$

接下来问题归结为：

$$\begin{aligned}&\text{最小化 } g(s(1),\cdots,s(m))=\mathbb{E}S\\&\text{满足}\sum_i q^{-s(i)}\leqslant 1(\text{Kraft})\\&\text{其中 } s(i)\text{ 是正整数}\end{aligned} \tag{1.1.8}$$

6 **定理 1.1.7** 问题(1.1.8)的最优值的下界如下：

$$\min\mathbb{E}S\geqslant h_q(p(1),\cdots,p(m)) \tag{1.1.9}$$

其中

$$h_q(p(1),\cdots,p(m))=-\sum_i p(i)\log_q p(i) \tag{1.1.10}$$

证明 算法(1.1.8)是一个整数最优化问题。如果我们忽略约束条件 $s(1)$，…，$s(m)\in(1,$

2，…}，而是用一个松弛条件 $s(i)>0$，$1\leqslant i\leqslant m$ 作为代替，那么 Lagrange 充分性定理就可以使用。Lagrange 算子为

$$\mathscr{L}(s(1),\cdots,s(m),z;\lambda)=\sum_i s(i)p(i)+\lambda\left(1-\sum_i q^{-s(i)}-z\right)$$

（这里，$z\geqslant 0$ 是一个松弛变量）。关于 s_1，…，s_m 和 z 最小化 $\mathscr{L}$，可得

$$\lambda<0,z=0,\quad \frac{\partial\mathscr{L}}{\partial s(i)}=p(i)+q^{-s(i)}\lambda\ln q=0$$

由此

$$-\frac{p(i)}{\lambda\ln q}=q^{-s(i)},\quad 即\ s(i)=-\log_q p(i)+\log_q(-\lambda\ln q),1\leqslant i\leqslant m$$

调整约束条件 $\sum_i q^{-s(i)}=1$（松弛变量 $z=0$）得到

$$\sum_i p(i)/(-\lambda\ln q)=1,\quad 即\ -\lambda\ln q=1$$

所以，根据式(1.1.10)给定 h_q 的值，

$$s(i)=-\log_q p(i),\quad 1\leqslant i\leqslant m$$

是这个松弛问题的(唯一)最优解。松弛问题求解是基于变量 $s(i)$ 的一个更大集，所以，它的最小值不会大于原始问题的最小值。 □

备注 1.1.8 式(1.1.10)定义的 h_q 在整个信息论中起到极其重要的作用，它被称作概率分布($p(x)$，$x\in I$)的 q 元熵，并将出现在许多情形中。q 的相关性由下式充分体现

$$h_q(p(1),\cdots,p(m))=\frac{1}{\log q}h_2(p(1),\cdots,p(m))$$

此处 h_2 代表二进制熵：

$$h_2(p(1),\cdots,p(m))=-\sum_i p(i)\log p(i) \tag{1.1.11}$$

7

举例 1.1.9 (a) 请给出一个符号集为 J_q 的无损但不满足 Kraft 不等式的编码例子。再给出一个无损编码的例子，令其码长严格小于 $h_q(X)$。

(b) 证明基于无损码的"Kraft 和" $\sum_i q^{-s(i)}$ 可以是任意大的(对于足够大的源符号集)。

解答 (a) 考虑符号集 $I=\{0,1,2\}$ 和一个无损编码 f，$f(0)=0$，$f(1)=1$，$f(2)=00$，码长 $s(0)=s(1)=1$，$s(2)=2$。明显地，$\sum_{x\in I}2^{-s(x)}=5/4$ 不满足 Kraft 不等式。对于一个随机变量 X，其 $p(0)=p(1)=p(2)=1/3$，平均码长 $Es(X)=4/3<h(X)=\log 3=1.585$。

(b) 假设对于某一个正整数 L，字母表的大小为 $m=\#I=2(2^L-1)$。考虑用 $x\in I$ 组成一个具有最大码长 L 的码字 0，1，00，01，11，000，…。Kraft 和是

$$\sum_{x\in I}2^{-s(x)}=\sum_{i\in\{L,\cdots,1\}}\sum_{s(x)=i}2^{-s(x)}=\sum_{i\in L}2^i\times 2^{-i}=L$$

它可以任意大。 □

定理 1.1.7 的结论进一步详细阐述如下。

定理 1.1.10 (Shannon 无损编码定理(NLCT))对于一个随机信源，它发送符号的概率为 $p(i)>0$，对于符号集为 J_q 的可译码，它的最小平均码长服从：

$$h_q\leqslant\min\mathbb{E}S<h_q+1 \tag{1.1.12}$$

其中 $h_q=-\sum_i p(i)\log_q p(i)$ 是信源的 q 元熵，见式(1.1.10)。

证明 式(1.1.9)给出了 LHS 不等式。对于 RHS 不等式，令 $s(i)$ 为一个整数，使得

$$q^{-s(i)} \leqslant p(i) < q^{-s(i)+1}$$

这里的非严格界表明 $\sum_i q^{-s(i)} \leqslant \sum_i p(i) = 1$，即 Kraft 不等式。

所以，这里存在一个码长为 $s(1)$，…，$s(m)$的可译码，严格界为

8

$$s(i) < -\frac{\log p(i)}{\log q} + 1$$

所以

$$\mathrm{ES} < -\frac{\sum_i p(i)\log p(i)}{\log q} + \sum_i p(i) = \frac{h}{\log q} + 1 \qquad \square$$

例子 1.1.11 以下给出一个有关 Shannon NLCT 的一个有指导意义的应用例子。令信源符号集的大小 m 等于 2^k，假设字母 $i=1$，…，m 等概率发送：$p(i)=2^{-k}$。假如我们使用编码符号集 $J_2=\{0, 1\}$(二元码)。二进制熵为 $h_2 = -\log 2^{-k} \sum_{1\leqslant i\leqslant 2^k} 2^{-k} = k$，对于可译码平均需要至少 k 个二进制数字。使用比特作为熵的单位，则平均意义上编码需要至少 k 个比特。

此外，NLCT 引出了一个 Shannon-Fano 编码过程：我们固定正整数码字长度 $s(1)$，…，$s(m)$使得 $q^{-s(i)} \leqslant p(i) < q^{-s(i)+1}$，或者等效为

$$-\log_q p(i) \leqslant s(i) < -\log_q p(i) + 1, \quad 即\ s(i) = \lceil -\log_q p(i) \rceil \qquad (1.1.13)$$

然后构造一个无前缀编码，从最短的 $s(i)$向上，确保前面的码字都是无前缀的。Kraft 不等式保证了足够的空间。得到的码字虽然可能不是最优，但是它与最优码一样具有满足不等式(1.1.13)的平均码长。

Huffman 编码 f_q^*：$I_m \to J_q^*$ 能够获得最优性。我们首先讨论二进制 Huffman 编码，当 $q=2$(即 $J=\{0, 1\}$)时，下列算法构造 f 一个二进制树：

(i) 首先，对 $i\in I$ 排序，得到 $p(1)\geqslant p(2)\geqslant\cdots\geqslant p(m)$。

(ii) 指定符号 0 给字母 $m-1$，1 给字母 m。

(iii) 构造一个减小的符号集 $I_{m-1}=\{1, \cdots, m-2, (m-1, m)\}$，概率为

$$p(1), \cdots, p(m-2), p(m-1)+p(m)$$

使用减小的符号集，重复步骤(i)和(ii)。我们获得一个二进制树。一个 $m=7$ 的 Huffman 编码如图 1-1 所示。

$m=7$

i	p_i	$f(i)$	s_i
1	0.5	0	1
2	0.15	100	3
3	0.15	101	3
4	0.1	110	3
5	0.05	1110	4
6	0.025	11110	5
7	0.025	11111	5

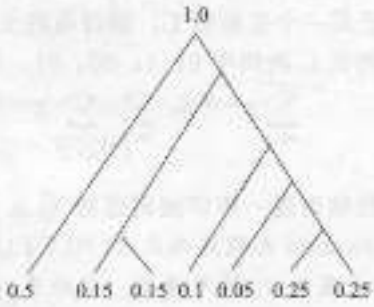


图 1-1

为了到达树的根节点 i，我们必须穿越的分支数量为 $s(i)$。树的结构和作为信源字母对根的识别一起，保证了编码是无前缀的。Huffman 二元码的最优性由下面两个简单的引

理来推出。 9

引理 1.1.12 任何最优的无前缀二元码，码长与其概率是倒序的

$$p(i) \geq p(i') \quad 意味着 \quad s(i) \leq s(i') \tag{1.1.14}$$

证明 如果上式不成立，我们可以对 i 和 i' 交换码字组成一个新编码。这样缩短了平均码长，保持了无前缀的性质。 □

引理 1.1.13 对于任何最优的无前缀二进制码，在所有最大长度的码字中，存在两个除最后一位数字以外其他部分完全相同的码。

证明 使之不成立的条件有两个：(i)存在单个最大长度的码字；(ii)存在两个或者多个最大长度的码字，它们在最后一个数字之前都不相同。在这两种情况下，我们可以从一些最大码长的码字中删除最后一个数字，而不影响无前缀性质。 □

引理 1.1.14 Huffman 编码在所有的无前缀二进制码中是最优的。

证明 用数学归纳法。对于 $m=2$，Huffman 编码 f_2^H 有 $f_2^H(1)=0$，$f_2^H(2)=1$，或者 $f_2^H(1)=1$，$f_2^H(2)=0$，并且是最优的。假设无论何种概率分布，Huffman 编码 f_{m-1}^H 对于 I_{m-1} 都是最优的。

进一步假设关于某些概率分布，Huffman 编码 f_m^H 对于 I_m 不是最优的。即对于 I_m 存在另一个具有更短平均码长的无前缀编码 f_m^*： 10

$$ES_m^* < ES_m^H \tag{1.1.15}$$

在此情况下概率分布可以假设为

$$p(1) \geq \cdots \geq p(m)$$

根据引理 1.1.12 和引理 1.1.13，在这两种码字中，我们可以重置码字，使得对应 $m-1$ 与 m 的码字有最大的长度，这两个码字唯一不同的只是最后一位数字。这容许我们把两个编码减少到 I_{m-1}。也就是说，在 Huffman 编码 f_m^H 中，我们能够从 $f_m^H(m)$ 和 $f_m^H(m-1)$ 中删除最后一位数字从而"粘合"这些码字。这样就构造了 Huffman 编码 f_{m-1}^H。在 f_m^* 中，执行相同的步骤就得到新的无前缀编码 f_{m-1}^*。

观察到在 Huffman 编码 f_m^H 中，从 $f_m^H(m)$ 和 $f_m^H(m-1)$ 中对 ES_m^H 的贡献为 $s^H(m)(p(m-1)+p(m))$；经过缩减以后，变为 $(s^H(m)-1)(p(m-1)+p(m))$。即 ES 减少了 $p(m-1)+p(m)$。在编码 f_m^* 中，从 $s^*(m)(p(m-1)+p(m))$ 减少到 $(s^*(m)-1)(p(m-1)+p(m))$ 有相似的贡献；差别也为 $p(m-1)+p(m)$。所有其他对 ES_{m-1}^H 和 ES_{m-1}^* 的贡献与对 ES_m^H 和 ES_m^* 的贡献都一样。所以，f_{m-1}^* 要优于 f_{m-1}^H：$ES_{m-1}^* < ES_{m-1}^H$，这和假设矛盾。 □

根据定理 1.1.14，我们可获得以下推论。

推论 1.1.15 Huffman 编码在所有可译的二元码中是最优的。

易得上述过程能够推广到 q 进制的 Huffman 编码(具有编码字母表 $J_q=\{0, 1, \cdots, q-1\}$)：不是合并两个具有最小概率的符号 $m-1$，$m \in I_m$，而是合并 q 个具有最小概率的符号，重复以上过程。事实上，Huffman 1952 年的原始论文已经描述了一般化的编码符号集，其中存在许多对 Huffman 编码的改进，包括不等编码代价(其中一些编码数字 $j \in J_q$ 的代价比其他的高)，这些不在本书中进行讨论。

举例 1.1.16 Huffman 编码的缺点是：码字长度是符号概率 $p(1), \cdots, p(m)$ 的复杂函数。然而，一些边界是可达的。假设 $p(1) \geq p(2) \geq \cdots \geq p(m)$。在任何二元码中：

(a) 如果 $p(1) \leq 1/3$，那么字母 1 必须用长度大于 2 的码字编码。

(b) 如果 $p(1) > 2/3$，那么字母 1 必须用长度为 1 的码字编码。 11

解答 (a) 有两种可能的情况：在构造一个 Huffman 编码的最后两步之前，字母 1 要么与

其他字母合并，要么不合并。在第一种情形下，$s(1)\geqslant 2$。否则，符号1，b和b'有

$$p(1)<1/3, p(1)+p(b)+p(b')=1, \quad 所以\ \max[p(b),p(b')]>1/3$$

那么在倒数第二步，字母1与b和b'中的一个合并，所以$s(1)\geqslant 2$。假设至少一个码字长度为1，这个码字分配给具有$p(1)<1/3$的字母1。所以，Huffman树的顶部如图1-2a所示，满足$0\leqslant p(b)$，$p(b')\leqslant 1-p(1)$，$p(b)+p(b')=1-p(1)$。

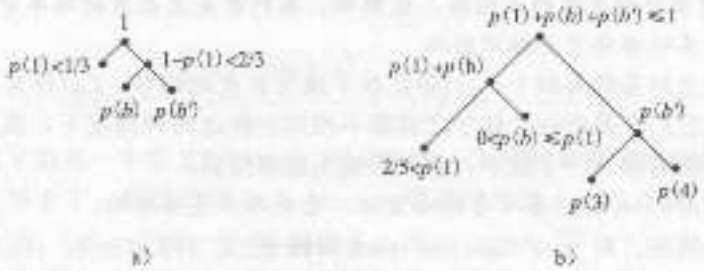


图 1-2

但是如果$\max[p(b),p(b')]>1/3$，$p(1)$应该和$\min[p(b),p(b')]$合并。所以，图1-2是不可能的，字母1的码长大于等于2。

当两种码字

$$\{0,01,110,111\}\ 和\ \{00,01,10,11\}$$

都是二进制Huffman编码，且概率分布为1/3，1/3，1/4，1/12，那么此时边界是清晰可辨的。

(b) 令$p(1)>2/5$，假设字母1在一个Huffman编码中的编码长度$s(1)\geqslant 2$。所以字母1在最后一步前和其他符号合并。换言之，在某一个阶段，我们让符号1，b和b'具有如下性质：

(i) $p(b')\geqslant p(1)>2/5$

(ii) $p(b')\geqslant p(b)$

(iii) $p(1)+p(b)+p(b')\leqslant 1$

(iv) $p(1)$，$p(b)\geqslant 1/2p(b')$

事实上，如果$p(b)<1/2p(b')$，那么当$p(b')$产生时，在前一步b应该被选择而不是$p(3)$或者$p(4)$。根据(iv)，$p(b)\geqslant 1/5$，那么(i)+(iii)是不可能的。

关于$p(1)$Huffman树的一部分如图1-2b所示，其中$p(3)+p(4)=p(b')$并且$p(1)+p(b')+p(b)\leqslant 1$。我们有

$$p(1)=2/5-\varepsilon, p(b')=2/5+\varepsilon+\delta, p(b)=2/5+\varepsilon+\delta-\eta$$

其中$\varepsilon>0$，δ，$\eta\geqslant 0$。那么

$$p(1)+p(b')+p(b)=6/5-3\varepsilon+2\delta-\eta\leqslant 1, \quad \eta\geqslant 1/5+3\varepsilon+2\delta$$

这使得

$$p(b)\leqslant 1/5-2\varepsilon-\delta<1/5$$

然而，因为

$$\max[p(3),p(4)]\geqslant p(b')/2\geqslant p(1)/2>1/5$$

概率$p(b)$应该被$\min[p(3),p(4)]$合并，即图1-2b是不可能的。因此，字符1的码字长度$s(1)=1$。 □

举例1.1.17 假设字符i_1，…，i_5的概率分别为0.45，0.25，0.2，0.05，0.05，计算Shannon-Fano编码和Huffman编码的平均码长。通过发现每种情况下的可译二进制码来

说明这两种方法。

解答 当 $q=2$ 时，

Shannon-Fano：

$p(i)$	$\lceil -\log_2 p(i) \rceil$	码字
0.45	2	00
0.25	2	01
0.2	3	100
0.05	4	11100
0.05	5	11111

E(码字长度)$=0.9+0.5+0.6+0.25+0.25=2.5$，

Huffman：

p_i	码字
0.45	1
0.25	01
0.2	000
0.05	0010
0.05	0011

E(码字长度)$=0.45+0.5+0.6+0.2+0.2=1.95$。 □ 13

举例 1.1.18 Shannon-Fano 编码一般并不是最优的，然而，它并不比 Huffman 编码长多少，试证明，如果 S_{SF} 是 Shannon-Fano 的码长，那么对于任意 $r=1,2,\cdots$ 以及码长为 S^* 的任意可译码 f^* 有

$$\mathbb{P}(S^* \leqslant S_{SF}-r) \leqslant q^{-r}$$

解答 记

$$\mathbb{P}(S^* \leqslant S_{SF}-r) = \sum_{i\in I:s^*(i)\leqslant s_{SF}(i)-r} p(i)$$

需要注意的是，$s_{SF}(i)<-\log_q p(i)+1$，因此，由 Kraft 不等式可得

$$\sum_{i\in I:s^*(i)\leqslant s_{SF}(i)-r} p(i) \leqslant \sum_{i\in I:s^*(i)\leqslant -\log_q p(i)-r} p(i) = \sum_{i\in I:s^*(i)-1\leqslant -\log_q p(i)-r} p(i) = \sum_{i\in I:p(i)\leqslant q^{-s^*(i)+r}} p(i)$$

$$\leqslant \sum_{i\in I} q^{-s^*(i)+r} = q^{-r}\sum_{i\in I} q^{-s^*(i)} \leqslant q^{-r}$$ □

现在，常见的做法不是将每个 $u\in I$ 单独编码，而是将信源消息划分为固定长度 n 的“区段”或者说“块”，并将这些段或块编码。很明显，这种方法增加了符号集中字母名义上的数量：这些块都是从 Cartesian 乘积 $I^n=I\times\cdots(n\text{ 次})\times I$ 中得来的，但最重要的是码块在典型信息中随机分布的熵：

$$h_q^{(n)} = -\sum_{i_1,\cdots,i_n} \mathbb{P}(U_1=i_1,\cdots,U_n=i_n)\log_q \mathbb{P}(U_1=i_1,\cdots,U_n=i_n) \tag{1.1.16}$$

（很显然，我们需要知道连续发送信源字母的联合分布。）$S^{(n)}$ 代表在可译段码字中的随机码长，定义每一信源比特的最小期望码长为 $e_n:=\min\frac{1}{n}\mathbb{E}S^{(n)}$，根据 Shannon 的 NLCT 准则，它满足

$$\frac{h_q^{(n)}}{n} \leqslant e_n \leqslant \frac{h_q^{(n)}}{n}+\frac{1}{n} \tag{1.1.17}$$

可以看到，当 n 较大时，$e_n\sim h_q^{(n)}/n$。 14

例子 1.1.19 对于一个发送字母 i 且概率为 $p(i)$ 的 Bernoulli 信源（如例子 1.1.2），等式(1.1.16)产生

$$h_q^{(n)} = -\sum_{(i_1,\cdots,i_n)} p(i_1)\cdots p(i_n)\log_q(p(i_1)\cdots p(i_n))$$
$$= -\sum_{j=1}^{n}\sum_{i_1,\cdots,i_n} p(i_1)\cdots p(i_n)\log_q p(i_j) = nh_q \tag{1.1.18}$$

其中，$h_q = -\sum p(i)\log_q p(i)$。这里 $e_n \sim h_q$。因此，对于较大的 n，在一个分段编码中，每个信源字母的最小期望码长最终会达到下界(1.1.13)，因此不会超过 $\min\mathcal{S}$，也就是逐字母编码的最小期望码长。这种现象在后面信源不独立的情况更为明显。在很多情况下，$h_q^{(n)} \ll nh_q$，即 $e_n \ll h_q$。这是数据压缩的核心思想。

所以，长字符串统计成为信源的重要特性。由定义可知，长度为 n 的字符串 $u^{(n)} = u_1\cdots u_n$ 满足 Cartesian 积 $I^{\times n}$；像这种总数是 m^n 的字符串，为了能够编码这些字符串，需要 $m^n = 2^{n\log m}$ 个不同的码字。如果码字有固定的长度(保证了无前缀属性)，那么这个长度介于 $\lfloor n\log m \rfloor$ 和 $\lceil n\log m \rceil$ 之间。那么对于较大的 n，编码速率接近 $\log m$ 比特每信源符号。但是如果有一些字符串很少出现，我们可以忽略它，并减少所使用码字的数量。这引出了以下的定义。

定义 1.1.20 如果一个信源以 $R>0$ 的速率编码，对于任意 n，我们能够得到一个集合 $A_n \subset I^{\times n}$，使得

$$\#A_n \leqslant 2^{nR} \quad 且 \quad \lim_{n\to\infty} P(U^{(n)} \in A_n) = 1 \tag{1.1.19}$$

换句话说，对于长信源字符串，我们能够在可忽略误差的情况下，以速率 R 来编码信息。

定义 1.1.21 一个给定信源的信息速率是可靠编码速率的下确界：

$$H = \inf[R: R \text{ 是可靠的}] \tag{1.1.20}$$

定理 1.1.22 对于具有符号集 I_m 的信源，

$$0 \leqslant H \leqslant \log m \tag{1.1.21}$$

15 两个界都是可达的。

证明 LHS 不等式是很简单的。它是由一个退化信源得到的(比如，例子 1.1.2c)；这里，A_n 包含小于或等于 m 个常量字符串。对于任意的 $R>0$，A_n 最终小于 2^{nR}。另一方面，$\#I^{\times n} = m^n = 2^{n\log m}$，因此 RHS 不等式成立。对于独立同分布信源，$p(u) = 1/m$；在这种情况下，$P(A_n) - (1/m^n)\#A_n$，当 $\#A_n \leqslant 2^{nR}$，$R < \log m$ 时，$P(A_n)$ 接近于零。 □

例子 1.1.23 (a)对于电报英语，$m = 27 \simeq 2^{4.76}$ 即 $H \leqslant 4.76$。幸运的是，$H \ll 4.76$ 使得以下各项成为可能：(i)数据压缩；(ii)错误校验；(iii)解码；(iv)译码。电报英语的精确值 H 还不知道(更不用说文学英语了)，准确地评估它是件很困难的事情。然而，假设它来自于通过大量“随机性”和“同质化”操作的信源，现代理论工具以及计算设备就能够计算出这些给定(长)文本的信息速率(见文献[36]的 8.3 节)。

数值分析的许多结果在文献[136]中能够找到，它解释了三种文本：(a)莎士比亚文集；(b)来自各种报刊的混合文本；(c)英皇欽定圣经。这些文本都取消了标点，单词之间的空格也都删除了。文本(a)和(b)分别得到的值是 1.7 和 1.25(这对现代新闻业是相当有吸引力的)。在情况(c)中，值是不定的；很显然在这个例子中，上述假设并不适合这种情况。(比如，创世纪的系谱枚举很难与 Paul 的哲学讨论信件相比较，所以信源的同质化很明显没有得以保持。)

更具有挑战性的是比较两种语言：哪一个更加适合互相通信？用上述实验来比较托尔斯泰和陀思妥耶夫斯基的文集也会很有趣。

为了便于说明，我们给出了莫尔斯电码创始人 Samuel Morse(1791—1872)的原始表，

它提供了在由相对少量的常用词占主导的电报英文中不同字母的频度。

E	T	A	I	N	O	S	H	R
12000	9000	8000	8000	8000	8000	8000	6400	6200
D	L	U	C	M	F	W	Y	G
4400	4000	3400	3000	3000	2500	2000	2000	1700
P	B	V	K	Q	J	X	Z	
1700	1600	1200	800	500	400	400	200	

(b) 相似的想法应用在给定数的十进制和二进制分解上。比如，取数字 π。如果对于二进制分解的信息速率接近 1(这是一个随机选择序列的信息速率)，我们也许会认为 π 表现得像随机数一样，或者说我们可以认为它是一个经特殊选择的数。同样的问题是关于 e、$\sqrt{2}$或者 Euler-Mascheroni 常量 $\gamma=\lim_{N\to\infty}\left(\sum_{1\leqslant n\leqslant N}\frac{1}{n}-\ln N\right)$。(Hilbert 的一个开放性问题就是证明或者反证 γ 是个超越数，这些超越数形成了在并发数 Bernoulli 信源下概率为 1 的集合。)就像多个实验结果显示的一样，对于在 $N\sim500\,000$ 中的数，上面提到的数字表现出像随机数一样的行为方式，见文献[26]。在 1.3 节中，我们将计算出 Bernoulli 和 Markov 信源的信息速率。 16

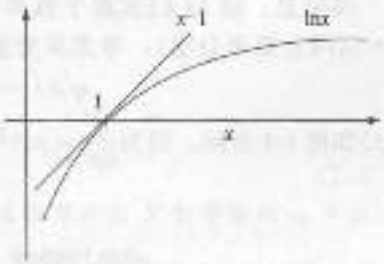


图 1-3

我们用下面简单但很基本的事实总结这一节。

定理 1.1.24 (Gibbs 不等式：参见 PSE II，421 页)假设$\{p(i)\}$和$\{p'(i)\}$是两个概率分布(在一个有限或者可数集 I 上的)。那么，对于任何 $b>1$，

$$\sum_i p(i)\log_b\frac{p'(i)}{p(i)}\leqslant 0,\quad 也就是\ -\sum_i p(i)\log_b p(i)\leqslant-\sum_i p(i)\log_b p'(i) \tag{1.1.22}$$

并且当且仅当 $p(i)=p'(i)$，$i\in I$ 时，等式成立。

证明 当 $x>0$，界

$$\log_b x\leqslant\frac{x-1}{\ln b}$$ 17

成立，当且仅当 $x=1$ 时，等式成立。令 $I'=\{i:\ p(i)>0\}$，我们有

$$\sum_i p(i)\log_b\frac{p'(i)}{p(i)}=\sum_{i\in I'}p(i)\log_b\frac{p'(i)}{p(i)}\leqslant\frac{1}{\ln b}\sum_{i\in I'}p(i)\left(\frac{p'(i)}{p(i)}-1\right)$$
$$=\frac{1}{\ln b}\left(\sum_{i\in I'}p'(i)-\sum_{i\in I'}p(i)\right)=\frac{1}{\ln b}\left(\sum_{i\in I'}p'(i)-1\right)\leqslant 0$$

如果等式成立，需要：(a)当 $p(i)=0$ 时，$\sum_{i\in I'}p'(i)=1$，即 $p'(i)=0$；(b)对于 $i\in I'$，$p'(i)/p(i)=1$。 □

1.2 熵：简介

只有熵变容易了。

Anton Chekhov(1860—1904)，俄罗斯作家及剧作家

这一节全部用来介绍熵的性质，为了简单起见，我们研究二进制的熵，这里的对数底

都是2，因此，我们将记号 h_2 的下标2省略。我们首先简单重复基本定义，重点强调其不同之处。

定义 1.2.1 给定事件 A 的概率为 $p(A)$，则 A 发生得到的信息量被定义为：

$$i(A) = -\log_2 p(A)$$

另外，假设 X 为一个随机变量，它从不同数值集合 $\{x_1, \cdots, x_m\}$ 中取有限个数，概率为 $p_i = p_X(x_i) = \mathbb{P}(X = x_i)$，二进制熵 $h(X)$ 被定义为从观察 X 中获得的信息期望：

$$h(X) = -\sum_{x_i} p_X(x_i)\log_2 p_X(x_i) = -\sum_i p_i \log_2 p_i = \mathbb{E}[-\log_2 p_X(X)] \quad (1.2.1)$$

（鉴于等式 $0 \cdot \log 0 = 0$，对于那些使 $p_X(x_i) > 0$ 的 x_i 来说和也许会减小。）

有时候换种思路会变得很有效：$i(A)$ 代表描述事件 A 所需要的信息量，$h(X)$ 给出了描述随机变量 X 的信息量的期望。

很明显，熵 $h(X)$ 依赖于概率分布而不是值 $x_1, \cdots x_m$：$h(X) = h(p_1, \cdots p_m)$。对于 $m=2$（两点概率分布），考虑单变量 $p \in [0, 1]$ 的函数 $\eta(p)$（$-\eta_2(p)$）更加方便：

18

$$\eta(p) = -p\log p - (1-p)\log(1-p) \quad (1.2.2a)$$

$\eta(p)$ 如图 1-4 所示，因为 $\frac{d^2}{dp^2}\eta(p) = -(\log e)/[p(1-p)] < 0$，其图像是凹形的，见图 1-4。

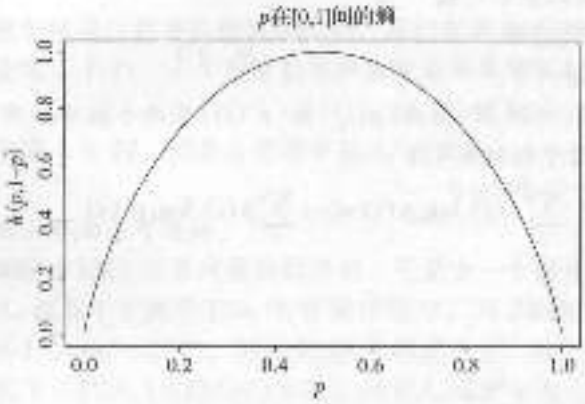


图 1-4

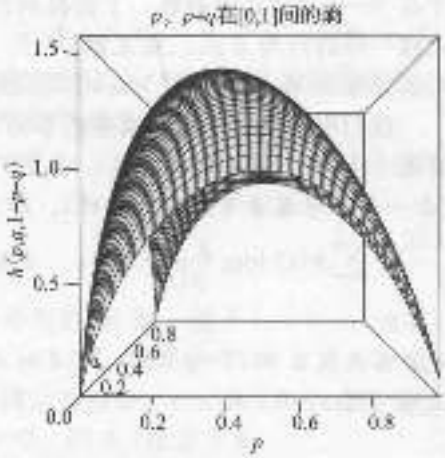


图 1-5

三点分布的熵

19

$$\eta_3(p, q) = -p\log p - q\log q - (1-p-q)\log(1-p-q)$$

如图 1-5 所示，其中变量 $p, q \in [0, 1]$ 且 $p+q \leqslant 1$，它也表现出凹函数的性质。

定义 1.2.1 表明对于两个独立的事件：A_1 和 A_2，

$$i(A_1 \cap A_2) = i(A_1) + i(A_2) \quad (1.2.2b)$$

且对于 $p(A) = 1/2$ 的事件 A，$i(A) = 1$。

定义 1.2.1 的验证来自一个事实：对于任何一个函数 $i^*(A)$，这个函数(i)取决于 A 的概率 $p(A)$（如果 $p(A) = p(A')$，则 $i^*(A) = i^*(A')$）；(ii)关于 $p(A)$ 是连续的；(iii)满足式(1.2.2b)，且与 $i(A)$ 一致（公理化定义的熵，参考下面的举例 1.2.24）。

定义 1.2.2 给定一对随机变量 X，Y，且取值分别为 x_i，y_j，联合熵 $h(x, y)$ 被定义为：

$$h(X,Y) = -\sum_{x_i, y_j} p_{X,Y}(x_i, y_j)\log p_{X,Y}(x_i, y_j) = \mathbb{E}[-\log_2 p_{X,Y}(X,Y)] \quad (1.2.3)$$

其中 $p_{X,Y}(x_i, y_j) = \mathbb{P}(X = x_i, Y = y_j)$ 是联合概率分布，换句话说，$h(X, Y)$ 是随机向量

(X, Y)取值为(x_i, y_j)的熵。

定义给定Y的X的条件熵$h(X|Y)$为给定Y的值时，从观察X获得的信息量的期望：

$$h(X|Y) = -\sum_{i,j} p_{X,Y}(x_i, y_j) \log p_{X|Y}(x_i|y_j) = \mathbb{E}[-\log p_{X|Y}(X|Y)] \quad (1.2.4)$$

在这里，$p_{X,Y}(i, j)$是$\mathbb{P}(X=x_i, Y=y_j)$的联合概率，$p_{X|Y}(x_i|y_j)$是$\mathbb{P}(X=x_i, Y=y_j)$的条件概率。很显然，式(1.2.3)和式(1.2.4)表明：

$$h(X|Y) = h(X,Y) - h(Y) \quad (1.2.5)$$

注意，一般来说，$h(X|Y) \neq h(Y|X)$。

对于在同一集合I中取值的随机变量X和Y，使得对于$p_Y(x)>0(x\in I)$，$h(X \| Y)$(也被称为X相对于Y的熵或者Kullback-Leibler距离$D(p_X \| p_Y)$)被定义为：

$$h(X \| Y) = \sum_x p_X(x) \log \frac{p_X(x)}{p_Y(x)} = \mathbb{E}_X\left\{-\log \frac{p_Y(X)}{p_X(X)}\right\} \quad (1.2.6)$$

其中$p_X(x)=\mathbb{P}(X=x)$，$p_Y(x)=\mathbb{P}(Y=x)$，$x\in I$。

熵的简单性质如下。

定理 1.2.3 (a) 如果一个随机变量X的最大值是m，则：

$$0 \leqslant h(X) \leqslant \log m \quad (1.2.7)$$

20

LHS成立当且仅当X取一个单一的值，且RHS等式成立当且仅当X取等概的m个值。

(b) 联合熵满足

$$h(X,Y) \leqslant h(X) + h(Y) \quad (1.2.8)$$

当且仅当X和Y是独立的，等号成立，即对于所有的$x, y\in I$，$\mathbb{P}(X=x, Y=y)=\mathbb{P}(X=x)\mathbb{P}(Y=y)$。

(c) 相对熵总是非负的：

$$h(X \| Y) \geqslant 0 \quad (1.2.5)$$

当且仅当X和Y是同分布时等式成立，即$p_X(x)=p_Y(x)$，$x\in I$。

证明 根据定理1.1.24，结论(c)与Gibbs不等式等价。接下来，(a)可以由(c)推导出，其中$\{p(i)\}$为X的概率分布且$p'(i)=1/m$，$1\leqslant i\leqslant m$。类似地，(b)也可以由(c)推导出，其中，i变为X，Y的数值对(i_1, i_2)，且$p(i)=p_{X,Y}(i_1, i_2)$为X，Y的联合概率分布，$p'(i)=p_X(i_1)p_Y(i_2)$表示它们边缘分布概率的乘积。形式上：

(a)
$$h(X) = -\sum_i p(i)\log p(i) \leqslant \sum_i p(i)\log m = \log m$$

(b)
$$\begin{aligned} h(X,Y) &= -\sum_{i_1,i_2} p_{X,Y}(i_1,i_2)\log p_{X,Y}(i_1,i_2) \\ &\leqslant -\sum_{i_1,i_2} p_{X,Y}(i_1,i_2) \times \log(p_X(i_1)p_Y(i_2)) \\ &= -\sum_{i_1} p_X(i_1)\log p_X(i_1) - \sum_{i_2} p_Y(i_2)\log p_Y(i_2) \\ &= h(X) + h(Y) \end{aligned}$$

这里我们用到了恒等式：$\sum_{i_2} p_{X,Y}(i_1,i_2) = p_X(i_1)$，$\sum_{i_1} p_{X,Y}(i_1,i_2) = p_Y(i_2)$。 □

举例 1.2.4 (a) 证明几何随机变量Y：$p_j=\mathbb{P}(Y=j)=(1-p)p^j$，$j=0, 1, 2, \cdots$，在所有具有相同均值且$\mathbb{Z}_+=\{0, 1, 2, \cdots\}$的分布中，它具有最大的熵。

(b) 设Z是一个从有限集合K中取值的随机变量，f是一个已知的实函数$f: K\to\mathbb{R}$，其

中 $f_*=\min[f(k):k\in K]$，且 $f^*=\max[f(k):k\in K]$. 设 $E(f)=\sum_{k\in K}f(k)/(\#K)$，并考虑随机变量 Z 的熵 $h(Z)$ 的最大化问题，约束条件为：

21

$$\mathbb{E}f(Z)\leqslant\alpha \tag{1.2.10}$$

证明：

(bi) 当 $f_*\geqslant\alpha\geqslant E(f)$ 时，那么最大化的概率分布是关于 K 的均匀分布，即 $\mathbb{P}(Z=k)=1/(\#K)$，$k\in K$。

(bii) 当 $f_*\leqslant\alpha<E(f)$，且 f 不是一个恒定变换时，那么最大化的概率分布为：

$$\mathbb{P}(Z=k)=p_k=e^{\lambda f(k)}/\sum_k e^{\lambda f(k)},k\in K \tag{1.2.11}$$

其中有 $\lambda=\lambda(\alpha)<0$，且满足：

$$\sum_k p_k f(k)=\alpha \tag{1.2.12}$$

此外，假设 Z 取可数的多个值，但是 $f\geqslant 0$，并且对于一个给定的 α 总存在 $\lambda<0$，$\sum_k e^{\lambda f(k)}<\infty$ 以及 $\sum_k p_k f(k)=\alpha$ 成立，其中 p_k 具有式(1.2.11)的形式。

(biii) 式(1.2.11)的概率分布仍然可以最大化式(1.2.10)中的 $h(Z)$。从(biii)中可以推出结论(a)。

(c) 证明 $h_Y(X)\geqslant 0$，并且对于所有 x，当且仅当 $\mathbb{P}(X=x)=\mathbb{P}(Y=x)$ 时，等式成立。考虑 Y 这个具有适当选择参数的属于 $\mathbb{Z}_+$ 的几何随机变量，证明如果均值 $\mathbb{E}X=\mu<\infty$，那么

$$h(X)\leqslant(\mu+1)\log(\mu+1)-\mu\log\mu \tag{1.2.13}$$

当且仅当 X 为几何随机变量时，等号成立。

解答 (a) 根据 Gibbs 不等式，对于所有均值为 $\sum_{i\geqslant 0}iq_i\leqslant\mu$ 的概率分布 $(q_0,q_1,\cdots)$，

$$h(q)=-\sum_i q_i\log q_i\leqslant-\sum_i q_i\log p_i=-\sum_i q_i(\log(1-p)+i\log p)$$
$$\leqslant-\log(1-p)-\mu\log p=h(Y)$$

其中 $p=\rho/1-\rho$。当且仅当 q 为均值为 μ 的几何随机变量，则等式成立。

(b) 首先观察关于 $p_k=1/(\#k)$ 的均匀分布，对应于式(1.2.11)中的 $\lambda=0$，使得 $h(Z)$ 获得全局最大化，在(bi)中，这个分布满足式(1.2.10)。因而在这个约束条件下，最大化

22

$h(Z)$。转到(bii)，假设 $p_k^*=e^{\lambda f(k)}/\sum_k e^{\lambda f(k)}$，$k\in K$，在这里选选 λ 来满足 $\mathbb{E}^*f(Z)=\sum_k p_k^* f(k)=\alpha$。假设 $q=\{q_k\}$ 是满足 $\mathbb{E}_q f=\sum_k q_k f(k)\leqslant\alpha$ 的任意分布。观察可得，由式(1.2.11)计算出的概率分布的期望是关于 λ 的非减函数。事实上，导数

$$\frac{\mathrm{d}\alpha}{\mathrm{d}\lambda}=\frac{\sum_k[f(k)]^2e^{\lambda f(k)}}{\sum_k e^{\lambda f(k)}}-\frac{\left(\sum_k f(k)e^{\lambda f(k)}\right)^2}{\left(\sum_k e^{\lambda f(k)}\right)^2}=\mathbb{E}[f(Z)]^2-[\mathbb{E}f(Z)]^2$$

是一个正数(它导出了随机变量 $f(Z)$ 的方差)；对于非恒定的 f，RHS 是实际上是非负的。因此对于非恒定的 f(即关于 $f_*<E(f)<f^*$)，对于在 $[f_*,f^*]$ 之间的任意 α，恰好存在一个满足式(1.2.12)且形如式(1.2.11)的概率分布，对于 $f_*\leqslant\alpha<E(f)$，有 $\lambda(\alpha)<0$。

接下来，我们利用这样一个事实：Kullback-Leibler 的距离 $D(q\mid p^*)$(如式(1.2.6))满足 $D(q\mid p^*)=\sum_k q_k\log(q_k/p_k^*)\geqslant 0$(Gibbs 不等式)以及 $\sum_k q_k f(k)\leqslant\alpha$，$\lambda<0$ 来获得

$$
\begin{aligned}
h(q) &= -\sum_k q_k \log q_k = -D(q \| p^*) - \sum_k q_k \log p_k^* \\
&\leqslant -\sum_k q_k \log p_k^* = -\sum_k q_k \left\{ -\log \sum_i e^{\lambda f(i)} + \lambda f(k) \right\} \\
&\leqslant -\sum_k q_k \left(-\log \sum_i e^{\lambda f(i)} \right) - \lambda \alpha = -\sum_k p_k^* \left(-\log \sum_i e^{\lambda f(i)} + \lambda f(k) \right) \\
&= -\sum_k p_k^* \log p_k^* = h(p^*)
\end{aligned}
$$

对于(biii)，如果从式(1.2.12)中得出的 $\lambda(\alpha)<0$，对于一个可数无限集 K，上述讨论仍然成立。

(c) 根据 Gibbs 不等式，$h_Y(X)\geqslant 0$，然后，通过 $f(k)=k$，$\alpha=\mu$ 以及 $\lambda=\log q$，我们使用(b)的结果，最大的熵分布可写为 $p_j^*=(1-p)p^j$，$j=0, 1, 2, \cdots$，这里 $\sum_k k p_j^* = \mu$，或 $\mu=p/1-p$，这个分布的熵等于

$$
\begin{aligned}
h(p^*) &= -\sum_j (1-p)p^j \log((1-p)p^j) \\
&= -\frac{p}{1-p}\log p - \log(1-p) = (\mu+1)\log(\mu+1) - \mu\log\mu
\end{aligned}
$$

其中，$\mu=p/(1-p)$。

或者：

$$
\begin{aligned}
0 \leqslant h_Y(X) &= \sum_i p(i) \log \frac{p(i)}{(1-p)p^i} \\
&= -h(X) - \log(1-p)\sum_i p(i) - (\log p)\left(\sum_i i p(i)\right) \\
&= -h(X) - \log(1-p) - \mu\log p
\end{aligned}
$$

p 的最佳选择是 $p=\mu/(\mu+1)$，那么，

$$
h(X) \leqslant -\log\frac{1}{\mu+1} - \mu\log\frac{\mu}{\mu+1} = (\mu+1)\log(\mu+1) - \mu\log\mu
$$

RHS 就是几何随机变量 Y 的熵 $h(Y)$)。当且仅当 $X\sim Y$，即 X 是几何随机变量时，等号成立。 □

Gibbs 不等式的一个简单但有用的推论是如下。

引理 1.2.5 (pooling 不等式)对于任意 q_1，$q_2\geqslant 0$，如果 $q_1+q_2>0$，则

$$
\begin{aligned}
-(q_1+q_2)\log(q_1+q_2) &\leqslant -q_1\log q_1 - q_2\log q_2 \\
&\leqslant -(q_1+q_2)\log\frac{q_1+q_2}{2}
\end{aligned}
\tag{1.2.14}
$$

当且仅当 $q_1q_2=0$ 即 q_1 或 q_2 为 0)时，第一个等号成立。当且仅当 $q_1=q_2$ 时，第二个等号成立。

证明 实际上，(1.2.14)等效为

$$
0 \leqslant h\left(\frac{q_1}{q_1+q_2}, \frac{q_2}{q_1+q_2}\right) \leqslant \log 2 (=1)
$$

□

根据引理 1.2.5，"粘合"一个随机变量的不同值可以减小对熵的相应贡献。另一方面，等概率的"重新分布"可以增加贡献。引理 1.2.5 的一个显而易见的推论如下。

定理 1.2.6 假设一个离散随机变量 X 是另一个离散随机变量 Y 的函数：$X=\phi(Y)$。那么

$$
h(X) \leqslant h(Y) \tag{1.2.15}
$$

当且仅当 ϕ 可逆时等号成立。

证明　实际上，如果 ϕ 可逆，那么 X 和 Y 概率分布的区别只在于概率的顺序，不会改变
24 熵值。如果 ϕ "粘合"某些值 y_i，那么我们可以反复利用 pooling 不等式的左值。 □

举例 1.2.7　假设 $p_1, \cdots, p_n$ 是一个概率分布，其中 $p^* = \max\{p_i\}$，证明下面的熵 $h = -\sum_i p_i \log p_i$ 的下界：

(i) $h \geqslant -p^* \log p^* - (1-p^*)\log(1-p^*) = \eta(p^*)$

(ii) $h \geqslant -\log p^*$

(iii) $h \geqslant 2(1-p^*)$

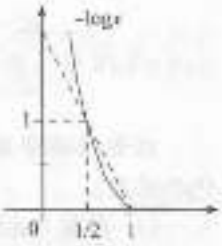


图 1-6

解答　(i)可由 pooling 不等式推导出。(ii)的成立是由于

$$h \geqslant -\sum_i p_i \log p^* = -\log p^*$$

对于(iii)，先假设 $p^* \geqslant 1/2$。由于函数 $p \mapsto \eta(p)$，$0 \leqslant p \leqslant 1$ 是凹的(参见式(1.2.3))，它在[1/2，1]上的图像在直线 $x \mapsto 2(1-p)$ 之上(如图 1-6 所示)。那么，由(i)可得

$$h \geqslant \eta(p^*) \geqslant 2(1-p^*) \tag{1.2.16}$$

另一方面，如果 $p^* \leqslant 1/2$，我们利用(ii)可得：

$$h \geqslant -\log p^*$$

并利用不等式：对于 $0 \leqslant p \leqslant 1/2$ 有 $-\log p \geqslant 2(1-p)$。 □

定理 1.2.8　(Fano 不等式)假设随机变量 X 取 $m>1$ 个值，其中一个值的概率为 $(1-\varepsilon)$，那么

$$h(X) \leqslant \eta(\varepsilon) + \varepsilon \log(m-1) \tag{1.2.17}$$

25 其中 η 是式(1.2.2a)定义的函数。

证明　假设 $p_1 = p(x_1) = 1-\varepsilon$，那么

$$\begin{aligned} h(X) = h(p_1, \cdots, p_m) &= -\sum_{i=1}^{m} p_i \log p_i \\ &= -p_1 \log p_1 - (1-p_1)\log(1-p_1) + (1-p_1)\log(1-p_1) - \sum_{2 \leqslant i \leqslant m} p_i \log p_i \\ &= h(p_1, 1-p_1) + (1-p_1) h\left(\frac{p_2}{1-p_1}, \cdots, \frac{p_m}{1-p_1}\right) \end{aligned}$$

右边第一项等于 $\eta(\varepsilon)$，第二项不超过 $\varepsilon \log(m-1)$。 □

定义 1.2.9　给定随机变量 X，Y，Z，我们称当给定 Z 的 X 和 Y 是条件独立的，如果对所有 x 和 y，以及对所有 z 满足 $\mathbb{P}(Z=z)>0$，有

$$\mathbb{P}(X=x, Y=y \mid Z=z) = \mathbb{P}(X=x \mid Z=z)\mathbb{P}(Y=y \mid Z=z) \tag{1.2.18}$$

对于条件熵我们立即得到如下定理。

定理 1.2.10　(a) 对于所有的随机变量 X，Y，有

$$0 \leqslant h(X \mid Y) \leqslant h(X) \tag{1.2.19}$$

当且仅当 X 是 Y 的函数时第一个等号成立，当且仅当 X 和 Y 相互独立时第二个等号成立。

(b) 对于所有的随机变量 X，Y，Z，

$$h(X \mid Y, Z) \leqslant h(X \mid Y) \leqslant h(X \mid \phi(Y)) \tag{1.2.20}$$

当且仅当 X 和 Z 在给定 Y 条件下相互独立时第一个等号成立，当且仅当 X 和 Z 在给定 $\phi(Y)$ 条件下相互独立时第二个等号成立。

证明　(a) 式(1.2.19)的左边可由式(1.2.4)推导出(因为 $h(X \mid Y)$ 是非负项的和)。右边可

由式(1.2.5)和式(1.2.8)推导出。式(1.2.19)的左边等效为等式 $h(X, Y)=h(Y)$ 或 $h(X, Y)=h(\phi(X, Y))$，其中 $\phi(X, Y)=Y$。通过定理1.2.6可知，当且仅当依概率1映射 $(X, Y)\mapsto Y$ 是可逆的，即 X 是 Y 的函数时，等式成立。当且仅当 $h(X, Y)=h(X)+h(Y)$，即 X 和 Y 相互独立时，式(1.2.19)的右边等号成立。

(b) 对于下界，使用与式(1.2.5)相似的公式

$$h(X|Y,Z)=h(X,Z|Y)-h(Z|Y) \tag{1.2.21}$$

和与式(1.2.10)相似的不等式

$$h(X,Z|Y)\leqslant h(X|Y)+h(Z|Y) \tag{1.2.22}$$

[26]

当且仅当 X 和 Z 在给定 Y 条件下相互独立时等号成立。对于右边的界，使用：

(i) 等式 $h(X|Y, \phi(Y))=h(X, Y|\phi(Y))-h(Y|\phi(Y))$，它是式(1.2.21)中的一个特例，同时注意到 $h(X|Y, \phi(Y))=h(X|Y)$。

(ii) 不等式 $h(X, Y|\phi(Y))\leqslant h(X|\phi(Y))+h(Y|\phi(Y))$，它是式(1.2.22)中的一个特例，当且仅当 X 和 Y 在给定 $\phi(Y)$ 条件下相互独立时等号成立。 □

上面的定理1.2.8和下面的定理1.2.11说明了当 X 近似为一个常数(相应地，近似为 Y 的一个函数)时熵 $h(X)$ 和条件熵 $h(X|Y)$ 如何受到制约。

定理1.2.11 (广义Fano不等式)对于一对分别从 $x_1, \cdots, x_m$ 和 $y_1, \cdots, y_n$ 取值的随机变量 X 和 Y，如果

$$\sum_{j=1}^{n}\mathbb{P}(X=x_1, Y=y_j)=1-\varepsilon \tag{1.2.23}$$

那么

$$h(X|Y)\leqslant \eta(\varepsilon)+\varepsilon\log(m-1) \tag{1.2.24}$$

其中 $\eta(\varepsilon)$ 在(1.2.3)中已定义。

证明 令 $\varepsilon_j=\mathbb{P}(X\neq x_1|Y=y_j)$，我们有

$$\sum_j p_Y(y_j)\varepsilon_j=\sum_j\mathbb{P}(X\neq x_1, Y=y_j)=\varepsilon \tag{1.2.25}$$

根据条件熵的定义，Fano不等式和函数 $\eta(\cdot)$ 的凹性可得，

$$\begin{aligned}h(X|Y)&\leqslant\sum_j p_Y(y_j)(\eta(\varepsilon_j)+\varepsilon_j\log(m-1))\\&\leqslant\sum_j p_Y(y_j)\eta(\varepsilon_j)+\varepsilon\log(m-1)\leqslant\eta(\varepsilon)+\varepsilon\log(m-1)\end{aligned}$$ □

如果随机变量 X 取可数的多个值 $\{x_1, x_2, \cdots\}$，那么上面的定义和大多数论断可以重复；而取不同值时，显著的区别在于(1.2.7)中右边的界以及式(1.2.17)和式(1.2.24)中的不等式。

目前所列出的熵的许多性质可以扩展到随机序列中。

定理1.2.12 对于一对随机序列 $\boldsymbol{X}^{(n)}=(X_1, \cdots, X_n)$ 和 $\boldsymbol{Y}^{(n)}=(Y_1, \cdots, Y_n)$， [27]

(a) 联合熵，即

$$h(\boldsymbol{X}^{(n)})=-\sum_{x^{(n)}}\mathbb{P}(\boldsymbol{X}^{(n)}=\boldsymbol{x}^{(n)})\log\mathbb{P}(\boldsymbol{X}^{(n)}=\boldsymbol{x}^{(n)})$$

满足

$$h(\boldsymbol{X}^{(n)})=\sum_{i=1}^{n}h(X_i|\boldsymbol{X}^{(i-1)})\leqslant\sum_{i=1}^{n}h(X_i) \tag{1.2.26}$$

当且仅当元素 $X_1, \cdots, X_n$ 相互独立时等号成立；

(b) 条件熵，即

$$h(X^{(n)}|Y^{(n)})=-\sum_{x^{(n)},y^{(n)}}\mathbb{P}(X^{(n)}=x^{(n)},Y^{(n)}=y^{(n)})\log\mathbb{P}(X^{(n)}=x^{(n)}|Y^{(n)}=y^{(n)})$$

满足

$$h(X^{(n)}|Y^{(n)})\leqslant\sum_{i=1}^{n}h(X_i|Y^{(n)})\leqslant\sum_{i=1}^{n}h(X_i|Y_i) \tag{1.2.27}$$

当且仅当 $X_1,\cdots,X_n$ 在给定 $Y^{(n)}$ 条件下相互独立时左边等号成立，当且仅当对于每个 $i=1,\cdots,n$，X_i 和 $\{Y_r:1\leqslant r\leqslant n,r\neq i\}$ 在给定 Y_i 条件下相互独立时右边等号成立。

证明 此证明重复之前标量情况下的过程即可。 □

定义 1.2.13 X 和 Y 的互信息或互熵 $I(X;Y)$ 定义为

$$\begin{aligned}I(X;Y)&=\sum_{x,y}p_{X,Y}(x,y)\log\frac{p_{X,Y}(x,y)}{p_X(x)p_Y(y)}=\mathbb{E}\log\frac{p_{X,Y}(X,Y)}{p_X(X)p_Y(Y)}\\&=h(X)+h(Y)-h(X,Y)=h(X)-h(X|Y)\\&=h(Y)-h(Y|X)\end{aligned} \tag{1.2.28}$$

从定义中可以看出，$I(X;Y)=I(Y;X)$。

直观上来看，$I(X;Y)$ 度量了 Y 承载的关于 X 的信息量(反之亦然)。定理 1.2.10(b) 表明了如下定理。

定理 1.2.14 如果随机变量 $\phi(Y)$ 是 Y 的一个函数，那么

$$0\leqslant I(X;\phi(Y))\leqslant I(X;Y) \tag{1.2.29}$$

当且仅当 X 和 $\phi(Y)$ 独立时第一个等号成立，当且仅当 X 和 Y 在给定 $\phi(Y)$ 条件下相互独立时第二个等号成立。

28

举例 1.2.15 假设两个非负的随机变量 X 和 Y，其中 $Y=X+N$，而 N 是在 $\mathbb{Z}_+$ 上取值的几何随机变量，并且独立于 X。请确定 Y 的分布，使得 X 和 Y 的互信息在期望 $\mathbb{E}X\leqslant K$ 的约束下最大，并且说明这个分布可以通过在某一概率下将 X 赋值为 0，在互补概率下让其服从几何分布来实现。

解答 由于 $Y=X+N$，X 和 N 相互独立，我们有

$$I(X;Y)=h(Y)-h(Y|X)=h(Y)-h(N)$$

此外 $\mathbb{E}(Y)=\mathbb{E}(X)+\mathbb{E}(N)\leqslant K+\mathbb{E}(N)$。所以如果我们能保证 Y 服从均值为 $K+\mathbb{E}(N)$ 的几何分布，那么它能给出 $I(X;Y)$ 最大值。基于此，写出概率生成函数的等式：

$$\mathbb{E}(z^Y)=\mathbb{E}(z^X)\mathbb{E}(z^N),z>0$$

其中 $\mathbb{E}(z^N)=(1-p)/(1-zp)$，$0<z<1/p$，以及

$$\mathbb{E}(z^Y)=\frac{1-p^*}{1-zp^*},0<z<\frac{1}{p^*}$$

其中 p^* 可以从下式中求得

$$\mu_Y=\frac{p^*}{1-p^*}=K+\frac{p}{1-p}=\frac{K(1-p)+p}{1-p}$$

这会得到

$$p^*=\frac{K(1-p)+p}{1+K(1-p)},E(z^Y)=\frac{1-p}{1-K(1-p)-z(p+K(1-p))}$$

以及

$$\mathbb{E}(z^X)=\frac{1-zp}{1-K(1-p)-z(p+K(1-p))} \tag{1.2.30}$$

在这个例子中提到的 X 的分布形式会导致

$$\mathbb{E}(z^X) = \kappa_0 + (1-\kappa_0)\frac{1-p_X}{1-zp_X} \tag{1.2.31}$$

其中 $\kappa_0 + (1-\kappa_0)(1-p_X) = \mathbb{P}(X=0)$。选取

$$p_X = \frac{p+K(1-p)}{1+K(1-p)}, \kappa_0 = \frac{p}{p+K(1-p)}$$

我们看到式(1.2.30)和式(1.2.31)相等。 □ 29

我只求消息……

Charles Dickens(1812—1870)，英国作家，引自《David Copperfield》

在定义 1.2.13 和定理 1.2.14 中，随机变量 X 和 Y 可以被随机序列取代。此外，重复上面对于序列 $\boldsymbol{X}^{(n)}$ 和 $\boldsymbol{Y}^{(n)}$ 的证明，我们得到以下定理。

定理 1.2.16 (a) 随机序列之间的互信息服从

$$I(\boldsymbol{X}^{(n)};\boldsymbol{Y}^{(n)}) \geqslant h(\boldsymbol{X}^{(n)}) - \sum_{i=1}^{n} h(X_i|\boldsymbol{Y}^{(n)}) \geqslant h(\boldsymbol{X}^{(n)}) - \sum_{i=1}^{n} h(X_i|Y_i) \tag{1.2.32}$$

(b) 如果 $X_1, \cdots, X_n$ 相互独立，那么

$$I(\boldsymbol{X}^{(n)};\boldsymbol{Y}^{(n)}) \geqslant \sum_{i=1}^{n} I(X_i;\boldsymbol{Y}^{(n)}) \tag{1.2.33}$$

观察得到

$$\sum_{i=1}^{n} I(X_i;\boldsymbol{Y}^{(n)}) \geqslant \sum_{i=1}^{n} I(X_i;Y_i) \tag{1.2.34}$$

举例 1.2.17 X，Z 是随机变量，$\boldsymbol{Y}^{(n)} = (Y_1, \cdots, Y_n)$ 是一个随机序列。

(a) 证明不等式

$$0 \leqslant I(X;Z) \leqslant \min\{h(X), h(Z)\}$$

(b) 证明或通过反例证伪下面不等式

$$I(X;\boldsymbol{Y}^{(n)}) \leqslant \sum_{j=1}^{n} I(X;Y_j) \tag{1.2.35}$$

首先在 $Y_1, \cdots, Y_n$ 相互独立的假设下证明或证伪，然后在 $Y_1, \cdots, Y_n$ 在给定 X 相互独立的假设下证明或证伪。

(c) 证明或通过反例证伪下面不等式

$$I(X;\boldsymbol{Y}^{(n)}) \geqslant \sum_{j=1}^{n} I(X;Y_j) \tag{1.2.36}$$

首先在 $Y_1, \cdots, Y_n$ 相互独立的假设下证明或证伪，然后在 $Y_1, \cdots, Y_n$ 在给定 X 条件下相互独立的假设下证明或证伪。 30

解答 (a) 根据 Gibbs 不等式，$I(X;Z) \geqslant 0$ 且

$$\begin{aligned} I(X;Z) &= -\sum_{x,z} \mathbb{P}(X=x, Z=z)\log\frac{\mathbb{P}(X=x, Z=z)}{\mathbb{P}(X=x)\mathbb{P}(Z=z)} \\ &= h(X) - h(X|Z) = h(Z) - h(Z|X) \end{aligned}$$

这里 $h(X|Z) \geqslant 0$，$h(Z|X) \geqslant 0$，因此 $I(X;Z) \leqslant h(X)$ 且 $I(X;Z) \leqslant h(Z)$，所以 $I(X;Z) \leqslant \min[h(X), h(Z)]$。

(b) 写出

$$I(X;\boldsymbol{Y}^{(n)}) = h(\boldsymbol{Y}^{(n)}) - h(\boldsymbol{Y}^{(n)}|X) \tag{1.2.37}$$

然后，如果 $Y_1, \cdots, Y_n$ 在给定 X 条件下相互独立，式(1.2.37)的右边等价于

$$h(\boldsymbol{Y}^{(n)})-\sum_{j=1}^{n}h(Y_j|X)\leqslant\sum_{j=1}^{n}[h(Y_j)-h(Y_j|X)]=\sum_{j=1}^{n}I(X;Y_j)$$

即得到了式(1.2.35)。

(c) 接下来，如果Y_1，…，Y_n相互独立，式(1.2.37)的右边等价于

$$\sum_{j=1}^{n}h(Y_j)-h(\boldsymbol{Y}^{(n)}|X)\geqslant\sum_{j=1}^{n}[h(Y_j)-h(Y_j|X)]=\sum_{j=1}^{n}I(X;Y_j)$$

即得到了式(1.2.36)的右值。

另一方面，性质(b)在独立条件下不成立。实际上，取$n=2$，$\boldsymbol{Y}^{(2)}=(Y_1,Y_2)$，令$Y_1$和$Y_2$以1/2概率独立地取值0或1，$j=1,2$。设$X=(Y_1+Y_2)\bmod 2$，那么有

$$h(X)=h(X|Y_j)=1,\quad 所以\ I(X;Y_j)=0,j=1,2$$

但是

$$h(X|\boldsymbol{Y}^{(2)})=0,\quad 所以\ I(X;\boldsymbol{Y}^{(2)})=1$$

此外，(c)在条件独立情况下也不成立。实际上，取关于状态±1的DTMC$(U_1,U_2,\cdots)$，初始分布为$(1/2,1/2)$，转移矩阵为$\begin{pmatrix}0&1\\1&0\end{pmatrix}$。设

$$Y_1=U_1,X=U_2,Y_2=U_3$$

那么Y_1，Y_2在给定X条件下相互独立：$Y_1=Y_2=-X$。另一方面，

$$\begin{aligned}1&=I(X;\boldsymbol{Y}^{(2)})=h(\boldsymbol{Y}^{(2)})=h(Y_1)=h(Y_2)\\&<h(Y_1)+h(Y_2)=I(X;Y_1)+I(X;Y_2)=2\end{aligned}$$

□

31

回忆一个定义在凸集$V\subseteq\mathbb{R}^n$上的实函数$f(\boldsymbol{y})$，它被称为凹函数，如果

$$f(\lambda_0\boldsymbol{y}^{(0)}+\lambda_1\boldsymbol{y}^{(1)})\geqslant\lambda_0f(\boldsymbol{y}^{(0)})+\lambda_1f(\boldsymbol{y}^{(1)})$$

对于任意$\boldsymbol{y}^{(0)}$，$\boldsymbol{y}^{(1)}\in V$，$\lambda_0$，$\lambda_1\in[0,1]$，$\lambda_0+\lambda_1=1$。如果等号仅在$\boldsymbol{y}^{(0)}=\boldsymbol{y}^{(1)}$或$\lambda_0\lambda_1=0$的时候成立，它就叫作严格凹函数。我们把$h(X)$看作变量$\boldsymbol{p}=(p_1,\cdots,p_m)$的函数：在这种情况下集合$V$是$\{\boldsymbol{y}=(y_1,\cdots,y_m)\in\mathbb{R}^m:y_i\geqslant0,1\leqslant i\leqslant m,y_1+\cdots+y_m=1\}$。

定理 1.2.18 熵是概率分布的严格凹函数。

证明 随机变量$X^{(i)}$服从概率分布$\boldsymbol{p}^{(i)}$，$i=0,1$，假设随机变量Λ分别以概率λ_0，λ_1取值0和1，并且独立于$X^{(0)}$，$X^{(1)}$。设$X=X^{(\Lambda)}$，那么不等式$h(\lambda_0\boldsymbol{p}^{(0)}+\lambda_1\boldsymbol{p}^{(1)})\geqslant\lambda_0h(\boldsymbol{p}^{(0)})+\lambda_1h(\boldsymbol{p}^{(1)})$等价于

$$h(X)\geqslant h(X|\Lambda)\tag{1.2.38}$$

它能从式(1.2.19)中推导出。如果我们假设式(1.2.38)的等号成立，那么X和Λ必须独立。另外假设$\lambda_0>0$，利用独立性可写出

$$\mathbb{P}(X=i,\Lambda=0)=\mathbb{P}(X=i)\mathbb{P}(\Lambda=0)=\lambda_0\mathbb{P}(X=i)$$

左边等于$\lambda_1\mathbb{P}(X=i,\Lambda=0)=\lambda_1p_i^{(0)}$，右边等于$\lambda_0(\lambda_0p_i^{(0)}+\lambda_1p_i^{(1)})$。我们可以消去$\lambda_1$得到

$$(1-\lambda_0)p_i^{(0)}=\lambda_0p_i^{(1)}$$

即概率分布$\boldsymbol{p}^{(0)}$和$\boldsymbol{p}^{(1)}$成比例。所以它们要么相等，要么$\lambda_0=0$，$\lambda_1=1$。假设$\lambda_1>0$也可以得到类似的结论。 □

举例 1.2.19 证明

$$\rho(X,Y)=h(X|Y)+h(Y|X)$$

服从

$$\rho(X,Y)=h(X)+h(Y)-2I(X;Y)=h(X,Y)-I(X;Y)=2h(X,Y)-h(X)-h(Y)$$

证明ρ是对称的，即$\rho(X,Y)=\rho(Y,X)\geqslant0$，并且满足三角不等式，即$\rho(X,Y)+\rho(Y,$

$Z)\geqslant\rho(X,Z)$。证明当且仅当 X 和 Y 是彼此的函数时 $\rho(X,Y)=0$。另外证明如果 X' 和 X 是彼此的函数时 $\rho(X,Y)=\rho(X',Y)$。所以 ρ 可以认为是随机变量 X 集合的一个度量，当且仅当 X' 和 X 是彼此的函数时，被认为有等价关系 $X\sim X'$。 32

解答 检验三角不等式

$$h(X|Z)+h(Z|X)\leqslant h(X|Y)+h(Y|X)+h(Y|Z)+h(Z|Y)$$

或者

$$h(X,Z)\leqslant h(X,Y)+h(Y,Z)-h(Y)$$

基于此，写出 $h(X,Z)\leqslant h(X,Y,Z)$。注意到

$$h(X,Z|Y)+h(Y)\leqslant h(X|Y)+h(Z|Y)+h(Y)=h(X,Y)+h(Y,Z)-h(Y)$$

当且仅当(i)$Y=\phi(X,Z)$和(ii)X，Z 是给定 Y 条件下相互独立时等号成立。 □

备注 1.2.20 性质 $\rho(X,Z)=\rho(X,Y)+\rho(Y,Z)$意味着“点”$Y$ 位于通过 X 和 Z 的“直线”上，换句话说，三个点 X，Y，Z 在同一条直线上。给定 Y，X 和 Z 的条件独立性可以用另一种(优雅的)方式表述：$X\to Y\to Z$ 满足 Markov 性(简称是 Markov 的)。然后假设我们有四个随机变量 X_1，X_2，X_3，X_4，使得对所有$1\leqslant i_1<i_2<i_3\leqslant 4$，随机变量 X_{i_1}，X_{i_3} 在给定 X_{i_2} 相互独立；这一性质意味着 $X_1\to X_2\to X_3\to X_4$ 是 Markov 的，或者从几何上看，所有四个点在一条直线上。下面的事实成立：如果 $X_1\to X_2\to X_3\to X_4$ 是 Markov 的，则互熵满足

$$I(X_1;X_3)+I(X_2;X_4)=I(X_1;X_4)+I(X_2;X_3) \tag{1.2.39}$$

等效地，对于联合熵，

$$h(X_1,X_3)+h(X_2,X_4)=h(X_1,X_4)+h(X_2,X_3) \tag{1.2.40}$$

实际上，对于如上的所有三元组 X_{i_1}，X_{i_2}，X_{i_3}，利用度量 ρ 我们有

$$\rho(X_{i_1},X_{i_3})=\rho(X_{i_1},X_{i_2})+\rho(X_{i_2},X_{i_3})$$

按照联合熵和独立熵，它可以重新写为

$$h(X_{i_1},X_{i_3})=h(X_{i_1},X_{i_2})+h(X_{i_2},X_{i_3})-h(X_{i_2})$$

那么式(1.2.39)变为下面的形式

$$\begin{aligned}&h(X_1,X_3)+h(X_2,X_4)-h(X_2)+h(X_2,X_3)+h(X_3,X_4)-h(X_3)\\&=h(X_1,X_2)+h(X_2,X_3)-h(X_2)+h(X_3,X_4)+h(X_3,X_4)-h(X_3)\end{aligned}$$

它是一个简单的恒等式。 33

举例 1.2.21 考虑如下的不等式，假设三变量 $X\to Y\to \mathbf{Z}$ 是 Markov 链，其中 $\mathbf{Z}$ 是随机字符串$(Z_1,\cdots,Z_n)$，可以得到

$$\sum_{1\leqslant i\leqslant n}I(X;Z_i)\leqslant I(X,Y)+I(\mathbf{Z}),\quad 其中\ I(\mathbf{Z}):=\sum_{1\leqslant i\leqslant n}h(Z_i)-h(\mathbf{Z})$$

解答 由 $X\to Y\to \mathbf{Z}$ 的 Markov 性质可以得到如下结果

$$I(X;\mathbf{Z})\leqslant I(X;Y)$$

因此，能够证明得到

$$\sum_{1\leqslant i\leqslant n}I(X;Z_i)-I(\mathbf{Z})\leqslant I(X;\mathbf{Z}) \tag{1.2.41}$$

如下所示，界(1.2.41)对任意 X 和 $\mathbf{Z}$(没有涉及 Markov 性质)都适用。事实上，式(1.2.41)等价于下面的式子

$$nh(X)-\sum_{1\leqslant i\leqslant n}h(X,Z_i)+h(\mathbf{Z})\leqslant h(X)+h(\mathbf{Z})-h(X,\mathbf{Z})$$

或

$$h(X,\mathbf{Z})-h(X)\leqslant\sum_{1\leqslant i\leqslant n}h(X,Z_i)-nh(X)$$

从而反过来可以得出不等式 $h(\mathbf{Z}|X)\leqslant\sum_{i\in[n]}h(Z_i|X)$。 □

举例 1.2.22 对于概率向量 $p=\begin{bmatrix}p_1\\ \vdots\\ p_n\end{bmatrix}$，可以有等式 $h(p)=-\sum_{i}^{n}p_i\log p_i$，其中每项有 $p_i\geqslant 0$ 且 $p_1+\cdots+p_n=1$。

(a) 如果 $P=(P_{ij})$ 是一个二重随机矩阵(即方阵中任一元素 $P_{ij}\geqslant 0$ 且每行和每列的和都为1)，试说明 $h(Pp)\geqslant h(p)$。此外，当且仅当 P 为置换矩阵时，试说明 $h(Pp)=h(p)$。

(b) 如果 P 是一个随机矩阵且 p 是满足 $P:Pp=p$ 的一个不变向量，试说明 $h(p)\geqslant -\sum_{j=1}^{n}\sum_{k=1}^{n}p_jP_{jk}\log P_{jk}$。

解答 (a) 对于所有的 λ_i，$c_i\geqslant 0$ 且 $\sum_{i}^{n}\lambda_i=1$，通过利用 log 函数 $x\mapsto\log x$ 的凹性，可以得到 $\log(\lambda_1c_1+\cdots+\lambda_nc_n)\geqslant\sum_{i}^{n}\lambda_i\log c_i$。应用到 $h(Pp)=-\sum_{ij}P_{ij}p_j\log\left(\sum_{k}P_{ik}p_k\right)\geqslant-\sum_{j}p_j\log\left(\sum_{i,k}P_{ik}P_{ik}p_k\right)=-\sum_{j}p_j\log((P^{\mathrm{T}}Pp)_j)$。利用 Gibbs 不等式有 RHS$\geqslant h(p)$。当且仅当 $P^{\mathrm{T}}Pp=p$ 时，也即 $P^{\mathrm{T}}P=I$ 为单位矩阵时，等式成立。当且仅当 P 为置换矩阵时这种情况发生。

31

(b) 当平稳 Markov 信源 $(U_1,U_2,\cdots)$ 有均匀分布 p 时，这时 LHS 等于 $h(U_n)$，然而等式右边就是 $h(U_n|U_{n-1})$。可以从一般不等式 $h(U_n|U_{n-1})\leqslant h(U_n)$ 得到结果。 □

举例 1.2.23 随机变量序列 $\{X_j:j=1,2,\cdots\}$ 形成了一个具有有限状态空间的 DTMC。

(a) 通过引用标准条件熵的属性，证明 $h(X_j|X_{j-1})\leqslant h(X_j|X_{j-2})$，且在平稳 DTMC 的情况下，证明有 $h(X_j|X_{j-2})\leqslant 2h(X_j|X_{j-1})$。

(b) 证明当 $1\leqslant m\leqslant n$ 时，互信息 $I(X_m;X_n)$ 不随 m 递减，不随 n 递增。

解答 (a) 利用 Markov 性质和平稳性，有

$$\begin{aligned}h(X_j|X_{j-2})&=h(X_j|X_{j-1},X_{j-2})\leqslant h(X_j|X_{j-1})\leqslant h(X_j,X_{j-1}|X_{j-2})\\&=h(X_j|X_{j-1},X_{j-2})+h(X_{j-1}|X_{j-2})=2h(X_j|X_{j-1})\end{aligned}$$

(b) 可以得出

$$\begin{aligned}I(X_m;X_n)-I(X_m;X_{n+1})&=h(X_n|X_{n+1})-h(X_n|X_m)\\&=h(X_n|X_{n+1})-h(X_n|X_m,X_{n+1})\end{aligned}$$

(在给定 X_n 的情况下，因为 X_m 和 X_{n+1} 是条件独立的)

上式是 $\geqslant 0$ 的，因此 $I(X_m;X_n)$ 是不随 n 增加的。

同样，

$$I(X_{m-1};X_n)-I(X_m;X_n)=h(X_m|X_{m-1})-h(X_m|X_n,X_{m-1})\geqslant 0$$

因此，$I(X_m;X_n)$ 不随 m 递减。

在这里没有使用平稳性假定。DTMC 甚至可能不是时间齐次的(也即转移概率可能不仅仅依赖 i，j，同样依赖于过渡的时间)。 □

举例 1.2.24 给定随机变量 Y_1，Y_2，Y_3，定义

$$I(Y_1;Y_2|Y_3)=h(Y_1|Y_3)+h(Y_2|Y_3)-h(Y_1,Y_2|Y_3)$$

现在令序列 X_n，$n=0,1,\cdots$ 是 DTMC，证明如下式子

$$I(X_{n-1};X_{n+1}|X_n)=0,\quad 因此\ I(X_{n-1};X_{n+1})\leqslant I(X_n;X_{n+1})$$

同时证明对于 $m=0, 1, 2, \cdots$，$I(X_n; X_{n+m})$ 是不随 m 增加的。 35

解答　根据 Markov 性质，在给定 X_n 时，X_{n-1} 和 X_{n+1} 满足条件独立。因此可得，

$$h(X_{n-1}, X_{n+1} \mid X_n) = h(X_{n+1} \mid X_n) + h(X_{n-1} \mid X_n)$$

且 $I(X_{n-1}; X_{n+1} \mid X_n) = 0$。同时，

$$\begin{aligned}
&I(X_n; X_{n+m}) - I(X_n; X_{n-m+1}) \\
&= h(X_{n+m}) - h(X_{n-m+1}) - h(X_n, X_{n-m+1}) + h(X_n, X_{n+m}) \\
&= h(X_n \mid X_{n+m-1}) - h(X_n \mid X_{n-m}) \\
&= h(X_n \mid X_{n+m-1}) - h(X_n \mid X_{n-m}, X_{n+m+1}) \geqslant 0
\end{aligned}$$

因为条件独立性能得到最后的等式，根据 1.2.21 得到最后一个不等式。 □

举例 1.2.25　(熵的公理化定义)(a) 考虑概率分布 $(p_1, \cdots, p_m)$ 和不确定性(熵)的相应测定 h

$$h(p_1q_1, p_1q_2, \cdots, p_1q_n, p_2, p_3, \cdots, p_m) = h(p_1, \cdots, p_m) - p_1 h(q_1, \cdots, q_n) \quad (1.2.42)$$

如果 $(q_1, \cdots, q_n)$ 是另外一个分布，也就是说，如果一个不确定事件(概率为 p_1)被分割为条件概率为 $(q_1, \cdots, q_n)$ 的子事件，根据上式知总的不确定性被附加地分割了。假定函数 h 在它的参数上是对称的，因此当随机事件 2，3，…，m 被分割时也可得到类似的关系。

假定函数 $F(m) := h(1/m, \cdots, 1/m)$ 关于 m 是单调递增的。证明，作为式(1.2.42)的结论，可得 $F(m^k) = kF(m)$，同时对一些常数 c，有 $F(m) = c\log m$。因此证明有

$$h(p_1, \cdots, p_m) = -c\sum_j p_j \log p_j \quad (1.2.43)$$

假定 p_j 是有理数。在满足连续性假定的条件下，式(1.2.43)对于任意集合 $\{p_j\}$ 都是正确的。

(b) 熵的另一个公理化描述如下所示。对于任意 $k<m$，如果一个对称函数 h 服从

$$\begin{aligned}
h(p_1, \cdots, p_m) &= h(p_1 + \cdots + p_k, p_{k+1}, \cdots, p_m) \\
&\quad + (p_1 + \cdots + p_k) h\left(\frac{p_1}{p_1 + \cdots + p_k}, \cdots, \frac{p_k}{p_1 + \cdots + p_k}\right) \quad (1.2.44)
\end{aligned}$$ 36

$h(1/2, 1/2) = 1$ 且当 $p \in [0, 1]$ 时，$h(p, 1-p)$ 是一个连续函数，那么有

$$h(p_1, \cdots, p_m) = -\sum_j p_j \log p_j$$

解答　(a) 由式(1.2.42)可知，对于函数 $F(m) = h(1/m, \cdots, 1/m)$，我们得到如下的恒等式：

$$\begin{aligned}
F(m^2) &= h\left(\frac{1}{m} \times \frac{1}{m}, \cdots, \frac{1}{m} \times \frac{1}{m}, \frac{1}{m^2}, \cdots, \frac{1}{m^2}\right) \\
&= h\left(\frac{1}{m}, \frac{1}{m^2}, \cdots, \frac{1}{m^2}\right) + \frac{1}{m} F(m) \\
&\quad \vdots \\
&= h\left(\frac{1}{m}, \cdots, \frac{1}{m}\right) + \frac{m}{m} F(m) = 2F(m)
\end{aligned}$$

归纳假设 $F(m^{k-1}) = (k-1)F(m)$。则有

$$\begin{aligned}
(m^k) &= h\left(\frac{1}{m} \times \frac{1}{m^{k-1}}, \cdots, \frac{1}{m} \times \frac{1}{m^{k-1}}, \frac{1}{m^2}, \cdots, \frac{1}{m^k}\right) \\
&= h\left(\frac{1}{m^{k-1}}, \frac{1}{m^k}, \cdots, \frac{1}{m^k}\right) + \frac{1}{m} F(m) \\
&\quad \vdots \\
&= h\left(\frac{1}{m^{k-1}}, \cdots, \frac{1}{m^{k-1}}\right) + \frac{m}{m} F(m) = (k-1)F(m) + F(m) = kF(m)
\end{aligned}$$

现在，对于给定的正实数 $b>2$，m，可以找出一个正整数 n 满足 $2^n \leqslant b^m \leqslant 2^{n+1}$，也就是

$$\frac{n}{m} \leqslant \log_2 b \leqslant \frac{n}{m}+\frac{1}{m}$$

利用 $F(m)$ 的单调性，可以得到 $nF(2) \leqslant mF(b) \leqslant (n+1)F(2)$，或者

$$\frac{n}{m} \leqslant \frac{F(b)}{F(2)} \leqslant \frac{n}{m}+\frac{1}{m}$$

[37] 可以推出如下式子 $\left|\log_2 b-\frac{F(b)}{F(2)}\right| \leqslant \frac{1}{m}$，然后令 $m\to\infty$，$F(b)=c\log b$ 且 $c=F(2)$。

现在取有理数 $p_1=\frac{r_1}{r}$，…，$p_n=\frac{r_n}{r}$，可以得到如下式子

$$\begin{aligned}
h\left(\frac{r_1}{r},\cdots,\frac{r_n}{r}\right) &= h\left(\frac{r_1}{r}\times\frac{1}{r_1},\cdots,\frac{r_1}{r}\times\frac{1}{r_1},\frac{r_2}{r},\cdots,\frac{r_n}{r}\right)-\frac{r_1}{r}F(r_1)\\
&\vdots\\
&= h\left(\frac{1}{r},\cdots,\frac{1}{r}\right)-c\sum_{1\leqslant i\leqslant n}\frac{r_i}{r}\log r_i\\
&= c\log r-c\sum_{1\leqslant i\leqslant n}\frac{r_i}{r}\log r_i=-c\sum_{1\leqslant i\leqslant n}\frac{r_i}{r}\log\frac{r_i}{r}
\end{aligned}$$

(b) 对于第二个定义，关键点是不需要假定函数 $F(m)=h(1/m,\cdots,1/m)$ 在参数 m 上的单调性。同样，通过利用式(1.2.44)，可以很容易得出可加性

$$F(mn)=F(m)+F(n)$$

上式对任意正整数 m，n 都成立。因此，对于一个规范素数分解 $m=q_1^{a_1}\cdots q_l^{a_l}$，我们可得

$$F(m)=a_1F(q_1)+\cdots+a_lF(q_l)$$

接下来，我们证明

$$\frac{F(m)}{m}\to 0,\ F(m)-F(m-1)\to 0 \tag{1.2.45}$$

其中 $m\to\infty$。事实上

$$F(m)=h\left(\frac{1}{m},\cdots,\frac{1}{m}\right)=h\left(\frac{1}{m},\frac{m-1}{m}\right)-\frac{m-1}{m}h\left(\frac{1}{m-1},\cdots,\frac{1}{m-1}\right)$$

即

$$h\left(\frac{1}{m},\frac{m-1}{m}\right)=F(m)-\frac{m-1}{m}F(m-1)$$

利用 $h(p,1-p)$ 的连续性和对称性可得

$$\lim_{m\to\infty}h\left(\frac{1}{m},\frac{m-1}{m}\right)=h(0,1)=h(1,0)$$

但是从

[38]

$$h\left(\frac{1}{2},\frac{1}{2},0\right)=h\left(\frac{1}{2},\frac{1}{2}\right)+\frac{1}{2}h(1,0)$$

和(又一次利用对称性)

$$h\left(\frac{1}{2},\frac{1}{2},0\right)=h\left(0,\frac{1}{2},\frac{1}{2}\right)=h(1,0)+h\left(\frac{1}{2},\frac{1}{2}\right)$$

我们得到 $h(1,0)=0$。因此

$$\lim_{m\to\infty}\left(F(m)-\frac{m-1}{m}F(m-1)\right)=0 \tag{1.2.46}$$

下一步，可以写出下式

$$mF(m)-\sum_{k=1}^{m}k\left(F(k)-\frac{k-1}{k}F(k-1)\right)$$

或者，可以恒等变形如下式

$$\frac{F(m)}{m}=\frac{m+1}{2m}\left[\frac{2}{m(m+1)}\sum_{k=1}^{m}k\left(F(k)-\frac{k-1}{k}F(k-1)\right)\right]$$

其中方括号中的式子是序列的 $m(m+1)/2$ 项的算术平均值

$$F(1),F(2)-F(1),F(2)-F(1),F(3)-\frac{2}{3}F(2),F(3)-\frac{2}{3}F(2),$$
$$F(3)-\frac{2}{3}F(2),\cdots,F(k)-\frac{k-1}{k}F(k-1),\cdots,$$
$$F(k)-\frac{k-1}{k}F(k-1),\cdots$$

上述序列趋于 0。因此，其最终值为 0 且 $F(m)/m\to 0$。与此同时，

$$F(m)-F(m-1)=\left(F(m)-\frac{m-1}{m}F(m-1)\right)-\frac{1}{m}F(m-1)\to 0$$

式(1.2.46)成立。现在定义下式

$$c(m)=\frac{F(m)}{\log m}$$

并证明 $c(m)$ 是常量。我们只需证明对于任意素数 p，$c(p)$ 是常量成立即可。首先，我们来证明序列 $\{c(p)\}$ 是有界的。事实上，根据上述推导假设序号 $c(p)$ 是无界的。然后，可以找出一个无穷的素数序列 p_1，p_2，…，p_n，…，其中 p_n 是最小的素数，即 $p_n>p_{n-1}$ 且 $c(p_n)>c(p_{n-1})$。通过构造这样的序列，如果一个素数 $q<p_n$，就可以得到 $c(q)<c(p_n)$。 [39]

考虑到把数 $p_n-1=q_1^{\alpha_1}\cdots q_k^{\alpha_k}$，其中 $q_1=2$，通过标准分解为质因子。我们将差值 $F(p_n)-F(p_n-1)$ 写成如下式子

$$F(p_n)-\frac{F(p_n)}{\log p_n}\log(p_n-1)=c(p_n)\log(p_n-1)-F(p_n-1)$$
$$=\frac{F(p_n)}{p_n}\frac{p_n}{\log p_n}\log\frac{p_n}{p_n-1}+\sum_{j=1}^{k}\alpha_j(c(p_n)-c(q_j))\log q_j$$

通过上面的推导得到

$$\sum_{j=1}^{k}\alpha_j(c(p_n)-c(q_j))\log q_j\geqslant(c(p_n)-c(2))\log 2=(c(p_n)-c(2))\quad(1.2.47)$$

而且，随着 $\lim_{p\to\infty}\frac{p}{\log p}\log\frac{p}{p-1}=0$，通过式(1.2.46)和式(1.2.47)可得 $c(p_n)-c(2)\leqslant 0$，这和 $c(p)$ 的构建相矛盾。因此，可知 $c(p)$ 有上界。同样地，我们也可得到 $c(p)$ 有下界，而且，通过上述证明，可以得到 $\sup_p c(p)$，$\inf_p c(p)$。

现在假设 $c(\hat{p})=\sup_p c(p)>c(2)$，给定一个正整数 m，分解为质因子 $\hat{p}^m-1=q_1^{\alpha_1}\cdots q_k^{\alpha_k}$，其中 $q_1=2$。像上述证明一样，写出差值 $F(\hat{p}^m)-F(\hat{p}^m-1)$ 为

$$F(\hat{p}^m)-\frac{F(\hat{p}^m)}{\log\hat{p}^m}\log(\hat{p}^m-1)+c(\hat{p})\log(\hat{p}^m-1)-F(\hat{p}^m-1)$$
$$=\frac{F(\hat{p}^m)}{\hat{p}^m}\frac{\hat{p}^m}{\log\hat{p}^m}\log\frac{\hat{p}^m}{\hat{p}^m-1}+\sum_{j=1}^{k}\alpha_j(c(\hat{p})-c(q_j))\log q_j$$
$$\geqslant\frac{c(\hat{p}^m)}{\hat{p}^m}\frac{\hat{p}^m}{\log\hat{p}^m}\log\frac{\hat{p}^m}{\hat{p}^m-1}+(c(\hat{p})-c(2))$$

同时取极限 $m\to\infty$ 会得到 $c(\hat{p})-c(2)\leqslant 0$，这会得出一个矛盾的结论。同样，我们可以证明 $\inf_p c(p)=c(2)$。因此，$c(p)=c$ 是个常数，且 $F(m)=c\log m$。根据条件 $F(2)=h\left(\frac{1}{2},\frac{1}{2}\right)=1$，我们得到 $c=1$。最后，就像在(a)中表述的，我们可以得到

$$h(p_1,\cdots,p_m)=-\sum_{i=1}^{m}p_i\log p_i \tag{1.2.48}$$

对任意有理数 $p_1,\cdots,p_m\geqslant 0$ 且 $\sum_{i=1}^{m}p_i=1$ 都成立。利用连续性，式(1.2.48)可以推广到无理数概率的情景。 □

举例 1.2.26 证明更均匀的分布可以得到更大的熵。这就是说，在集合$\{1,\cdots,n\}$上有 $\boldsymbol{p}=(p_1,\cdots,p_n)$和 $\boldsymbol{q}=(q_1,\cdots,q_n)$两个概率分布，如果按递减顺序重新排列值 $p_1,\cdots,p_n$ 和 $q_1,\cdots,q_n$，

40

$$p_1\geqslant\cdots\geqslant p_n,\quad q_1\geqslant\cdots\geqslant q_n$$

有

$$\sum_{i=1}^{k}p_i\leqslant\sum_{i=1}^{k}q_i,\quad k=1,\cdots,n$$

则称 $\boldsymbol{p}$ 比 $\boldsymbol{q}$ 分布更均匀($\boldsymbol{p}\preceq\boldsymbol{q}$，参见文献[108])。

可以得到

$$h(\boldsymbol{p})\geqslant h(\boldsymbol{q}),\quad 当\ \boldsymbol{p}\preceq\boldsymbol{q}\ 时$$

解答 我们将概率分布 $\boldsymbol{p}$ 和 $\boldsymbol{q}$ 写成离散变量的非递增函数，

$$\boldsymbol{p}\sim p^{(1)}\geqslant\cdots\geqslant p^{(n)}\geqslant 0,\boldsymbol{q}\sim q^{(1)}\geqslant\cdots\geqslant q^{(n)}\geqslant 0,\quad 其中\sum_i p^{(i)}=\sum_i q^{(i)}=1$$

假如 $\boldsymbol{p}\neq\boldsymbol{q}$，就存在着 i_1，i_2 满足(a)$1\leqslant i_1\leqslant i_2\leqslant n$；(b)$q^{(i_1)}>p^{(i_1)}\geqslant p^{(i_2)}>q^{(i_2)}$；(c)$q^{(i)}\geqslant p^{(i)}$，其中 $1\leqslant i\leqslant i_1$，$q^{(i)}\leqslant p^{(i)}$，其中 $i\geqslant i_2$，则有 $\boldsymbol{p}\preceq\boldsymbol{q}$ 成立。

现在在 s 中应用归纳法，其中 s 是满足 $q^{(i)}\neq p^{(i)}$ 值 i 的数目，其中 $i=1,\cdots,n$。如果 $s=0$，可得 $\boldsymbol{p}=\boldsymbol{q}$ 且熵一致。做出归纳假设且 s 的值每次增加 1。如上取一对数 i_1，i_2，增加 $q^{(i_1)}$，减少 $q^{(i_2)}$，因此 $q^{(i_1)}+q^{(i_2)}$ 的和保持不变，一直到 $q^{(i_1)}$ 达到 $p^{(i_1)}$ 或者 $q^{(i_2)}$ 达到 $p^{(i_2)}$(见图 1-7)。性质(c)保证修改后的分布 $\boldsymbol{p}\preceq\boldsymbol{q}$。因为函数 $x\to\eta(x)=-x\log x-(1-x)\log(1-x)$在$[0,1/2]$上严格递增。因此，修改后的分布的熵也是严格递增的。在这个过程的最后减小 s。然后我们使用归纳假设方法即可证明。

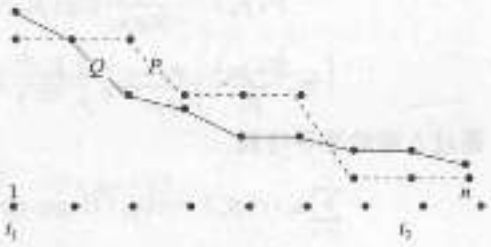


图 1-7

1.3 Shannon 第一编码定理，Markov 信源的熵率

信源的信息速率的一个有用的意义是，它表明样本字符串传输的最小增长速率渐近地接近全概率。

引理 1.3.1 令 H 为一个信源(见式(1.1.20))的信息速率。定义

41

$$D_n(R)=-\max[\mathbb{P}(U^{(n)}\in A):A\subset I^n,\#A\leqslant 2^{nR}] \tag{1.3.1}$$

随着 $n\to\infty$，对于任意 $\varepsilon>0$，有

$$\lim D_n(H+\varepsilon)=1,\quad 且若\ H>0,D_n(H-\varepsilon)\not\to 1 \tag{1.3.2}$$

证明 通过定义可知，$R_\varepsilon := H+\varepsilon$ 是一个可靠的编码率。因此，存在一个集合 $A_n \subset I^{\times n}$ 的序列，随着 $n\to\infty$，有 $\#A_n \leqslant 2^{nR}$，$\mathbb{P}(U^{(n)} \in A_n) \to 1$。因为 $D_n(R) \geqslant \mathbb{P}(U^{(n)} \in A_n)$，故 $D_n(R) \to 1$。

现在假设 $H>0$，令 $R_\varepsilon := H-\varepsilon$；其中当 ε 足够小时，有 $R>0$。然而，R 不是一个可靠速率。也就是说不存在一个拥有上述性质的序列 A_n。取一个集合 C_n，使得式(1.3.1)取得最大值。然后有 $\#C_n \leqslant 2^{nR}$，但是 $\mathbb{P}(C_n) \not\to 1$。 □

给定一个字符串 $u^{(n)} = u_1 \cdots u_n$，考虑到它的每个信源字符的对数似然函数：

$$\xi_n(u^{(n)}) := -\frac{1}{n}\log_+ p_n(u^{(n)}), u^{(n)} \in I^{\times n} \tag{1.3.3a}$$

其中 $p_n(u^{(n)}) := \mathbb{P}(U^{(n)} = u^{(n)})$ 是分配给字符串 $u^{(n)}$ 的概率。此处及以下满足，如果 $x>0$，则 $\log_+ x = \log x$，如果 $x=0$，则它的值为 0。对于一个随机字符串，$U^{(n)} = u_1, \cdots, u_n$，有

$$\xi_n(U^{(n)}) = -\frac{1}{n}\log_+ p_n(U^{(n)}) \tag{1.3.3b}$$

是一个随机变量。

引理 1.3.2 对于所有的 R，$\varepsilon>0$

$$\mathbb{P}(\xi_n \leqslant R) \leqslant D_n(R) \leqslant \mathbb{P}(\xi_n \leqslant R+\varepsilon) + 2^{-n\varepsilon} \tag{1.3.4}$$ 42

证明 为了简便，在符号 $u^{(n)}$ 和 $U^{(n)}$ 中省略上标(n)。令

$$\begin{aligned} B_n &:= \{u \in I^{\times n} : p_n(u) \geqslant 2^{-nR}\} \\ &= \{u \in I^{\times n} : -\log p_n(u) \leqslant nR\} \\ &= \{u \in I^{\times n} : \xi_n(u) \leqslant R\} \end{aligned}$$

那么

$$1 \geqslant \mathbb{P}(U \in B_n) = \sum_{u \in B_n} p_n(u) \geqslant 2^{-nR} \# B_n, \quad 由于\ \# B_n \leqslant 2^{nR}$$

因此，

$$D_n(R) = \max[\mathbb{P}(U \in A_n) : A_n \subset I^{\times n}, \# A \leqslant 2^{nR}] \geqslant \mathbb{P}(U \in B_n) = \mathbb{P}(\xi_n \leqslant R)$$

式(1.3.4)的 LHS 得证。

另一方面，存在一个集合 $C_n \subseteq I^{\times n}$ 使得式(1.3.1)取到最大值。对于这样一个集合，$D_n R = \mathbb{P}(U \in C_n)$ 可以分解为如下式：

$$\begin{aligned} D_n(R) &= \mathbb{P}(U \in C_n, \xi_n \leqslant R+\varepsilon) + \mathbb{P}(U \in C_n, \xi_n > R+\varepsilon) \\ &\leqslant \mathbb{P}(\xi_n \leqslant R+\varepsilon) + \sum_{u \in C_n} p_n(u) 1(p_n(u) < 2^{-n(R-\varepsilon)}) \\ &< \mathbb{P}(\xi_n \leqslant R+\varepsilon) + 2^{-n(R-\varepsilon)} \# C_n \\ &= \mathbb{P}(\xi_n \leqslant R+\varepsilon) + 2^{-n(R-\varepsilon)} 2^{nR} \\ &= \mathbb{P}(\xi_n \leqslant R+\varepsilon) + 2^{-n\varepsilon} \end{aligned}$$

□

定义 1.3.3 (参见 PSE II，367 页)对于所有的 $\varepsilon>0$，如果

$$\lim_{n\to\infty} \mathbb{P}(|\eta_n - r| \geqslant \varepsilon) = 0 \tag{1.3.5}$$

则随机变量 $\{\eta_n\}$ 序列依概率收敛于常数 r。

在这个定义中，令一个随机变量 η 替换 r，我们获得了一个更加一般的依概率收敛于某个随机变量的定义。

依概率收敛以后表示为 $\eta_n \xrightarrow{\mathbb{P}} r$(另外，$\eta_n \xrightarrow{\mathbb{P}} \eta$)

备注 1.3.4 依概率收敛(到一个期望值)恰恰在大数定理中占据重要位置(参见下面的式(1.3.8))。见 PSE I，78 页。

定理 1.3.5 (Shannon 第一编码定理(FCT)) 如果 ξ_n 依概率收敛于常数 γ，那么 $\gamma=H$，H 为信源的信息速率。

43

证明 令 $\xi_n \xrightarrow{P} \gamma$，由于 $\xi_n \geqslant 0$，$\gamma \geqslant 0$，根据引理 1.3.2，对任意的 $\varepsilon>0$，有

$$D_n(\gamma+\varepsilon) \geqslant \mathbb{P}(\xi_n \leqslant \gamma+\varepsilon) \geqslant \mathbb{P}(\gamma-\varepsilon \leqslant \xi_n \leqslant \gamma+\varepsilon)$$
$$= \mathbb{P}(|\xi_n-\gamma| \leqslant \varepsilon) = 1-\mathbb{P}(|\xi_n-\gamma|>\varepsilon) \to 1 (n \to \infty)$$

因此，$H \leqslant \gamma$。特别地，如果 $\gamma=0$，那么 $H=0$。如果 $\gamma>0$，根据引理 1.3.2，我们再次得到

$$D_n(\gamma-\varepsilon) \leqslant \mathbb{P}(\xi_n \leqslant \gamma-\varepsilon/2)+2^{-n\varepsilon/2} \leqslant \mathbb{P}(|\xi_n-\gamma| \geqslant \varepsilon/2)+2^{-n\varepsilon/2} \to 0$$

根据引理 1.3.1，$H \geqslant \gamma$。因此，$H=\gamma$。 □

备注 1.3.6 (a) $\xi_n \xrightarrow{P} \gamma=H$ 的收敛性等价于下述的渐近均分性：对任意的 $\varepsilon>0$，

$$\lim_{n\to\infty} \mathbb{P}(2^{-n(H+\varepsilon)} \leqslant p_n(U^{(n)}) \leqslant 2^{-n(H-\varepsilon)}) = 1 \tag{1.3.6}$$

事实上，

$$\mathbb{P}(2^{-n(H+\varepsilon)} \leqslant p_n(U^{(n)}) \leqslant 2^{-n(H-\varepsilon)}) = \mathbb{P}\left(H-\varepsilon \leqslant -\frac{1}{n}\log p_n(U^{(n)}) \leqslant H+\varepsilon\right)$$
$$= \mathbb{P}(|\xi_n-H| \leqslant \varepsilon) = 1-\mathbb{P}(|\xi_n-H|>\varepsilon)$$

换言之，对任意的 $\varepsilon>0$，总存在 $n_0=n_0(\varepsilon)$，使得对任意的 $n>n_0$，集合 $I^{(n)}$ 可分解为不相交的子集 Π_n 和 T_n，满足

(i) $\mathbb{P}(U^{(n)} \in \Pi_n)<\varepsilon$。

(ii) 对任意的 $u^{(n)} \in T_n$，$2^{-n(H+\varepsilon)} \leqslant \mathbb{P}(U^{(n)}=u^{(n)}) \leqslant 2^{-n(H-\varepsilon)}$ 成立。

具体而言，T_n 是"典型"的字符串，而 Π_n 是剩余集合。可以认为，对于具有渐近均分性的信源，用等长的码字来编码典型字符串是值得的，而剩余部分则没必要考虑。于是，可以得出有效编码速率为 $H+o(1)$ 比特/信源符号，而信源速率为 $\log m$ 比特/信源符号。

(b) 注意到

$$\mathbb{E}\xi_n = -\frac{1}{n}\sum_{u^{(n)} \in I^{(n)}} p_n(u^{(n)})\log p_n(u^{(n)}) = \frac{1}{n}h^{(n)} \tag{1.3.7}$$

最简单(且最有启发性)的信源例子为 Bernoulli 信源。

定理 1.3.7 对于一个 Bernoulli 信源 U_1，U_2，…，$\mathbb{P}(U_i=x)=P(x)$，有

$$H = -\sum_x p(x)\log p(x)$$

44

证明 对于一个 IID 序列 U_1，U_2，…，其字符串的概率为

$$p_n(u^{(n)}) = \prod_{i=1}^{n} p(u_i), u^{(n)} = u_1 \cdots u_n$$

因此，$-\log p_n(u) = \sum_i -\log p(u_i)$，设 $\sigma_i = -\log p(U_i)$，$i=1, 2, \cdots$，则 σ_1，σ_2，…形成了一个 IID 随机变量的序列。对于一个随机字符串 $U^{(n)}=U_1 \ldots U_n$，$-\log p_n(U^{(n)}) = \sum_{i=1}^{n} \sigma_i$，其中随机变量 $\sigma_i = -\log p(U_i)$ 满足 IID。

接下来，设 $\xi_n = \frac{1}{n}\sum_{i=1}^{n} \sigma_i$。注意到 $\mathbb{E}\sigma_i = -\sum_j p(j)\log p(j) = h$ 且

$$\mathbb{E}\xi_n = \mathbb{E}\left(\frac{1}{n}\sum_{i=1}^{n}\sigma_i\right) = \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}\sigma_i = \frac{1}{n}\sum_{i=1}^{n} h = h$$

最后的等式和式(1.3.7)一致，即对于 Bernoulli 信源，$h^{(n)}=nh$(参见式(1.1.18))。因此，$\mathbb{E}\xi_n=h$。根据大数定理，我们立即得出 $\xi_n\xrightarrow{P}h$。于是根据定理 1.3.5(FCT)有 $H=h$。□

定理 1.3.8 (IID 随机变量的大数定理)对任意方差有限且均值为 $\mathbb{E}\eta_i=r$ 的 IID 随机变量 $\eta_1,\eta_2,\cdots$组成的序列，对任意的 $\varepsilon>0$，有

$$\lim_{n\to\infty}\mathbb{P}\left(\left|\frac{1}{n}\sum_{i=1}^{n}\eta_i-r\right|\geqslant\varepsilon\right)=0 \tag{1.3.8}$$

证明　定理 1.3.8 的证明基于著名的 Chebyshev 不等式，参见 PSE II，368 页。□

引理 1.3.9　对任意随机变量 η 和任意的 $\varepsilon>0$，有

$$\mathbb{P}(\eta\geqslant\varepsilon)\leqslant\frac{1}{\varepsilon^2}\mathbb{E}\eta^2$$

证明　参见 PSE I，75 页。□

接下来，考虑 Markov 信源 $U_1,U_2,\cdots$，其中码字符号取自符号集 $I_m=\{1,\cdots,m\}$。假定转移矩阵 $P(u,v)$(或其功率)服从

$$\min_{u,v}P^{(r)}(u,v)=\rho>0,\quad 对某些\ r\geqslant 1 \tag{1.3.9}$$

45

这个条件说明 DTMC 是不可约和非周期的。于是(参见 PSE II，71 页)DTMC 有一个唯一的不变(平衡)分布 $\pi(1),\cdots,\pi(m)$：

$$0\leqslant\pi(u)\leqslant 1,\sum_{u=1}^{m}\pi(u)=1,\pi(v)=\sum_{u=1}^{m}\pi(u)P(u,v) \tag{1.3.10}$$

同时 n 步转移概率 $P^{(n)}(u,v)$收敛于 $\pi(v)$，且对所有初始分布$(\lambda(u),u\in I)$，$(\lambda P^{n-1})(v)=\mathbb{P}(U_n=v)$：

$$\lim_{n\to\infty}P^{(n)}(u,v)=\lim_{n\to\infty}\mathbb{P}(U_n=v)=\lim_{n\to\infty}\sum_{u}\lambda(u)P^{(n)}(u,v)=\pi(v) \tag{1.3.11}$$

此外，式(1.3.11)的收敛速度呈指数(几何指数)增长。

定理 1.3.10　假设 $r=1$ 时，条件(1.3.9)仍满足，则 DTMC $U_1,U_2,\cdots$有一个唯一的不变分布(1.3.10)，且对任意的 $u,v\in I$ 和任意的 I 中的初始分布 λ，有

$$|P^{(n)}(u,v)-\pi(v)|\leqslant(1-\rho)^n\quad 和\quad |\mathbb{P}(U_n=v)-\pi(v)|\leqslant(1-\rho)^{n-1} \tag{1.3.12}$$

在 $r\geqslant 1$ 的情形下，我们将式(1.3.12)的右端的$(1-\rho)^n$ 和$(1-\rho)^{n-1}$替换为$(1-\rho)^{\lfloor n/r\rfloor}$和$(1-\rho)^{\lfloor(n-1)/r\rfloor}$。

证明　参见举例 1.3.13。□

现在，我们引入 Markov 信源的信息速率 H。

定理 1.3.11　对于一个 Markov 信源，在条件(1.3.9)下有，

$$H=-\sum_{1\leqslant u,v\leqslant m}\pi(u)P(u,v)\log P(u,v)=\lim_{n\to\infty}h(U_{n-1}|U_n) \tag{1.3.13}$$

如果信源是稳定的，则 $H=h(U_{n-1}|U_n)$。

证明　参见式(1.3.3b)，我们再次使用 Shannon FCT 来验证 $\xi_n\xrightarrow{P}H$，其中 H 由式(1.3.13)定义且 $\xi_n=-\frac{1}{n}\log p_n(U^{(n)})$，换言之，条件(1.3.9)说明了 Markov 信源的渐近均分性。

Markov 特性意味着对任意的字符串 $u^{(n)}=u_1\cdots u_n$，有

$$p_n(u^{(n)})=\lambda(u_1)P(u_1,u_2)\cdots P(u_{n-1},u_n) \tag{1.3.14a}$$

从而 $-\log p_n(u^{(n)})$ 可以写为求和的形式，

$$-\log\lambda(u_1)-\log P(u_1,u_2)-\cdots-\log P(u_{n-1},u_n) \tag{1.3.14b}$$

对于一个任意字符串 $U^{(n)}=U_1\cdots U_n$，随机变量 $-\log p_n(U^{(n)})$ 有类似的形式，

46

$$-\log\lambda(U_1)-\log P(U_1,U_2)-\cdots-\log P(U_{n-1},U_n) \tag{1.3.15}$$

在信源为 Bernoulli 信源的情况下，可以导出

$$\sigma_1(U_1):=-\log\lambda(U_1),\sigma_i(U_{i-1},U_i):=-\log P(U_{i-1},U_i),i\geqslant 2 \tag{1.3.16}$$

假设

$$\xi_n=\frac{1}{n}\left(\sigma_1+\sum_{i=1}^{n-1}\sigma_{i+1}\right) \tag{1.3.17}$$

则 σ 的期望值为

$$\mathbb{E}\sigma_1=-\sum_u\lambda(u)\log\lambda(u) \tag{1.3.18a}$$

并且，由于 $\mathbb{P}(U_i=v)=\lambda P^{i-1}(v)=\sum_u\lambda(u)P^{(i-1)}(u,v)$，可以得到

$$\begin{aligned}\mathbb{E}\sigma_{i+1}&=-\sum_{u,u'}\mathbb{P}(U_i=u,U_{i+1}=u')\log P(u,u')\\&=-\sum_{u,u'}(\lambda P^{i-1})(u)P(u,u')\log P(u,u'),i\geqslant 1\end{aligned} \tag{1.3.18b}$$

定理 1.3.10 说明 $\lim\limits_{i\to\infty}\mathbb{E}\sigma=H$，因此

$$\lim_{n\to\infty}\mathbb{E}\xi_n=\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}\sigma_i=H$$

且 $\xi_n\xrightarrow{\mathbb{P}}H$ 的收敛性符合大数定律，对于 (σ_i) 序列，有

$$\lim_{n\to\infty}\mathbb{P}\left(\left|\frac{1}{n}\sum_{i=1}^{n}\sigma_i-H\right|\geqslant\varepsilon\right)=0 \tag{1.3.19}$$

然而，这种情形并没有 Bernoulli 信源中的情形那么简单。有两个需要克服的难点：(i)只有当 $i\to\infty$ 时，$\mathbb{E}\sigma_i$ 等于 H；(ii)σ_1，σ_2，…不再相互独立。在更坏的情形下，它们甚至不能构成 DTMC 或是高阶的 Markov 链。(一个序列 U_1，U_2，…构成 k 阶的 DTMC，如果对于所有的 $n\geqslant 1$，有

$$\begin{aligned}&\mathbb{P}(U_{n+k+1}=u'|U_{n+k}=u_k,\cdots,U_{n+1}=u_1,\cdots)\\&=\mathbb{P}(U_{n+k+1}=u'|U_{n+k}=u_k,\cdots,U_{n+1}=u_1)\end{aligned}$$

显然，在一个 k 阶的 DTMC 中，向量 $\overline{U}_n=(U_n,U_{n+1},\cdots,U_{n+k-1})$，$n\geqslant 1$ 构成了一个普通的 DTMC。)某种意义上，序列 σ_1，σ_2，…的“记忆”无限长。然而，它呈指数衰减，其

47

准确的含义在举例 1.3.14 中给出。

总之，通过使用 Chebyshev 不等式，我们得到

$$\mathbb{P}\left(\left|\frac{1}{n}\sum_{i=1}^{n}\sigma_i-H\right|\geqslant\varepsilon\right)\leqslant\frac{1}{n^2\varepsilon^2}\mathbb{E}\left(\sum_{i=1}^{n}(\sigma_i-H)\right)^2 \tag{1.3.20}$$

引理 1.3.12 和定理 1.3.11 如下。 □

引理 1.3.12 *式(1.3.20)的右端的期望满足边界条件*

$$\mathbb{E}\left(\sum_{i=1}^{n}(\sigma_i-H)\right)^2\leqslant Cn \tag{1.3.21}$$

其中，$C>0$ 为一个不依赖于 n 的常量。

证明 参见举例 1.3.14。 □

根据式(1.3.21)，可以推出式(1.3.20)的右端不大于$\frac{C}{n\varepsilon^2}$且随着$n\to\infty$趋于0。

举例 1.3.13 证明如下边界条件(参见式(1.3.12))：

$$|P^{(n)}(u,v)-\pi(v)|\leqslant(1-\rho)^n \tag{1.3.22}$$

证明 (与PSE II，72页相比)首先，注意到式(1.3.12)揭示了定理1.3.10和式(1.3.10)的第二个边界。实际上，将$\pi(v)$看成是极限

$$\lim_{n\to\infty}P^{(n)}(u,v)=\lim_{n\to\infty}\sum_{\tilde{u}}P^{(n-1)}(u,\tilde{u})P(\tilde{u},v)=\sum_{\tilde{u}}\pi(\tilde{u})P(\tilde{u},v) \tag{1.3.23}$$

从而导出了式(1.3.10)。如果$\pi'(1)$，$\pi'(2)$，…$\pi'(m)$是另一个不变概率向量，即

$$0\leqslant\pi'(u)\leqslant1,\quad \sum_{u=1}^{m}\pi'(u)=1,\quad \pi'(v)=\sum_{u}\pi'(u)P(u,v)$$

则对于所有的$n\geqslant1$，有$\pi'(v)=\sum_{u}\pi'(u)P^{(n)}(u,v)$，当$n\to\infty$时其极限为

$$\pi'(v)=\sum_{u}\pi'(u)\lim_{n\to\infty}P^{(n)}(u,v)=\sum_{u}\pi'(u)\pi(v)=\pi(v)$$

即不变概率向量是唯一的。

为了证明式(1.3.22)，设

$$m_n(v)=\min_{u}P^{(n)}(u,v),\quad M_n(v)=\max_{u}P^{(n)}(u,v) \tag{1.3.24}$$

48

于是，

$$m_{n+1}(v)=\min_{u}P^{(n+1)}(u,v)=\min_{u}\sum_{\tilde{u}}P(u,\tilde{u})P^{(n)}(\tilde{u},v)$$

$$\geqslant\min_{u}P^{(n)}(u,v)\sum_{\tilde{u}}P(u,\tilde{u})=m_n(v)$$

类似地，

$$M_{n+1}(v)=\max_{u}P^{(n+1)}(u,v)=\max_{u}\sum_{\tilde{u}}P(u,\tilde{u})P^{(n)}(\tilde{u},v)$$

$$\leqslant\max_{u}P^{(n)}(u,v)\sum_{\tilde{u}}P(u,\tilde{u})=M_n(v)$$

由于$0\leqslant m_n(v)\leqslant M_n(v)\leqslant1$，$m_n(v)$和$M_n(v)$都存在极限

$$m(v)=\lim_{n\to\infty}m_n(v)\leqslant\lim_{n\to\infty}M_n(v)=M(v)$$

进一步将$M(v)-m(v)$的差写为极限的形式

$$\lim_{n\to\infty}(M_n(v)-m_n(v))=\lim_{n\to\infty}\max_{u,u'}(P^{(n)}(u,v)-P^{(n)}(u',v))$$

所以，若能证明

$$\max_{u,u',v}|P^{(n)}(u,v)-P^{(n)}(u',v)|\leqslant(1-\rho)^n \tag{1.3.25}$$

则对于任意v有$M(v)=m(v)$。此外，用$\pi(v)$来表示$M(v)=m(v)$的共有的值，可得

$$|P^{(n)}(u,v)-\pi(v)|\leqslant M_n(v)-m_n(v)\leqslant(1-\rho)^n$$

为了证明式(1.3.25)，考虑一个$I\times I$的DTMC，设(u_1,u_2)为状态，转移概率如下

$$\boldsymbol{P}((u_1,u_2),(v_1,v_2))=\begin{cases}P(u_1,v_1)P(u_2,v_2), & u_1\neq u_2\\ P(u,v), & u_1=u_2=u;v_1=v_2=v\\ 0, & u_1=u_2 \text{ and } v_1\neq v_2\end{cases} \tag{1.3.26}$$

容易验证，$\boldsymbol{P}((u_1,u_2),(v_1,v_2))$实际上是一个$m^2\times m^2$的转移概率矩阵。如果$u_1=u_2=u$，则

$$\sum_{v_1,v_2}\boldsymbol{P}((u_1,u_2),(v_1,v_2))=\sum_{v}P(u,v)=1$$

若 $u_1 \neq u_2$，则

49

$$\sum_{v_1,v_2} \boldsymbol{P}((u_1,u_2),(v_1,v_2)) = \sum_{v_1} P(u_1,v_1) \sum_{v_2} P(u_2,v_2) = 1$$

（不等式 $0 \leqslant \boldsymbol{P}((u_1, u_2), (v_1, v_2)) \leqslant 1$ 可从式(1.3.26)直接得到。）

这就是所谓的 $l \times l$ 的耦合 DTMC。用 (V_n, W_n) 来表示，其中 $n \geqslant 1$。注意到元素 V_n 和 W_n 都是转移概率为 $P(u, v)$ 的 DTMC。更准确地说，元素 V_n 和 W_n 独立地移动，直到它们第一次在随机时间 τ 时相遇，将该时间称之为耦合时间。经过 τ 时间后，V_n 和 W_n "合在一块"同步移动，其转移概率仍为 $P(u, v)$。

假设我们从状态 (u, u') 开始该耦合链，则

$$\begin{aligned}
&|P^{(n)}(u,v) - P^{(n)}(u',v)| \\
&= |\mathbb{P}(V_n = v \mid V_1 = u, W_1 = u') - \mathbb{P}(W_n = v \mid V_1 = u, W_1 = u')|
\end{aligned}$$

（由于 (V_n, W_n) 中的每个元素以相同的转移概率移动。）

$$\begin{aligned}
&= |\mathbb{P}(V_n = v, W_n \neq v \mid V_1 = u, W_1 = u') - \mathbb{P}(V_n \neq v, W_n = v \mid V_1 = u, W_1 = u')| \\
&\leqslant \mathbb{P}(V_n \neq W_n \mid V_1 = u, W_1 = u') = \mathbb{P}(\tau > n \mid V_1 = u, W_1 = u') \qquad (1.3.27)
\end{aligned}$$

现在，概率服从

$$\mathbb{P}(\tau = 1 \mid V_1 = u, W_1 = u') \geqslant \sum_v P(u,v)P(u',v) \geqslant \rho \sum_v P(u',v) = \rho$$

即互补概率满足

$$\mathbb{P}(\tau > 1 \mid V_1 = u, W_1 = u') \leqslant 1 - \rho$$

由于耦合链的强 Markov 性质，

$$\mathbb{P}(\tau > n \mid V_1 = u, W_1 = u') \leqslant (1-\rho)^n \qquad (1.3.28)$$

边界(1.3.28)和边界(1.3.27)共同说明了式(1.3.25)。 □

举例 1.3.14 在满足式(1.3.9)且 $r=1$ 的条件下，证明如下边界：

58

$$|\mathbb{E}[(\sigma_i - H)(\sigma_{i+k} - H)]| \leqslant (H + |\log\rho|)^2 (1-\rho)^{k-1} \qquad (1.3.29)$$

证明 简单起见，假设 $i>1$；$i=1$ 的情形仅需要少量改动。回顾随机变量 σ_i，$i>1$ 的定义，得到

$$\begin{aligned}
\mathbb{E}[(\sigma_i - H)(\sigma_{i+k} - H)] = &\sum_{u,u'} \sum_{v,v'} \mathbb{P}(U_i = u, U_{i+1} = u'; U_{i+k} = v, U_{i+k+1} = v') \\
&\times (-\log P(u,u') - H)(-\log P(v,v') - H) \qquad (1.3.30)
\end{aligned}$$

我们的目标是将该表达式和下述表达式相比较

$$\begin{aligned}
&\sum_{u,u'} \sum_{v,v'} (\lambda P^{i-1})(u) P(u,u') [-\log P(u,u') - H] \\
&\times \pi(v) P(v,v') [-\log P(v,v') - H]
\end{aligned} \qquad (1.3.31)$$

注意到，根据 H 在式(1.3.13)中的定义，求和 $\sum_{u,v}$ 收敛，于是式(1.3.31)收敛。

求和(1.3.30)和(1.3.31)的不同在于，概率

$$\mathbb{P}(U_i = u, U_{i+1} = u'; U_{i+k} = v, U_{i+k+1} = v') = (\lambda P^{i-1})(u) P(u,u') P^{(k-1)}(u',v) P(v,v')$$

和

$$(\lambda P^{i-1})(u) P(u,u') \pi(v) P(v,v')$$

并不一致。然而，这些概率的差的绝对值满足如下边界

$$|P^{(k-1)}(u',v) - \pi(v)| \leqslant (1-\rho)^{k-1}$$

由于 $-\log P(\cdot,\cdot) - H| \leqslant H - |\log\rho|$，我们得到式(1.3.29)。 □

定理 1.3.11 的证明是为了证明式(1.3.21)，平方展开，利用期望的可加性，可得

$$\mathbb{E}\Big(\sum_{i=1}^{n}(\sigma_i - H)\Big)^2 = \sum_{1\le i\le n}\mathbb{E}[(\sigma_i - H)^2] + 2\sum_{1\le i<j\le n}\mathbb{E}[(\sigma_i - H)(\sigma_j - H)] \quad (1.3.32)$$

式(1.3.32)中的第一个求和比较容易，该和包含了 n 个 $\mathbb{E}(\sigma_i - H)^2$，其中每个以常量为界(设 C' 为 $(H+|\log_2 p|)^2$)。因此，该求和最多为 $C'n$。第二个求和有困难，它包含了 $n(n-1)/2$ 个元素，我们将其限制范围如下： 51

$$\Big|\sum_{1\le i<j\le n}\mathbb{E}[(\sigma_i - H)(\sigma_j - H)]\Big| \le \sum_{i=1}^{n}\Big(\sum_{k=1}^{n}|\mathbb{E}[(\sigma_i - H)(\sigma_{i+k} - H)]|\Big) \quad (1.3.33)$$

通过使用式(1.3.29)便完成了证明。

接下来的定理说明了在概率渐近分析中(相对)熵的作用：参见 PSE I，82 页。

定理 1.3.15 设 $\zeta_1, \zeta_2, \cdots$ 为 IID 随机变量序列，其取值为 0 或 1 的概率分别为 $1-p$ 和 p，$0<p<1$。则对于任意正整数序列 k_n，随着 $n\to\infty$，$k_n\to\infty$ 且 $n-k_n\to\infty$，有

$$\mathbb{P}\Big(\sum_{i=1}^{n}\zeta_i = k_n\Big) \sim (2\pi np^*(1-p^*))^{-1/2}\exp(-nD(p\,\|\,p^*)) \quad (1.3.34)$$

其中，$\sim$ 表示随着 $n\to\infty$，左式和右式的比值趋于 1。$p^*(=p_n^*)$ 表示比率 $\frac{k_n}{n}$，$D(p\,\|\,p^*)$ 表示相对熵 $h(X\,\|\,Y)$，其中 X 服从分布 ζ_i(取值为 0 或 1 的概率分别为 $1-p$ 和 p)，Y 取相同的值且概率分别为 $1-p^*$ 和 p^*。

证明 使用 Stirling 公式(参考 PSE I，72 页)：

$$n! \sim \sqrt{2\pi n}n^n e^{-n} \quad (1.3.35)$$

(实际上，该公式有一个更准确的形式：$n! = \sqrt{2\pi n}n^n e^{-n+\theta(n)}$，其中 $\frac{1}{12n+1}<\theta(n)<\frac{1}{12n}$，但对于我们的目的来说式(1.3.35)已足够。)于是，式(1.3.34)的左端的概率为(简单起见，k_n 的下标 n 省略)

$$\binom{n}{k}p^k(1-p)^{n-k} \sim \Big(\frac{n}{2\pi k(n-k)}\Big)^{1/2}\frac{n^n}{k^k(n-k)^{n-k}}p^k(1-p)^{n-k}$$

$$= (2\pi np^*(1-p^*))^{-1/2}\times\exp[-k\ln k/n-(n-k)\ln(n-k)/n+k\ln p+(n-k)\ln(1-p)]$$

但是最后一个公式的右端和式(1.3.34)的右端相一致。 □

如果 p^* 接近于 p，我们可以得到

$$D(p\,\|\,p^*) = \frac{1}{2}\Big(\frac{1}{p}+\frac{1}{1-p}\Big)(p^*-p)^2 + O(|p^*-p|^3) \quad (1.3.36)$$

由于 $D(p\,\|\,p^*)|_{p^*=p} = \Big(\frac{\mathrm{d}}{\mathrm{d}p^*}D(p\,\|\,p^*)\Big)\Big|_{p^*=p} = 0$，我们立即可以得到以下推论。 52

推论 1.3.16 (局部 De Moivre-Laplace 定理；参见 PSE I，81 页)如果 $n(p^*-p)=k_n-np=o(n^{1/2})$，那么

$$\mathbb{P}\Big(\sum_{i=1}^{n}\zeta_i = k_n\Big) \sim \frac{1}{\sqrt{2\pi np(1-p)}}\exp\Big(-\frac{n}{2p(1-p)}(p^*-p)^2\Big) \quad (1.3.37)$$

举例 1.3.17 每个单位时间内，一台设备读取有 N 个字符的字符串序列的当前版本，字符串中只有 0 或 1。然后发送字符串中字符 1 的个数。在每次读取之前字符串都会被重新打乱(从 0 变到 1 或者从 1 变到 0，每个字符等概率变化)，求这个信源的信息速率表达式。

解答 信源是 Markov 信源，状态空间为 $\{0, 1, \cdots, N\}$，传输概率矩阵为

$$\begin{bmatrix} 0 & 1 & 0 & 0 & \cdots & 0 & 0 \\ 1/N & 0 & (N-1)/N & 0 & \cdots & 0 & 0 \\ 0 & 2/N & 0 & (N-2)/N & \cdots & 0 & 0 \\ & & \cdots & & \cdots & & \\ 0 & 0 & 0 & 0 & \cdots & 0 & 1/N \\ 0 & 0 & 0 & 0 & \cdots & 1 & 0 \end{bmatrix}$$

该 DTMC 是不可约并且是周期的。它拥有一个唯一的不变分布

$$\pi_i = 2^{-N}\binom{N}{i}, 0 \leqslant i \leqslant N$$

根据定理 1.3.11

$$H = -\sum_{i,j}\pi_i P(i,j)\log P(i,j) = 2^{1-N}\frac{1}{N}\sum_{j=1}^{N-1}\binom{N}{j} j\log\frac{N}{j} \qquad \square$$

举例 1.3.18 一个平稳信源发送符号 0，1，…，$m(m\geqslant 4$，m 是偶数)，DTMC 的传输概率为 $p_{ij}=P(U_{n+1}=k\mid U_n=j)$

$$p_{j+1} = 1/3, 0 \leqslant j \leqslant m-2, p_{j-2} = 1/3, 2 \leqslant j \leqslant m$$

$$p_{j0} = 1/3, 2 \leqslant j \leqslant m-2, p_{00} = p_{11} = p_{m-1,m-1} = p_{mm} = 2/3$$

[53] 第一个符号的分布是等概率的，求信源的信息速率。结果是否和 Shannon FCT 相矛盾？

如果 m 是奇数的话，结果会怎么变化？对于奇数 m，怎么使用 Shannon FCT 来得到上面源的信息速率？

解答 对于 m 是偶数的情况，DTMC 是可约的：有两种互通类，$I_1=\{0, 2, \cdots, m\}$，共有 $m/2+1$ 个状态，$I_2=\{1, 3, \cdots, m-1\}$，共有 $m/2$ 个状态。

相对地，对于任意 n 长字符串的 A_n 集合，

$$\mathbb{P}(A_n) = q\mathbb{P}_1(A_{n1}) + (1-q)\mathbb{P}_2(A_{n2}) \tag{1.3.38}$$

其中 $A_{n1}=A_n\cap I_1$，$A_{n2}=A_n\cap I_2$；$\mathbb{P}_i$ 指的是在 I_i 中的 DTMC，$i=1, 2$，$q=\mathbb{P}(U_1\in I_1)$。

式(1.3.3b)中的随机变量是 $\xi_n=-\frac{1}{n}\log p_n(U^{(n)})$；根据式(1.3.38)，

$$\begin{aligned} \xi_n &= -\frac{1}{n}\log p_{n1}(U^{(n)}), \quad \text{概率为 } q \\ &= -\frac{1}{n}\log p_{n2}(U^{(n)}), \quad \text{概率为 } 1-q \end{aligned} \tag{1.3.39}$$

DTMC 是不可约的，并且在其互通类中非周期，它们的不变分布是均匀分布：

$$\pi_i^{(1)} = \frac{2}{m+2}, i\in I_1, \pi_i^{(2)} = \frac{2}{m}, i\in I_2$$

它们的信息速率是相等的，分别是：

$$H^{(1)} = \log 3 - \frac{8}{3(m+2)}, \quad H^{(2)} = \log 3 - \frac{8}{3m} \tag{1.3.40}$$

根据式(1.3.38)，整个 DTMC 的信息速率等于：

$$H_{\infty} = \begin{cases} H^{(1)} = \max[H^{(1)}, H^{(2)}], & 0<q\leqslant 1 \\ H^{(2)}, & q=0 \end{cases} \tag{1.3.41}$$

对于 $0<q<1$，ShannonFCT 是不适用的：

$$-\frac{1}{n}\log p_{n1}(U^{(n)}) \xrightarrow{\mathbb{P}_1} H^{(1)}, \quad \text{但是} -\frac{1}{n}\log p_{n2}(U^{(n)}) \xrightarrow{\mathbb{P}_2} H^{(2)}$$

也就是 ξ_n 收敛到一个非常数的极限。然而，如果 $q(1-q)$，那么式(1.3.41)减少到一个单行，并且 ShannonFCT 是可以使用的：ξ_n 收敛到一个对应的常数 $H^{(1)}$。

如果 m 是奇数，那么也同样有两个互通类，$I_1=\{0, 2, \cdots, m-1\}$ 和 $I_2=\{1, 3, \cdots, m\}$。每一个都含有 $(m+1)/2$ 个状态。就像之前的 DTMC$\mathbb{P}_1$ 和 $\mathbb{P}_2$ 是不可约的以及非周期的，并且不变分布是均匀分布： 54

$$\pi_i^{(1)}=\frac{2}{m+1}, i\in I_1,\quad \pi_i^{(2)}=\frac{2}{m+1}, i\in I_2$$

它们的通用信息率等于

$$H_{\mathrm{univ}}=\log 3-\frac{8}{3(m+1)} \tag{1.3.42}$$

同样给出了整个 DTMC 的信息率。它与 Shannon 的 FCT 契合，因为现在

$$\xi_n=-\frac{1}{n}\log p_n(U^{(n)})\xrightarrow{\mathcal{P}}H_{\mathrm{univ}} \tag{1.3.43}$$

□

举例 1.3.19 设 a 是符号集 A 的大小，b 是符号集 B 的大小。从符号集 $A+B$ 中选中一个符号源，并且假定没有两个出自于 A 的符号连续出现。

(a) 假定一个信息符合 DTMC，所有被允许在指定位置出现的字符等可能出现，证明这个源的信息率是：

$$H=\frac{a\log b+(a+b)\log(a+b)}{2a+b} \tag{1.3.44}$$

(b) 通过解决一个递归关系，或相反地，求出多少个长度为 n 的字符串满足没有两个出自于 A 的字母连续出现的约束。假定这些字符串是等可能并且 $n\to\infty$，证明极限信息率变为：

$$H=\log\left(\frac{b+\sqrt{b^2+4ab}}{2}\right)$$

为什么答案会不同？

解答 (a) DTMC 的传输概率通过下面的式子给定：

$$P(x,y)=\begin{cases}0, & x,y\in\{1,\cdots,a\}\\ 1/b, & x\in\{1,\cdots,a\}, y\in\{a+1,\cdots,a+b\}\\ 1/(a+b), & x\in\{a+1,\cdots,a+b\}, y\in\{1,\cdots,a+b\}\end{cases}$$

这个链是不可约且非周期的，并且满足 $\min P^{(2)}(x, y)>0$；所以，不变分布：$\pi=(\pi(x), x\in\{1, \cdots, a+b\})$ 是唯一的。我们可以从详细平衡方程(DBE)$\pi(x)P(x, y)=\pi(y)$ $P(y, x)$(参见 PSE II，82 页)中找到 π，其中 55

$$\pi(x)=\begin{cases}1/(2a+b), & x\in\{1,\cdots,a\}\\ (a+b)/[b(2a+b)], & x\in\{a+1,\cdots,a+b\}\end{cases}$$

DBE 意味着 π 是一个不变量，$\pi(y)=\sum_x\pi(x)P(x, y)$，但是反之就不成立，所以我们可以得到式(1.3.44)。

(b) 设 M_n 为符合约束的长度为 n 的字符串数量，A_n 是符合约束且结尾字母来源于 A 的 n 长字符串的数量，B_n 是符合约束且最后结尾字母来源于 B 的 n 长字符串的数量，那么

$$M_n=A_n+B_n, A_{n+1}=aB_n,\quad B_{n+1}=b(A_n+B_n)$$

即得

$$B_{n+1} = bB_n + abB_{n-1}$$

最后的迭代可以通过下式解决：

$$B_n = c_+\lambda_+^n + c_-\lambda_-^n$$

其中$\lambda_\pm$是矩阵$\begin{bmatrix} 0 & ab \\ 1 & b \end{bmatrix}$的特征值，

$$\lambda_\pm = \frac{b \pm \sqrt{b^2 + 4ab}}{2}$$

并且$c_\pm$是常数，$c_+ > 0$。所以

$$M_n = a(c_+\lambda_+^{n-1} + c_-\lambda_-^{n-1}) + (c_+\lambda_+^n + c_-\lambda_-^n) = \lambda_+^n\left(c_+\left(a\frac{\lambda_+^{n-1}}{\lambda_+^n} + \frac{\lambda_+^n}{\lambda_+^n}\right) + c_-\left(a\frac{1}{\lambda_+} - 1\right)\right)$$

$\frac{1}{n}\log M_n$可以被表示为和

$$\log\lambda_+ + \frac{1}{n}\log\left(c_+\left(a\frac{\lambda_+^{n-1}}{\lambda_+^n} + \frac{\lambda_+^n}{\lambda_+^n}\right) + c_-\left(a\frac{1}{\lambda_+} + 1\right)\right)$$

注意$\left|\frac{\lambda_-}{\lambda_+}\right| < 1$。所以，极限信息率等于

$$\lim_{n\to\infty}\frac{1}{n}\log M_n = \log\lambda_+$$

56 两个答案之所以不同，是因为条件均匀分布导致了子序列字母间的强依赖关系：它们没有形成一个DTMC。 □

举例 1.3.20 令$\{U_j: j=1, 2, \cdots\}$是一个不可约且非周期的有限状态空间的DTMC。给定$n \geq 1$，$\alpha \in (0, 1)$，根据它们的概率$(\mathbb{P}(U^{(n)} = u_1^{(n)}) \geq \mathbb{P}(U^{(n)} = u_2^{(n)}) \geq \cdots)$排序字符串$u^{(n)}$，并且按排序选择它们，直到剩余集合的概率$\leq 1-\alpha$。令$M_n(\alpha)$表示选中字符串的数量。证明信源的信息率为$\lim_{n\to\infty}\frac{1}{n}M_n(\alpha) = H$。

(a) 当转移概率矩阵P的行都相等时(也就是说$\{U_j\}$是Bernoulli序列)。

(b) 当P的行是彼此的置换，且在一般情况下时，评估这个结果对于编码理论的重要性。

解答 (a) 令$\mathbb{P}$表示IID序列(U_n)的概率分布，$H = -\sum_{j=1}^{m} p_j\log p_j$(信源的二元熵)。固定$\varepsilon > 0$并分割所有$n$字符串的集合$I^{(n)}$为三个不相交的子集：

$$\mathscr{K}_+ = \{u^{(n)}: p(u^{(n)}) \geq 2^{-n(H-\varepsilon)}\}, \mathscr{K}_- = \{u^{(n)}: p(u^{(n)}) \leq 2^{-n(H+\varepsilon)}\}$$

和

$$\mathscr{K} = \{u^{(n)}: 2^{-n(H+\varepsilon)} < p(u^{(n)}) < 2^{-n(H-\varepsilon)}\}$$

通过大数定理(或者渐近均分定理)，$-\frac{1}{n}\log\mathbb{P}(U^{(n)})$收敛到$H(-h)$，也就是$\lim_{n\to\infty}\mathbb{P}(K_+ \cup K_-) = 0$，$\lim_{n\to\infty}\mathbb{P}(K) = 1$。所以，为了使概率$\geq\alpha$，对于足够大的$n$，在(i)中不能局限于$\mathscr{K}_+$，必须从$\mathscr{K}$中取字符串，(ii)不需要$\mathscr{K}_-$中的字符串，即得到$\mathscr{K}$中最后一个选中的字符串。通过$\mathscr{M}_n(\alpha)$定义选定的字符串，以及通过$M_n$定义的$\#\mathscr{M}_n(\alpha)$，可以得到两个双边界：

$$\alpha \leq \mathbb{P}(\mathscr{M}_n(\alpha)) \leq \alpha + 2^{-n(H-\varepsilon)}$$

以及

$$2^{-n(H-\varepsilon)}M_n(\alpha) \leqslant P(\mathscr{M}_n(\alpha)) \leqslant P(K_+) + 2^{-n(H-\varepsilon)}M_n(\alpha)$$

将第一个式子 $P(\mathscr{M}_n(\alpha))$ 的界代入第二个式子得

$$2^{-n(H-\varepsilon)}M_n(\alpha) \leqslant \alpha + 2^{-n(H-\varepsilon)} \quad 和 \quad 2^{-n(H-\varepsilon)}M_n(\alpha) \geqslant \alpha - P(K_-)$$

[57]

这些不等式意味着

$$\limsup_{n\to\infty} \frac{1}{n}\log M_n(\alpha) \leqslant H+\varepsilon \quad 和 \quad \liminf_{n\to\infty} \frac{1}{n}\log M_n(\alpha) \geqslant H-\varepsilon$$

因为 ε 是任意的，故极限是 H。

(b) 在置换时，这个论证可能会在没有任何变化的情况下重复，因为排序的概率像(a)中一样形成了一个相同的集合，并且通常是通过对 $(1/n)\xi_n$ 应用大数定律，参见式(1.3.3b)和式(1.3.19)。最后对于编码理论的性能：如果我们准备处理以使错误概率 $\leqslant \alpha$，那么不需要给所有的 m^n 序列编码 $u^{(n)}$，仅需编码最频繁的 $\sim 2^{nH}$ 个。因为 $H \leqslant \log m$(在大多数情况下 $< \log m$)，它得到了存储空间上的巨大节约(数据压缩)。 □

举例 1.3.21 一个二元信源根据以下规则发送数字 0 或 1：

$$P(X_n = k \mid X_{n-1} = j, X_{n-2} = i) = q_r$$

其中 k，j，i 和 r 取 0 或 1，$r = k - j - i \bmod 2$，并且 $q_0 + q_1 = 1$。求信源的信息率。

也得出二元 Bernoulli 信源的信息率，以概率 q_0 和 q_1 发送数字 0 和 1，阐述这两种结果的关系。

解答 这个信源是一个第二序列的 DTMC，就是 (X_n, X_{n-1}) 产生一个四状态的 DTMC

$$P(00,00) = q_0, P(00,01) = q_1, P(01,10) = q_0, P(01,11) = q_1$$
$$P(10,00) = q_0, P(10,01) = q_1, P(11,10) = q_0, P(11,11) = q_1$$

剩下 8 个转移概率矩阵的元为空。这就给出了

$$H = -q_0\log q_0 - q_1\log q_1$$

对于 Bernoulli 信源答案是一样的。 □

举例 1.3.22 找到随机走步在 3×3 棋盘 DTMC 的频率：

$$\begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix} \tag{1.3.45}$$

找到车、象、皇后和国王的频率。 [58]

解答 我们仅考虑国王的 DTMC，其他情形相似。转移矩阵可以表示为：

$$\begin{bmatrix}
0 & 1/3 & 0 & 1/3 & 1/3 & 0 & 0 & 0 & 0 \\
1/5 & 0 & 1/5 & 1/5 & 1/5 & 1/5 & 0 & 0 & 0 \\
0 & 1/3 & 0 & 0 & 1/3 & 1/3 & 0 & 0 & 0 \\
1/5 & 1/5 & 0 & 0 & 1/5 & 0 & 1/5 & 1/5 & 0 \\
1/8 & 1/8 & 1/8 & 1/8 & 0 & 1/8 & 1/8 & 1/8 & 1/8 \\
0 & 1/5 & 1/5 & 0 & 1/5 & 0 & 0 & 1/5 & 1/5 \\
0 & 0 & 0 & 1/3 & 1/3 & 0 & 0 & 0 & 1/3 \\
0 & 0 & 0 & 1/5 & 1/5 & 1/5 & 1/5 & 0 & 1/5 \\
0 & 0 & 0 & 0 & 1/3 & 1/3 & 0 & 1/3 & 0
\end{bmatrix}$$

通过对称不变分布：$\pi_1 = \pi_3 = \pi_6 = \pi_7 = \lambda$，$\pi_2 = \pi_4 = \pi_8 = \pi_9 = \mu$，$\pi_5 = \nu$，通过 DBE

$$\lambda/3 = \mu/5, \lambda/3 = \nu/8, 4\lambda + 4\mu + \nu = 1$$

推出 $\lambda=\frac{3}{40}$，$\mu=\frac{1}{8}$，$\nu=\frac{1}{5}$，我们有

$$H=-4\lambda\frac{1}{3}\log\frac{1}{3}-4\mu\frac{1}{5}\log\frac{1}{5}-\nu\frac{1}{8}\log\frac{1}{8}-\frac{1}{10}\log 15=\frac{3}{40} \qquad \square$$

1.4 信道，解码规则，Shannon 第二编码定理

本节我们证明 Shannon 理论的一个核心思想——第二编码定理(SCT)，也就是大家所知道的有噪编码定理(NCT)。Shannon 证明了他的想法并且在他 20 世纪 40 年代的论文和著作中给出了证明的扩展。他的论点受到了一些专业数学家的批评(也不是完全不合理)。数学界花费了大约十年的时间给出了 SCT 严谨和完整的证明。然而，事后看来，Shannon 的直觉以及他对熵和编码等基础概念及其与长随机序列的数学统计关系的扎实掌握是令人敬佩的，本书将重点研究这个话题很多方面的内容，因而不可避免地会带入我们的个人偏好。

至此，我们已经考虑了一个发送随机文本 $U_1U_2\cdots$ 的信源，和一个使用码 $f_N: I^{\times N}\to J^{\times N}$，$J=\{0,1\}$ 的二元码本 $x^{(N)}$ 编码的信息 $u^{(N)}$。现在我们关注信息 u 的长度和码字长度 N：它是通过发送信号的信道特征决定的。码 f_N 应该被接收端获知，这很重要。

39

通常，信道易受到噪声的影响，噪声可以使得传输的信息失真，在输出端的信息通常与输入端的信息不同。形式上，一个信道可以通过条件分布来表征：

$$\boldsymbol{P}_{\mathrm{ch}}(\text{接收信息 } y^{(N)}\mid\text{码字 } x^{(N)}\text{ 发送}) \tag{1.4.1}$$

我们再次假设这对于输入端和输出端都是已知的。(我们用一个符号 $\boldsymbol{P}_{\mathrm{ch}}(\cdot\mid\text{码字 } x^{(N)}\text{ 发送})$，或者简明地用 $\boldsymbol{P}_{\mathrm{ch}}(\cdot\mid x^{(N)})$。为了强调这个概率分布在码字 $x^{(N)}$ 发送的情况下是由信道生成的，)提到下面的信道，我们参考一个条件概率(1.4.1)(或者说一组相互联系的条件概率，取决于 N)。因此，我们用符号 $\boldsymbol{Y}^{(N)}$ 代表信道输出的随机字符串，假定码字 $x^{(N)}$ 发送的情况下

$$\boldsymbol{P}_{\mathrm{ch}}(\boldsymbol{Y}^{(N)}=y^{(N)}\mid x^{(N)})=\boldsymbol{P}_{\mathrm{ch}}(y^{(N)}\mid x^{(N)})$$

一个重要的例子是无记忆二元信道(MBC)，其中

$$P_{\mathrm{ch}}(y^{(N)}\mid x^{(N)})=\prod_{i=1}^{N}P(y_i\mid x_i) \tag{1.4.2}$$

如果 $y^{(N)}=y_1\cdots y_N$，$x^{(N)}=x_1\cdots x_N$。这里，$P(y\mid x)$，$x,y=0,1$ 是一个符号到符号的信道概率(也就是说考虑到符号 x 已经被发送，条件概率是为了在信道输出端获得符号 y)。很显然，$\{P(y\mid x)\}$ 是一个 2×2 的统计矩阵(一般被称为信道矩阵)。特别地，如果 $P(1\mid 0)=P(0\mid 1)=p$，这个信道被称为对称的(MBSC)。这个信道矩阵有一个形式

$$\Pi=\begin{pmatrix}1-p & p\\ p & 1-p\end{pmatrix}$$

p 被称为行错误概率(或者符号错误概率)。

例子 1.4.1 考虑一个无记忆信道，其中 $Y=X+Z$，加性噪声 Z 以 1/2 的概率取 0 或者 1；a 是一个给定实数。输入的符号集是 $\{0,1\}$，Z 是独立于 X 的。

这个信道的特征依赖于 a 的取值，实际上，如果 $a\neq\pm 1$，这个信道是唯一可译码。换句话说，如果我们必须用这个信道来传送长度为 n(总长是 2^n)的信息(字符串)，那么所有信息可以马上被发送，并且接收者可以将它恢复。但是如果 $a=\pm 1$，那么就有可能发生错误，为了保证接收者可以恢复信息，那么我们必须对信息进行编码，也就导致了需要增加

发送到信道里的信息的长度，从 n 到 N。 60

换句话说，发送到信道的长度为 N 的字符串将变成代表长度为 n 的信源信息的码字。能保证接收者恢复原始信号的 n/N 最大比是一个非常重要的信道特征，称为信道容量。正如我们将看到的，$a\neq\pm1$ 变为 $a=\pm1$ 导致信道容量从 1(无需编码)到 1/2(需要的码字长度为信源信息长度的两倍)。

所以，我们需要介绍一个译码规则 $\hat{f}_N$：$J^{\times N}\to I^{\times n}$ 以使得总体错误概率

$$\begin{aligned}\epsilon &= \sum_{u^{(n)}} \mathbb{P}(\hat{f}_N(\boldsymbol{Y}^{(N)})\neq \boldsymbol{u}^{(n)}, \boldsymbol{u}^{(n)}\text{ 发出}) \\ &= \sum_{u^{(n)}} \mathbb{P}(\boldsymbol{U}^{(n)}=\boldsymbol{u}^{(n)})\boldsymbol{P}_A(\hat{f}_N(\boldsymbol{Y}^{(N)})\neq \boldsymbol{u}^{(n)} \mid f_n(\boldsymbol{u}^{(n)})\text{ 发送})\end{aligned} \tag{1.4.3}$$

很小。我们将试着(在一定的条件下才会成功)令错误概率随着 $n\to\infty$ 趋近于 0。

这个想法是基于以下的事实：

(1) 对于有渐近均分性的信源，发射的不同 n 长字符串的数量是 $2^{n(H+o(1))}$，其中 $H\leq\log m$ 是信源的信息率。因此，我们需要编码的不是 $m^n=2^{n\log m}$ 信息，仅仅是可能小一些的 $2^{n(H+o(1))}$。也就是说，码 f_n 也许被定义在 $I^{\times n}$ 的子集上，码字长度是 $N=\lceil nH\rceil$。

(2) 我们可能试一个更大的 N：$N=\lceil \bar{R}^{-1}nH\rceil$，$\bar{R}$ 是一个常数，$0<\bar{R}<1$，也就是说，码字长度从 $\lceil nH\rceil$ 到 $\lceil \bar{R}^{-1}nH\rceil$ 的增长将允许我们引出一个在码 f_n 上的冗余，也许我们希望可以用这个冗余去减少整体错误概率(1.4.3)(前提是附加的解码规则是“好的”)。当然，最小化 $\bar{R}^{-1}$ 是理想的，也就是说最大化 $\bar{R}$：它将给出有最优参数的码。信道将决定 $\bar{R}$ 的上界。

引入符号标记是有帮助的。因为码字长度是一个关键参数，我们写 N 而不是 $\bar{R}^{-1}nH$，$\bar{R}N$ 而不是 Hn。由信源发送不同字符串的数目变为 $2^{N(\bar{R}+o(1))}$。在未来，指数 $n\sim\frac{N\bar{R}}{H}$ 可能将在一些地方被忽略(除了被 N 代替)。考虑利用 $\#\mathscr{U}=2^{N(\bar{R}+o(1))}$，一个典型的由信源发射的离散序列子集 $\mathscr{U}_N$ 是很方便的。形式上，$\mathscr{U}_N$ 可以包含不同长度的字符串；只有对数渐近的 $\#\mathscr{U}_N$ 需要考虑。因此，我们忽略 $\boldsymbol{u}^{(n)}$ 中的上标 (n)。 61

定义 1.4.2 $\bar{R}\in(0,1)$ 被称作可靠的传输速率(对于一个给定的信道)，成立基于以下条件：假设信源的字符串从符合 $\#\mathscr{U}_N=2^{N(\bar{R}+o(1))}$ 的集合 $\mathscr{U}_N$ 等概取值，那么存在一个编码规则 f_N：$\mathscr{U}_N\to\mathscr{X}_N\subseteq I^{\times N}$ 和一个译码规则 $\hat{f}_N$：$J^{\times N}\to\mathscr{U}_N$，误码率

$$\sum_{u\in\mathscr{U}_N}\frac{1}{\#\mathscr{U}_N}\boldsymbol{P}_B(\hat{f}_N(\boldsymbol{Y}^{(N)})\neq \boldsymbol{u}\mid f_N(\boldsymbol{u})\text{ 发送}) \tag{1.4.4}$$

当 $N\to\infty$ 时取值为 0。也就是说，对于每个满足 $\lim\limits_{N\to\infty}\frac{1}{N}\log\#\mathscr{U}_N=\bar{R}$ 的序列 $\mathscr{U}_N$，存在一个编码规则 f_N：$\mathscr{U}_N\to\mathscr{X}_N$，$\mathscr{X}_N\subseteq I^{\times N}$ 的序列和一个译码规则 $\hat{f}_N$：$J^{\times N}\to\mathscr{U}_N$ 的序列，使得

$$\lim_{N\to\infty}\frac{1}{\#\mathscr{U}_N}\sum_{u\in\mathscr{U}_N}\sum_{y^{(N)}}\mathbf{1}(\hat{f}_N(\boldsymbol{Y}^{(N)})\neq \boldsymbol{u})\boldsymbol{P}_B(\boldsymbol{Y}^{(N)}\mid f_N(\boldsymbol{u}))=0 \tag{1.4.5}$$

定义 1.4.3 信道容量的上确界定义如下

$$C=\sup[\bar{R}\in(0,1),\bar{R}\text{ 是一个可靠传输速率}] \tag{1.4.6}$$

备注 1.4.4 (a) 从物理意义上，信道容量可以认为是一个极限 $\lim\limits_{N\to\infty}\frac{1}{N}\log n(N)$，其中 $n(N)$ 是字符串长度为 N 的最大数目，此字符串通过信道后可以无差错译码。

(b) $\mathscr{U}_N$ 的等概分布式是由于它是最坏的情况，参见下面的定理 1.4.6。

(c) 如果编码规则 f_N 是一对一的(无损)，那么它满足译码规则，其映射为 $J^{(N)}\to\mathscr{X}_N$，而不是 $J^{(N)}\to\mathscr{U}_N$：如果我们正确地猜测了已经发射的码字 $x^{(N)}$，我们令 $u=f_N^{-1}(x^{(N)})$。除此之外，如果基于$\mathscr{U}$的信源的分布是等概的，那么错误率 ε 可以写为码字集合$\mathscr{X}_N$ 的一个平均：

$$\varepsilon=\frac{1}{\#\mathscr{X}}\sum_{x\in\mathscr{X}_N}[1-P_{ch}(\hat{f}_N(Y^{(N)})=x|x\ 发送)]$$

因此，$\varepsilon=\varepsilon^{av}$ 是有意义的，它表明了平均错误概率。另一个形式是最大错误概率

$$\varepsilon^{max}=\max\{1-P_{ch}(\hat{f}_N(Y^{(N)})=x|x\ 发送);x\in\mathscr{X}_N\}$$

明显可得，$\varepsilon^{av}\leqslant\varepsilon^{max}$。在本节中，我们基于 $\varepsilon^{av}\to0$ 考虑问题，但是留下了一个问题：$\varepsilon^{max}\to0$ 是否正确。然而，在 2.2 节中，我们减少考虑用 ε^{av} 代替 ε^{max} 这个问题，所以如果 ε^{max} 代替了 ε^{av}，本节信道容量的函数也是成立的。

62

备注 1.4.5 (a) 根据下面的定理 1.4.17，可以给出 MBC 信道的信道容量

$$C=\sup_{p_{X_k}}I(X_k;Y_k) \tag{1.4.7}$$

在这里，$I(X_k;Y_k)$是输入符号 X_k 和输出符号 Y_k 之间的互信息(下标 k 可以省略)，联合分布为

$$\mathbb{P}(X=x,Y=y)=p_X(x)P(y|x),x,y=0,1 \tag{1.4.8}$$

这里 $p_X(x)=\mathbb{P}(X=x)$。式(1.4.7)中的上确界是全局可能性分布 $p_X(x)=(p_X(0),p_X(1))$。一个有用的公式即 $I(X;Y)=h(Y)-h(Y|X)$(见式(1.3.12))。事实上，对于 MBSC 来说

$$\begin{aligned}h(Y|X)&=-\sum_{x=0,1}p_X(x)\sum_{y=0,1}P(y|x)\log P(y|x)\\&=-\sum_{y=0,1}P(y|x)\log P(y|x)=h_2(p,1-p)=\eta(p)\end{aligned} \tag{1.4.9}$$

下标 2 可以省略。

因此 $h(Y|X)=\eta(p)$不依赖于输入分布 p_X，并且对于 MBSC

$$C=\sup_{p_X}h(Y)-\eta(p) \tag{1.4.10}$$

但是当 $p_X(0)=p_X(1)=1/2$ 并且 $p_Y(0)=p_Y(1)=1/2(p+1-p)=1/2$ 时，$\sup\limits_{p_X}h(Y)$ 等于 $\log2-1$。因此对于一个行错误概率为 p 的 MBSC，有

$$C=1-\eta(p) \tag{1.4.11}$$

(b) 假设信源 $U_1U_2\cdots$具有渐近均分性，并且信息速率为 H。为了通过一个信道容量为 C 的信道发送一段信源产生的信号，我们需要对长度为 n 的信源信息用长度为 $n(H+\varepsilon)/C$ 的码字进行编码，以确保当 $n\to\infty$ 时整体错误概率趋近于零，数值 $\varepsilon>0$ 可以取任意小的值。因此，如果 $H/C<1$，可以使一段信息的编码速度比它的产生速度快：信道可以被用来可靠地从信源传送信息。相反，如果 $H/C>1$，信源信息的产生速度太快以至于我们无法对其进行编码并通过信道可靠地传送出去。这种情况下传输是不可靠的。对于一个 Bernoulli 或者静态 Markow 源和一个 MBSC，条件 $H/C<1$ 等价于 $h(U)+\eta(p)<1$ 或者 $h(U_{n+1}|U_n)+\eta(p)<1$。

63

事实上，Shannon 的想法并没有因为他主导同时代的数学界而被轻易地接收。到底谁被认为是信息论的发明者这个观点就变得很有意思了。

定理 1.4.6 确定一个信道(即式(1.4.1)中的条件概率 P_{ch})和信源序列集合$\mathscr{U}$。用 $\varepsilon(\mathbb{P})$来

表示对于在$\mathscr{U}$上概率分布为$\mathbb{P}$的集合$U^{(n)}$的全局错误概率(1.4.3)，在所有的编码解码规则下都是最小化的。那么

$$\varepsilon(\mathbb{P}) \leqslant \varepsilon(\mathbb{P}^0) \tag{1.4.12}$$

其中$\mathbb{P}^0$表示在$\mathscr{U}$上的均匀分布。

证明　确定编码解码规则f和$\hat{f}$，并且让一个序列$u\in\mathscr{U}$的概率分布为$\mathbb{P}(u)$。当u发散时定义错误概率为

$$\beta(u) := \sum_{y:\hat{f}(y)\neq u} P_{n,u}(y \mid f(u))$$

全局错误概率为

$$\varepsilon(=\varepsilon(\mathbb{P},f,\hat{f}) = \sum_{u\in\mathscr{U}} \mathbb{P}(u)\beta(u)$$

如果我们变换码字的配置(也就是用$f(u')$编码u，其中$u'=\lambda(u)$，并且λ是程度$\#\mathscr{U}$的一个变换)。我们可以得到全局错误概率$\varepsilon(\lambda)=\sum_{u\in\mathscr{U}}\mathbb{P}(u)\beta(\lambda(u))$。在$\mathbb{P}(u)=(\#\mathscr{U})^{-1}$的情况下(均匀分布)，$\varepsilon(\lambda)$不依赖于$\lambda$，并且可以得出

$$\bar{\varepsilon} = \frac{1}{\#\mathscr{U}}\sum_{u\in\mathscr{U}}\beta(u) = \varepsilon(\mathbb{P}^0,f,\hat{f})$$

对于每一个概率分布$\{\mathbb{P}(u),\ u\in\mathscr{U}\}$，存在$\lambda$可以使得$\varepsilon(\lambda)\leqslant\bar{\varepsilon}$。事实上，进行一个随机排列$\Lambda$，令它均匀分布在$(\#\mathscr{U})!$个关于$\#\mathscr{U}$的全排列上。那么

$$\begin{aligned}\min_{\lambda}\varepsilon(\lambda) &\leqslant \mathbb{E}\varepsilon(\Lambda) = \mathbb{E}\sum_{u\in\mathscr{U}}\mathbb{P}(u)\beta(\Lambda u)\\ &= \sum_{u\in\mathscr{U}}\mathbb{P}(u)\mathbb{E}\beta(\Lambda u) = \sum_{u\in\mathscr{U}}\mathbb{P}(u)\frac{1}{\#\mathscr{U}}\sum_{\bar{u}\in\mathscr{U}}\beta(\bar{u}) = \bar{\varepsilon}\end{aligned}$$

因此，对于给定的任何f和$\hat{f}$，我们都可以找到新的编码解码规则来使得全局错误概率$\leqslant\varepsilon(\mathbb{P}^0,f,\hat{f})$。对$f$和$\hat{f}$的最小化会引出式(1.4.12)。 □ 64

举例 1.4.7　随机变量X和Y，取值限定在"符号集"I和J，分别代表传输信道的输入和输出符号。信道的条件概率为$P(x|y)=\mathbb{P}(X=x|Y=y)$，用$h(P(\cdot\mid y))$表示条件分布$P(\cdot\mid y)$，$y\in J$的熵

$$h(P(\cdot\mid y)) = -\sum_{x} P(X|y)\log P(x|y)$$

用$h(X|Y)$表示给定Y情况下X的条件熵。定义理想观测解码规则为一个映射$f^{(I)}: J\to I$，这样对于所有的$y\in J$都有$P(f(y)\mid y)=\max_{x\in I}P(x|y)$。试证明

(a) 在这样的准则下，错误概率为

$$\pi_{cr}^{(I)}(y) = \sum_{x\in I}\mathbb{1}(x\neq f(y))P(x|y)$$

满足$\pi_{cr}^{(I)}(y)\leqslant\frac{1}{2}h(P(\cdot\mid y))$。

(b) 错误概率的期望值满足$\pi_{cr}^{(I)}(Y)\leqslant\frac{1}{2}h(X|Y)$。

解答　的确，(a)满足举例1.2.7中的(iii)，因为

$$\pi_{cr}^{(I)} = 1-P(f(y)|y) = 1-P_{max}(\cdot\mid y)$$

上式小于或者等于$\frac{1}{2}h(P(\cdot\mid y))$。最终，因为$h(X|Y)=\mathbb{E}h(P(\cdot\mid Y))$，可以通过求期望由(a)推出(b)。 □

我们在前面提到，一个一般的解码规则(或者译码器)是映射$\hat{f}_N: \mathcal{Y}^{\times N} \to \mathscr{U}_N$；在无损编码准则$f_N$条件下，$\hat{f}_N$是一个映射$\mathcal{Y}^{\times N} \to \mathscr{X}_N$。其中$\mathscr{X}$是码字的集合，有时通过对每个码字$x^{(N)}$固定一个集合$A(x^{(N)}) \subset \mathcal{Y}^{\times N}$来确定解码规则会变得很方便，这样$A(x_1^{(N)})$和$A(x_2^{(N)})$在$x_1^{(N)} \neq x_2^{(N)}$时是互斥的，并且并集$\bigcup_{x^{(N)} \in \mathscr{X}_N} A(x^{(N)})$给出全部的$\mathcal{Y}^{\times N}$。考虑到$y^{(N)} \in A(x^{(N)})$，我们将其译码为$\hat{f}_N(y^{(N)}) = x^{(N)}$。

虽然在信道容量的定义中我们假设信源信息是均匀分布的(正如我们提到的，就定理1.4.6的而言，这给出了最坏的情况)，但是在实际过程中信源并不总是符合这种假设。最后，我们需要分辨这两种情况：(i)接收端知道信源符号串的概率(因此也知道码字$x^{(N)} \in \mathscr{X}_N$的概率分布$p_N(x^{(N)})$)

$$p(u) = \mathrm{P}(U = u) \tag{1.4.13}$$

65 和(ii)接收端不知道$p_N(x^{(N)})$，两种自然解码规则分别如下。

(i) 理想观测(Ideal Observer，IO)准则将接收到的信号$y^{(N)}$译为最大后验概率的码字$x^{(N)*}$

$$\mathrm{P}(x^{(N)}\ 发送 \mid y^{(N)}\ 接收) = \frac{p_N(x^{(N)})P_{ch}(y^{(N)} \mid x^{(N)})}{p_Y^{(N)}(y^{(N)})} \tag{1.4.14}$$

其中

$$p_Y^{(N)}(y^{(N)}) = \sum_{\tilde{x}^{(N)} \in \mathscr{X}_N} p_N(\tilde{x}^{(N)})P_{ch}(y^{(N)} \mid \tilde{x}^{(N)})$$

(ii) 最大似然(Maximum Likelihood，ML)准则通过将接收到的信号$y^{(N)}$译为最大先验概率的码字$x_*^{(N)}$

$$P_{ch}(y^{(N)} \mid x^{(N)}) \tag{1.4.15}$$

定理 1.4.8 假设编码准则f定义在具有非零发生概率的消息集合上，并且是一一对应的，那么：

(a) 对于任何编码准则，IO译码器可以实现最小的全局误码率。

(b) 如果信源信息U在集合$\mathscr{U}$上是等概分布的，那么对于任何上面提到的编码准则$f: \mathscr{U} \to \mathscr{X}_N$，随机码字$X^{(N)} = f(U)$在$\mathscr{X}_N$上是等概的，IO译码器和ML译码器的效果相同。

证明 为了简化，我们再次忽略上标(N)。

(a) 注意对于一个接收到的字符y，IO显著最大化了联合概率$p(x)P_{ch}(y|x)$(当y确定时，式(1.4.14)的分母就确定了)。如果我们使用编码准则f和解码规则$\hat{f}$，全局错误概率(见式(1.4.3))为

$$\begin{aligned}
&\sum_u \mathcal{P}(U = u)P_{ch}(\hat{f}(y) \neq u \mid f(u)\ 发送) \\
&= \sum_x p(x) \sum_y \mathbf{1}(\hat{f}(y) \neq x)P_{ch}(y|x) = \sum_y \sum_x \mathbf{1}(x \neq \hat{f}(y))p(x)P_{ch}(y|x) \\
&= \sum_y \sum_x p(x)P_{ch}(y|x) - \sum_y p(\hat{f}(y))P_{ch}(y|\hat{f}(y)) = 1 - \sum_y p(\hat{f}(y))P_{ch}(y|\hat{f}(y))
\end{aligned}$$

要注意当$\hat{f}$符合IO准则时，和$\sum_y p(\hat{f}(y))P_{ch}(y|\hat{f}(y))$中的每个符号的取值都最大化了。因此，可以实现整体和的最大化和全局错误概率的最小化。

66 (b) 第一个描述已经明显指出了证明方向，事实上，第二个也很清楚。 □

假设在信道容量的定义中信源信息是等概的，那么就很自然要去研究ML译码器。当

使用 ML 译码器的时候，可能因为其中一个译码器选择了一个错误的码字或者一个编码准则 f 不是一一对应的而发生一个错误。这样的概率在定理 1.4.8 中做出了估计。为了更进一步的简化，我们用 $\boldsymbol{P}$ 代替 $\boldsymbol{P}_{X|Y}$ 符号 $\mathbb{P}$ 主要用来表示联合输入/输出分布。

引理 1.4.9 如果信源信息在集合$\mathscr{U}$是等概的，当使用 ML 译码器和编码准则 f 时，全局错误概率符合

$$\varepsilon(f) \leqslant \frac{1}{\#\mathscr{U}} \sum_{u \in \mathscr{U}} \sum_{u' \in \mathscr{U}, u' \neq u} \mathbb{P}(\boldsymbol{P}(\boldsymbol{Y}|f(u')) \geqslant \boldsymbol{P}(\boldsymbol{Y}|f(u)) | \boldsymbol{U} = u) \tag{1.4.16}$$

证明 如果信源发送 u 并且使用 ML 译码器，我们得到

(a) 会有一个错误，对于某些 $u' \neq u$，当 $\boldsymbol{P}(\boldsymbol{Y}|f(u')) > \boldsymbol{P}(\boldsymbol{Y}|f(u))$时。

(b) 可能有一个错误，对于某些 $u' \neq u$(这包含 $f(u)=f(u')$的情况)，当 $\boldsymbol{P}(\boldsymbol{Y}|f(u'))=\boldsymbol{P}(\boldsymbol{Y}|f(u))$时。

(c) 没有错误，对于任何 $u' \neq u$，当 $\boldsymbol{P}(\boldsymbol{Y}|f(u')) < \boldsymbol{P}(\boldsymbol{Y}|f(u))$时。

因此，概率的上界可以表示为：

$$\begin{aligned}\mathbb{P}(\text{error}|\boldsymbol{U}=u) &\leqslant \mathbb{P}(\boldsymbol{P}(\boldsymbol{Y}|f(u')) \geqslant \boldsymbol{P}(\boldsymbol{Y}|f(u)), \quad \text{对某些 } u' \neq u | \boldsymbol{U}=u) \\ &\leqslant \sum_{u' \in \mathscr{U}} \mathbf{1}(u' \neq u) \mathbb{P}(\boldsymbol{P}(\boldsymbol{Y}|f(u')) \geqslant \boldsymbol{P}(\boldsymbol{Y}|f(u)) | \boldsymbol{U}=u)\end{aligned}$$

乘上 $\frac{1}{\#\mathscr{U}}$并且对所有 u 求和可以得到结果。 □

备注 1.4.10 假如用 $p(u)$来取代$\frac{1}{\#\mathscr{U}}$，上界(1.4.16)对于所有的概率分布 $p(u)=\mathbb{P}(\boldsymbol{U}=u)$都应当是不变的。

正如我们已经提到的，随机编码在确定性编码准则中是一项有用的工具。确定性编码准则是映射 $f: \mathscr{U} \to J^{\times N}$；如果 $\#\mathscr{U}=r$，那么 f 表示为码字$\{f(u_1), \cdots, f(u_r)\}$的集合，或者相当于级联的超字符串(或者码本)

$$f(u_1)\cdots f(u_r) \in (J^{\times N})^{\times r} = \{0,1\}^{\times Nr}$$

其中 $u_1, \cdots, u_r$ 是信源序列(不是字母)构成集合$\mathscr{U}$，如果 f 是无损的，那么当 $i \neq j$ 时 $f(u_i) \neq f(u_j)$。随机编码准则是一个$(J^{\times N})^r$ 上的随机元素 F，概率分布为 $\mathbb{P}(F=f)$，$f \in (J^{\times N})^r$。 67

相应地，F 可能被认为是随机码字 $F(u_i)(i=1, \cdots, r)$的集合或者相应地被认为是一个随机码本

$$F(u_1)F(u_2)\cdots F(u_r) \in \{0,1\}^{Nr}$$

一个典型的例子是码字 $F(u_1)$，$F(u_2)$，$\cdots$，$F(u_r)$是相互独立的，并且构成 $F(u_i)$的(随机)符号 W_{i1}，$\cdots$，W_{iN}也是相互独立的。

考虑随机编码准则的原因是：

(1) 一个好的确定性码的存在通常是由于存在一个好的随机码；

(2) 随机码的计算一般比最优确定性码简单，因为离散最优化被关于概率分布的最优化替代了。

随机码的缺点在于它并不总是一一对应的(对于 $u \neq u'$，$F(u)$也可能与 $F(u')$相同)。然而，当 N 较大时，这样的情况发生的概率微不足道。

随机码的思想回归了 Shannon，正如在数学历史中经常发生的，一个聪明的主意解决了一个问题，但是同时也打开了一个关于其他问题的潘多拉魔盒。在这方面，随机码的余波引出了一个特殊问题，那就是寻找好的非随机码。现代信息与编码理论的一个主要部分就是围绕着这个问题，并且到目前为止没有发现一个普适的解决方案。然而，也实现了一

系列卓越的成果，在本书中将会讨论其中的部分问题。

继续随机码，随机码的编码准则 F 的误差概率期望如下：

$$E_r = \mathscr{E}\varepsilon(F) = \sum_f \varepsilon(f)\,\mathscr{P}(F=f) \tag{1.4.17}$$

定理 1.4.11

(1) 存在一个确定编码准则 f 使得 $\varepsilon(f) \leqslant E_r$。

(2) 对于任意 $\rho \in (0,1)$，$\mathscr{P}\left(\varepsilon(F) < \frac{E_r}{1-\rho}\right) \geqslant \rho$。

证明 (i) 很明显可知。对于(ii)，用 Chebyshev 不等式(参见 PSE I，75 页)：

68

$$\mathscr{P}\left(\varepsilon(F) \geqslant \frac{E_r}{1-\rho}\right) \leqslant \frac{1}{E_r}\frac{\rho E_r}{1-\rho} = 1-\rho \qquad \square$$

定义 1.4.12 对于随机信息 $\boldsymbol{X}^{(N)} = X_1 \cdots X_N$ 和 $\boldsymbol{Y}^{(N)} = Y_1 \cdots Y_N$，定义

$$C_N := \sup\left\{\frac{1}{N} I(\boldsymbol{X}^{(N)}; \boldsymbol{Y}^{(N)}),\ 对于输出概率分布\ P_{\boldsymbol{X}^{(N)}}\right\} \tag{1.4.18}$$

其中 $I(\boldsymbol{X}^{(N)}; \boldsymbol{Y}^{(N)})$ 是如下给定的互信息

$$h(\boldsymbol{X}^{(N)}) - h(\boldsymbol{X}^{(N)} \mid \boldsymbol{Y}^{(N)}) = h(\boldsymbol{Y}^{(N)}) - h(\boldsymbol{Y}^{(N)} \mid \boldsymbol{X}^{(N)})$$

备注 1.4.13 一个简单的启发式论证(在 2.2 节中将会给出严格证明)表明信道容量不可能超过其输入和输出信号的互信息。确实，对于每个典型的输入 N 序列，有

$$大约\ 2^{h(\boldsymbol{Y}^{(N)} \mid \boldsymbol{X}^{(N)})}\ 个可能的\ \boldsymbol{Y}^{(N)}\ 序列$$

所有的这些序列等可能出现。我们无法知道发送的是哪个序列 $\boldsymbol{X}$ 除非没有两个 $\boldsymbol{X}^{(N)}$ 序列产生相同的 $\boldsymbol{Y}^{(N)}$ 输出序列。典型 $\boldsymbol{Y}^{(N)}$ 序列的数目是 $2^{h(\boldsymbol{Y}^{(N)})}$。这个集合分出大小为 $2^{h(\boldsymbol{Y}^{(N)} \mid \boldsymbol{X}^{(N)})}$ 的子集，分别对应着不同的输入 $\boldsymbol{X}^{(N)}$。那么非联合的集合数

$$\leqslant 2^{h(\boldsymbol{Y}^{(N)}) - h(\boldsymbol{Y}^{(N)} \mid \boldsymbol{X}^{(N)})} = 2^{I(\boldsymbol{X}^{(N)}; \boldsymbol{Y}^{(N)})}$$

因此，长度为 N 的可区分的信号总数不能比 $2^{I(\boldsymbol{X}^{(N)}; \boldsymbol{Y}^{(N)})}$ 更大。这两个论证稍有不同。典型序列 $\boldsymbol{X}^{(N)}$ 的数量是 $2^{h(\boldsymbol{X}^{(N)})}$。然而仅有 $2^{h(\boldsymbol{X}^{(N)}, \boldsymbol{Y}^{(N)})}$ 个联合典型序列 $(\boldsymbol{X}^{(N)}, \boldsymbol{Y}^{(N)})$。所以随机抽取一堆序列为联合典型序列的概率为 $2^{-I(\boldsymbol{X}^{(N)}; \boldsymbol{Y}^{(N)})}$。所以可区分的信号数量的上界是 $2^{h(\boldsymbol{Y}^{(N)}) + h(\boldsymbol{X}^{(N)}) - h(\boldsymbol{X}^{(N)}, \boldsymbol{Y}^{(N)})}$。

定理 1.4.14 (Shannon SCT：相反部分)信道容量 C 遵循

$$C \leqslant \limsup_{N\to\infty} C_N \tag{1.4.19}$$

证明 考虑一个码 $f = f_N: \mathscr{U}_N \to \mathscr{X}_N \subseteq I^{\times N}$，其中 $\#\mathscr{U}_N = 2^{N(\overline{R}+o(1))}$，$\overline{R} \in (0,1)$。我们想要证明对于任何解码规则，都有

59

$$\varepsilon(f) \geqslant 1 - \frac{C_N + o(1)}{\overline{R} + o(1)} \tag{1.4.20}$$

定理的要求立即遵从了式(1.4.20)和信道容量的定义，因为

$$\liminf_{N\to\infty} \varepsilon(f) \geqslant 1 - \frac{1}{\overline{R}} \limsup_{N\to\infty} C_N$$

当 $\overline{R} > \limsup_{N\to\infty} C_N$ 时，上式 > 0。

让我们针对一一对应的 f(否则 $\varepsilon(f)$ 会更大)来验证式(1.4.20)。当序列 U 也是均匀分布的时候，码字 $\boldsymbol{X}^{(N)} = f(\boldsymbol{U})$ 也是均匀分布的，并且如果解码规则是 $\hat{f}: I^{\times N} \to \mathscr{X}$，那么当 N 足够大时，我们有，

$$
\begin{aligned}
NC_N &\geqslant I(\boldsymbol{X}^{(N)};\boldsymbol{Y}^{(N)}) \geqslant I(\boldsymbol{X}^{(N)};\hat{f}(\boldsymbol{Y}^{(N)}))(\text{参见定理 1.2.5}) \\
&= h(\boldsymbol{X}^{(N)}) - h(\boldsymbol{X}^{(N)} \mid \hat{f}(\boldsymbol{Y}^{(N)})) \\
&= \log r - h(\boldsymbol{X}^{(N)} \mid \hat{f}(\boldsymbol{Y}^{(N)}))(\text{通过等分布}) \\
&\geqslant \log r - \varepsilon(f)\log(r-1) - 1
\end{aligned}
$$

这里和下面都有 $r=\#\mathscr{X}$。最后的界限遵循广义 Fano 不等式(1.2.25)。确实，观察(随机)码字 $\boldsymbol{X}^{(N)}=f(\boldsymbol{U})$需要从码字集合$\mathscr{X}(=\mathscr{X}_N)$中取 r 个值 $x_1^{(N)}, \cdots, x_r^{(N)}$，错误概率为

$$
\varepsilon(f) = \sum_{i=1}^{r} \mathbb{P}(\boldsymbol{X}^{(N)} = x_i^{(N)}, \hat{f}(\boldsymbol{Y}^{(N)}) \neq x_i^{(N)})
$$

所以式(1.2.25)表明

$$
h(\boldsymbol{X}^{(N)} \mid \hat{f}(\boldsymbol{Y}^{(N)})) \leqslant h_2(\varepsilon) - \varepsilon\log(r-1) \leqslant 1 + \varepsilon(f)\log(r-1)
$$

并且我们得到 $NC_N \geqslant \log r - \varepsilon(f)\log(r-1) - 1$。最终可得，$r=2^{N(\overline{R}+o(1))}$，并且

$$
NC_N \geqslant N(\overline{R}+o(1)) - \varepsilon(f)\log(2^{N(\overline{R}+o(1))}-1)
$$

也就是

$$
\varepsilon(f) \geqslant \frac{N(\overline{R}+o(1)) - NC_N}{\log(2^{N(\overline{R}+o(1))}-1)} - 1 = \frac{C_N + o(1)}{\overline{R} - o(1)} \qquad \square
$$

用 $p(\boldsymbol{X}^{(N)}, \boldsymbol{Y}^{(N)})$作为分配的随机变量的输入和输出为随机码字 $\boldsymbol{X}^{(N)}$ 和 $\boldsymbol{Y}^{(N)}$ 的信道联合概率。为了简化，$p_X(\boldsymbol{X}^{(N)})$和 $p_Y(\boldsymbol{Y}^{(N)})$分别表示随机变量 $\boldsymbol{X}^{(N)}$和 $\boldsymbol{Y}^{(N)}$的边缘概率。 70

定理 1.4.15 (Shannon SCT：直接部分)假设我们可以找到一个恒定的常数 $c\in(0,1)$，以至于对于任意的 $\overline{R}\in(0,c)$和 $N\geqslant 1$ 都存在一个随机码 $F(\boldsymbol{u}_1), \cdots, F(\boldsymbol{u}_r)$，其中 $r=2^{N(\overline{R}+o(1))}$，码字 $F(\boldsymbol{u}_i)\in\mathscr{T}^{\wedge N}$是独立同分布的，那么随机输入输出信号的互信息

$$
\theta_N := \frac{1}{N}\log\frac{p(\boldsymbol{X}^{(N)},\boldsymbol{Y}^{(N)})}{p_X(\boldsymbol{X}^{(N)})\,p_Y(\boldsymbol{Y}^{(N)})} \tag{1.4.21}
$$

当 $N\to\infty$时依概率收敛于 c。那么信道容量 $C\geqslant c$。

定理 1.4.15 的证明在举例 1.4.24 和举例 1.4.25 之后给出(后者在技巧上更加复杂)。开始，我们解释 Shannon 在他 1948 年的论文中证明概述的策略。(这个想法变成一个正式的论证花了大概十年的时间。)

首先，生成一个随机码本$\mathscr{X}$。码本由 $r=2^{N\overline{R}}$ 个码字构成 $\boldsymbol{X}^{(N)}(1), \cdots, \boldsymbol{X}^{(N)}(r)$。假设收发双方都知道码字 $\boldsymbol{X}^{(N)}(1), \cdots, \boldsymbol{X}^{(N)}(r)$和信道转移概率矩阵 $\boldsymbol{P}_{\mathrm{ch}}(y|x)$。接着根据一个统一分布选择信息，对应的码字通过信道发送出去。接收用户使用最大似然(ML)准则来译码，也就是选择一个后验概率最大的信息。但是这道程序很难进行分析。我们可以换一种方法，使用一种次优但是直接的典型序列译码。如果仅有一个输入，那么信息 w 对应的码字和信道输出信号是联合型的，接收端宣布信息 w 被发送。如果没有此码字存在或者它并不是独一无二的，那么接收端提示一个错误。出人意料的是，这种程序是渐近最优的。最终可得，一个好的随机码本的存在暗示了一个好的非随机编码的存在。

换句话说，信道容量 C 不会比 c 值的上界小，c 是式(1.4.21)对于一个合适的随机编码依概率收敛的值。

推论 1.4.16 对于在定理 1.4.15 的假设中的 c，我们有

$$
\sup c \leqslant C \leqslant \limsup_{N\to\infty} C_N \tag{1.4.22}
$$

所以，如果式(1.4.22)的 LHS 和 RHS 重合，那么它们的共有值将给出信道容量。

接下来我们用 Shannon SCT 来计算一个 MBC 的容量。回想一下(参见式(1.4.2))，对于一个 MBC，

71

$$P(\boldsymbol{y}^{(N)}|\boldsymbol{x}^{(N)}) = \prod_{i=1}^{N} P(y_i|x_i) \tag{1.4.23}$$

定理 1.4.17 对于一个 MBC

$$I(\boldsymbol{X}^{(N)};\boldsymbol{Y}^{(N)}) \leqslant \sum_{j=1}^{N} I(X_j;Y_j) \tag{1.4.24}$$

如果输入符号 X_1，…，X_N 是互相独立的，那么等号成立。

证明 因为 $P(\boldsymbol{y}^{(N)}|\boldsymbol{x}^{(N)}) = \prod_{i=1}^{N} P(y_i|x_i)$，条件熵 $h(\boldsymbol{Y}^{(N)}|\boldsymbol{X}^{(N)})$ 等于和 $\sum_{j}^{N} h(Y_j|X_j)$。那么互信息为

$$I(\boldsymbol{X}^{(N)};\boldsymbol{Y}^{(N)}) = h(\boldsymbol{Y}^{(N)}) - h(\boldsymbol{Y}^{(N)}|\boldsymbol{X}^{(N)}) = h(\boldsymbol{Y}^{(N)}) - \sum_{1\leqslant j\leqslant N} h(Y_j|X_j)$$

$$\leqslant \sum_{j} (h(Y_j) - h(Y_j|X_j)) = \sum_{j} I(X_j;Y_j)$$

当且仅当 Y_1，…，Y_N 互相独立时等号成立。但是当 X_1，…，X_N 互相独立的时候 Y_1，…，Y_N 也是互相独立的。 □

备注 1.4.18 比较不等式(1.4.24)和(1.2.27)，注意这些界中相反的不等式。

定理 1.4.19 一个 MBC 的容量为

$$C = \sup_{p_{X_1}} I(X_1;Y_1) \tag{1.4.25}$$

上确界是在符号 X_1 所有可能的分布 p_{X_1} 上取得。

证明 根据 C_N 的定义，NC_N 不会超过

$$\sup_{p_X} I(\boldsymbol{X}^{(N)};\boldsymbol{Y}^{(N)}) \leqslant \sum_{j} \sup_{p_{X_j}} I(X_j;Y_j) = N \sup_{p_{X_1}} I(X_1;Y_1)$$

所以根据 Shannon SCT(相反部分)

$$C \leqslant \limsup_{N\to\infty} C_N \leqslant \sup_{p_{X_1}} I(X_1;Y_1)$$

另一方面，采取随机编码 F，码字为 $F(u_i) = V_{i1}\cdots V_{iN}$，$1\leqslant i\leqslant r$，包含独立同分布依概率 p^* 分布的符号 V_{ij}，概率分布 p^* 最大化了 $I(X_1;Y_1)$。(这样的随机编码定义适合任何的 r，也就是，对于任何 R(即使 $R>1$!)。)对这样的随机编码，随机互信息 Θ_N 等于

72

$$\frac{1}{N}\log\frac{p(\boldsymbol{X}^{(N)},\boldsymbol{Y}^{(N)})}{p_X(\boldsymbol{X}^{(N)})p_Y(\boldsymbol{Y}^{(N)})} = \frac{1}{N}\sum_{j=1}^{N}\log\frac{p(X_j,Y_j)}{p^*(X_j)p_Y(Y_j)} = \frac{1}{N}\sum_{j=1}^{N}\zeta_j$$

其中 $\zeta_j = \log\dfrac{p(X_j,Y_j)}{p^*(X_j)p_Y(Y_j)}$

随机变量 ζ_j 是独立同分布的，并且

$$E\zeta_j = E\log\frac{p(X_j,Y_j)}{p^*(X_j)p_Y(Y_j)} = I_{p^*}(X_1;Y_1)$$

根据针对独立同分布的随机变量的大数定理(见定理 1.3.5)，对于上面提到的随机编码

$$\Theta_N \xrightarrow{P} I_{p^*}(X_1;Y_1) = \sup_{p_{X_1}} I(X_1;Y_1)$$

根据 Shannon SCT(直接部分)

$$C \geqslant \sup_{p_{X_1}} I(X_1;Y_1)$$

因此，$C = \sup_{p_{X_1}} I(X_1;Y_1)$。 □

备注 1.4.20 (a) (X_1,Y_1) 对可以替换为任意 (X_j,Y_j)，$j\geqslant 1$。

(b) 回想 X_i 和 Y_i 的联合分布是通过 $\mathcal{P}(X_i=x, Y_i=y)=p_{X_i}(x)p(y|x)$ 定义的，其中 $(P(y|x))$ 是信道转移矩阵。

(c) 虽然我们要到对于每个 r（也就是对每个 $\overline{R}\geq 0$）上面的结构都成立，但是事实上仅限于 $\overline{R}\leq C$ 的情况。

例子 1.4.21 一位统计者对一个转移概率为 $P(y|x)$，信道容量 $C=\max_{p_X} I(X;Y)$ 的无记忆信道（MBC）进行预处理，他声明通过形成 $Y'=g(Y)$ 会严格地提高信道容量。他是对的吗？当然不是，因为预处理（或者篡改）不会提升信道容量。的确，

$$I(X;Y)=h(X)-h(X|Y)\geq h(X)-h(X|g(Y))=I(X;g(Y)) \tag{1.4.26}$$

在何种情况下他不能严格地降低信道容量？式(1.4.26)成立当且仅当分布 p_X 最大化 $I(X;Y)$，给出 $g(Y)$，随机变量 X 和 Y 条件独立时等号成立。（例如，$g(y_1)=g(y_2)$ 当且仅当对于任何 x，$P_{X|Y}(x|y_1)=P_{X|Y}(x|y_2)$；也就是说只有当条件概率 $P_{X|Y}(\cdot|y_1)$ 对应的 y 的值是相同的时，g 会融合在一起。）对一个 MBC 来说，当且仅当 g 是一一对应或者 $p=P(0|1)=P(1|0)=1/2$ 的时候，等号成立。 73

当信道是对称的（MBSC），式(1.4.25)可以进一步地简化，比如 $P(1|0)=P(0|1)=p$。更精确地，按照备注 1.4.5(a)（见式(1.4.11)），我们可以得到

定理 1.4.22 对于一个错误率为 p 的 MBSC，

$$C=1-h(p,1-p)=1-\eta(p) \tag{1.4.27}$$

见式(1.4.11)，0，1 等概分布，独立同分布的符号 V_i 的随机编码可以实现信道容量。

举例 1.4.23

(a) 考虑一个无记忆信道，A 和 B 是两个输入符号，输出为 A，B，$*$。假设每个输入符号被阻拦的概率为 1/2，传输为 $a*$ 的概率是 1/2。写出信道矩阵。

(b) 如果输出被一个不能区分 A 和 $*$ 的信宿处理，计算新信道的容量。信道转移矩阵为

$$\begin{bmatrix} 1 & 0 \\ 1/2 & 1/2 \end{bmatrix}$$

解答 (a) 信道转移矩阵为

$$\begin{bmatrix} 1/2 & 0 & 1/2 \\ 0 & 1/2 & 1/2 \end{bmatrix}$$

它是对称的（行与行之间是置换的关系。所以，$h(Y|X=x)=-2\times\frac{1}{2}\log\frac{1}{2}=1$ 并不取决于 $x=A$，B 的值。那么 $h(Y|X)=1$，并且

$$I(X;Y)=h(Y)-1 \tag{1.4.28}$$

如果 $\mathcal{P}(X=A)=\alpha$，那么 Y 的输出分布为

$$\left(\frac{1}{2}\alpha,\frac{1}{2}(1-\alpha),\frac{1}{2}\right)$$

$h(Y|X)$ 在 $\alpha=1/2$ 时取得最大值。那么容量等于

$$h(1/4,1/4,1/2)-1=\frac{1}{2} \tag{1.4.29}$$

(b) 此处，信道是不对称的。如果 $\mathcal{P}(X=A)=\alpha$，那么条件熵可分解为

$$h(Y|X)=\alpha h(Y|X=A)+(1-\alpha)h(Y|X=B)=\alpha\times 0+(1-\alpha)\times 1=(1-\alpha)$$ 74

那么

$$h(Y)=-\frac{1+\alpha}{2}\log\frac{1+\alpha}{2}-\frac{1-\alpha}{2}\log\frac{1-\alpha}{2}$$

且

$$I(X;Y)=-\frac{1+\alpha}{2}\log\frac{1+\alpha}{2}-\frac{1-\alpha}{2}\log\frac{1-\alpha}{2}-1+\alpha$$

当 $\alpha=3/5$ 时取最大值，信道容量为

$$(\log 5)-2=0.321\,928 \qquad \square$$

我们的下一个目标是证明 Shannon 的 SCT(定理 1.4.15)的直接的部分。先前声明过，证明是基于以下两个实例的。

举例 1.4.24 令 F 为一个随机编码，独立于信源 U，那么码字 $F(\boldsymbol{u}_1)$，…，$F(\boldsymbol{u}_r)$是独立同分布的，概率分布是 p_F：

$$p_F(\boldsymbol{v})=\mathbb{P}(F(\boldsymbol{u})=\boldsymbol{v}),\quad \boldsymbol{v}(=\boldsymbol{v}^{(N)})\in J^{\times N}$$

这里，$\boldsymbol{u}_j$，$j=1$，…，r，是信源序列，$r=2^{N(\mathbb{H}+\alpha(\varepsilon))}$。定义随机码字 $\boldsymbol{V}_1$，…，$\boldsymbol{V}_{r-1}$

$$\text{如果 } \boldsymbol{U}=\boldsymbol{u}_i,\text{那么 } \boldsymbol{V}_{i-1}=F(\boldsymbol{u}_i)\qquad i<j$$

$$\text{且 } \boldsymbol{V}_{i-1}=F(\boldsymbol{u}_{i+1})\quad i\geqslant j,\quad 1\leqslant j\leqslant r,\quad 1\leqslant i\leqslant r-1 \tag{1.4.30}$$

$\boldsymbol{U}$(消息序列)，$\boldsymbol{X}=F(\boldsymbol{U})$(随机码字)和 $\boldsymbol{V}_1$，…，$\boldsymbol{V}_{r-1}$是独立的码，$\boldsymbol{X}$ 的每一个 $\boldsymbol{V}_1$，…，$\boldsymbol{V}_{r-1}$有以下分布 p_F。

解答 证明是直接的，符合下面的联合概率

$$\mathbb{P}(\boldsymbol{U}=\boldsymbol{u}_j,\boldsymbol{X}=\boldsymbol{x},\boldsymbol{V}_1=\boldsymbol{v}_1,\cdots,\boldsymbol{V}_{r-1}=\boldsymbol{v}_{r-1})=\mathbb{P}(\boldsymbol{U}=\boldsymbol{u}_j)p_F(\boldsymbol{x})p_F(\boldsymbol{v}_1)\cdots p_F(\boldsymbol{v}_{r-1}) \quad \square \tag{1.4.31}$$

举例 1.4.25 对于任意 $\kappa>0$，检验举例 1.4.24 中的随机编码

$$E-\mathbb{E}\varepsilon(F)\leqslant\mathbb{P}(\Theta_N\leqslant\kappa)+r2^{-N\kappa} \tag{1.4.32}$$

75 这里，将式(1.4.21)中的随机变量定义为 Θ_N，且 $\mathbb{E}\Theta_N=\frac{1}{N}I(\boldsymbol{X}^{(N)},\boldsymbol{Y}^{(N)})$。

解答 对于给定码 $\boldsymbol{x}(=\boldsymbol{x}^{(N)})$和 $\boldsymbol{y}(=\boldsymbol{y}^{(N)})\in J^{\times N}$，证

$$S_r(\boldsymbol{x}):=\{\boldsymbol{x}'\in J^{\times N};\boldsymbol{P}(\boldsymbol{y}|\boldsymbol{x}')\geqslant\boldsymbol{P}(\boldsymbol{y}|\boldsymbol{x})\} \tag{1.4.33}$$

也就是说，当 $\boldsymbol{x}$ 传输，$\boldsymbol{y}$ 接收时，$S_r(\boldsymbol{x})$包括了 ML 译码器可能产生的所有码，对于一个给定的非随机编码规则 f 和一个信源序列 $\boldsymbol{u}$，对于一些 $\boldsymbol{u}'\neq\boldsymbol{u}$，如果 $f(\boldsymbol{u}')\in S_r(f(\boldsymbol{u}))$，那么 $\delta(f,\boldsymbol{u},\boldsymbol{y})=1$，其他情况 $\delta(f,\boldsymbol{u},\boldsymbol{y})=0$。清楚地看到，$\delta(f,\boldsymbol{u},\boldsymbol{y})$等于

$$1-\prod_{\boldsymbol{u}'\neq\boldsymbol{u}}\mathbf{1}(f(\boldsymbol{u}')\notin S_r(f(\boldsymbol{u})))=1-\prod_{\boldsymbol{u}'\neq\boldsymbol{u}}[1-\mathbf{1}(f(\boldsymbol{u}')\in S_r(f(\boldsymbol{u})))]$$

对于所有的非随机编码 f，$\varepsilon(f)\leqslant\mathbb{E}\delta(f,\boldsymbol{U},\boldsymbol{Y})$，对于所有随机编码 F，$E-\mathbb{E}\varepsilon(F)\leqslant\mathbb{E}\delta(F,\boldsymbol{U},\boldsymbol{Y})$，进一步，对于举例 1.4.24 中的随机编码，期望值 $\mathbb{E}\delta(F,\boldsymbol{U},\boldsymbol{Y})$不超过

$$\mathbb{E}\Big(1-\prod_{i=1}^{r-1}[1-\mathbf{1}(\boldsymbol{V}_i\in S_r(\boldsymbol{X}))]\Big)=\sum_{\boldsymbol{x}}p_{\boldsymbol{X}}(\boldsymbol{x})\sum_{\boldsymbol{y}}\boldsymbol{P}(\boldsymbol{y}|\boldsymbol{x})$$

$$\times\mathbb{E}\Big[\Big(1-\prod_{i=1}^{r-1}[1-\mathbf{1}(\boldsymbol{V}_i\in S_r(\boldsymbol{X}))]\Big)\Big|\boldsymbol{X}=\boldsymbol{x},\boldsymbol{Y}=\boldsymbol{y}\Big]$$

根据独立性，上式等价于

$$\sum_{\boldsymbol{x}}p_{\boldsymbol{X}}(\boldsymbol{x})\sum_{\boldsymbol{y}}\boldsymbol{P}(\boldsymbol{y}|\boldsymbol{x})\Big(1-\prod_{i=1}^{r-1}\mathbb{E}(1-\mathbf{1}_{\{\boldsymbol{v}_i\in S_r(\boldsymbol{x})\}})\Big)$$

进一步地，由独立同分布的性质(同举例 1.4.24 解释的一样)

$$\prod_{i=1}^{r-1}\mathbb{E}(1-\mathbf{1}_{\{\boldsymbol{v}_i\in S_r(\boldsymbol{x})\}})=(1-Q_r(\boldsymbol{x}))^{r-1}$$

其中

$$Q_y(x) := \sum_{x'} 1(x' \in S_y(x)) p_X(x')$$

所以，期望错误率为 $E \leqslant 1 - \mathbb{E}(1-Q_Y(X))^{r-1}$。

用 $\mathbb{T}=\mathbb{T}(\kappa)$ 标记，码字 x，y 对的集合

$$\Theta_N = \frac{1}{N}\log\frac{p(x,y)}{p_X(x)p_Y(y)} > \kappa$$

利用等式

$$1-(1-Q_y(x))^{r-1} = \sum_{j=0}^{r-1}(1-Q_y(x))^j Q_y(x) \tag{1.4.34}$$

76

接下来观察

$$1-(1-Q_y(x))^{r-1} \leqslant 1, \quad 当 (x,y) \notin \mathbb{T} \tag{1.4.35}$$

当 $(x, y) \in \mathbb{T}$ 时

$$\begin{aligned}(1-(1-Q_y(x))^r) &= \sum_{j=1}^{r-1}((1-Q_y(x))^j)Q_y(x) \\ &\leqslant (r-1)Q_y(x)\end{aligned}$$

得到

$$E \leqslant \mathbb{P}((X,Y) \notin \mathbb{T}) + (r-1)\sum_{(x,y)\in\mathbb{T}} p_X(x)P(y|x)Q_y(x) \tag{1.4.36}$$

现在观察

$$\mathbb{P}((X,Y) \notin \mathbb{T}) = \mathbb{P}(\Theta_N \leqslant \kappa) \tag{1.4.37}$$

最后，对于 $(x, y) \in \mathbb{T}$ 和 $x' \in S_y(x)$

$$P(y|x') \geqslant P(y|x) \geqslant p_Y(y)2^{N\kappa}$$

两边同乘 $\frac{p_X(x')}{p_Y(y)}$，得到 $\mathbb{P}(X=x'|Y=y) \geqslant p_X(x')2^{N\kappa}$。两边同加 $x' \in S_y(x)$，得到 $1 \geqslant \mathbb{P}(S_Y(x)|Y=y) \geqslant Q_y(x)2^{N\kappa}$，或者

$$Q_y(x) \leqslant 2^{-N\kappa} \tag{1.4.38}$$

把式(1.4.37)和式(1.4.38)代入式(1.4.36)得到式(1.4.32)。 □

证明　定理 1.4.15 的证明现在可以容易地完成。$\overline{R}=c-2\varepsilon$，$\kappa=c-\varepsilon$。因为 $r=2^{N(\overline{R}+\varepsilon)}$，$E$ 不会超过下式

$$\mathbb{P}(\Theta_N \leqslant c-\varepsilon) + 2^{N(\overline{R}-c+\varepsilon)} = \mathbb{P}(\Theta_N \leqslant c-\varepsilon) + 2^{-N\varepsilon}$$

当 $N\to\infty$ 时，等式趋于 0，因为在条件 $\Theta_N \xrightarrow{\mathscr{P}} c$ 下，$\mathbb{P}(\Theta_N \leqslant c-\varepsilon) \to 0$。所以，随机编码 F 当 $N\to\infty$ 时，期望错误率趋于 0。

由于定理 1.4.11(i)，对于任何 $N \geqslant 1$，存在一个确定编码 $f=f_N$，比如 $\overline{R}=c-2\varepsilon$，$\lim_{N\to\infty}\varepsilon(f)=0$。所以，$\overline{R}$ 是一个可靠传输速率。对于任何 $\varepsilon>0$ 都成立，因此 $C \geqslant c$。 □

证明中的论证形式是 P. Whittle 提出的，且呈现在文献[52]114～117 页中。我们感谢 C. Goldie 提供的信息。另一种方法是基于联合典型性的概念；这个办法在 2.2 节中用过，我们讨论了连续分布噪声的信道。 77

定理 1.4.17 和定理 1.4.19 可以扩展到一个无记忆信道，此信道由任意有限个输出符号集 $J_q=\{0, \cdots, q-1\}$ 构成。也就是说，信道的输入端，有一个码 $Y^{(N)}=Y_1\cdots Y_N$，其中的 Y_j 随机地从 J_q 中取值。所以无记忆的性质为

$$P_{\mathbb{P}}(y^{(N)}|x^{(N)}) = \prod_{i=1}^{N} P(y_i|x_i) \tag{1.4.39}$$

一个点对点的信道概率 $P(y|x)$ 组成一个 $2\times q$ 的随机矩阵。一个无记忆的信道是对称的，当矩阵的行是其他行的置换，当列也是其他列的置换时，矩阵是双重对称的。可靠传输速率和信道容量的定义是固定的。无记忆二进制信道的容量在图 1.8 中展示。

定理 1.4.26 一个无记忆的对称信道的容量为

$$C \leqslant \log q - h(p_0, \cdots, p_{q-1}) \tag{1.4.40}$$

其中 $(p_0, \cdots, p_{q-1})$ 是矩阵的行。双重堆成信道可以实现相等性，最大化随机编码有独立从 J_q 取值的同分布的符号 V_i，其概率为 $1/q$。

证明 证明采用二进制情况，利用 $I(X_1;Y_1)=h(Y_1)-h(Y_1|X_1)\leqslant\log q-h(Y_1|X_1)$。但是在对称的情形中

$$\begin{aligned} h(Y_1|X_1) &= -\sum_{x,y}\mathbb{P}(X_1=x)P(y|x)\log P(y|x) \\ &= -\sum_{x}\mathbb{P}(X_1=x)\sum_{y}p_y\log p_y = h(p_0,\cdots,p_{q-1}) \end{aligned} \tag{1.4.41}$$

而且，如果矩阵的列是其他列的置换，$h(Y_1)$ 达到 $\log q$。实际上，采用随机编码 $\mathbb{P}(Y=y)=\sum_{x=0}^{q-1}\mathbb{P}(X_1=x)P(y|x)=\frac{1}{q}\sum_{x}P(y|x)$，$\sum_{x}P(y|x)$ 取自信道矩阵的列，它不取决于 y。所以，$\mathbb{P}(Y=y)$ 不取决于均匀分布 $y\in J_q$。 □

备注 1.4.27 (a) 举例 1.4.24、1.4.25 和定理 1.4.6、1.4.15、1.4.17 采用了随机编码 F，当 $N\to\infty$ 时，$E\to 0$。这不仅保证了一个优良的非随机编码——当 $N\to\infty$ 时，误码率 E 趋于 0(参见定理 1.4.11(i))，而且几乎所有的编码都是渐近最优的。实际上，由于定理 1.4.11(ii) $\rho=1-\sqrt{E}$，$\mathbb{P}(\varepsilon(F)<\sqrt{E})\geqslant 1-\sqrt{E}\to 1$，当 $N\to\infty$ 时。但是，这对找到一个好的编码没有帮助：构造一个好的编码是信息论中具有挑战性的工作，我们将会在后面回到这个问题。

78

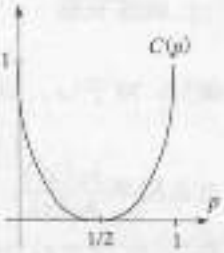


图 1-8

举例 1.4.28 比特通过信道传输，一个概率为 μ 的比特可能会被改变，一个概率为 λ 的比特可能会被忽略。连续比特是独立的。计算信道的最优编码和容量。

解答 信道矩阵是 2×3 的

$$\Pi = \begin{pmatrix} 1-\lambda-\mu & \lambda & \mu \\ \lambda & 1-\lambda-\mu & \mu \end{pmatrix}$$

行是其他行的置换，所以有相同的熵。因此，条件熵 $h(Y|X)$ 等于

$$h(1-\lambda-\mu,\lambda,\mu) = -(1-\lambda-\mu)\log(1-\lambda-\mu)-\lambda\log\lambda-\mu\log\mu$$

这与输入符号 X 的分布无关。

所以，当 $h(Y)$ 最大化时，$I(X;Y)$ 也最大化。如果 $p_Y(0)=p$，$p_Y(1)=q$，那么

79

$$h(Y) = -\mu\log\mu - p\log p - q\log q$$

上式在 $p=q=(1-\mu)/2$(通过注水法)时是最大的，即 $p_X(0)=p_X(1)=1/2$ 时。容量用下面的表达式表示：

$$\begin{aligned} &-(1-\mu)\log\frac{1-\mu}{2}+(1-\lambda-\mu)\log(1-\lambda-\mu)+\lambda\log\lambda \\ &= (1-\mu)\left(1-h\left(\frac{1-\lambda-\mu}{1-\mu},\frac{\lambda}{1-\mu}\right)\right) \end{aligned}$$

□

举例 1.4.29 (a)(数据处理定理)考虑到两个连续的独立信道。在信道 1 中传输一个随机变量 X，在 Y 中接收。接着在信道 2 中传输并在 Z 中接收，证明

$$I(X;Z) \leqslant I(X;Y)$$

所以在第二个信道中的进一步处理只能减少相互的信息量。

信道独立意味着，给定 Y，随机变量 X 和 Z 条件独立，可以推断出

$$h(X,Z|Y) = h(X|Y) + h(Z|Y)$$

以及

$$h(X,Y,Z) + h(Z) = h(X,Z) + h(Y,Z)$$

定义 $I(X;Z|Y)$ 为 $h(X|Y)+h(Z|Y)-h(X,Z|Y)$，可以看到

$$I(X;Z|Y) = I(X;Y) - I(X;Z)$$

那么在数据处理不等式中，不等式 $I(X;Z) \neq I(X;Y)$ 成立吗？

(b) 离散时间信道的输入输出在符号集中表示出来，这个字符集的字母是整数除上确定 r 的余数。传输的字母 $[x]$ 以概率 p_j 作为 $[j+x]$ 被接收，在这里 x 和 j 以及 $[c]$ 表示为 c 除 r 的余数。计算出信道容量。

解答 (a) 给定 Y，随机变量 X 和 Z 是条件独立的。因此

$$h(X|Y) = h(X|Y,Z) \leqslant h(X|Z)$$

并且

$$I(X;Y) = h(X) - h(X|Y) \geqslant h(X) - h(X|Z) = I(X;Z)$$ 80

给定 Z，当且仅当 X 和 Y 条件独立时，等式是成立的，即如果第二信道是无误差的 $(Y,Z) \mapsto Z$ 是一一对应的或者说第一信道是安全带噪的，即 X，Y 是独立的。

(b) 信道矩阵的每行都是其他行的置换，因此 $h(Y|X)=h(p_0,\cdots,p_{r-1})$ 并不依赖于 p_X。当 $p_X(i)=1/r$ 时，$h(Y)$ 最大，这时，

$$C = \log r - h(p_0,\cdots,p_{r-1})$$ □

举例 1.4.30 求有 n 个完全相同的级联独立二进制对称信道(MBSC)的错误率，每一级的错误率是 $0<p<1$(如图 1-9 所示)。

图 1-9

证明级联容量在 $n\to\infty$ 时趋向于 0。

解答 有着 n 个级联信道的信道参数是 Π^n，这里

$$\Pi = \begin{bmatrix} 1-p & p \\ p & 1-p \end{bmatrix}$$

计算特征向量/值导出

$$\Pi^n = \frac{1}{2}\begin{bmatrix} 1+(1-2p)^n & 1-(1-2p)^n \\ 1-(1-2p)^n & 1+(1-2p)^n \end{bmatrix}$$

它的误差概率是 $1/2(1-(1-2p)^n)$。如果 $0<p<1$，Π^n 收敛于

$$\begin{bmatrix} 1/2 & 1/2 \\ 1/2 & 1/2 \end{bmatrix}$$

并且信道容量达到

$$1 - h(1/2,1/2) = 1 - 1 = 0$$

如果 $p=0$ 或者 1，那么信道是无差错的，且 $C=1$。 □ 81

举例 1.4.31 考虑两个独立 MBC，它们的容量是 C_1，C_2。证明或者提供一个相类似的例子，使下面每一个关于混合信道容量 C 的要求成为规定。

(a) 如果信道是连续的，并且信道的输出并没有进行任何编码成进入另一个信道，那么 $C=\min[C_1, C_2]$。

(b) 假设信道是并行使用的，换句话说就是在每一秒中，一个符号(来自于它的输入字母中)在信道 1 中传输，下一个符号在信道 2 中传输；因此每个信道每秒发送 1 个符号，那么 $C=C_1+C_2$。

(c) 如果信道拥有同样的输入符号集，而且在每一秒中，都会选择一个符号，并且同时在两个信道上发送，那么 $C=\max[C_1, C_2]$。

(d) 如果信道 $i=1, 2$ 的矩阵是 Π_i，复合矩阵有

$$\Pi=\begin{pmatrix}\Pi_1 & 0\\ 0 & \Pi_2\end{pmatrix}$$

那么 C 就由 $2^C=2^{C_1}+2^{C_2}$ 给定。对于哪种操作的模式，这是一致的？

解答 (a)

$$\xrightarrow{X}\boxed{\text{信道1}}\xrightarrow{Y}\boxed{\text{信道2}}\xrightarrow{Z}$$

就像在举例 1.4.29a 中一样，

$$I(X;Z)\leqslant I(X;Y),\quad I(X;Z)\leqslant I(Y;Z)$$

因此

$$C=\sup_{p_X} I(X;Z)\leqslant \sup_{p_X} I(X;Y)=C_1$$

类似地

$$C=\sup_{p_Y} I(Y;Z)=C_2$$

即 $C\leqslant\min[C_1, C_2]$，但许会出现一个严格不等式：取 $\delta\in(0, 1/2)$，并且矩阵

$$\mathrm{ch1}\sim\begin{pmatrix}1-\delta & \delta\\ \delta & 1-\delta\end{pmatrix},\mathrm{ch2}\sim\begin{pmatrix}1-\delta & \delta\\ \delta & 1-\delta\end{pmatrix}$$

以及

$$\mathrm{ch}[1+2]\sim\frac{1}{2}\begin{pmatrix}(1-\delta)^2+\delta^2 & 2\delta(1-\delta)\\ 2\delta(1-\delta) & (1-\delta)^2+\delta^2\end{pmatrix}$$

62

在这里，$1/2>2\delta(1-\delta)>\delta$，

$$C_1=C_2=1-h(\delta,1-\delta)$$

以及

$$C=1-h(2\delta(1-\delta),1-2\delta(1-\delta))<C_1$$

这是因为 $h(\varepsilon, 1-\varepsilon)$ 在 $\varepsilon\in[0, 1/2]$ 上严格递增。

(b)

$$X_1\longrightarrow\boxed{\text{信道1}}\longrightarrow Y_1$$

$$X_2\longrightarrow\boxed{\text{信道2}}\longrightarrow Y_2$$

联合信道容量是

$$C = \sup_{p(x_1,x_2)} I((X_1,X_2);(Y_1,Y_2))$$

但是

$$\begin{aligned} I((X_1,X_2);(Y_1,Y_2)) &= h(Y_1,Y_2) - h(Y_1,Y_2 \mid X_1,X_2) \\ &\leqslant h(Y_1) + h(Y_2) - h(Y_1 \mid X_1) - h(Y_2 \mid X_2) \\ &= I(X_1;Y_1) + I(X_2;Y_2) \end{aligned}$$

当且仅当 X_1，X_2 相互独立时，等号成立。因此 $C=C_1+C_2$，并且 $p_{(X_1,X_2)}$ 的最大值是 $p_{X_1} \times p_{X_2}$，这里，p_{X_1} 和 p_{X_2} 都是 $I(X_1;Y_1)$ 和 $I(X_2;Y_2)$ 最大时的值。

(c)

在这里，

$$C = \sup_{p_X} I(X;(Y_1;Y_2))$$

并且

$$\begin{aligned} I((Y_1;Y_2);X) &= h(X) - h(X \mid Y_1,Y_2) \\ &\geqslant h(X) - \min_{i=1,2} h(X \mid Y_i) = \min_{i=1,2} I(X;Y_i) \end{aligned}$$

83

所以，$C \geqslant \max[C_1, C_2]$。可能产生一个严格的不等式：参照在(a)中举的例子，这里，$C_i = 1-h(\delta, 1-\delta)$。并且，

$$\begin{aligned} I((Y_1;Y_2);X) &= h(Y_1,Y_2) - h(Y_1,Y_2 \mid X) \\ &= h(Y_1,Y_2) - h(Y_1 \mid X) - h(Y_2 \mid X) \\ &= h(Y_1,Y_2) - 2h(\delta,1-\delta) \end{aligned}$$

如果我们令 $p_X(0)=p_X(1)=1/2$，那么

$(Y_1,Y_2)=(0,0)$ 的概率为 $[(1-\delta)^2+\delta^2]/2$

$(Y_1,Y_2)=(1,1)$ 的概率为 $[(1-\delta)^2+\delta^2]/2$

$(Y_1,Y_2)=(1,0)$ 的概率为 $\delta(1-\delta)$

$(Y_1,Y_2)=(0,1)$ 的概率为 $\delta(1-\delta)$

且

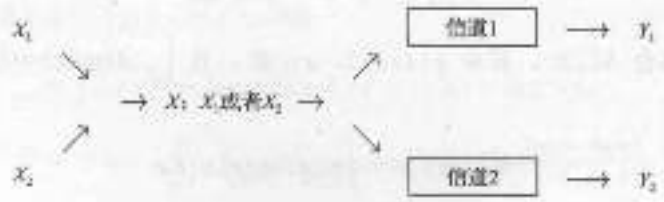
$$h(Y_1,Y_2) = 1 + h(2\delta(1-\delta), 1-2\delta(1-\delta))$$

并且

$$\begin{aligned} I((Y_1,Y_2);X) &= 1 + h(2\delta(1-\delta), 1-2\delta(1-\delta)) - 2h(\delta,1-\delta) \\ &> 1 - h(\delta,1-\delta) = C_i \end{aligned}$$

所以，$C>C_i$，$i=1,2$。

(d)



它与(c)的不同之处在于每一秒只有一个符号被发送到信道1或信道2。如果一个给定的符号通过某个特殊的通道发送，我们固定其概率为 α 和 $1-\alpha$，那么：

$$I(X;Y)=h(\alpha,1-\alpha)+\alpha I(X_1;Y_1)+(1-\alpha)I(X_2;Y_2) \tag{1.4.42}$$

实际上，$I(X;Y)=h(Y)-h(Y|X)$，其中

$$\begin{aligned}h(Y)&=-\sum_y \alpha p_{Y_1}(y)\log\alpha p_{Y_1}(y)-\sum_y(1-\alpha)p_{Y_2}(y)\log(1-\alpha)p_{Y_2}(y)\\&=-\alpha\log\alpha-(1-\alpha)\log(1-\alpha)+\alpha h(Y_1)+(1-\alpha)h(Y_2)\end{aligned}$$

84

且

$$\begin{aligned}h(Y|X)&=-\sum_{x,y}\alpha p_{X_1Y_1}(x,y)\log p_{Y_1|X_1}(y|x)-\sum_{x,y}(1-\alpha)p_{X_2,Y_2}(y|x)\log p_{Y_2|X_2}(y|x)\\&=\alpha h(Y_1|X_1)+(1-\alpha)h(Y_2|X_2)\end{aligned}$$

证明了式(1.4.42)。这就产生了

$$C=\max_{0\leqslant\alpha\leqslant1}[h(\alpha,1-\alpha)+\alpha C_1+(1-\alpha)C_2]$$

下式给出了最大值

$$\alpha=2^{C_1}/(2^{C_1}+2^{C_2}),1-\alpha=2^{C_2}/(2^{C_1}+2^{C_2})$$

且 $C=\log(2^{C_1}+2^{C_2})$。 □

举例 1.4.32 一名间谍按下述方式与他的联系人通信。每个小时中，要么他不打电话，要么他打电话并只允许电话响一定的次数——不多于 N 次，因为害怕被侦察。他的联系人不接听，只记录他的电话是否响起，并记录响铃次数。由于电话系统的不足，来电也许不能实时连接；能够正常连接的概率为 p，其中 $0<p<1$，且在不同通话中是独立的。但是间谍不知道哪个电话连通了。如果连接成功，那么电话号码就准确地传送出去。没有打电话时，间谍与另一个用户连接失败的概率可以忽略不计。写出这个信道的信道矩阵，且详细地计算出它的信道容量。在 N 的条件下确定 p，在最优编码的同时，使间谍总能够接通电话。

解答 信道的字母表为 $\{0,1,\cdots,N\}$，其中 0～未响铃(一个小时内)，且 $j\geqslant1\sim j$，信道矩阵是 $P(0|0)=1$，$P(0|j)=1-p$，$P(j|j)=p$，$1\leqslant j\leqslant N$，而且 $h(Y|X)=-q(p\log p+(1-p)\log(1-p))$，其中 $q=p_X(X\geqslant1)$。此外，当下式成立时，给出的 q 使 $h(Y)$ 达到最大值

$$p_Y(0)=1-pq,p_Y(k)=\frac{pq}{N},1\leqslant k\leqslant N$$

使 $I(X;Y)=h(Y)-h(Y|X)$ 取最大值，则 q 引出 $p(1-p)^{(1-p)/p}\times(1-pq)=pq/N$，或者

85

$$q=\min\left(\frac{1}{p}\left(1+\frac{1}{Np}\left(\frac{1}{1-p}\right)^{(1-p)/p}\right)^{-1},1\right)$$

条件 $q=1$ 等价于 $\log N\geqslant-\frac{1}{p}\log(1-p)$，即 $N\geqslant\frac{1}{(1-p)^{1/p}}$。 □

1.5 微分熵及其性质

定义 1.5.1 假设随机变量 X 存在概率密度(PDF) $p(x)$，$x\in\mathbb{R}^n$：

$$\mathbf{P}\{X\in A\}=\int_A p(x)\mathrm{d}x$$

对于任何(可测)集合 $A\subseteq\mathbb{R}^n$，其中 $p(x)\geqslant0$，$x\in\mathbb{R}^n$，且 $\int_{\mathbb{R}^n}\mathrm{d}xp(x)=1$。微分熵 $h_{\mathrm{diff}}(X)$ 定义为

$$h_{\mathrm{diff}}(X)=-\int p(x)\log p(x)\mathrm{d}x \tag{1.5.1}$$

假设积分绝对可积。在离散的情况中，$h_{\mathrm{diff}}(X)$可被认为是概率密度的一个函数 p：$x\in\mathbb{R}^n\mapsto\mathbb{R}_+=[0,\infty)$。区别在于 $h_{\mathrm{diff}}(X)$可以为负，例如对于在$[0,a]$上的均匀分布，其中 $a<1$。（当 $x\in\mathbb{R}$ 时，我们将 x 写成 x_n。）微分熵的相对熵、联合熵和条件微分熵都可以类似于离散的情况进行定义：

$$h_{\mathrm{diff}}(X\|Y)=D_{\mathrm{diff}}(p\|p')=-\int p(x)\log\frac{p'(x)}{p(x)}\mathrm{d}x \tag{1.5.2}$$

$$h_{\mathrm{diff}}(X,Y)=-\int p_{X,Y}(x,y)\log p_{X,Y}(x,y)\mathrm{d}x\mathrm{d}y \tag{1.5.3}$$

$$h_{\mathrm{diff}}(X|Y)=-\int p_{X,Y}(x,y)\log p_{X|Y}(x|y)\mathrm{d}x\mathrm{d}y=h_{\mathrm{diff}}(X,Y)-h_{\mathrm{diff}}(Y) \tag{1.5.4}$$

同样假设积分绝对可积。在这里，$p_{X,Y}$是联合概率密度，$p_{X|Y}$是条件概率密度（条件分布的 PDF），所以，当我们清楚在讨论哪种熵时我们会忽略下标 diff，定理 1.2.3(b)(c)、1.2.12 和 1.2.18 同样适用于微分熵，证明完全类似，在这里就不再重复。

备注 1.5.2 令 $0\leqslant x\leqslant 1$，那么 x 可以被写成和 $\sum_{n\geqslant 1}a_n2^{-n}$，其中 $a_n(=a_n(x))$等于 0 或 1。对于“大多数”x 来说序列不会收敛到有限和（也就是说，存在无数多个 n 使得 $a_n=1$；正规的说法是对于 $x\in(0,1)$使无数的 $a_n(x)=1$ 的集合，它的(Lebesgue)测度等于 1）。因此，如果我们想通过二进制数字去“解码”x，我们需要无限长的码字。换句话说，对于在 $0\leqslant X\leqslant 1$ 上均匀分布的变量 X，它的一个典型值需要无限多比特对它进行“精确”描述。在一般情况下当存在 PDF $f_X(x)$我们很容易得到相同结论。 86

然而，如果我们希望精确地用前 n 个二进制数字代表随机变量 X 的输出，那么我们平均需要 $n+h(X)$比特，其中 $h(X)$是 X 的微分熵。微分熵可为正可为负，甚至可以是$-\infty$。由于 $h(X)$可以取正取负，$n+h(X)$可以大于或小于 n。在离散情况下，熵具有平移和伸缩不变性，因为它只取决于概率 $p_1,\cdots,p_m$，不取决于随机变量的值。然而，微分熵具有平移不变性但不具有伸缩不变性，从下面恒等式中可以看出来（参见定理 1.5.7）

$$h(aX+b)=h(X)+\log|a|$$

然而相对熵，即 Kullback-Leibler 距离 $D(p\|q)$具有伸缩不变性。

举例 1.5.3 考虑在 $0\leqslant x\leqslant e^{-1}$上的一个 PDF，有

$$f_r(x)=C_r\frac{1}{x(-\ln x)^{r+1}},\quad 0<r<1$$

那么微分熵 $h(X)=-\infty$。

解答 经 $y=-\ln x$ 替换后，我们得到

$$\int_0^{e^{-1}}\frac{1}{x(-\ln x)^{r+1}}\mathrm{d}x=\int_1^{\infty}\frac{1}{y^{r+1}}\mathrm{d}y=\frac{1}{r}$$

所以，$C_r=r$。进一步，利用 $z=\ln(-\ln x)$

$$\int_0^{e^{-1}}\frac{\ln(-\ln x)}{x(-\ln x)^{r+1}}\mathrm{d}x=\int_0^{\infty}ze^{-rz}\mathrm{d}z=\frac{1}{r^2}$$

所以，

$$\begin{aligned}h(X)&=-\int f_r(x)\ln f_r(x)\mathrm{d}x\\&=-\int f_r(x)(-\ln r+\ln x+(r+1)\ln(-\ln x))\mathrm{d}x\\&=-\ln r-\int_0^{e^{-1}}\left(\frac{r}{x(-\ln x)^r}-r(r+1)\frac{\ln(-\ln x)}{x(-\ln x)^{r+1}}\right)\mathrm{d}x\end{aligned}$$

[87] 对于$0<r<1$，第二项是无穷的，其他两项是有限的。 □

定理 1.5.4 $X=(X_1, \cdots, X_d)\sim N(\mu, C)$是多变量正态分布向量，均值$\mu=(\mu_1, \cdots, \mu_d)$，协方差矩阵$C=(c_{ij})$，即$\mathbb{E}X_i=\mu_i$，$\mathbb{E}(X_i-\mu_i)(X_j-\mu_j)=c_{ij}=c_{ji}$，$1\leqslant i, j\leqslant d$。那么

$$h(X)=\frac{1}{2}\log[(2\pi e)^d\det C] \tag{1.5.5}$$

证明 PDF $p_X(x)$是

$$p(x)=\frac{1}{((2\pi)^d\det C)^{1/2}}\exp\left(-\frac{1}{2}(x-\mu,C^{-1}(x-\mu))\right), x\in\mathbb{R}^d$$

然后$h(X)$有如下形式

$$\begin{aligned}
&\int_{\mathbb{R}^d}p(x)\left(-\frac{1}{2}\log((2\pi)^d\det(C)-\frac{\log e}{2}(x-\mu,C^{-1}(x-\mu))\right)dx\\
&=\frac{\log e}{2}\mathbb{E}\left(\sum_{i,j}(x_i-\mu_i)(x_j-\mu_j)(C^{-1})_{ij}\right)+\frac{1}{2}\log((2\pi)^d\det C)\\
&=\frac{\log e}{2}\sum_{i,j}(C^{-1})_{ij}\mathbb{E}(x_i-\mu_i)(x_j-\mu_j)+\frac{1}{2}\log((2\pi)^d\det C)\\
&=\frac{\log e}{2}\sum_{i,j}(C^{-1})_{ij}C_{ji}+\frac{1}{2}\log((2\pi)^d\det C)\\
&=\frac{d\log e}{2}+\frac{1}{2}\log((2\pi)^d\det C)=\frac{1}{2}\log((2\pi e)^d\det C)
\end{aligned}$$ □

定理 1.5.5 对于均值为μ，协方差矩阵为$C=(C_{ij})$(即$C_{ij}=\mathbb{E}[(X_i-\mu_i)(X_j-\mu_j)]=C_{ji}$)的随机变量向量$X=(X_1, \cdots, X_d)$有

$$h(X)\leqslant\frac{1}{2}\log((2\pi e)^d\det C) \tag{1.5.6}$$

当且仅当X符合多变量正态分布时等号成立。

证明 令$p(x)$是X的PDF，$p^*(x)$是均值为μ，协方差矩阵为C的正态分布密度。为了不失一般性，假设$\mu=0$，观察到$\log p^*(x)$除去多余的常数项，是x_i的二次型。此外，对于每个单项式x_ix_j，$\int dxp^*(x)x_ix_j=\int dxp(x)x_ix_j=C_{ij}=C_{ji}$，二次型$\log p^*(x)$的矩相等。我们有

[88]

$$\begin{aligned}
0&\leqslant D(p\parallel p^*)(\text{通过 Gibbs 不等式})=\int p(x)\log\frac{p(x)}{p^*(x)}dx\\
&=-h(p)-\int p(x)\log p^*(x)dx=-h(p)-\int p^*(x)\log p^*(x)dx\\
&(\text{通过上面的备注})=-h(p)+h(p^*)
\end{aligned}$$

当且仅当$p=p^*$时等号成立。 □

举例 1.5.6 (a) 证明在给定均值的情况下，对于在$[0, \infty)$上分布的PDF中，指数分布能最大化熵分量；在给定方差的情况下，对于在$\mathbb{R}$上分布的PDF中，正态分布能最大化熵分量。

此外，设$\boldsymbol{X}=(X_1, \cdots, X_d)^{\mathrm{T}}$是一个$\mathbb{E}\boldsymbol{X}=\boldsymbol{0}$的随机向量，$\mathbb{E}X_iX_j=C_{ij}$，$1\leqslant i, j\leqslant d$。那么$h_{\mathrm{diff}}(\boldsymbol{X})\leqslant\frac{1}{2}\log((2\pi e)^d\det(C_{ij}))$，当且仅当$\boldsymbol{X}\sim N(0, C)$时等号成立。

(b) 证明对于取值不超过m个的随机变量的熵$h(X)\leqslant\log m$(参见式(1.2.7))在当类一离散随机变量X在$\mathbb{Z}_+$上取无限多值时存在下面的推广：

$$h(X)\leqslant\frac{1}{2}\log\left(2\pi e\left(\mathrm{Var}X+\frac{1}{12}\right)\right)$$

证明 (a) 对于 Gauss 的情况，参见定理 1.5.5。对于指数分布的情况，通过 Gibbs 定理，我们知道对于任意 PDF 为 $f(y)$ 的随机变量 Y 有 $\int f(y)\log[f(y)e^{\lambda y}/\lambda]\mathrm{d}y \geq 0$ 或者

$$h(Y) \leq (\lambda \mathbb{E}Y\log e - \log\lambda) = h(\mathrm{Exp}(\lambda))$$

当且仅当 $Y\sim \mathrm{Exp}(\lambda)$，$\lambda=(\mathbb{E}Y)^{-1}$ 时等号成立。

(b) 设 X_0 是一个离散随机变量，满足 $\mathbb{P}(X_0=i)=p_i$，$i=1,2,\cdots$，随机变量 U 独立于 X_0，并在 $[0,1]$ 上均匀分布。设 $X=X_0+U$。对于满足 $\mathrm{Var}X=\mathrm{Var}Y$ 的正态随机变量 Y，

$$h_{\mathrm{diff}}(X) \leq h_{\mathrm{diff}}(Y) = \frac{1}{2}\log(2\pi e\,\mathrm{Var}\,Y) = \frac{1}{2}\log\left(2\pi e\left(\mathrm{Var}X+\frac{1}{12}\right)\right) \qquad \square$$ 39

$\mathbb{E}X$ 的值对 $h(X)$ 没有影响，就像下面的定理所说的那样。

定理 1.5.7 (a) 微分熵平移不变，即对于所有的 $y\in\mathbb{R}^d$，

$$h(X+y) = h(X)$$

(b) 微分熵乘性可变：

$$h(aX) = h(X)+\log|a|,\text{对所有的 } a\in\mathbb{R}$$

此外，如果 $A=(A_{ij})$ 是一个 $d\times d$ 的非退化矩阵，考虑仿射变换 $x\in\mathbb{R}^d\mapsto Ax+y\in\mathbb{R}^d$。

(c) 那么

$$h(AX+y) = h(X)+\log|\det A| \tag{1.5.7}$$

证明 此证明十分简单，留作读者练习。 □

举例 1.5.8 (相对熵的数据处理不等式)S 为有限集，$\Pi=(\Pi(x,y),\ x,y\in S)$ 是统计核(也就是说，对于 $x,y\in S$，$\Pi(x,y)\geq 0$，$\sum_{y\in S}\Pi(x,y)=1$；换句话说，$\Pi(x,y)$ 是某个 Markov 链的转移概率。)证明 $D(p_1\Pi\,\|\,p_2\Pi)\leq D(p_1\,|\,p_2)$，其中 $p_i\Pi(y)=\sum_{x\in S}p_i(x)\Pi(x,y)$，$y\in S$(也就是说，同时对两个概率分布采用 Markov 算子不能增加相对熵)。

将这一事实推广到微分熵的情况。

解答 在离散的情况下 Π 由随机矩阵 $(\Pi(x,y))$ 定义。通过 log-sum 不等式(参见 PSE II，426 页)，对于所有的

$$\sum_x p_1(x)\Pi(x,y)\log\frac{\sum_w p_1(w)\Pi(w,y)}{\sum_z p_2(z)\Pi(z,y)} \leq \sum_x p_1(x)\Pi(x,y)\log\frac{p_1(x)\Pi(x,y)}{p_2(x)\Pi(x,y)}$$

$$= \sum_x p_1(x)\Pi(x,y)\log\frac{p_1(x)}{p_2(x)}$$

对 y 作加法我们得到

$$D(p_1\Pi\,|\,p_2\Pi) = \sum_x\sum_y p_1(x)\Pi(x,y)\log\frac{\sum_w p_1(w)\Pi(w,y)}{\sum_z p_2(Z)\Pi(Z,y)}$$

$$\leq \sum_x\sum_y p_1(x)\Pi(x,y)\log\frac{p_1(x)}{p_2(x)} = D(p_1\,|\,p_2)$$

在连续的情况中，如果我们用积分替换相加，类似的不等式也可以得到。 □ 40

微分熵的概念能够在许多意想不到的情况下发挥作用。在这里我们考虑正定矩阵的行列式及其比例的不等式(参见文献[39]、[36])。已知一个随机向量 $\boldsymbol{X}=(X_1,\cdots,X_n)$ 的协方差矩阵 $C=(C_{ij})$ 是正定的，即对于任一复向量 $\boldsymbol{y}=(y_1,\cdots,y_n)$，标量积 $(\boldsymbol{y},C\boldsymbol{y})=\sum_{i,j}C_{ij}y_i\overline{y_j}$

可写成

$$\sum_{i,j}\mathbf{E}(X_i-\mu_i)(X_j-\mu_j)y_i\overline{y_j}=\mathbf{E}\left|\sum_j(X_j-\mu_j)y_j\right|^2\geqslant 0$$

相反地，对于任一正定矩阵 C 都存在一个协方差矩阵为 C 的 PDF，例如一个多变量正态分布(如果 C 不是严格正定的话，则为退化概率分布)。

举例 1.5.9 如果 C 是正定的，那么 $\log[\det C]$ 在 C 上是凹的。

解答 取两个正定矩阵 $C^{(0)}$ 和 $C^{(1)}$，$\lambda\in[0,1]$。设 $\boldsymbol{X}^{(0)}$ 和 $\boldsymbol{X}^{(1)}$ 是两个多变量正态分布向量，$\boldsymbol{X}^{(\lambda)}\sim N(0,C^{(\lambda)})$。根据定理 1.2.18，设 $X=\boldsymbol{X}^{\Lambda}$，其中随机变量 Λ 各以概率 λ 和 $1-\lambda$ 取 0 和 1，并且与 $\boldsymbol{X}^{(0)}$ 和 $\boldsymbol{X}^{(1)}$ 独立。那么随机变量 X 有协方差 $C=\lambda C^{(0)}+(1-\lambda)C^{(1)}$，$X$ 不必是正态分布，所以

$$\begin{aligned}
&\frac{1}{2}\log(2\pi e)^d+\frac{1}{2}\log[\det(\lambda C^{(0)}+(1-\lambda)C^{(1)})]\\
&=\frac{1}{2}\log((2\pi e)^d\det C)\geqslant h(X)(\text{由定理 1.5.5})\\
&\geqslant h(X\mid\Lambda)(\text{由定理 1.2.11})\\
&=\frac{\lambda}{2}\log((2\pi e)^d\det C^{(0)})+\frac{1-\lambda}{2}\log((2\pi e)^d\det C^{(1)})\\
&=\frac{1}{2}[\log(2\pi e)^d+\lambda\log(\det C^{(0)})+(1-\lambda)\log(\det C^{(1)})]
\end{aligned}$$

□

这一性质通常被称作 Ky Fan 不等式，在 1950 年被第一次证明，当时使用了复杂得多的方法。另一个著名的不等式源于 Hadamard。

举例 1.5.10 对于一个正定矩阵 $C=(C_{ij})$，

$$\det C\leqslant\prod_i C_{ii}\tag{1.5.8}$$

91 当且仅当 C 是对角阵时等号成立。

解答 如果 $X=(X_1,\cdots,X_n)\sim N(0,C)$，可得

$$\frac{1}{2}\log[(2\pi e)^d\det C]=h(X)\leqslant\sum_i h(X_i)=\sum_i\frac{1}{2}\log(2\pi e C_{ii})$$

当且仅当 $X_1,\cdots,X_n$ 是独立的，等式成立，即 C 是对角的。 □

下面我们讨论所谓的熵-功率不等式(EPI)。存在 EPI 的情况是相当有趣的：它考虑了信息论中一个缺少简单解释的难解事实。它由 Shannon 提出，文献[141]包含了这种不等式的一个简单论证。然而，第一个关于 EPI 的严格证明差不多在 20 年之后才被提出，但在一些限制的条件下，它仍然是需要努力改进的主题。Shannon 使用 EPI 是为了在 Gauss 信道中，限制有连续噪声加性信道的容量；详见第 4 章。同时 EPI 和熵的单调性这个重要性质有关系，其中一个例子是后面的定理 1.5.15。

关于 EPI 的现存证明是不容易理解的，可以阅读文献[82]中那个更易懂的证明。

定理 1.5.11 (熵-功率不等式)对于概率分布函数分别为 $f_X(x)$ 和 $f_Y(x)$，$x\in\mathbf{R}^1$ 的两个独立随机变量 X 和 Y，

$$h(X+Y)\geqslant h(X'+Y')\tag{1.5.9}$$

其中 X'，Y' 是独立正态随机变量，且 $h(X)=h(X')$，$h(Y)=h(Y')$。

在 d-维空间的情况下熵-功率不等式有如下表现形式。

对于概率分布函数分别为 $f_X(x)$，$f_Y(x)$，$x\in\mathbf{R}^d$ 的两个独立随机变量 X，Y，有

$$e^{2h(X+Y)/d}\geqslant e^{2h(X)/d}+e^{2h(Y)/d}\tag{1.5.10}$$

很容易看出当 $d=1$ 时，式(1.5.9)和式(1.5.10)是相等的。通常情况下，不等式(1.5.9)通过利用下面的式(1.5.13)可以得到式(1.5.10)，其中式(1.5.13)可以独立地确定。注意到在离散随机变量的情况下，不等式(1.5.10)不一定都是正确的，考虑下面的例子：令 $X\sim Y$ 是独立的且 $P_X(0)=1/6$，$P_X(1)=2/3$，$P_X(2)=1/6$，那么

$$h(X)=h(Y)=\ln 6-\frac{2}{3}\ln 4, h(X+Y)=\ln 36-\frac{16}{36}\ln 8-\frac{18}{36}\ln 18$$

经检验，$e^{2h(X+Y)}=e^{2h(X)}+e^{2h(Y)}$。如果 X，Y 是非随机常量，则 $h(X)=h(Y)=h(X+Y)=0$，这明显不满足 EPI。由此我们推断出概率分布函数的存在是必要条件且不能被省略，在一个不同形式中，EPI 可以扩展到离散型随机变量，但是现在我们不讨论这个理论。 92

有时微分熵被定义为 $h(X)=-\mathbb{E}\log_2 p(X)$；则式(1.5.10)会有如下表现形式 $2^{h(X+Y)/d}\geqslant 2^{h(X)/d}+2^{h(Y)/d}$。

熵-功率不等式不仅在信息论和概率方面占据重要地位，而且在几何和分析方面同样很重要。为此我们提出了下面著名的 Brunn-Minkowski 定理，它是 EPI 的一个特殊情况。根据定义 $A-\varnothing=A$，定义两个集合的集总和为

$$A_1+A_2=\{x_1+x_2: x_1\in A_1, x_2\in A_2\}$$

定理 1.5.12 (Brunn-Minkowski)

(a) 令 A_1，A_2 为一个可测集，则体积满足

$$V(A_1+A_2)^{1/d}\geqslant V(A_1)^{1/d}+V(A_2)^{1/d} \tag{1.5.11}$$

(b) 两个集合 A_1，A_2 的集总和体积大于两个球 B_1，B_2 的集总和体积，其中 B_1，B_2 分别和 A_1，A_2 有相同的体积：

$$V(A_1+A_2)\geqslant V(B_1+B_2) \tag{1.5.12}$$

其中 B_1，B_2 是球体且满足 $V(A_1)=V(B_1)$，$V(A_2)=V(B_2)$。

举例 1.5.13 令 C_1，C_2 为正定 $d\times d$ 矩阵，则

$$[\det(C_1+C_2)]^{1/d}\geqslant[\det C_1]^{1/d}+[\det C_2]^{1/d} \tag{1.5.13}$$

解答 令 $X_1\sim N(0, C_1)$，$X_2\sim N(0, C_2)$，则有 $X_1+X_2\sim N(0, C_1+C_2)$。根据熵-功率不等式可得

$$\begin{aligned}(2\pi e)(\det(C_1+C_2))^{1/d}&=e^{2h(X_1+X_2)/d}\\&\geqslant e^{2h(X_1)/d}+e^{2h(X_2)/d}=(2\pi e)(\det C_1)^{1/d}+(2\pi e)(\det C_2)^{1/d}\end{aligned}$$

□

举例 1.5.14 一个 Töplitz $n\times n$ 矩阵 C 有如下性质：如果 $|i-j|=|r-s|$，则 $C_{ij}=C_{rs}$。令 $C_k=C(1, 2, \cdots, k)$ 表示由行和列 $1, \cdots, k$ 构造的 Töplitz 正定矩阵的主子式。证明对于 $|C|=\det C$，有 93

$$|C_1|\geqslant|C_2|^{1/2}\geqslant\cdots\geqslant|C_n|^{1/n} \tag{1.5.14}$$

$|C_n|/|C_{n-1}|$ 随着 n 增加而递减，且

$$\lim_{n\to\infty}\frac{|C_n|}{|C_{n-1}|}=\lim_{n\to\infty}|C_n|^{1/n} \tag{1.5.15}$$

解答 令 $(X_1, X_2, \cdots, X_n)\sim N(0, C_n)$，因为

$$h(X_k|X_{k-1},\cdots,X_1)=h(X_{k+1}|X_k,\cdots,X_2)\geqslant h(X_{k+1}|X_k,\cdots,X_1)$$

则数量 $h(X_k|X_{k-1}, \cdots, X_1)$ 随着 k 增加而递减，其中等式可以从 Töplitz 假定中得到，不等式可以从增加条件会减少熵这个事实中得到。下一步，我们利用 1.6 节中问题 1.8b 的结果，得到运行平均值会随 k 递减。

$$\frac{1}{k}h(X_1,\cdots X_k)=\frac{1}{k}\sum_{i=1}^{k}h(X_i|X_{i-1},\cdots X_1)$$

式(1.5.14)从下式得到

$$h(X_1,\cdots,X_n)=\frac{1}{2}\log[(2\pi e)^n|C_n|]$$

因为 $h(X_n|X_{n-1},\cdots,X_1)$ 是一个递减序列，所以它会存在一个极限。因此，利用 Cesáro 均值定理

$$\lim_{n\to\infty}\frac{h(X_1,X_2,\cdots,X_n)}{n}=\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}h(X_k|X_{k-1},\cdots,X_1)=\lim_{n\to\infty}h(X_n|X_{n-1},\cdots,X_1)$$

将这个转换为行列式形式，可以得到式(1.5.15)。 □

熵 功率不等式可以扩展到几个被加数的情况下

$$e^{2h(X_1+\cdots+X_n)/d}\geq\sum_{i=1}^{n}e^{2h(X_i)/d}$$

但是更有趣的是，下面的中间不等式是正确的。令 X_1，X_2，…，X_{n+1} 是独立同分布平方可积的随机变量。则

$$e^{2h(X_1+\cdots+X_n)/d}\geq\frac{1}{n}\sum_{j=1}^{n+1}e^{2h(\sum_{i\neq j}X_i)/d} \tag{1.5.16}$$

正如我们已知的，当存在如下限制条件时：考虑随机变量的方差受限于上式的情况，当随机变量为 Gauss 分布时微分熵最大。我们将给出如下的一个重要结论，但不给出在中心极限定理上每一次求和都会使熵增加的证明。

94

定理 1.5.15 令 X_1，X_2，…为独立同分布平方可积的随机变量，且 $EX_i=0$，$\mathrm{Var}X_i=1$。则

$$h\left(\frac{X_1-\cdots+X_n}{\sqrt{n}}\right)\leq h\left(\frac{X_1+\cdots+X_{n+1}}{\sqrt{n+1}}\right) \tag{1.5.17}$$

1.6 本章附加问题

问题 1.1 令 Σ_1，Σ_2 分别为大小是 m 和 q 的符号集。如果 $f:\Sigma_1\to\Sigma_2^*$ 是可译码则意味着什么？从 Kraft 和 Gibbs 不等式推断出，如果以概率 p_1，…，p_m 从 Σ_1 中取出字符，则期望的字长度至少为 $h(p_1,\cdots,p_m)/\log q$。

寻找一个可译二元码，它包含码字 011，0111，01111，11111，且有另外三个长度为 2 的码字。如何检验你得到的码是可译的？

解答 引入 $\Sigma^*=\bigcup_{n\geq 1}\Sigma^n$，其中所有数字字符串集合都来自于 Σ。我们发送的一个信息 $x_1x_2\cdots x_n\in\Sigma_1^*$ 作为 $f(x_1)f(x_2)\cdots f(x_n)\in\Sigma_2^*$ 的级联，即将 f 扩展为一个函数 $f^*:\Sigma_1^*\to\Sigma_2^*$。假如 f^* 是单射的话，就说这个码是可译的。

Kraft 不等式说明当且仅当

$$\sum_{i=1}^{m}q^{-s_i}\leq 1 \tag{1.6.1}$$

码字长为 s_1，…，s_m 的前缀码 $f:\Sigma_1\to\Sigma_2^*$ 是存在的。

事实上，每一个可译码都满足这个不等式。

Gibbs 不等式说明，如果 p_1，…，p_m 和 $\hat{p}_1$，…，$\hat{p}_m$ 是两个概率分布，则

$$h(p_1,\cdots,p_m)=-\sum_{j=1}^{m}p_j\log p_j\leq-\sum_{j=1}^{m}p_j\log\hat{p}_j \tag{1.6.2}$$

当且仅当 $p_j=\hat{p}_j$ 等式成立。

假设 f 是码字长为 $s_1, \cdots, s_m$ 的可译码。令 $\hat{p}_i = q^{-s_i}/c$，其中 $c = \sum_{i=1}^{m} q^{-s_i}$。则根据 Gibbs 不等式可得：

$$h(p_1, \cdots, p_m) \leqslant -\sum_{i=1}^{m} p_i \log \hat{p}_i = -\sum_{i=1}^{m} p_i(-s_i \log q - \log c) = \left(\sum_i p_i s_i\right)\log q + \left(\sum_i p_i\right)\log c$$ 95

利用 Kraft 不等式，$c \leqslant 1$，即 $\log c \leqslant 0$。可以得到

$$\text{期望码字长度为} \sum_i p_i s_i \geqslant h(p_1, \cdots, p_m)/\log q$$

在例子中，另外的三个码字必须是 00，01，10(不可以使用 11，因为 1s 序列是不可译的)。反转每个码字的顺序可以得到一个前缀码，而前缀码是可译的，因此，码是可译的。

总结得出，我们提出了 Kraft 不等式必要性的另一个证明。令 $s = \max s_i$，我们将 $\mathscr{H}$ 中的任意字通过增加一些固定的符号扩展到长度 s。如果 $x = x_1 x_2 \cdots x_{s_i} \in \mathscr{H}$，则满足 $x_1 x_2 \cdots x_{s_i} y_{s_i+1} \cdots y_s$ 形式的任意一个字都不属于 $\mathscr{H}$，因为 x 是有前缀的。但是至多有 q^{s-s_i} 个这样的字。当在 i 上总结时，我们可知排除在外的字总数是 $\sum_{i=1}^{m} q^{s-s_i}$。但是它不可能超过字的总数 q^s。因此，式(1.6.1)满足：

$$q^s \sum_{i=1}^{m} q^{-s_i} \leqslant q^s$$ □

问题 1.2 考虑一个有 m 个字母的字母表，其中每个字母出现的概率为 $1/m$。为了使期望码字长度 $(s_1 + \cdots + s_m)/m$ 最短，使用二元 Huffman 码对这些字母进行编码，其中 s_i 表示字母 i 的码字的长度。令 $s = \max\{s_i: 1 \leqslant i \leqslant m\}$，记 n_ℓ 是长度为 ℓ 的码字数目。

(a) 说明 $2 \leqslant n_s \leqslant m$。

(b) 当 m 取何值时有 $n_s = m$?

(c) 依据 m 决定 s。

(d) 证明 $n_{s-1} + n_s = m$，即任意两个码字长度最多相差 1。

(e) 判定 n_{s-1}，n_s。

(f) 描述对于一个理想的英文字母模型($m = 27$)的码字长度，其中所有符号都是等概率的。

(g) 二元 Huffman 码被用来编码概率分别为 $p_1 \geqslant \cdots \geqslant p_m > 0$ 的符号 1，…，m，其中 $\sum_{1 \leqslant i \leqslant m} p_i = 1$。令 s_1 为最短码字的长度，s_m 为最长的。判定 s_m，s_1 的最大和最小值，并且找出得到它们的二叉树。

解答 (a) 从 Huffman 码是树状结构可得到界 $n_s \geqslant 2$。更精确地说，假定 $n_s = 1$，即一个最大长度的码字是独一无二的，比如说是字母 i 所对应的码字。在不违反无前缀的条件下，可以在最后修改通向字母 i 长度为 s 的分支，但这与最小化相矛盾。显然可以得到 $n_s \leqslant m$。(从下面所述可得 n_s 总是偶数。) 96

(b) 当 $n_s = m$ 时，说明所有的码字都是等长的。这只有在 $m = 2^k$ 时才会发生，在这种情况下 $s = k$(有 2^k 个叶子的完全二叉树 T_k)。

(c) 通常情况下

$$s = \begin{cases} \log m, & m = 2^k \\ \lceil \log m \rceil, & m \neq 2^k \end{cases}$$

$m=2^k$ 的情况在(b)中讨论过了，故我们现在假定 $m\neq 2^k$。那么，$2^k<m<2^{k+1}$，其中 $k=\lfloor \log m\rfloor$。从观察中可以明显得到概率为 $1/m$ 的二叉树(也被称为二元 m-tree $\mathbb{B}_m$)包含完全二叉树 T_k，但是包含在 T_{k+1} 中。因此，s 就是上式所表达的。

(d) 事实上，当在均匀分布 $1/m$，…，$1/m$ 的情况下，树的分支的长度和最大值 s 不可能相差2或更多。实际上，假定有一个通向字母 i 的二叉树分支 B_i，且选定通向字母 j 的分支 M_j，它存在最大长度 s。在传统术语中，字母 j 在 s 上占用且字母 i 在 $i\leqslant s-2$ 占用。最后，分支 B_i，M_j 必须合并，这就会产生一个矛盾。例如，最小争议的图形依然是不正确的，见图1-10。在这里，概率为 $1/m$ 的顶点 i 应该和概率为 $2/m$ 的顶点 a 或 b 合并，而不是将 a，b 一块合并(如图所示)，如果这样它会产生一个概率为 $4/m$ 的顶点 c。

(e) 总之，有：(i)当 $m=2^k$，m-tree $\mathbb{B}_m$ 与 T_k 一致。(ii)当 $m\neq 2^k$ 时，我们可以通过

97 如下方法得到 $\mathbb{B}_m$。首先取一棵二叉树 T_k，其中 $k=[\log m]$，$1\leqslant m-2^k<2^k$。然后 T_k 的 $m-2^k$ 个叶子再往下分一步：这在树 T_{k+1} 上生成了 $2(m-2^k)=2m-2^{k+1}$ 个叶子。如图1-11所示，剩下的 T_k 中的 $2^k-(m-2^k)=2^{k+1}-m$ 个叶子保持不变。因此，

$$n_{k+1}=2^{k+1}-m,n_k=2m-2^{k+1},\quad 其中\ k=[\log m]$$

(f) 在英语的例子中，在服从均匀分布的 $m=27=16+11$ 个符号中，我们有5个长度为4的码字和22个长度为5的码字。平均码长为

$$\frac{5\times 4+22\times 5}{27}=\frac{130}{27}\approx 4.8$$

(g) s_1 的极小值为1(这很显然)，s_1 的极大值为 $\lceil \log_2 m\rceil$，即满足 $2^i<m\leqslant 2^{i+1}$ 的正整数 i，s_m 的极大值为 $m-1$(这也很显然)，s_m 的极小值为 $\lfloor \log_2 m\rfloor$，即满足 $2^{i-1}<m\leqslant 2^i$ 的自然数 i。

图1-12给出了对应 $s_1=1$ 和 $s_m=m-1$ 的树。

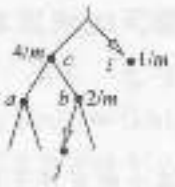


图 1-10

图 1-11

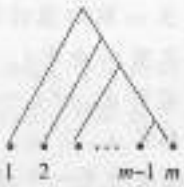


图 1-12

其特征如下：

i	$f(i)$	s_i
1	0	1
2	10	2
⋮	⋮	⋮
$m-1$	11…10	$m-1$
m	11…11	$m-1$

且满足以下条件

98 $$p_1>p_2+\cdots+p_m>2(p_3+\cdots+p_m)>\cdots>2^{m-1}p_m$$

当 $p_1=\cdots=p_m=1/m$ 服从均匀分布时，我们可以得到一棵最大化 s_1 和最小化 s_m 的树。当 $m=2^l$ 时，所有树枝的长度相等且满足 $l=\log_2 m$(一棵完全二叉树)：参考图1-13。

否则，比如当 $2^l<m<2^{l+1}$ 时，树中的第 l 层有 $2^{l+1}-m$ 个叶子，而第 $l+1$ 层有 $2(m-2^l)$ 个叶子：参考图1-14。

图 1-13

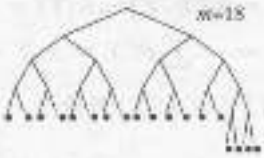


图 1-14

实际上，通过 Huffman 构造树，这里所考虑的树总是一棵完全二叉树的子树，树中的最短树枝长度不会大于$\lceil \log_2 m \rceil$而最长树枝长度不会小于$\lfloor \log_2 m \rfloor$。 □

问题 1.3 一个二元擦除信道，其擦除概率为 p，其离散无记忆二元信道(MBC)的信道矩阵为

$$\begin{bmatrix} 1-p & p & 0 \\ 0 & p & 1-p \end{bmatrix}$$

描述 Shannon 第二编码定理(SCT)并利用该定理来计算该信道的容量。 99

解答 Shannon SCT 指出，对于一个 MBC

$$(\text{容量}) = (\text{每信息符号传送的最大信息量})$$

这里容量是指所有可靠的信息速率的上确界，等式的右边定义为

$$\max_{X} I(X;Y)$$

其中随机变量 X 和 Y 分别表示输入和输出。

二元擦除信道输入字符为 0 或 1，且正确传输的概率为 $1-p$，错误传输(将输出记为 $*$)的概率为 p。设输入随机变量 X 取 0 的概率为 α，取 1 的概率为 $1-\alpha$。则输出随机变量 Y 的取值为

$$\mathbb{P}(Y=0) = (1-p)\alpha$$
$$\mathbb{P}(Y=1) = (1-p)(1-\alpha)$$
$$\mathbb{P}(Y=*) = p$$

将 Y 作为条件，得到

$$\left.\begin{aligned} &h(X|Y=0)=0 \\ &h(X|Y=1)=0 \\ &h(X|Y=*)=h(\alpha) \end{aligned}\right\} \text{意味着 } h(X|Y)=ph(\alpha)$$

因此，

$$\begin{aligned} \text{容量} &= \max_X I(X;Y) = \max_X [h(X)-h(X|Y)] \\ &= \max_X [h(\alpha)-ph(\alpha)] = (1-p)\max_\alpha h(\alpha) = 1-p \end{aligned}$$ 100

因为 $h(\alpha)=-\alpha\log\alpha-(1-\alpha)\log(1-\alpha)$在 $\alpha=1/2$ 时取到最大值 1，所以上式成立。 □

问题 1.4 令 X 和 Y 为两个离散随机变量，其累积分布函数(CDF)分别为 $\mathbb{P}_X$ 和 $\mathbb{P}_Y$。

(a) 定义条件熵 $h(X|Y)$，证明其满足以下不等式

$$h(X|Y) \leqslant h(X)$$

且给出等式成立的充要条件。

(b) 对于每个 $\alpha\in[0,1]$，混合随机变量 $W(\alpha)$的 PDF 形式为

$$\mathbb{P}_{W(\alpha)}(x) = \alpha\mathbb{P}_X(x) + (1-\alpha)\mathbb{P}_Y(x)$$

证明对于所有的 α，$W(\alpha)$的熵满足：

$$h(W(\alpha)) \geqslant \alpha h(X) + (1-\alpha)h(Y)$$

(c) 令 $h_{\mathbb{P}_\lambda}(\lambda)$为 Poisson 随机变量 $\mathrm{Po}(\lambda)$的熵，证明 $h_{\mathbb{P}_\lambda}(\lambda)$在 $\lambda>0$ 上是一个非减函数。

解答 (a) 依据定义，

$$h(X|Y)=h(X,Y)-h(Y)$$
$$=-\sum_{x,y}\mathbb{P}(X=x,Y=y)\log\mathbb{P}(X=x,Y=y)+\sum_{y}\mathbb{P}(Y=y)\log\mathbb{P}(Y=y)$$

不等式 $h(X|Y)\leqslant h(X)$ 等价于

$$h(X,Y)\leqslant h(X)+h(Y)$$

且服从 Gibbs 不等式 $\sum_i p_i\log\frac{p_i}{q_i}\geqslant 0$。事实上，令 $i=(x,y)$ 且

$$p_i=\mathbb{P}(X=x,Y=y),q_i=\mathbb{P}(X=x)\mathbb{P}(Y=y)$$

则

$$0\leqslant\sum_{x,y}\mathbb{P}(X=x,Y=y)\log\frac{\mathbb{P}(X=x,Y=y)}{\mathbb{P}(X=x)\mathbb{P}(Y=y)}$$
$$=\sum_{x,y}\mathbb{P}(X=x,Y=y)\log\mathbb{P}(X=x,Y=y)$$
$$-\sum_{x,y}\mathbb{P}(X=x,Y=y)[\log\mathbb{P}(X=x)+\log\mathbb{P}(Y=y)]$$
$$=-h(X,Y)+h(X)+h(Y)$$

101 当 X 和 Y 相互独立时，等式成立。

(b) 定义随机变量 T，其取 0 的概率为 α，取 1 的概率为 $1-\alpha$，则满足分布为 $W(\alpha)$ 的随机变量 Z 为

$$Z=\begin{cases}X, & T=0\\ Y, & T=1\end{cases}$$

根据(a)得到

$$h(Z|T)\leqslant h(Z)$$

其中等式左端为 $\alpha h(X)+(1-\alpha)h(Y)$，等式右端为 $h(W(\alpha))$。

(c) 注意到对独立随机变量 X 和 Y，$h(X+Y|X)=h(Y|X)=h(Y)$，因此，再一次根据(a)，

$$h(X+Y)\geqslant h(X+Y|X)=h(Y)$$

根据这一事实，对于所有的 $\lambda_1<\lambda_2$，取两个相互独立的随机变量 $X\sim\mathrm{Po}(\lambda_1)$，$Y\sim\mathrm{Po}(\lambda_2-\lambda_1)$，则

$$h(X+Y)\geqslant h(X)\quad\text{意味着}\quad h_{\mathrm{Po}}(\lambda_2)\geqslant h_{\mathrm{Po}}(\lambda_1)$$ □

问题 1.5 通过二元对称信道(MBSC)，以差错概率 p 在速率 R 上进行可靠传输的含义是什么？考虑 Shannon 第二编码定理(SCT)，计算 MBSC 中所有可能的可靠传输速率的上确界。讨论下列条件成立时的结果：(i) p 非常小；(ii) $p=1/2$；(iii) $p>1/2$。

解答 一个 MBSC 能在速率 R 上可靠传输，需要有一个 N 长码字构成的序列 $\mathscr{X}_N$，$N=1,2,\cdots$，其中一共有 $\lfloor 2^{NR}\rfloor$ 个码字，且差错概率渐近趋于零，即

$$\bar{e}(\mathscr{X}_N)=\max_{x\in\mathscr{X}_N}\mathbb{P}(\text{错误}\,|\,x\text{ 发送})\to 0\text{ 当 }N\to\infty$$

根据 SCT，可实现的信道容量为 $\sup R=\max_x I(X;Y)$，即每个输入符号所能传输的最大信息量。这里 X 为一个 Bernoulli 随机变量，取值为 0 和 1 且概率分别为 α 和 $1-\alpha$，$\alpha\in[0,1]$，而 Y 为对应输入 X 所产生的输出随机变量。然后计算互信息 $I(X;Y)$：

102 $$I(X;Y)=h(X)-h(X|Y)=h(Y)-h(Y|X)$$

注意到二元熵函数 $h(x)\leqslant 1$ 且等式在 $x=1/2$ 时成立。当选择 $\alpha=1/2$，则错误概率为 p 的 MBSC 的信道容量为

$$\max_{\varepsilon} I(X;Y) = \max_{\varepsilon}[h(Y)-h(Y|X)] = \max_{\varepsilon}[h(\varepsilon p+(1-\varepsilon)(1-p))-\eta(p)]$$
$$= 1+p\log p+(1-p)\log(1-p)$$

(i) 当 p 很小时，信道容量近似为 1(无噪信道容量)。

(ii) 当 $p=1/2$ 时，信道容量为 0(信道无法使用)。

(iii) 当 $p>1/2$ 时，可以将输出符号的标签交换，即将 p 和 $1-p$ 对换，从而信道容量非 0。 □

问题 1.6 (i) π^e 和 e^π 哪个更大？

(ii) 证明对数和不等式：对非负数 $a_1, a_2, \cdots, a_n$ 和 $b_1, b_2, \cdots, b_n$，有

$$\sum_i a_i\log\frac{a_i}{b_i} \geqslant \left(\sum_i a_i\right)\log\left[\frac{\sum_i a_i}{\sum_i b_i}\right] \tag{1.6.3}$$

当 a_i/b_i 为常量时，等式成立。

(iii) 考虑两个离散概率分布函数 $p(x)$ 和 $q(x)$，定义相对熵(Kullback-Leibler 距离)并证明 Gibbs 不等式

$$D(p\|q) = \sum_x p(x)\log\left(\frac{p(x)}{q(x)}\right) \geqslant 0 \tag{1.6.4}$$

当对所有的 x 有 $p(x)=q(x)$ 时，等式成立。

利用式(1.6.4)，证明对任意正函数 $f(x)$ 和 $g(x)$，以及任意有限集合 A，

$$\sum_{x\in A} f(x)\log\left(\frac{f(x)}{g(x)}\right) \geqslant \left(\sum_{x\in A} f(x)\right)\log\left[\frac{\sum_{x\in A} f(x)}{\sum_{x\in A} g(x)}\right]$$

验证对任意 $0\leqslant p, q\leqslant 1$，

$$p\log\left(\frac{p}{q}\right)+(1-p)\log\left(\frac{1-p}{1-q}\right) \geqslant (2\log_2 e)(q-p)^2 \tag{1.6.5}$$

并证明对任意概率分布 $p=(p(x))$ 和 $q=(q(x))$，

$$D(p\|q) \geqslant \frac{\log_2 e}{2}\left(\sum_x |p(x)-q(x)|\right)^2 \tag{1.6.6}$$ [103]

解答 (i) 令 $x=\ln\pi$，并且执行两次取对数运算可以得到不等式 $x-1>\ln x$。当 $x>1$ 时是成立的，所以 $e^\pi>\pi^e$。

(ii) 不失一般性地假定 $a_i>0$ 并且 $b_i>0$。函数 $g(x)=x\log x$ 为严格凸函数。所以通过 Jensen 不等式，对于任意的参数 $\sum c_i=1$，$c_i\geqslant 0$，有以下不等式

$$\sum c_i g(x_i) \geqslant g\left(\sum c_i x_i\right)$$

选择 $c_i=b_i\left(\sum_j b_j\right)^{-1}$ 和 $x_i=a_i/b_i$，我们可以得到

$$\sum_i \frac{a_i}{\sum b_j}\log\frac{a_i}{b_i} \geqslant \left(\sum_i \frac{a_i}{\sum b_j}\right)\log\left[\frac{\sum_i a_i}{\sum_i b_i}\right]$$

显然这是一个对数-和不等式。

(iii) 存在一个常数 $c>0$ 使得

$$\log y \geqslant c\left(1-\frac{1}{y}\right)，当且仅当\ y=1\ 时等式成立$$

记 $B=\{x:\ p(x)>0\}$，

$$D(p\|q)=\sum_{x\in B}p(x)\log\frac{p(x)}{q(x)}\geqslant c\sum_{x\in B}p(x)\left(1-\frac{q(x)}{p(x)}\right)=c[1-q(B)]\geqslant 0$$

当且仅当$q(x)=p(x)$时等式成立。接着，记

$$f(A)=\sum_{x\in A}f(x),\quad p(x)=\frac{f(x)}{f(A)}\mathbf{1}(x\in A)$$

$$g(A)=\sum_{x\in A}g(x),\quad q(x)=\frac{g(x)}{g(A)}\mathbf{1}(x\in A)$$

然后可以得到

$$\begin{aligned}\sum_{x\in A}f(x)\log\frac{f(x)}{g(x)}&=f(A)\sum_{x\in A}p(x)\log\frac{f(A)p(x)}{g(A)q(x)}\\&=f(A)\underbrace{\sum_{x\in A}p(x)\log\frac{p(x)}{q(x)}}_{\text{非负的相对熵}}+f(A)\log\frac{f(A)}{g(A)}\\&\geqslant f(A)\log\frac{f(A)}{g(A)}\end{aligned}$$

104

通过观察不难得出式(1.6.5)中的不等式关系。最后，考虑$A=\{x:p(x)\leqslant q(x)\}$。因为

$$\sum_x|p(x)-q(x)|=2[q(A)-p(A)]=2[p(A^c)-q(A^c)]$$

进而可以得到

$$\begin{aligned}D(p\|q)&=\sum_{x\in A}p(x)\log\frac{p(x)}{q(x)}+\sum_{x\in A^c}p(x)\log\frac{p(x)}{q(x)}\\&\geqslant p(A)\log\frac{p(A)}{q(A)}+p(A^c)\log\frac{p(A^c)}{q(A^c)}\\&\geqslant(2\log_2 e)[p(A)-q(A)]^2=\frac{\log_2 e}{2}\Big(\sum_x|p(x)-q(x)|\Big)^2\end{aligned}$$

□

问题1.7 (a) 给出条件熵的定义，并证明对于随机变量U和V，其联合熵满足

$$h(U,V)=h(V|U)+h(U)$$

给定随机变量X_1，…，X_n，通过归纳法或其他方法证明以下链式法则：

$$h(X_1,\cdots,X_n)=\sum_{i=0}^{n}h(X_i|X_1,\cdots,X_{i-1})\tag{1.6.7}$$

(b) 定义k-元素子集合上的子集平均(熵)为

$$h_k^{(n)}=\sum_{S:|S|=k}\frac{h(X_S)}{k}\Big/\binom{n}{k}$$

其中对于(下标集合)$S=\{s_1,\cdots,s_k\}$定义联合熵$h(X_S)=h(X_{s_1},\cdots,X_{s_k})$。假设对于任意$i$，当$T\subseteq S$，$i\notin S$时有$h(X_i|X_S)\leqslant h(X_i|X_T)$。

考虑具有如下形式的项

$$h(X_1,\cdots,X_n)-h(X_1,\cdots,X_{i-1},X_{i+1},\cdots,X_n)$$

证明$h_k^{(n)}\leqslant h_{k-1}^{(n)}$。

基于$h_k^{(n)}\leqslant h_{k-1}^{(n)}$，进一步证明$h_k^{(n)}\leqslant h_{k-1}^{(n)}$，其中$k=2$，…，$n$。

(c) 令$\beta>0$，并且定义

$$t_k^{(n)}=\sum_{S:|S|=k}e^{\beta h(X_S)/k}\Big/\binom{n}{k}$$

105

证明$t_1^{(n)}\geqslant t_2^{(n)}\geqslant\cdots\geqslant t_n^{(n)}$。

解答 (a) 条件熵可以定义为

$$h(V|U)=h(U,V)-h(U)=\sum_u \mathbb{P}(U=u)h(V|U=u)$$

其中 $h(V|U=u)$ 是 $U=u$ 时 V 的条件分布熵：

$$h(V|U=u)=-\sum_v \mathbb{P}(V=v|U=u)\log\mathbb{P}(V=v|U=u)$$

链式法则(1.6.7)可以通过对变量 n 进行归纳获得。

(b) 根据链式法则

$$h(X_1,\cdots,X_n)=h(X_1,\cdots,X_{n-1})+h(X_n|X_1,\cdots,X_{n-1}) \tag{1.6.8}$$

及其更一般化的表达，

$$\begin{aligned}h(X_1,\cdots,X_n)&=h(X_1,\cdots,X_{i-1},X_{i+1},\cdots,X_n)+h(X_i|X_1,\cdots,X_{i-1},X_{i+1},\cdots,X_n)\\&\leqslant h(X_1,\cdots,X_{i-1},X_{i+1},\cdots,X_n)+h(X_i|X_1,\cdots,X_{i-1})\end{aligned} \tag{1.6.9}$$

上式中不等式基于下面的关系得出

$$h(X_i|X_1,\cdots,X_{i-1},X_{i+1},\cdots,X_n)\leqslant h(X_i|X_1,\cdots,X_{i-1})$$

从 $i=1$ 到 n 诸加式(1.6.9)

$$nh(X_1,\cdots,X_n)\leqslant\sum_{i=1}^{n}h(X_1,\cdots,X_{i-1},X_{i+1},\cdots,X_n)+\sum_{i=1}^{n}h(X_i|X_1,\cdots,X_{i-1})$$

根据链式法则(1.6.7)可知，上式右边第二项(求和项)的值等于 $h(X_1,\cdots,X_n)$。进而可以得到

$$(n-1)h(X_1,\cdots,X_n)\leqslant\sum_{i=1}^{n}h(X_1,\cdots,X_{i-1},X_{i+1},\cdots,X_n)$$

这意味着 $h_n^{(n)}\leqslant h_{n-1}^{(n)}$，因为

$$\frac{1}{n}h(X_1,\cdots,X_n)\leqslant\frac{1}{n}\sum_{i=1}^{n}\frac{h(X_1,\cdots,X_{i-1},X_{i+1},\cdots,X_n)}{n-1} \tag{1.6.10}$$

一般来说，考虑在$\{1,\cdots,n\}$中选取一个具有 k 个元素的子集 S。记 $S\setminus\{i\}$ 为 $S(i)$，我们可以通过上面的讨论得到

$$\frac{1}{k}h[X(S)]\leqslant\frac{1}{k}\sum_{i\in S}\frac{h(X[S(i)])}{k-1}$$

106

通过上述表达式，可以进一步得到

$$\binom{n}{k}h_k^{(n)}=\sum_{S\subset\{1,\cdots,n\},|S|=k}\frac{h[X(S)]}{k}\leqslant\sum_{S\subset\{1,\cdots,n\},|S|=k}\sum_{i\in S}\frac{h(X[S(i)])}{k(k-1)} \tag{1.6.11}$$

最后，我们观察到每一个具有 $k-1$ 个元素的子集 $S(i)$，在式(1.6.11)中的加和项中出现$[n-(k-1)]$次。所以我们可以把 $h_k^{(n)}$ 记为

$$\left(\sum_{T\subset\{1,\cdots,n\},|T|=k-1}\frac{h[X(T)]}{k-1}\right)\frac{n-(k-1)}{k}\Big/\binom{n}{k}=\sum_{T\subset\{1,\cdots,n\},|T|=k-1}\frac{h[X(T)]}{k-1}\Big/\binom{n}{k-1}=h_{k-1}^{(n)}$$

(c) 针对集合 $S_0=\{1,2,\cdots,n\}$，首先对(1.6.11)式两边取幂，然后应用算术或者几何平均不等式可以得到

$$e^{h(X(S_0))/n}\leqslant e^{[h(S_0(1))+\cdots+h(S_0(n))]/(n(n-1))}\leqslant\frac{1}{n}\sum_{i=1}^{n}e^{h(S_0(i))/(n-1)}$$

上式等价于 $t_n^{(n)}\leqslant t_{n-1}^{(n)}$。现在，我们采用证明(b)问题所采用的方法，用所有子集进行平均来证明对于所有 $k\leqslant n$ 均有 $t_k^{(n)}\leqslant t_{k-1}^{(n)}$。 □

问题 1.8 令 $p_1,\cdots,p_n$ 为概率分布且 $p^*=\max_i[p_i]$，证明

(i) $-\sum_i p_i \log_2 p_i \geqslant -p^* \log_2 p^* - (1-p^*)\log_2(1-p^*)$

(ii) $-\sum_i p_i \log_2 p_i \geqslant \log_2(1/p^*)$

(iii) $-\sum p_i \log_2 p_i \geqslant 2(1-p^*)$

随机变量 X，Y 和值 x，y 来源于有限"字母表" I 和 J，代表着传输信道的输入和输出，条件分布 $P(x|y)=\mathbb{P}(X=x|Y=y)$。令 $h(P(\cdot|y))$ 表示条件分布 $P(\cdot|y)(y\in J)$ 的熵，并且 $h(X|Y)$ 代表 X 和 Y 的条件熵。定义理想观测者译码准则为映射 $f: J\to I$，使得对于所有 $y\in J$ 均满足 $P(f(y)|y)=\max_{x\in I}P(x|y)$。证明在这种译码准则下的差错概率为

$$\pi_e(y)=\sum_{x\in I,x\neq f(y)}P(x|y)$$

其中差错概率满足 $\pi_e(y)\leqslant\frac{1}{2}h(P(\cdot|y))$，并且差错概率的均值满足

$$\mathbb{E}\pi_e(Y)\leqslant\frac{1}{2}h(X|Y)$$

107

解答 (i)式右边下界由聚合不等式(参见引理 1.2.5)得出。(ii)式右边下界可由下面关系式导出

$$-\sum_i p_i\log p_i\geqslant\sum_i p_i\log\frac{1}{p^*}=\log\frac{1}{p^*}$$

为了证明(iii)式，可以在 $p^*\geqslant 1/2$ 时使用(i)式，在 $p^*\leqslant 1/2$ 时使用(ii)式。首先假定 $p^*\geqslant 1/2$，通过(i)式可得，

$$h(p_1,\cdots,p_n)\geqslant h(p^*,1-p^*)$$

函数 $x\in(0,1)\mapsto h(x,1-x)$ 是上凸函数，其在 $(1/2,1)$ 处的图形严格地处于线性函数 $x\mapsto 2(1-x)$ 之上。

所以，

$$h(p_1,\cdots,p_n)\geqslant 2(1-p^*)$$

另一方面，当 $p^*\leqslant 1/2$ 时使用(ii)式可得：

$$h(p_1,\cdots,p_n)\geqslant\log\frac{1}{p^*}$$

对于 $0\leqslant x\leqslant 1/2$，

$$\log\frac{1}{x}\geqslant 2(1-x)\quad 当且仅当\quad x=\frac{1}{2}$$

最后，我们使用(iii)式来推导最终结论。记

$$\pi_e(y)=1-\mathbb{P}_h(f(y)|y)=1-p_{\max}(\cdot|y)$$

其值小于 $h(P(\cdot|Y))/2$。最后注意到 $h(X|Y)=\mathbb{E}h(P(\cdot|Y))$，通过取均值的方法可以证明 $\mathbb{E}\pi_e(Y)$ 的界。 □

问题 1.9 给出信息速率 H 和信源渐近均分性的定义。计算 Bernoulli 源信息速率。给定一个无记忆二元信道，给出该信道容量 C 的定义。根据 Shannon 第二编码定理(SCT)，推断 $C=\sup_{p_X}I(X;Y)$。

一个擦除信道保持传输符号无误的概率是 $1-p$，将其变为不可读符号的概率为 p。试给出擦除信道的容量。

解答 对于取自有限符号集 I 的信源 U_1，U_2，…，其信息速率 H 是对于所有 $R>0$ 的上确界，其中 R 满足约束，即存在一系列集合 $A_n\in I\times\cdots\times I$($n$ 次)使得 $|A_n|\leqslant 2^{nR}$ 并

且 $\lim_{n\to\infty}\mathbb{P}(U_1^n\in A_\varepsilon)=1$。

渐近均分特性意味着，当 $n\to\infty$ 时，在某种意义上（这里我们指依概率收敛）

$$-\frac{1}{n}\log p_n(U_1^n)\to H$$

108

在这里 $U_1^n=U_1\cdots U_n$，$p_n(u_1^n)=\mathbb{P}(U_1^n=u_1^n)$。Shannon 第二定理 SCT 指出如果随机变量 $-\log p_n(U_1^n)/n$ 收敛到一个极限，那么这个极限就是 H。

一个无记忆二元信道（MBC）具有如下的条件（转移）概率：

$$\mathbb{P}_{\text{ch}}(\boldsymbol{Y}^{(N)}\mid\boldsymbol{X}^{(N)}\text{ 发送})=\prod_{1\le i\le N}P(y_i\mid x_i)$$

其发生传输错误的概率是

$$\varepsilon^{(N)}=\sum_{u}\mathbb{P}_{\text{source}}(U=u)\mathbb{P}_{\text{ch}}(\hat{f}_N(\boldsymbol{Y}^{(N)})\neq u\mid f_N(u)\text{ 发送})$$

其中 $\mathbb{P}_{\text{source}}$ 代表使用 f_N 码字和 $\hat{f}_N$ 的解码规则的信源分布概率。如果 $\overline{R}\in(0,1)$ 是可靠传输速率，$\mathbb{P}_{\text{source}}$ 在源字符串为 u 且满足 $\#\mathscr{U}_N=2^{N(\overline{R}+o(1))}$ 的集合 $\mathscr{U}_N$ 上是均匀分布的，那么就存在 f_N 和 $\hat{f}_N$ 使得

$$\lim_{N\to\infty}\frac{1}{\#\mathscr{U}_N}\sum_{u\in\mathscr{U}_N}\mathbb{P}_{\text{ch}}(\hat{f}_N(\boldsymbol{Y}^{(N)})\neq u\mid f_N(u)\text{ 发送})=0$$

信道容量是所有可靠传输速率的上确界。

对于一个擦除信道，信道转移矩阵是

$$\begin{array}{c|ccc} & 0 & 1 & \times \\ \hline 0 & 1-p & 0 & p \\ 1 & 0 & 1-p & p \end{array}$$

条件熵是 $h(Y\mid X)=h(p,1-p)$ 并不依赖于 p_X。所以，容量

$$C=\sup_{p_X}I(X;Y)=\sup_{p_X}h(Y)-h(Y\mid X)$$

在 $p_X(0)=p_X(1)=1/2$ 上可达并且

$$h(Y)=-(1-p)\log[(1-p)/2]-p\log p=h(p,1-p)+(1-p)$$

因此，该信道的容量是 $C=1-p$。 □

问题 1.10 给出 Huffman 编码规则的定义并证明它在可译码中的最优性。针对符号概率 $\frac{1}{5},\frac{1}{5},\frac{1}{6},\frac{1}{10},\frac{1}{10},\frac{1}{10},\frac{1}{10},\frac{1}{30}$ 计算 Huffman 编码码字长度。

证明如下推断或给出一个反例：如果使用 Huffman 编码所得的一个码字长度是 l，那么在该编码中存在另一个长度为 l' 的码字，且满足 $|l-l'|\le 1$。

109

解答 第一部分的答案：

概率	码字	长度
1/5	00	2
1/5	100	3
1/6	101	3
1/10	110	3
1/10	010	3
1/10	011	3
1/10	1110	4
1/30	1111	4

第二部分，一个反例如下：

概率	码字	长度
1/2	0	1
1/8	100	3
1/8	101	3
1/8	110	3
1/8	111	3

□

问题 1.11 一个无记忆信道的输入符号为{0，1}，正确输出符号的概率为$(n-1)/n^2$，转换输出的概率为$1/n^2$。(因此，对于$n=1$而言信道是二元无噪声信道。)对于$n\geq 2$的情况，会产生$2(n-1)$种的"污点"，按照惯例用α_i和$\beta_i(i=1,\cdots,n-1)$来表示这些"污点"，它们的概率相似：$P(\alpha_i|0)=(n-1)/n^2$，$P(\beta_i|0)=1/n^2$，$P(\beta_i|1)=(n-1)/n^2$，$P(\alpha_i|1)=1/n^2$。证明信道容量C_n会随着n单调增加，并且有$\lim\limits_{n\to\infty}C_n=\infty$。如果我们简单地将污点$\alpha_i$当作0并且将$\beta_i$当作1，那么信道的容量会受到怎样的影响？

解答 信道的概率转移矩阵为

$$\begin{array}{c} 0 \\ 1 \\ \\ \end{array}\begin{bmatrix} \frac{n-1}{n^2} & \frac{1}{n^2} & \frac{n-1}{n^2} & \frac{1}{n^2} & \cdots & \frac{n-1}{n^2} & \frac{1}{n^2} \\ \frac{1}{n^2} & \frac{n-1}{n^2} & \frac{1}{n^2} & \frac{n-1}{n^2} & \cdots & \frac{1}{n^2} & \frac{n-1}{n^2} \\ 0 & 1 & \alpha_1 & \beta_1 & \cdots & \alpha_{n-1} & \beta_{n-1} \end{bmatrix}$$

信道是双对称的(矩阵的行和列互为置换)，因此可实现信道容量的输入分布为

110

$$p_X(0)=p_X(1)=\frac{1}{2}$$

信道容量C_n为

$$\begin{aligned} C_n &= \log(2n)+n\frac{1}{n^2}\log(n^2)+n\frac{n-1}{n^2}\log\frac{n^2}{n-1} \\ &= 1+3\log n-\frac{n-1}{n}\log(n-1)\to+\infty,\quad 当\ n\to\infty \end{aligned}$$

此外，可以推算出

$$C(x)=1+3\log x-\left(1-\frac{1}{x}\right)\log(x-1),x\geq 1$$

我们发现

$$\begin{aligned} \frac{\mathrm{d}C(x)}{\mathrm{d}x} &= \frac{3}{x}-\frac{1}{x-1}+\frac{1}{x(x-1)}-\frac{1}{x^2}\log(x-1) \\ &= \frac{2}{x}-\frac{1}{x^2}\log(x-1) \\ &= \frac{1}{x^2}[2x-\log(x-1)]>0,\quad x>1 \end{aligned}$$

因此，C_n对于$n\geq 1$会随着n增长。当α_i和β_i分别被当作0或1时，信道容量不变。 □

问题 1.12 $X_i(i=1,2,\cdots)$是独立同分布的随机变量，取值为1，0的概率分别为p和$(1-p)$。基于如下条件证明局部De Moivre-Laplace定理

$$P(S_n=k)=\frac{1}{\sqrt{2\pi y(1-y)n}}\exp[-nh_y(y)+\theta_n(k)],\quad k=1,\cdots,n-1 \qquad (1.6.12)$$

在这里

$$S_n = \sum_{1\le i\le n} X_i, \quad y = k/n, \quad h_p(y) = y\ln\left(\frac{y}{p}\right) + (1-y)\ln\left(\frac{1-y}{1-p}\right)$$

余项 $\theta_n(k)$ 遵循

$$|\theta_n(k)| < \frac{1}{6ny(1-y)}, \quad y = k/n$$

提示：对余项使用 Stirling 公式

$$n! = \sqrt{2\pi n}\, n^n e^{-n+\theta(n)}$$

其中

$$\frac{1}{12n+1} < \theta(n) < \frac{1}{12n}$$

找到 k^+ 和 k^- 满足 $0\le k^-, k^+ \le n$（依赖于 n），那么当 $n\to\infty$ 时，$\mathbb{P}(S_n = k^+)$ 是渐近最大的并且 $\mathbb{P}(S_n = k^-)$ 是渐近最小的。之后，进一步给出相应的渐近性公式。 111

解答 可以写出

$$\begin{aligned}
\mathbb{P}(S_n = k) &= \binom{n}{k}(1-p)^{n-k}p^k = \frac{n!}{k!(n-k)!}(1-p)^{n-k}p^k \\
&= \sqrt{\frac{n}{2\pi k(n-k)}}\frac{n^n}{k^k(n-k)^{n-k}}(1-p)^{n-k}p^k \times \exp[\theta(n)-\theta(k)-\theta(n-k)] \\
&= \frac{1}{\sqrt{2\pi ny(1-y)}}\exp[-k\ln y-(n-k)\ln(1-y) \\
&\quad + k\ln p + (n-k)\ln(1-p)]\exp[\theta(n)-\theta(k)-\theta(n-k)] \\
&= \frac{1}{\sqrt{2\pi ny(1-y)}}\exp[-nh_p(y)] \times \exp[\theta(n)-\theta(k)-\theta(n-k)]
\end{aligned}$$

因为

$$|\theta(n)-\theta(k)-\theta(n-k)| < \frac{1}{12n}+\frac{1}{12k}+\frac{1}{12(n-k)} < \frac{2n^2}{12nk(n-k)}$$

式(1.6.12)成立，$\theta_n(k)=\theta(n)-\theta(k)-\theta(n-k)$。根据 Gibbs 不等式可知 $h_p(y)\ge 0$ 和当且仅当 $y=p$ 时 $h_p(y)=0$。而且由

$$\frac{\mathrm{d}h_p(y)}{\mathrm{d}y} = \ln\frac{y}{p} - \ln\frac{1-y}{1-p} \quad 和 \quad \frac{\mathrm{d}^2h_p(y)}{\mathrm{d}y^2} = \frac{1}{y}+\frac{1}{1-y} > 0$$

可以得到

$$\left.\frac{\mathrm{d}h_p(y)}{\mathrm{d}y}\right|_{y=p} = 0, \quad \frac{\mathrm{d}h_p(y)}{\mathrm{d}y} < 0, \quad 0<y<p \quad 和 \quad \frac{\mathrm{d}h_p(y)}{\mathrm{d}y} > 0, \quad p<y<1$$

因此

$$\underline{h}_p = \min h_p(y) = 0, \quad 在\ y=p\ 时取得$$

$$\overline{h}_p = \max h_p(y) = \min\left(\ln\frac{1}{p}, \ln\frac{1}{1-p}\right), \quad 在\ y=0\ 或\ y=1\ 时取得$$

于是，对于 $n\gg 1$ 的最大概率在 $y^*=p$ 时取得，即 $k^+=\lfloor np\rfloor$：

$$\mathbb{P}(S_n = \lfloor np\rfloor) \simeq \frac{1}{\sqrt{2\pi np(1-p)}}\exp(\theta_n(\lfloor np\rfloor))$$

其中

$$|\theta_n(\lfloor np\rfloor)| \le \frac{1}{6np(1-p)}$$

类似地，最小概率是

$$\mathbb{P}(S_n=0)=p^n,\qquad 0<p\leqslant 1/2$$

112 $$\mathbb{P}(S_n=n)=(1-p)^n,\quad 1/2\leqslant p<1$$ □

问题1.13 (a) 证明一个概率分布为 $p=(p(1),\cdots,p(n))$ 的离散随机变量 X 的熵 $h(X)=-\sum_{i=1}^{n}p(i)\log p(i)$ 是向量 p 的下凸函数。

证明概率分布为 $\mathbb{P}(X=i,Y=k)=p_X(i)P_{Y|X}(k|i)(i,k=1,\cdots,n)$ 的随机变量 X 和 Y 之间的互信息 $I(X;Y)=h(Y)-h(Y|X)$，对于固定的条件概率 $\{P_{Y|X}(k|i)\}$，是向量 $p_X=(p_X(1),\cdots,p_X(n))$ 的下凸函数。

(b) 试证明

$$h(X)\geqslant -p^*\log_2 p^*-(1-p^*)\log_2(1-p^*)$$

其中 $p^*=\max_x\mathbb{P}(X=x)$，然后推出当 $p^*\geqslant 1/2$ 时

$$h(X)\geqslant 2(1-p^*)\tag{1.6.13}$$

说明不等式(1.6.13)在 $p^*\leqslant 1/2$ 时也成立。

解答 (a) $h(p)$ 是下凸的意味着

$$h(\lambda_1 p_1+\lambda_2 p_2)\geqslant \lambda_1 h(p_1)+\lambda_2 h(p_2)\tag{1.6.14}$$

对于所有概率向量 $p_j=(p_j(1),\cdots,p_j(n))(j=1,2)$ 以及 $\lambda_1,\lambda_2\in(0,1)$ 成立，其中 $\lambda_1+\lambda_2=1$。令 X_1 的分布为 p_1，X_2 的分布为 p_2。同时令

$$Z=1(\text{以概率}\ \lambda_1)\quad Z=2(\text{以概率}\ \lambda_2)$$

$Y=X_Z$。那么 Y 的概率分布是 $\lambda_1 p_1+\lambda_2 p_2$。根据定理1.2.11(a)，有

$$h(Y)\geqslant h(Y|Z)$$

根据条件熵的定义

$$h(Y|Z)=\lambda_1 h(X_1)+\lambda_2 h(X_2)$$

这就推出式(1.6.14)。现在

$$I(X;Y)=h(Y)-h(Y|X)=h(Y)-\sum p_X(i)h(P_{Y|X}(\cdot|i))\tag{1.6.15}$$

如果 $P_{Y|X}(\cdot|\cdot)$ 是固定的，由于上式的第二部分是 p_X 的线性函数，因此是下凸函数。其中，第一项 $h(Y)$ 是 p_Y 的下凸函数，p_Y 也是 p_X 的线性函数。因此 $h(Y)$ 是 p_X 的下凸函数，$I(X;Y)$ 也是 p_X 的下凸函数。

(b) 考虑两种情况：(i) $p^*\geqslant 1/2$；(ii) $p^*\leqslant 1/2$。考虑情况(i)，通过聚合不等式可得

113 $$h(X)\geqslant h(p^*,1-p^*)\geqslant(1-p^*)\log\frac{1}{p^*(1-p^*)}\geqslant(1-p^*)\log\frac{1}{\frac{1}{2}}=2(1-p^*)$$

在情况(ii)下我们针对 n（随机变量 X 可能取值的个数）采用归纳法。首先取初始值 $n=3$；不失一般性，假设 $p^*=p_1\geqslant p_2\geqslant p_3$，那么 $1/3\leqslant p_1\leqslant 1/2$，$(1-p_1)/2\leqslant p_2\leqslant p_1$。可以进一步写出

$$h(p_1,p_2,p_3)=h(p_1,1-p_1)+(1-p_1)h(q,1-q),\quad 其中\ q=\frac{p_2}{1-p_1}$$

因为 $1/2\leqslant q\leqslant p_1/(1-p_1)\leqslant 1$，

$$h(q,1-q)\geqslant h(p_1/(1-p_1),(1-2p_1)/(1-p_1))$$

也就是

$$h(p_1,p_2,p_3)\geqslant h(p_1,p_1,1-2p_1)=2p_1+h(2p_1,1-2p_1)$$

根据(a)可以得出不等式 $2p_1+h(2p_1,1-2p_1)\geqslant 2(1-p_1)$ 等价于

$$h(2p_1,1-2p_1)\geqslant 2-4p_1,\ 1/3\leqslant p_1\leqslant 1/2$$

或等价于

$$h(p,1-p)\geqslant 2-2p,2/3\leqslant p\leqslant 1$$

因此，对于 $n=3$，$h(p_1,p_2,p_3)\geqslant 2(1-p^*)$ 与 p^* 的值无关。归纳法的第一步完成了。

下一步我们假设对于 X 取值 $\leqslant n-1$ 时 $h(X)\geqslant 2(1-p^*)$ 成立。然后取 $\boldsymbol{p}=(p_1,\cdots,p_n)$，并不失一般性地假设 $p^*=p_1\geqslant\cdots\geqslant p_n$。可以进一步写出 $\boldsymbol{q}=(p_2/(1-p_1),\cdots,p_{n-1}/(1-p_1))$ 和

$$h(\boldsymbol{p})=h(p_1,1-p_1)+(1-p_1)h(\boldsymbol{q})\geqslant h(p_1,1-p_1)+(1-p_1)2(1-q_1)\quad(1.6.16)$$

不等式 $h(\boldsymbol{p})\geqslant 2(1-p^*)$ 可以根据如下式子得出

$$h(p_1,1-p_1)+(1-p_1)(2-2q_1)\geqslant(2-2p_1)$$

该式子等价于

$$h(p_1,1-p_1)\geqslant 2(1-p_1)(1-1+q_1)=2(1-p_1)q_1=2p_2$$

其中 $1/n\leqslant p_1<1/2$，$(1-p_1)/(n-1)\leqslant p_2<p_1$。但很明显可以观察到

$$h(p_1,1-p_1)\geqslant 2(1-p_1)\geqslant 2p_2$$

（当 $p_1=0$，$1/2$ 时等号成立）。因此，不等式(1.6.16)可以根据归纳假设得出。 □

问题 1.14 考虑一个概率分布 p_i，$i\in I=\{1,2,\cdots,n\}$，该分布可以使 $\log_2(1/p_i)$ 对于所有的 $p_i>0$ 的 i 都是一个整数。可以将 I 看作一个符号集，其中的字母将通过二进制字符编码。一个 Shannon-Fano(SF)编码分配给字母 i 一个长度为 $\ell_i=\lceil\log_2(1/p_i)\rceil$ 的码字；根据 Kraft 不等式可以构建唯一可译码。证明 SF 码的竞争最优性：如果 $\ell_i'(i\in I)$，是任意唯一可译码的二元码码字长度，那么 114

$$\mathbb{P}(\ell_i<\ell_i')\geqslant\mathbb{P}(\ell_i'<\ell_i)\quad(1.6.17)$$

当且仅当 $\ell_i=\ell_i'$ 时等号成立。

线索：你可能会用到不等式 $\operatorname{sgn}(\ell-\ell')\leqslant 2^{\ell'-\ell}-1$，$\ell,\ell'=1,\cdots,n$。 □

解答 可以得到

$$\mathbb{P}(\ell_i'<\ell_i)-\mathbb{P}(\ell_i'>\ell_i)=\sum_{i:\ell_i'<\ell_i}p_i-\sum_{i:\ell_i'>\ell_i}p_i$$

$$=\sum_i p_i\operatorname{sign}(\ell_i-\ell_i')=\mathbb{E}\operatorname{sgn}(\ell-\ell')\leqslant\mathbb{E}(2^{\ell'-\ell}-1)$$

其中利用了如下事实：对于整数 x 有 $\operatorname{sign}x\leqslant 2^x-1$。进而可以利用 Kraft 不等式获得

$$\mathbb{E}(2^{\ell'-\ell}-1)=\sum_i p_i(2^{\ell_i'-\ell_i}-1)=\sum_i 2^{-\ell_i}(2^{\ell_i'-\ell_i}-1)=\sum_i 2^{-\ell_i'}-\sum_i 2^{-\ell_i}$$

$$\leqslant 1-\sum_i 2^{-\ell_i}=1-1=0$$

这就推出了不等式

$$\mathbb{P}(\ell_i<\ell_i')\geqslant\mathbb{P}(\ell_i'<\ell_i)$$

为了让等号成立，我们必须使得：(a) $2^{\ell_i'-\ell_i}-1=0$ 或 1，$i\in I$（因为只有当 $x=0$ 或者 1 时才有 $\operatorname{sign}x=2^x-1$）；(b) $\sum_i 2^{-\ell_i'}=1$。因为 $\sum_i 2^{-\ell_i}=1$，唯一的可能就是 $2^{\ell_i'-\ell_i}=1$，即 $\ell_i=\ell_i'$。 □

问题 1.15 定义一个二元信道的信道容量 C。令 $C_N=(1/N)\sup I(\boldsymbol{X}^{(N)};\boldsymbol{Y}^{(N)})$，其中 $I(\boldsymbol{X}^{(N)};\boldsymbol{Y}^{(N)})$ 表示通过信道发送的长度为 N 的随机码字 $\boldsymbol{X}^{(N)}$ 和对应接收码字 $\boldsymbol{Y}^{(N)}$ 之间的互信息，而上确界定义在 $\boldsymbol{X}^{(N)}$ 的概率分布上。证明 $C\leqslant\limsup_{N\to\infty}C_N$。

解答 一个二元信道可以定义为如下的一个条件概率分布序列

$$\mathbb{P}_{\mathrm{ch}}^{(N)}(\boldsymbol{y}^{(N)}\mid\boldsymbol{x}^{(N)}),N=1,2,\cdots$$ 115

其中 $\boldsymbol{x}^{(N)}=x_1\cdots x_N$ 是输入端的二进制码字序列，$\boldsymbol{y}^{(N)}=y_1\cdots y_N$ 是输出端的二进制码字序

列，信道容量 C 是一个概率族 $\{P_{\mathbf{X}}^{(N)}(\cdot\mid\cdot)\}$ 的渐近参数，定义

$$C=\sup[\overline{R}\in(0,1):\overline{R}\text{ 是可靠的传输速率}] \tag{1.6.18}$$

$\overline{R}\in(0,1)$ 被称为可靠传输速率（对于一个给定的信道）。如果假定随机信源序列在集合 $\mathscr{U}^{(N)}$ 上是等概率分布的且集合 $\mathscr{U}^{(N)}$ 的势为 $\#\mathscr{U}^{(N)}=2^{N(\overline{R}-o(1))}$，那么存在一个编码规则 $f^{(N)}:\mathscr{U}^{(N)}\to\mathscr{X}_N\subseteq\{0,1\}^N$ 和一个译码规则 $\hat{f}^{(N)}:\{0,1\}^N\to\mathscr{U}^{(N)}$，当 $N\to\infty$ 时错误概率

$$e^{(N)}:=\sum_{u\in\mathscr{U}^{(N)}}\frac{1}{\#\mathscr{U}^{(N)}}P_{\mathbf{X}}^{(N)}(\{\mathbf{y}^{(N)}:\hat{f}^{(N)}(\mathbf{y}^{(N)})\neq u\}\mid f^{(N)}(u)) \tag{1.6.19}$$

趋于零，即 $e^{(N)}\to 0$。

$$e^{(N)}=e^{(N)}(f^{(N)},\hat{f}^{(N)})$$

Shannon 第二编码定理（SCT）的逆定理表明

$$C\leqslant\limsup_{N\to\infty}\frac{1}{N}\sup_{P_{\mathbf{X}}^{(N)}}I(\mathbf{X}^{(N)};\mathbf{Y}^{(N)}) \tag{1.6.20}$$

其中 $I(\mathbf{X}^{(N)};\mathbf{Y}^{(N)})$ 是随机输入和输出符号 $\mathbf{X}^{(N)}$ 和 $\mathbf{Y}^{(N)}$ 之间的互信息，$P_{\mathbf{X}}^{(N)}$ 是 $\mathbf{X}^{(N)}$ 的一个分布。

为了证明本题，只需要验证当 $\#\mathscr{U}^{(N)}=2^{N(\overline{R}+o(1))}$ 时，对于所有的 $f^{(N)}$ 与 $\hat{f}^{(N)}$，都有

$$e^{(N)}\geqslant 1-\frac{C_N+o(1)}{\overline{R}+o(1)} \tag{1.6.21}$$

其中

$$C_N=\frac{1}{N}\sup_{P_{\mathbf{X}}^{(N)}}I(\mathbf{X}^{(N)};\mathbf{Y}^{(N)})$$

实际上，如果 $\overline{R}>\limsup_{N\to\infty}C_N$，那么根据式(1.6.21)可以得出 $\liminf_{N\to\infty}\inf_{f^{(N)},\hat{f}^{(N)}}e^{(N)}>0$，$\overline{R}$ 是不可靠的。

为了不失一般性地证明式(1.6.21)，假设 $f^{(N)}$ 是无损的，那么输入码 $x^{(N)}$ 是等概率的，概率值为 $1/(\#\mathscr{U}^{(N)})$。对于所有译码规则 $\hat{f}^{(N)}$ 和任何足够大的 N，

$$\begin{aligned}NC_N&\geqslant I(\mathbf{X}^{(N)};\mathbf{Y}^{(N)})\geqslant I(\mathbf{X}^{(N)};\hat{f}(\mathbf{Y}^{(N)}))\\&\quad-h(\mathbf{X}^{(N)})-h(\mathbf{X}^{(N)}\mid\hat{f}(\mathbf{Y}^{(N)}))\\&=\log(\#\mathscr{U}^{(N)})-h(\mathbf{X}^{(N)}\mid\hat{f}(\mathbf{Y}^{(N)}))\\&\geqslant\log(\#\mathscr{U}^{(N)})-1-e^{(N)}\log(\#\mathscr{U}^{(N)}-1)\end{aligned} \tag{1.6.22}$$

118

这里最后的一个界由广义 Fano 不等式导出

$$\begin{aligned}h(\mathbf{X}^{(N)}\mid\hat{f}(\mathbf{Y}^{(N)}))&\leqslant-e^{(N)}\log e^{(N)}-(1-e^{(N)})\log(1-e^{(N)})+e^{(N)}\log(\#\mathscr{U}^{(N)}-1)\\&\leqslant 1+e^{(N)}\log(\#\mathscr{U}^{(N)}-1)\end{aligned}$$

现在，从式(1.6.22)中可得

$$NC_N\geqslant N[\overline{R}+o(1)]-1-e^{(N)}\log(2^{N(\overline{R}+o(1))}-1)$$

例如

$$e^{(N)}\geqslant\frac{N[\overline{R}-o(1)]-NC_N-1}{\log(2^{N(\overline{R}+o(1))}-1)}=1-\frac{C_N+o(1)}{\overline{R}+o(1)}$$

证毕。 □

问题 1.16 一个无记忆信道具有输入符号 0，1 以及输出符号 0，1 和 *（无法辨别）。给定信道矩阵

$$\boldsymbol{P}(0\mid 0)=1,\boldsymbol{P}(0\mid 1)=\boldsymbol{P}(1\mid 1)=\boldsymbol{P}(*\mid 1)=1/3$$

计算信道容量以及对应的输入概率 $p_X(0)$ 和 $p_X(1)$（在算出信道容量的基础上）。

有人建议，只有在输出为 1 时可能出现符号 *，亦可以把 * 当作 1；你可以从输出序列中得到更多的信息，它提高了信道的容量。你同意吗？证明你的答案。

解答　用公式

$$C=\sup_{p_X} I(X;Y)=\sup_{p_X}[h(Y)-h(Y|X)]$$

其中 p_X 是输入符号的分布：

$$p_X(0)=p, p_X(1)=1-p, 0\leqslant p\leqslant 1$$

所以，用 p 的函数计算 $I(X;Y)$，其中

$$h(Y)=-p_Y(0)\log p_Y(0)-p_Y(1)\log p_Y(1)-p_Y(*)\log p_Y(*)$$

其中

$$p_Y(0)=p+(1-p)/3=(1+2p)/3$$
$$p_Y(1)=p_Y(*)=(1-p)/3$$

并且

$$h(Y)=-\frac{1+2p}{3}\log\frac{1+2p}{3}-\frac{2(1-p)}{3}\log\frac{1-p}{3}$$

117

同样，

$$h(Y|X)=-\sum_{x=0,1}p_X(x)\sum_y P(y|x)\log P(y|x)$$
$$=-p_X(1)\log 1/3=(1-p)\log 3$$

所以

$$I(X;Y)=-\frac{1+2p}{3}\log\frac{1+2p}{3}-\frac{2(1-p)}{3}\log\frac{1-p}{3}-(1-p)\log 3$$

求导可得

$$\frac{\mathrm{d}}{\mathrm{d}p}I(X;Y)=-2/3(\log(1/3+2p/3)+2/3\log(1/3-p/3)+\log 3$$

所以，最大值 $\max I(X;Y)$ 可以从下面关系中得到：

$$\frac{2}{3}\log\frac{1-p}{1+2p}+\log 3=0$$

可进一步推出

$$\log\frac{1-p}{1+2p}=-\frac{3}{2}\log 3:=b$$

和

$$\frac{1-p}{1+2p}=2^b,\quad 即\ 1-2^b=p(1+2^{b+1})$$

（解该方程可以获得的）答案是

$$p=\frac{1-2^b}{1+2^{b+1}}$$

对于证明的最后一个部分，记

$$I(X;Y)=h(X)-h(X|Y)\leqslant h(X)-h(X|Y')=I(X;Y')$$

上式对于任意 Y'（Y 的函数）均成立；其中的等式当 Y 与 X 条件独立时成立。我们考虑的信道就是这种情况，所以建议中的容量并不会增加。 □

问题 1.17　(a) 给定一对离散的变量 X，Y，定义其联合熵 $h(X,Y)$ 和条件熵 $h(X|Y)$。

(b) 证明 $h(X,Y)\geqslant h(X|Y)$，并且解释等式成立的条件。

(c) 令 $0<\delta<1$，证明

$$h(X|Y) \geqslant (\log(\delta^{-1}))\mathbb{P}(q(X,Y) \leqslant \delta)$$

118 其中 $q(x, y)=\mathbb{P}(X=x|Y=y)$。讨论当 δ，X，Y 分别为何值时等式成立？

解答 (a) 条件熵的定义为

$$h(X|Y)=-\mathbb{E}\log q(x,y)=-\sum_{x,y}\mathbb{P}(X=x,Y=y)\log q(x,y)$$

其中

$$q(x,y)=\mathbb{P}(X=x|Y=y)$$

联合熵的定义为

$$h(X,Y)=-\sum_{x,y}\mathbb{P}(X=x,Y=y)\log\mathbb{P}(X=x,Y=y)$$

(b) 从定义中可得

$$h(X,Y)=h(X|Y)-\sum_{y}\mathbb{P}(Y=y)\log\mathbb{P}(Y=y)\geqslant h(X|Y)$$

(b)中的等式当 $h(Y)=0$ 时成立，即 Y 几乎肯定是一个常数。

(c) 利用 Chebyshev 不等式可得

$$\begin{aligned}\mathbb{P}(q(X,Y)\leqslant\delta)&=\mathbb{P}(-\log q(X,Y)\geqslant\log 1/\delta))\\&\leqslant\frac{1}{\log 1/\delta}\mathbb{E}[-\log q(X,Y)]=\frac{1}{\log 1/\delta}h(X|Y)\end{aligned}$$

这里等式基于下列条件成立

$$\mathbb{P}(q(X,Y)=\delta)=1$$

这要求：(i)$\delta=1/m$，其中 m 是一个正整数；(ii)对于所有 $y\in Y$ 的支集，存在一个势为 m 的集合 A_y，使得

$$\mathbb{P}(X=x|Y=y)=\frac{1}{m},\quad x\in A_y \qquad \square$$

问题 1.18 一个文本由 Bernoulli 信源生成，该信源包含字母 1，2，…，m，概率分别为 p_1，p_2，…，p_m。希望通过一个无记忆二元对称信道(MBSC)可靠地传输这个文本，其中行误码率为 p^*。解释信道容量的含义，并且证明

$$C=1-h(p^*,1-p^*)$$

解释为什么在如下条件下可靠通信是可以实现的：

$$h(p_1,p_2,\cdots,p_m)+h(p^*,1-p^*)<1$$

而在如下条件下可靠通信是不可实现的：

$$h(p_1,p_2,\cdots,p_m)+h(p^*,1-p^*)>1$$

119 其中 $h(p_1,p_2,\cdots,p_m)=-\sum_{i=1}^{m}p_i\log_2 p_i$。

解答 Bernoulli 信源的渐近均分性质表明，由信源发射的长度为 n 的离散典型序列的数量是 $2^{nH+o(n)}$，它们的概率近似相同，都为 $2^{-nH+o(n)}$：

$$\lim_{n\to\infty}\mathbb{P}(2^{-n(H-\varepsilon)}\leqslant\mathbb{P}_n(\boldsymbol{U}^{(n)})\leqslant 2^{-n(H-\varepsilon)})=1$$

其中 $H=h(p_1,p_2,\cdots,p_m)$。

记

$$T_n(=T_n(\varepsilon))=\{\boldsymbol{u}^{(n)}:2^{-n(H-\varepsilon)}\leqslant\mathbb{P}_n(\boldsymbol{u}^{(n)})\leqslant 2^{-n(H-\varepsilon)}\}$$

可观察到

$$\lim_{n\to\infty}\frac{1}{n}\log\#T_\varepsilon = H,\quad 即\lim_{n\to\infty}\sup\frac{1}{n}\log\#T_\varepsilon < H+\varepsilon$$

由信道容量的定义可得，信源符号 $u^{(n)}\in T_n(\varepsilon)$ 可由二进制长度为 $\overline{R}^{-1}(H+\varepsilon)$ 的码字编码，并能可靠地通过一个无记忆的对称信道发送，该信道矩阵为

$$\begin{bmatrix} 1-p^* & p^* \\ p^* & 1-p^* \end{bmatrix}$$

对于任意 $\overline{R}<C$，其中

$$C=\sup_{p_X} I(X;Y)=\sup_{p_X}[h(Y)-h(Y|X)]$$

此处的上确界考虑对于所有输入为二进制符号的分布

$$p_X=(p_X(0),p_X(1))$$

而输出符号 Y 的条件分布为

$$\mathbb{P}(Y=y|X=x)=\begin{cases}1-p^*, & y=x\\ p^*, & y\neq x\end{cases}$$

可得

$$h(Y|X)=-p_X(0)[-(1-p^*)\log(1-p^*)-p^*\log p^*]$$
$$+p_X(1)[-p^*\log p^*-(1-p^*)\log(1-p^*)]=h(p^*,1-p^*)$$

独立于 p_X，所以

$$C=\sup_{p_X}h(Y)-h(p^*,1-p^*)=1-h(p^*,1-p^*)$$

因为对于

$$p_Y(0)=p_Y(1)=1/2，当\ p_X(0)=p_X(1)=1/2\ 时$$

$h(Y)$ 可得，所以，如果

$$H<C\quad\Leftrightarrow\quad h(p_1,\cdots,p_n)+h(p^*,1-p^*)<1$$

那么对于足够小的 $\varepsilon>0$ 和接近容量 C 的 $\overline{R}<C$，$\overline{R}^{-1}(H+\varepsilon)$ 可小于 1，这表明存在一个长度为 n 的码字 f_n，当利用编码的 f_n 和响应的 ML 译码器时，误码率是

$$\leqslant \mathbb{P}(u^{(n)}\notin T_n)+\mathbb{P}(u^{(n)}\in T_n;当利用\ f_n(u^{(n)})\ 和\ ML\ 译码器时的误码率)$$

当 $n\to\infty$ 时，误码率 $\to 0$。

另一方面，

$$H>C\quad\Leftrightarrow\quad h(p_1,\cdots,p_n)+h(p^*,1-p^*)>1$$

那么对于所有 $\overline{R}<C$ 均有 $\overline{R}^{-1}H>1$，也就意味着我们不能用码字长度为 n 的码字 $u^{(n)}\in T_n$ 进行编码以使误码率趋于 0。所以，可靠传输不可能实现。 □

问题 1.19 一个 Markov 信源有一个包含 m 个字符的字母表，状态转移矩阵为 P_m，其中的元素 p_{jk} 如下定义

$$p_{11}=p_{mm}=2/3,\quad p_{jj}=1/3(1<j<m)$$
$$p_{j,j+1}=1/3(1\leqslant j<m),\quad p_{j,j-1}=1/3(1<j\leqslant m)$$

除了上述元素外，其他元素均为 0。计算信源的信息速率。

给出如上定义的状态转移矩阵 P_m 的具体形式。考虑另一个信源，其字母表包含 $m+n$ 个字符，其状态转移矩阵为 $\begin{bmatrix}P_m & 0\\ 0 & P_n\end{bmatrix}$，该矩阵中的零表示具有合适尺寸的全零矩阵。假设初始字符在字母表内均匀分布。试计算信源的信息速率是多少？ 121

解答 注意到状态转移矩阵

$$P_m = \begin{bmatrix} 2/3 & 1/3 & 0 & 0 & \cdots & 0 \\ 1/3 & 1/3 & 1/3 & 0 & \cdots & 0 \\ 0 & 1/3 & 1/3 & 1/3 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & 2/3 \end{bmatrix}$$

是 Hermitian 矩阵，所以具有平稳分布 $\pi=(\pi_i)$，$\pi_i=1/m$，$1\leqslant i\leqslant m$(均匀分布)。该信源的信息速率等于

$$H_m = -\sum_{i,k}\pi_i p_{ik}\log p_{ik} = -\frac{1}{m}\left(2\left(\frac{2}{3}\log\frac{2}{3}+\frac{1}{3}\log\frac{1}{3}\right)+3(m-2)\frac{1}{3}\log\frac{1}{3}\right)$$
$$= \log 3 - \frac{4}{3m}$$

状态转移矩阵为 $\begin{pmatrix} P_m & 0 \\ 0 & P_n \end{pmatrix}$ 的信源不是遍历的，其信息速率是如下两个速率的最大值

122

$$\max\{H_m, H_n\} = H_{\max\{m,n\}}.$$ □

问题 1.20 考虑一个取自有限字母表的信源。定义 $J_n=n^{-1}h(U^{(n)})$，$K_n=h(U_{n+1}|U^{(n)})$，$n=1,2,\cdots$。这里的 U_n 是序列的第 n 个符号，$U^{(n)}$ 是由前 n 个符号组成的一个序列，$h(U^{(n)})$ 是该序列的熵而 $h(U_{n+1}|U^{(n)})$ 是给定序列的条件符号熵。证明如果信源是平稳的，那么 J_n 和 K_n 是非递增的且有一个公共极限。

假设一个信源是 Markov 信源但不一定平稳。证明 U_1 与 U_3 的互信息不小于 U_1 与 U_2 的互信息。

解答 对于第二部分，由信源的 Markov 性可知

$$\mathbb{P}(U_1=u_1|U_2=u_2,U_3=u_3)=\mathbb{P}(U_1=u_1|U_2=u_2)$$

因此

$$I(U_1;(U_2,U_3)) = \mathbb{E}\left(-\log\frac{\mathbb{P}(U_1=u_1|U_2=u_2,U_3=u_3)}{\mathbb{P}(U_1=u_1)}\right)$$
$$= \mathbb{E}\left(-\log\frac{\mathbb{P}(U_1=u_1|U_2=u_2)}{\mathbb{P}(U_1=u_1)}\right) = I(U_1;U_2)$$

因为

$$I(U_1;(U_2,U_3))\geqslant I(U_1;U_3)$$

可得结果。 □

问题 1.21 针对概率如下的五个信息构造一个 Huffman 编码：

信息	1	2	3	4	5
概率	0.1	0.15	0.2	0.26	0.29

解答

信息	1	2	3	4	5
概率	0.1	0.15	0.2	0.026	0.029
码字	101	100	11	01	00

平均码长为 2.4。 □

问题 1.22 阐述第一编码定理(FCT)，该定理能度量一个具有某种长期性质的信源的信息速率。请从渐近均分性的角度解释 FCT，并且给出 Bernoulli 信源的信息速率。

考虑这样一个 Bernoulli 信源，其发射符号 0，1 的概率分别为 $1-p$ 和 p，其中 $0<p<1$。

令 $\eta(p)=-p\log p-(1-p)\log(1-p)$，而 $\varepsilon>0$ 为某一定值。令 $\boldsymbol{U}^{(n)}$ 为信源发射的前 n 个符号构成的序列。证明存在这样一个集合 S_n，它包含了 $\boldsymbol{U}^{(n)}$ 可能的取值，且满足下面的概率不等式

$$\mathbb{P}(\boldsymbol{U}^{(n)}\in S_n)\geqslant 1-\left(\log\frac{p}{1-p}\right)^2\frac{p(1-p)}{n\varepsilon^2}$$

所以对于每个 $\boldsymbol{u}^{(n)}\in S_n$，概率 $\mathbb{P}(\boldsymbol{U}^{(n)}=\boldsymbol{u}^{(n)})$ 的取值在 $2^{-n(h+\varepsilon)}$ 和 $2^{-n(h-\varepsilon)}$ 之间。

解答　对于 Bernoulli 信源下列关系成立：

$$-\frac{1}{n}\log P_n(\boldsymbol{U}^{(n)})=-\frac{1}{n}\sum_{1\leqslant j\leqslant n}\log P(U_j)\to\eta(p)$$

即对于所有 $\varepsilon>0$，由 Chebyshev 不等式有

$$\mathbb{P}\left(\left|-\frac{1}{n}\log P_n(\boldsymbol{U}^{(n)})-h\right|>\varepsilon\right)\leqslant\frac{1}{\varepsilon^2n^2}\mathrm{Var}\Big(\sum_{1\leqslant j\leqslant n}\log P(U_j)\Big)$$

$$=\frac{1}{\varepsilon^2n}\mathrm{Var}[\log P(U_1)]\qquad(1.6.23)$$ 123

这里

$$P(U_j)=\begin{cases}1-p, & U_j=0,\\ p, & U_j=1,\end{cases}\quad P_n(\boldsymbol{U}^{(n)})=\prod_{1\leqslant j\leqslant n}P(U_j)$$

和

$$\mathrm{Var}\Big(\sum_{1\leqslant j\leqslant n}\log P(U_j)\Big)=\sum_{1\leqslant j\leqslant n}\mathrm{Var}(\log P(U_j))$$

其中

$$\begin{aligned}\mathrm{Var}(\log P(U_j))&=\mathbb{E}[\log P(U_j)]^2-[\mathbb{E}\log P(U_j)]^2\\&=p(\log p)^2+(1-p)(\log(1-p))^2-(p\log p+(1-p)\log(1-p))^2\\&=p(1-p)\left(\log\frac{p}{1-p}\right)^2\end{aligned}$$

所以，由(1.6.23)式中的界可得

$$\mathbb{P}(2^{-n(h+\varepsilon)}\leqslant P_n(\boldsymbol{U}^{(n)})\leqslant 2^{-n(h-\varepsilon)})\geqslant 1-\frac{1}{n\varepsilon^2}p(1-p)\left(\log\frac{p}{1-p}\right)^2$$

基于上述推导，现在可以设

$$S_n=\{\boldsymbol{u}^{(n)}=u_1\cdots u_n:2^{-n(h+\varepsilon)}\leqslant\mathbb{P}(\boldsymbol{U}^{(n)}=\boldsymbol{u}^{(n)})\leqslant 2^{-n(h-\varepsilon)}\}$$

由此可以得到结果。 □

问题 1.23　考虑用具有 q 个码字的码本对字母表 $\{1,2,\cdots,m\}$ 中的符号进行编码($q<m$)。针对一个码字长度为 $s_1,\cdots,s_q$ 的可译码阐述 Kraft 不等式的含义。假设一个信源从字母表 $\{1,2,\cdots,m\}$ 中选择发射符号，已知每个符号发射的概率为 $P_i>0$。记 S 为这符号的信源编码时信源输出的随机码长。期望找到一个能最小化 q^S 平均值的可译码。尝试证明下界 $\mathbb{E}(q^S)\geqslant\Big(\sum_{1\leqslant i\leqslant m}\sqrt{p_i}\Big)^2$，并解释在什么条件下等式成立。

同时，证明对于上述准则的最优码必须满足 $\mathbb{E}(q^S)<q\Big(\sum_{1\leqslant i\leqslant m}\sqrt{p_i}\Big)^2$。

提示：利用 Cauchy-Schwarz 不等式，对于所有的正值 x_i，y_i，

$$\sum_{1\leqslant i\leqslant n}x_iy_i\leqslant\Big(\sum_{1\leqslant i\leqslant n}x_i^2\Big)^{1/2}\Big(\sum_{1\leqslant i\leqslant n}y_i^2\Big)^{1/2}$$

当对于所有 i 都有 $x_i=cy_i$ 时，上式中的等式成立。 124

解答　由 Cauchy-Schwarz 不等式，

$$\sum_{1\le i\le m} p_i^{1/2} = \sum_{1\le i\le m} p_i^{1/2} q^{s_i/2} q^{-s_i/2} \le \Big(\sum_{1\le i\le m} p_i q^{s_i}\Big)^{1/2} \Big(\sum_{1\le i\le m} q^{-s_i}\Big)^{1/2} \le \Big(\sum_{1\le i\le m} p_i q^{s_i}\Big)^{1/2}$$

由 Kraft 不等式 $\sum_{1\le i\le m} q^{-s_i} \le 1$，可得

$$\mathrm{E}q^{s} = \sum_{1\le i\le m} p_i q^{s_i} \ge \Big(\sum_{1\le i\le m} p_i^{1/2}\Big)^2$$

现在把概率 p_i 设为

$$p_i = (\alpha q^{-x_i})^2, x_i > 0$$

其中 $\sum_{1\le i\le m} q^{-x_i} = 1$（所以，$\sum_{1\le i\le m} p_i^{1/2} = c$）。令 s_i 为大于等于 x_i 的最小整数。基于 $\sum_{1\le i\le m} q^{-s_i} \le 1$ 和 Kraft 不等式可知，存在一个长度为 s_i 的可译码。对于这个码，$q^{s_i-1} < q^{x_i} = c/p_i^{1/2}$，由此可得

$$\mathrm{E}q^{s} = \sum_{1\le i\le m} p_i q^{s_i} = q\sum_{1\le i\le m} p_i q^{s_i-1} < q\sum_{1\le i\le m} p_i q^{x_i} = qc\sum_{1\le i\le m} p_i^{1/2} = q\Big(\sum_{1\le i\le m} p_i^{1/2}\Big)^2 \qquad \square$$

问题 1.24 一个信息速率为 H 的 Bernoulli 信源通过一个传输畅通或不通的字符对字符的线路。如果线路畅通，那么当一个字符被传输时，接收成功。如果线路不畅通，那么接收机只知道此线路不通。线路在畅通和不通的两个状态中转变，那么此线路服从 Markov 链（DTMC），转移概率是常数，且独立于已经传输的文本。

证明由接收信号组成的信源的信息速率为 $H_L + \pi_\perp H_S$，其中 H_S 是信号，$H_\perp$ 是由线路的函数限制的 DTMC 的信息速率，$\pi_\perp$ 是畅通线路的平稳概率分布。 □

解答 以概率 p_j 发送符号 $j=1, 2, \cdots$ 的 Bernoulli 信源的信息速率是 $H = -\sum_j p_j \log p_j$。

125 线路的状态是一个 DTMC，其 2×2 的状态转移矩阵表示如下

$$\begin{array}{c} \text{不通} \\ \text{畅通} \end{array}\begin{pmatrix} 1-\alpha & \alpha \\ \beta & 1-\beta \end{pmatrix}$$

并且平稳分布为

$$1-\pi_\perp(\text{不通}) = \frac{\beta}{\alpha+\beta}, \quad \pi_\perp(\text{畅通}) = \frac{\alpha}{\alpha+\beta}$$

（假设 $\alpha+\beta>0$）。接收信号序列服从 DTMC，其状态记为 0（不通），1，2，…，状态转移概率为

$$q_{00} = 1-\alpha, \qquad q_{0j} = \alpha p_j, \qquad j,k \ge 1$$
$$q_{j0} = \beta, \qquad q_{jk} = (1-\beta)p_k$$

上述状态转移链具有唯一的平稳分布

$$\pi_{RS}(0) = \frac{\beta}{\alpha+\beta}, \quad \pi_{RS}(j) = \frac{\alpha}{\alpha+\beta}p_j, \quad j \ge 1$$

那么接收信号的信息速率等于

$$\begin{aligned} H_{RS} &= -\sum_{j,k\ge 0} \pi_{RS}(j) q_{jk} \log q_{jk} = -\frac{\beta}{\alpha+\beta}\Big((1-\alpha)\log(1-\alpha) + \sum_{j\ge 1} \alpha p_j \log(\alpha p_j)\Big) \\ &\quad -\frac{\alpha}{\alpha+\beta}\Big(\sum_{j\ge 1} p_j\Big(\beta\log\beta + (1-\beta)\sum_{k\ge 1} p_k \log(1-\beta)p_k\Big)\Big) \\ &= H_\perp + \frac{\beta}{\alpha+\beta}H_S \end{aligned}$$

其中，$H_\perp$ 是线路状态 DTMC 的熵速率

$$H_\perp = -\frac{\beta}{\alpha+\beta}[(1-\alpha)\log(1-\alpha) + \alpha\log\alpha] - \frac{\alpha}{\alpha+\beta}[(1-\beta)\log(1-\beta) + \beta\log\beta]$$

$\pi=\alpha/(\alpha+\beta)$。 □

问题 1.25 考虑一个 Bernoulli 信源，其中一个信源符号取值为 i 的概率为 $p_i(i=1,\cdots,m)$。令 n_i 为 n 长序列 $u^{(n)}=u_1u_2\cdots u_n$ 中出现取值为 i 的符号个数。令 A_n 是至少以概率 $1-\varepsilon$ 包含序列 $u^{(n)}$ 的最小集合。证明 A_n 中的每一个序列都满足下面不等式

$$-\sum_i n_i\log p_i \leqslant nh+(nk/\varepsilon)^{1/2}$$ 128

其中 k 是一个独立于 n 或者 ε 的常数。对于 Markov 信源阐述(不需要证明)类似定理。

解答 对于一个包含信源符号 1，2，…，m 的 Bernoulli 信源，一个给定信源序列 $u^{(n)}=u_1u_2\cdots u_n$ 的概率是

$$\mathbb{P}(U^{(n)}=u^{(n)})=\prod_{1\leqslant i\leqslant m}p_i^{n_i}$$

集合 A_n 由出现概率最大的那些序列组成(这些序列依照概率值递减的顺序选出)，即具有最大值的 $\log\mathbb{P}(U^{(n)}=u^{(n)})=\sum_{1\leqslant i\leqslant m}n_i\log p_i$。因此，对某些(实数)$c$，

$$A_n=\{u^{(n)}:-\sum_i n_i\log p_i\leqslant c\}$$

为了确定 c，我们利用

$$\mathbb{P}(A_n)\geqslant 1-\varepsilon$$

所以，c 的取值将满足

$$\mathbb{P}\left(u^{(n)}:-\sum_{1\leqslant i\leqslant m}n_i\log p_i\geqslant c\right)<\varepsilon$$

现在，对于随机序列 $U^{(n)}=U_1\cdots U_n$，令 N_i 表示其中 i 出现的次数。那么

$$-\sum_{1\leqslant i\leqslant m}N_i\log p_i=\sum_{1\leqslant j\leqslant n}\theta_j,\quad 其中\ \theta_j=-\log p_i,\quad 当\ U_j=i\ 时$$

因为 U_j 是独立同分布的，所以随机变量 θ_j 也是独立同分布的。接着有

$$\mathbb{E}\theta_j=-\sum_{1\leqslant i\leqslant m}p_i\log p_i=h$$

并且

$$\mathrm{Var}\theta_j=\mathbb{E}(\theta_j)^2-(\mathbb{E}\theta_j)^2=\sum_{1\leqslant i\leqslant m}p_i(\log p_i)^2-\left(\sum_{1\leqslant i\leqslant m}p_i\log p_i\right)^2=v$$

那么

$$\mathbb{E}\left(\sum_{1\leqslant j\leqslant n}\theta_j\right)=nh\quad 且\quad \mathrm{Var}\left(\sum_{1\leqslant j\leqslant n}\theta_j\right)=nv$$

$h=H$ 是信源的信息速率。 127

根据 Chebyshev 不等式，对于所有的 $b>0$，

$$\mathbb{P}\left(\left|-\sum_{1\leqslant i\leqslant m}N_i\log p_i-nh\right|>b\right)\leqslant\frac{nv}{b^2}$$

在 $b=\sqrt{nk/\varepsilon}$ 时，我们可以得到

$$\mathbb{P}\left(\left|-\sum_{1\leqslant i\leqslant m}N_i\log p_i-nh\right|>\sqrt{\frac{nk}{\varepsilon}}\right)\leqslant\varepsilon$$

因此，对于所有的 $u^{(n)}\in A_n$，

$$-\sum_{1\leqslant i\leqslant m}n_i\log p_i\leqslant nh+\sqrt{\frac{nk}{\varepsilon}}:=c$$

对于一个不可约非周期的 Markov 源，结论是相似的，且

$$H=-\sum_{1\leqslant i,j\leqslant m}\pi_ip_{ij}\log p_{ij}$$

并且 $v_0 \geq 0$ 是由 $v = \lim_{n\to\infty} \sup \frac{1}{n} \mathrm{Var}\left(\sum_{1\leq i\leq n} \theta_i\right)$ 给定的一个常量。 □

问题 1.26 证明一个有效可译的无噪编码过程可引出满足一可达成能度量。

选择符号集 $\mathcal{C}_a = \{0, 1, \cdots, a-1\}$ 中长度为 $s_i (i=1, 2, \cdots, n)$ 的码字来最小化平均码长 $\sum_{i=1}^{n} p_i s_i$。这种方法不仅受到可译性的约束，还受到 $\sum_{i=1}^{n} q_i s_i$ 不能超过一个给定界的约束。在这里，q_i 对于原始符号集的公设概率分布 $\{p_i\}$ 特性来说，是一个可行的选择。给出 $\sum_{i=1}^{n} p_i s_i$ 最小值的界。

解答 如果我们忽略 $s_1, \cdots, s_n$ 是正整数的条件，这个最小化问题就变成了

$$\begin{aligned} &\text{最小化} \sum_i s_i p_i \\ &\text{约束于 } s_i \geq 0 \quad \text{和} \quad \sum_i a^{-s_i} \leq 1 (\text{Kraft}) \end{aligned} \tag{1.6.24}$$

该问题可以通过 Lagrange 方法解决，Lagrange 式子为

128

$$\mathscr{L}(s_1, \cdots, s_n, \lambda) = \sum_{1\leq i\leq n} s_i p_i - \lambda\left(1 - \sum_{1\leq i\leq n} a^{-s_i}\right)$$

这个松弛后的优化问题其有唯一解，由下式给出

$$s_i = -\log_a p_i, 1 \leq i \leq n \tag{1.6.25}$$

这里，松弛最优值

$$v_{\mathrm{rel}} = -\sum_{1\leq i\leq n} p_i \log_a p_i = h$$

为最优平均码长 $\sum_i s_i^* p_i$ 提供了一个下界

$$h \leq \sum_i s_i^* p_i$$

现在考虑另一个约束条件

$$\sum_{1\leq i\leq n} q_i s_i \leq b \tag{1.6.26}$$

式(1.6.24)中松弛后的优化问题加上式(1.6.26)中的约束条件同样能够通过 Lagrange 方法解决。在这里，如果

$$-\sum_i q_i \log_a p_i \leq b$$

那么添加新的约束并不影响式(1.6.24)中的极小化变量，即最优正数 $s_1, \cdots, s_n$ 依然由式(1.6.25)给定，并且最优值是 h。另外，当 $-\sum_i q_i \log_a p_i > b$ 时，新的极小化变量 $\tilde{s}_1, \cdots, \tilde{s}_n$ 仍然是唯一的(因为问题仍然是强 Lagrange 的)并且同时满足

$$\sum_i a^{-\tilde{s}_i} = 1, \sum_i q_i \tilde{s}_i = b$$

在这两种情况下，对于新的松弛问题的最优值 $\tilde{v}_{\mathrm{rel}}$ 满足 $h \leq \tilde{v}_{\mathrm{rel}}$。

最后，对于整数值码字长问题

$$\begin{aligned} &\text{最小化} \sum_i s_i p_i \\ &\text{约束于 } s_i \geq 1 \text{ 且为整数}, \sum_i a^{-s_i} \leq 1, \quad \sum_i q_i s_i \leq b \end{aligned} \tag{1.6.27}$$

的解 $\tilde{s}_1^*, \cdots, \tilde{s}_n^*$ 将满足

$$h \leqslant \bar{\mathcal{V}}^{(n)} \leqslant \sum_i \bar{s}_i p_i, \quad \sum_i \bar{s}_i q_i \leqslant b \qquad \square$$

问题 1.27 假设一个离散 Markov 信源$\{X_i\}$的状态转移概率是

$$p_{jk} = \mathbb{P}(X_{i+1} = k \mid X_i = j)$$

129

其平稳分布为(π_j)。假设信源符号可以被噪声以$\beta = 1 - \alpha$的概率擦除(在这个方案中，我们仅能看到事件“擦除”)，该事件与之前的符号或噪声独立。试证明被噪声影响的信源有如下的信息速率

$$-\alpha\log\alpha - \beta\log\beta - \alpha^2 \sum_j \sum_k \sum_{s\geqslant 1} \pi_j \beta^{s-1} p_{jk}^{(s)} \log p_{jk}^{(s)}$$

其中$p_{jk}^{(s)}$是原始 DTMC 的s步转移概率。

解答 将受噪声影响的信源序列用$\{\tilde{X}_i\}$表示，当擦除发生时$\tilde{X}_i = *$(一个污点)。相应地，可以通过将原始 Markov 信源序列x_1^n中被噪声影响的字符用污点置换来获得有噪声影响的信源序列$\tilde{x}_1^n$。该序列的概率$p_*(\tilde{x}) = \mathbb{P}(\tilde{X}_1^n = \tilde{x}_1^n)$可表示为

$$\sum_{x_1^n : \tilde{x}_1^n = *} \mathbb{P}(X_1^n = x_1^n)\mathbb{P}(\tilde{X}_1^n \mid X_1^n = x_1^n) \tag{1.6.28}$$

并作为乘积展开计算，其初始因子为

$$\lambda_{x_1}\alpha \text{ 或者 } \sum_j \lambda_j p_{x_1 j}^{(s)} \beta^{s-1}\alpha, \quad \text{其中 } 1 < s \leqslant n \text{ 或者 } 1$$

这依赖于在$\tilde{x}_1^n$中第一个未擦除字符的位置(如果有的话)。接下来对式(1.6.28)有贡献的因子都有相似结构

$$p_{x_{s-1}x_s}\beta \text{ 或者 } p_{x_{s-1}x_s}^{(s)}\beta^{s-1}\alpha \text{ 或者 } 1$$

因此，字符串$\tilde{x}_1^n$携带的信息$-\log p_*(\tilde{x}_1^n)$可计算为

$$\begin{aligned}
&-\log\mathbb{P}(X_{s_1} = x_{s_1}) - (s_1 - 1)\log\beta - \log\alpha \\
&-\log p_{x_{s_1}x_{s_2}}^{(s_2-s_1)} - (s_2 - s_1 - 1)\log\beta - \log\alpha - \cdots \\
&-\log p_{x_{s_{N-1}}x_{s_N}}^{(s_N - s_{N-1})} - (s_N - s_{N-1} - 1)\log\beta - \log\alpha
\end{aligned}$$

其中$1 \leqslant s_1 < \cdots < s_N \leqslant n$表示在$\tilde{x}_1^n$中未擦除符号连续出现的时间长度。

现在取$-\frac{1}{n}\log p_*(\tilde{X}_1^n)$，即由随机字符串$\tilde{X}_1^n$产生的信息速率。忽略初始比特，我们可以得到

$$-\frac{1}{n}\log p_*(\tilde{X}_1^n) = -\frac{N(\beta)}{n}\log\beta - \frac{N(\alpha)}{n}\log\alpha - \sum_i \frac{M(i,j;s)}{n}\log p_{ij}^{(s)}$$

130

这里

$$\begin{aligned}
N(\alpha) &= \text{在 } \tilde{X}_1^n \text{ 中未擦除数据的数量} \\
N(\beta) &= \text{在 } \tilde{X}_1^n \text{ 中被擦除数据的数量} \\
M(i,j;s) &= \text{在 } \tilde{X}_1^n \text{ 中长度为 } s+1 \text{ 的一系列数据 } i * \cdots * j
\end{aligned}$$

当$n \to \infty$时，极限频率将收敛(满足大数定理)：

$$\frac{N(\alpha)}{n} \to \alpha, \frac{N(\beta)}{n} \to \beta, \frac{M(i,j;s)}{n} \to \alpha\beta^{s-1}\pi_i p_{ij}^{(s)}\alpha$$

这使得

$$-\frac{1}{n}\log p_*(\tilde{X}_1^n) \to -\alpha\log\alpha - \beta\log\beta - \alpha^2 \sum_{i,j} \pi_i \sum_{s\geqslant 1} \beta^{s-1} p_{ij}^{(s)} \log p_{ij}^{(s)}$$

(上述收敛在概率意义上说是几乎肯定的)，根据 SCT，该极限值给出了受噪声影响的信源信息速率。 □

问题 1.28 一个二进制信源根据如下转移概率

$$P(X_i = k \mid X_{i-1} = j, X_{i-2} = i) = q_r$$

发出符号 0 或者 1，这里，k，j，i 以及 r 取值为 0 或者 1，$r = k - j - i \bmod 2$，$q_0 + q_1 = 1$。请给出上述信源的信息速率。

同时也请给出信源符号为 0 和 1 而发送概率为 q_0 和 q_1 的 Bernoulli 信源的信息速率。解释这两种信源信息速率之间的关联。

解答 将条件概率详尽地重写如下：

$$P(X_i = 0 \mid X_{i-1} = j, X_{i-2} = i) = \begin{cases} q_0, & i = j \\ q_1, & i \neq j \end{cases}$$

$$P(X_i = 1 \mid X_{i-1} = j, X_{i-2} = i) = \begin{cases} q_1, & i = j \\ q_0, & i \neq j \end{cases}$$

这个信源是在{0，1}上的二阶 Markov 链，即有着四种状态{00，01，10，11}的 DTMC。4×4 的转移矩阵为

$$\begin{array}{c} 00 \\ 01 \\ 10 \\ 11 \end{array} \begin{bmatrix} q_0 & q_1 & 0 & 0 \\ 0 & 0 & q_1 & q_0 \\ q_1 & q_0 & 0 & 0 \\ 0 & 0 & q_0 & q_1 \end{bmatrix}$$

[131]

该信源的平稳分布是

$$\pi_{11} = \pi_{01} = \pi_{10} = \pi_{11} = \frac{1}{4}$$

而信息速率的计算可以采用如下的标准方法

$$H = -\sum_{\alpha,\beta=0,1} \pi_{\alpha\beta} \sum \sum_{\alpha=0,1} \pi_{\alpha\beta} p_{\alpha\beta,\beta\gamma} \log p_{\alpha\beta,\beta\gamma}$$

最终结果是

$$\frac{1}{2} \sum_{\alpha,\beta} h(q_0, q_1) = -q_0 \log q_0 - q_1 \log q_1$$ □

问题 1.29 一个连接离散无记忆信道的输入端有三个信源符号，分别是 1，2 和 3。字符 j 被接收为 $j-1$ 的概率是 p，被接收为 $j+1$ 的概率是 p，被接收为 j 的概率是 $1-2p$，输出端的输出符号取自包含从 0 到 4 的符号集。针对一般化的 p，尽可能明晰地确定最优输入分布。此外，当 $p=0$，$p=1/3$，$p=1/2$ 时，分别计算出信道容量。

解答 信道矩阵是 3×5 的：

$$\begin{array}{c} 1 \\ 2 \\ 3 \end{array} \begin{bmatrix} p & (1-2p) & p & 0 & 0 \\ 0 & p & (1-2p) & p & 0 \\ 0 & 0 & p & (1-2p) & p \end{bmatrix}$$

矩阵中的每一行都是其他行的置换，所以容量等于

$$C = \max_{P_X} [h(Y) - h(Y \mid X)] = (\max_{P_X} h(Y)) + [2p\log p + (1-2p)\log(1-2p)]$$

针对输入字符分布 P_X 的最大化仅对输出符号的熵 $h(Y)$ 起作用。

接下来，

$$h(Y) = -\sum_{y=0,1,2,3,4} P_Y(y) \log P_Y(y)$$

在这里

$$\left.\begin{aligned}
P_Y(0) &= P_X(1)p \\
P_Y(1) &= P_X(1)(1-2p)+P_X(2)p \\
P_Y(2) &= P_Y(1)p+P_X(2)(1-2p)+P_X(3)p \\
P_Y(3) &= P_X(3)(1-2p)+P_X(2)p \\
P_Y(4) &= P_X(3)p
\end{aligned}\right\} \tag{1.6.29}$$

132

上式中的对称性表明当 $P_X(0)=P_X(2)=q$ 以及 $P_X(1)=1-2q$ 时，$h(Y)$ 可取最大值，所以：

$$\begin{aligned}
\max h(Y) = \max\big[&-2qp\log(qp)-2[q(1-2p)+(1-2q)p] \\
&\times\log[q(1-2p)+(1-2q)p] \\
&-[2qp+(1-2q)(1-2p)]\log[2qp+(1-2q)(1-2q)]
\end{aligned}$$

为了找到该最大值，可以通过求导数解决：

$$\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}q}h(Y) &= -2p\log(qp)-2p-2(1-4p)\log[q(1-2p)+(1-2q)p] \\
&\quad -2(1-4p)-(2p-2)\log[2qp+(1-2q)(1-2p)]-(2p-2) \\
&= 4p-2p\log(qp)-2(1-4p)\log[q(1-2p)+(1-2q)p] \\
&\quad -2(1-4p)-2(p-1)\log[2qp+(1-2q)(1-2p)]=0
\end{aligned}$$

$p=0$ 对应一个完美的无差错信道，它的容量是 $\log 3$，并且当输入分布为 $P_X(1)=P_X(2)=P_X(3)=1/3$ 即 $q=1/3$ 时达到容量，而且

$$P_Y(1)=P_Y(2)=P_Y(3)=1/3, P_Y(0)=P_Y(4)=0$$

当 $p=1/3$ 时，输出符号概率是

$$P_Y(0)=P_Y(4)=q/3, P_Y(1)=(1-q)/3, P_Y(2)=1/3$$

$h(Y)$ 经过简化后可得

$$h(Y)=-2\frac{q}{3}\log\frac{q}{3}-2\frac{1-q}{3}\log\frac{1-q}{3}-\frac{1}{3}\log\frac{1}{3}$$

导数 $\mathrm{d}h(Y)/\mathrm{d}q$ 对应

$$-\frac{2}{3}\log\frac{q}{3}-\frac{2}{3}+\frac{2}{3}\log\frac{1-q}{3}+\frac{2}{3}$$

当 $q=1/2$ 时就消失了，即

$$P_X(1)=P_X(3)=1/2, P_X(2)=0,$$
$$P_Y(0)=P_Y(1)=P_Y(3)=P_Y(4)=1/6, P_Y(2)=1/3$$

接下来，条件熵为

$$h(Y|X)=\log 3$$

基于上式，可以得到信道的容量

$$C=-\frac{2}{3}\log\frac{1}{6}-\frac{1}{3}\log\frac{1}{3}-\log 3=\frac{2}{3}$$

最后，对于 $p=1/2$，我们有 $h(Y|X)=1$ 以及

$$P_Y(0)=P_Y(4)=\frac{q}{2}, P_Y(1)=P_Y(3)=\frac{1-2q}{2}, P_Y(2)=q$$ 133

输出熵为

$$h(Y)=-q\log\frac{q}{2}-\frac{1-2q}{2}\log\frac{1-2q}{2}-q\log q$$

$$= q - 2q\log_{\mathbb{R}} q - \frac{1-2q}{2}\log_{\mathbb{R}}\frac{1-2q}{2}$$

当 $q=1/6$ 时，熵最大化，即

$$P_X(1)=P_X(2)=\frac{1}{6}, P_X(3)=\frac{2}{3},$$

$$P_Y(0)=P_Y(4)=\frac{1}{12}, P_Y(1)=P_Y(3)=\frac{1}{3}, P_Y(2)=\frac{1}{6}$$

这个方案里，容量

$$C=\log 3-\frac{1}{2}$$

□

问题 1.30 输入非负整数值 X，无记忆离散时间信道输出为 Y，则

$$Y=\varepsilon X$$

这里 ε 与 X 独立，$\mathbb{P}(\varepsilon=1)=p$，$\mathbb{P}(\varepsilon=0)=1-p$，并且输入时受到条件 $\mathbb{E}X\leqslant 1$ 的制约。

考虑形式为 $a_i=cq^i(i=1,2,\cdots)$ 的输入概率分布 $\{a_i, i=0,1,\cdots\}$，试确定最优的输入分布，并给出一个信道容量的表达式。

解答 信道矩阵是

$$\begin{pmatrix} 1 & 0 & 0 & \cdots & 0 \\ 1-p & p & 0 & \cdots & 0 \\ 1-p & 0 & p & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \end{pmatrix}$$

对于概率为 $q_i=\mathbb{P}(X=i)$ 的输入分布，我们有

$$\mathbb{P}(Y=0)=q_0+(1-p)(1-q_0)=1-p+pq_0$$
$$\mathbb{P}(Y=i)=pq_i, i\geqslant 1$$

因此

134

$$h(Y)=-(1-p+pq_0)\log(1-p+pq_0)-\sum_{i\geqslant 1}pq_i\log(pq_i)$$

基于条件熵

$$h(Y|X)=-(1-q_0)[(1-p)\log(1-p)+p\log p]$$

互熵等于

$$I(Y;X)=-(1-p+pq_0)\log(1-p+pq_0)$$
$$-\sum_{i\geqslant 1}pq_i\log(pq_i)+(1-q_0)[(1-p)\log(1-p)+p\log p]$$

我们需要在 q_0，q_1，…受限于 $q_i\geqslant 0$，$\sum_i q_i=1$，$\sum_i iq_i\leqslant 1$ 的条件下最大化 $I(Y;X)$。首先，固定 q_0，并针对 $q_i(i\geqslant 1)$，最大化和项 $-\sum_{i\geqslant 1}pq_i\log(pq_i)$。利用 Gibbs 不等式，对于所有非负数 a_1，a_2，…，$\sum_{i\geqslant 1}a_i=1-q_0$，有

$$-\sum_{i\geqslant 1}q_i\log q_i\leqslant -\sum_{i\geqslant 1}q_i\log a_i,\quad \text{当且仅当 } q_i=a_i \text{ 时等式成立}$$

对于 $a_i=cd^i$，$\sum_i ia_i=1$，从 $\sum_i icd^i=1$，$cd/(1-d)=1-a_0$ 和 $1-d=a_0$，$c=(1-a_0)^2/a_0$ 三个条件出发，上式右边的项变成了

$$-(1-q_0)\log c-(\log d)\sum_{i\geqslant 1}ia_i=-(1-q_0)\log c-\log d$$

接下来，我们在 $a_0\in[0,1]$ 上最大化函数

$$f(a_0)=-(1-p+pa_0)\log(1-p+pa_0)-p(1-a_0)\log\frac{(1-a_0)^2}{a_0}$$
$$-\log a_0+(1-a_0)[(1-p)\log(1-p)+p\log p]$$

利用求导法，需令上式的一阶导数为零，即

$$f'(a_0)=0 \tag{1.6.30a}$$

二阶导数小于等于零

$$f''(a_0)=\frac{p^2}{q+pa_0}-\frac{2p}{1-a_0}-\frac{p}{a_0}\leqslant 0 \tag{1.6.30b}$$

值得指出的是，式(1.6.30a)中方程的求解可以采用数值方法。方程(1.6.30a)的某个根能满足式(1.6.30b)，将这个根用 a_0^* 表示。然后，我们获得如下最优输入分布：

$$a_i=\begin{cases}a_0^*, & i=0\\ (1-a_i^*)^2(a_0^*)^{i-1}, & i\geqslant 1\end{cases}$$

与之对应的信道容量是 $C=f(a_0^*)$。 □

问题 1.31 考虑二元编码的十进制数，该编码方案将十进制数 0 编码为二进制码字 0000，将 1 编码为 0001，如此直到将 9 编码为 1001。编码不使用其他 4 个字符串的二进制序列。请尝试证明，通过用块进行编码，我们可以逼近每十进制数位所需码长的下界。(提示：假设所有整数是等概率的。)

135

解答 问题中的编码是可译的(该编码甚至是前缀码，对于所有定长可译码都可以这么认为)。标准的块编码过程是对 n 长的原始信源(U_n，具有符号集 $\mathscr{A}$)序列进行编码。与之相对应，这个 n 长序列可被视为 n 次扩展信源(U_n，具有符号集 $\mathscr{A}^n$)输出的一个字符。给定典型消息中块的联合概率 $p_n(u_i^{(n)})=\mathbb{P}(U_1=i_1,\cdots,U_n=i_n)$，我们能看到二进制熵为

$$h^{(n)}=-\sum_{i_1,\cdots,i_n}\mathbb{P}(U_1=i_1,\cdots,U_n=i_n)\log\mathbb{P}(U_1=i_1,\cdots,U_n=i_n)$$

用 $S^{(n)}$ 表示块上编码的随机码字长度。每信源符号的最小平均码长是 e_n：$=\min\frac{1}{n}\mathbb{E}S^{(n)}$。通过Shannon NC定理

$$\frac{h^{(n)}}{n\log q}\leqslant e_n\leqslant\frac{h^{(n)}}{n\log q}+\frac{1}{n}$$

其中 q 是原始符号集 $\mathscr{A}$ 的大小。我们能看到，对于较大的 n，$e_n\sim\frac{h^{(n)}}{n\log q}$。在本题中，$q=10$ 而

$$h^{(n)}=hn,\quad h=\log 10(\text{等概率})$$

因此，最小期望字长 e_n 能够任意接近 1。 □

问题 1.32 $\{U_i\}$ 是一个离散时间过程，取值为 u_i。设字符串 $u^{(n)}=u_1\cdots u_n$ 产生的概率是 $\mathbb{P}(u^{(n)})$。试着证明如果概率 $-\log\mathbb{P}(U^{(n)})/n$ 收敛到一个数 γ，那么此 γ 就是过程信息速率。

写出 m 状态 DTMC 的信息速率公式。考虑转移矩阵有元素 p_{jk} 的速率，这里

$$p_{jj}=p,p_{j,j+1}=1-p(j=1,\cdots,m-1),p_{m1}=1-p$$

将上述结论和两状态的信源(转移概率为 p 和 $1-p$)的信息速率联系起来讨论。

解答 考虑具有 m 状态的平稳 DTMC，其状态转移矩阵为 $P=(p_{ij})$，平稳分布 $\pi=(\pi_i)$，

计算可知该信源的信息速率为

$$h=-\sum_{i,j}\pi_i p_{ij}\log p_{ij}$$

如果矩阵 P 是不可约的(即具有唯一的相通类)，上述结论对于任意初始 λ 均成立(这里平稳分布是唯一的)。

136

在这个例子中，转移矩阵是

$$\begin{bmatrix} p & 1-p & 0 & \cdots & 0 \\ 0 & p & 1-p & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1-p & 0 & 0 & \cdots & p \end{bmatrix}$$

每一行都是其他行的置换，每一行都有熵

$$-p\log p-(1-p)\log(1-p)$$

其平稳分布为 $\pi=(1/m,\cdots,1/m)$：

$$\sum_{1\le i\le m}\frac{1}{m}p_{ij}=\frac{1}{m}(p+1-p)=\frac{1}{m}$$

且唯一，因为该 Markov 链具有唯一的相通类。因此，信息速率等于

$$h=\frac{1}{m}\sum_{1\le i\le m}[-p\log p-(1-p)\log(1-p)]=-p\log p-(1-p)\log(1-p)$$

对于 $m=2$，我们可以精确地获得矩阵 $\begin{bmatrix} p & 1-p \\ 1-p & p \end{bmatrix}$，所以基于平稳分布 $\pi=(1/2,1/2)$，信息速率同样为 $h=\eta(p)$。 □

问题 1.33 定义一个对称信道并计算它的容量。

美洲土著战士用烟来传递信号。信号用吹出烟雾的长度编码：短，中等，长。每单位时间内发送一缕烟。假设一缕烟被准确观察到的概率是 p，而以下三种情况发生的概率都是 $1-p$：(a)接收者将一缕短的烟视为中等长的烟，(b)中等长的烟被视为长的烟，(c)长的烟被视为短的烟。假设接收者知道所采用的编码系统，这个战士能可靠地传输信号的最大速率是多少？

假设一缕短烟完全消散而非变为中等长度更加合理。如果该假设成立，你推导的信道容量的公式将会怎样改变？

解答 假设我们用一个有 m 个字符的符号集 $\mathscr{I}$ 来选择输入无记忆信道符号，该信道的输出符号取自符号数为 n(包含非法字符)的输出符号集 $\mathscr{J}$。这个信道可以用它的 $m\times n$ 矩阵来描述，元素 p_{ij} 表示的是 $i\in\mathscr{I}$ 转变成 $j\in\mathscr{J}$ 的概率。信道矩阵的行组成随机 n 维向量(在 $\mathscr{J}$ 上的概率分布)：

137

$$\begin{bmatrix} p_{11} & \cdots & p_{1j} & \cdots & p_{1n} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ p_{i1} & \cdots & p_{ij} & \cdots & p_{in} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ p_{m1} & \cdots & p_{mj} & \cdots & p_{mn} \end{bmatrix}$$

如果信道矩阵每行都是其他行的置换，该信道就称为对称信道(或者更一般地，对所有 $i\in\mathscr{I}$ 有相同的熵 $E=h(p_{i1},\cdots,p_{in})$)。如果除了每行是其他行的置换之外，每一列也都是其他列的置换(或者更一般地说，对所有的 $j\in\mathscr{J}$，每行有相同的和 $\Sigma=\sum_{1\le i\le m}p_{ij}$)，则

信道被叫作双对称信道。

对于无记忆信道，容量(可靠传输速率的上确界)定义为

$$C=\max_{P_X} I(X;Y)$$

这里，最大值取自输入字符的概率分布 $P_X=(P_X(i),\ i\in\mathscr{I})$，输入及输出随机字符 X 和 Y 通过信道矩阵联系后的互信息熵 $I(X,Y)$ 定义为

$$I(X;Y)=h(Y)-h(Y|X)=h(X)-h(X|Y)$$

对于对称信道，条件熵为

$$h(Y|X)=-\sum_{i,j}P_X(i)p_{ij}\log p_{ij}=h$$

无论输入符号概率 $p_X(i)$ 具有任何形式该表达式都成立。因此

$$C=(\max_{P_X}h(Y))-h(Y|X)$$

仅需让输出熵最大化

$$h(Y)=-\sum_j P_Y(j)\log P_Y(j),\quad 其中\ P_Y(j)=\sum_i P_X(i)p_{ij}$$

对于双对称信道，接下来的问题就比较易于求解了：因为在这个方案中 P_Y 是均匀的，$h(Y)$ 在输入符号为均匀分布 $P_X^*(i)=1/m$ 时取得最大化，

$$因为并不依赖于\ j\in\mathscr{I},\quad P_Y(j)=\frac{1}{m}\sum_i p_{ij}=\frac{1}{m}$$

因此，对于双对称信道，容量为

$$C=\log n-h(Y|X)$$

138

此例中，信道矩阵是 3×3 双对称的：

$$\begin{array}{l} 1\sim 短 \\ 2\sim 中等 \\ 3\sim 长 \end{array}\begin{bmatrix} p & 1-p & 0 \\ 0 & p & 1-p \\ 1-p & 0 & p \end{bmatrix}$$

可导出

$$C=\log 3+p\log p+(1-p)\log(1-p)$$

在这个修改的例子中，矩阵变成 3×4 的：

$$\begin{bmatrix} p & 0 & 0 & 1-p \\ 0 & p & 1-p & 0 \\ 1-p & 0 & p & 0 \end{bmatrix}$$

其中第四列和无信号输出状态(一个“污点”)一致。这里的极大值问题失去对称性：

$$\max\left[-\sum_{j=1,2,3,4}\Big(\sum_{i=1,2,3}P_X(i)p_{ij}\Big)\log\Big(\sum_{i=1,2,3}P_X(i)p_{ij}\Big)-\sum_{i=1,2,3}P_X(i)\sum_{j=1,2,3,4}p_{ij}\log p_{ij}\right]$$

$$约束于\ P_X(1),P_X(2),P_X(3)\geqslant 0,以及\ \sum_{i=1,2,3}P_X(i)=1$$

该问题的求解需要全面的分析。 □

问题 1.34 熵功率不等式(EPI，参见式(1.5.10))指出：对于相互独立的 d 维随机向量 $\boldsymbol{X}$，$\boldsymbol{Y}$，下列不等式成立：

$$2^{2h(\boldsymbol{X}+\boldsymbol{Y})/d}\geqslant 2^{2h(\boldsymbol{X})/d}+2^{2h(\boldsymbol{Y})/d} \tag{1.6.31}$$

当且仅当 $\boldsymbol{X}$ 和 $\boldsymbol{Y}$ 同是 Gauss 的且协方差矩阵成比例时等号成立。

令 X 为一个具有实数值的随机变量，其 PDF 是 f_X 且具有有限的微分熵 $h(X)$，令函数

g 是一阶具有严格的正导数 g'。假设 $\mathbb{E}\log_2 g'(X)$ 是有限的，证明变量 $g(X)$ 的微分熵满足

$$h(g(X)) = h(X) + \mathbb{E}\log_2 g'(X)$$

令 Y_1 和 Y_2 彼此独立，并假设它们的概率密度严格为正。试讨论乘积 Y_1Y_2 的微分熵满足

$$2^{2h(Y_1Y_2)} \geqslant a_1 2^{2h(Y_1)} + a_2 2^{2h(Y_2)}$$

139 其中 $\log_2(a_1) = 2\mathbb{E}\log_2 Y_2$ 和 $\log_2(a_2) = 2\mathbb{E}\log_2 Y_1$。

解答 随机变量 $g(X)$ 的 CDF 满足

$$F_{g(X)}(y) = \mathbb{P}(g(X) \leqslant y) = \mathbb{P}(X \leqslant g^{-1}(y)) = F_X(g^{-1}(y))$$

即 PDF $f_{g(X)}(y) = \dfrac{\mathrm{d}F_{g(X)}(y)}{\mathrm{d}y}$ 形式为：

$$f_{g(X)}(y) = f_X(g^{-1}(y))(g^{-1}(y))' = \frac{f_X(g^{-1}(y))}{g'(g^{-1}(y))}$$

于是

$$\begin{aligned} h(g(X)) &= -\int f_{g(X)}(y)\log_2 f_{g(X)}(y)\mathrm{d}y \\ &= -\int \frac{f_X(g^{-1}(y))}{g'(g^{-1}(y))}\log_2 \frac{f_X(g^{-1}(y))}{g'(g^{-1}(y))}\mathrm{d}y \\ &= -\int \frac{f_X(x)}{g'(x)}[\log_2 f_X(x) - \log_2 g'(x)]g'(x)\mathrm{d}x \\ &= h(X) - \mathbb{E}[\log_2 g'(X)] \end{aligned} \tag{1.6.32}$$

当 $g(t) = e^t$ 时，有

$$\log_2 g'(t) = \log_2 e^t = t\log_2 e$$

所以，$Y_i = e^{X_i} = g(X_i)$，而式(1.6.32)说明

$$h(e^{X_i}) = h(g(X_i)) = h(X_i) + \mathbb{E}X_i\log_2 e, \quad i = 1,2,3$$

其中 $X_3 = X_1 + X_2$，于是

$$h(Y_1Y_2) = h(e^{X_1+X_2}) = h(X_1 + X_2) + (\mathbb{E}X_1 + \mathbb{E}X_2)\log_2 e$$

所以，在熵-能量不等式中

$$\begin{aligned} 2^{2h(Y_1Y_2)} &= 2^{2h(X_1+X_2)+2(\mathbb{E}X_1+\mathbb{E}X_2)\log_2 e} \\ &\geqslant (2^{2h(X_1)} + 2^{2h(X_2)})2^{2(\mathbb{E}X_1+\mathbb{E}X_2)\log_2 e} \\ &= 2^{2\mathbb{E}X_2\log_2 e}(2^{2h(X_1)+2\mathbb{E}X_1\log_2 e}) \\ &\quad + 2^{2\mathbb{E}X_1\log_2 e}(2^{2h(X_2)+2\mathbb{E}X_2\log_2 e}) \\ &= a_1 2^{2h(Y_1)} + a_2 2^{2h(Y_2)} \end{aligned}$$

在这里，$a_1 = 2^{2\mathbb{E}X_2\log_2 e}$，即

$$\log_2 a_1 = 2\mathbb{E}X_2\log_2 e = 2\mathbb{E}\ln Y_2\log_2 e = 2\mathbb{E}\log_2 Y_2$$

140 类似地，$\log_2 a_2 = 2\mathbb{E}\log_2 Y_1$。 □

问题 1.35 在这个问题中，我们处理下面这个定义在 $0<a<b$ 的函数：

$$G(a,b) = \sqrt{ab}, \quad L(a,b) = \frac{b-a}{\log(b/a)}, \quad I(a,b) = \frac{1}{\mathrm{e}}(b^b/a^a)^{1/(b-a)}$$

验证

$$0 < a < G(a,b) < L(a,b) < I(a,b) < A(a,b) = \frac{a+b}{2} < b \tag{1.6.33}$$

接下来，对于 $0<a<b$ 定义

$$\Delta(a,b)=L(a,b)I(a,b)/G^2(a,b)$$

记随机变量 X 和 Y 的概率分布 $\boldsymbol{p}=(p_i)$ 和 $\boldsymbol{q}=(q_i)$：

$$\mathrm{P}(X=i)=p_i>0,\quad \mathrm{P}(Y=i)=q_i,i=1\cdots,r,\quad \sum p_i=\sum q_i=1$$

令 $m=\min[q_i/p_i]$，$M=\max[q_i/p_i]$，$\mu=\min[p_i]$，$\nu=\max[p_i]$。证明下列有关熵 $h(X)$ 和 Kullback-Leibler 散度 $D(\boldsymbol{p}\|\boldsymbol{q})$ 的界（参见 PSE II，419 页）：

$$0\leqslant \log r-h(X)\leqslant \log\Delta(\mu,\nu) \tag{1.6.34}$$

$$0\leqslant D(\boldsymbol{p}\|\boldsymbol{q})\leqslant \log\Delta(m,M) \tag{1.6.35}$$

解答 不等式(1.6.33)的证明较为容易，所以留作练习。对于 $a\leqslant x_i\leqslant b$，设 $\mathscr{A}(\boldsymbol{p},\boldsymbol{x})=\sum p_ix_i$，$\mathscr{G}(\boldsymbol{p},\boldsymbol{x})=\prod x_i^{p_i}$。下面这个一般性的不等式成立：

$$1\leqslant \frac{\mathscr{A}(\boldsymbol{p},\boldsymbol{x})}{\mathscr{G}(\boldsymbol{p},\boldsymbol{x})}\leqslant \Delta(a,b) \tag{1.6.36}$$

它说明

$$0\leqslant \log(\sum p_ix_i)-\sum p_i\log x_i\leqslant \log\Delta(a,b)$$

选择 $x_i=q_i/p_i$，我们立即得到式(1.6.35)。选取 $\boldsymbol{q}$ 为均匀分布，我们可以从式(1.6.35)得到式(1.6.34)，因为

$$\Delta\left(\frac{1}{r\nu},\frac{1}{r\mu}\right)=\Delta\left(\frac{1}{\nu},\frac{1}{\mu}\right)=\Delta(\mu,\nu)$$

接下来我们概述式(1.6.36)的证明，细节参考文献[144]、[50]。令 f 为凸函数，p，$q\geqslant 0$，$p+q=1$。然后对于 $x_i\in[a,b]$，我们有

$$0\leqslant \sum p_if(x_i)-f(\sum p_ix_i)\leqslant \max_x[pf(a)+qf(b)-f(pa+qb)] \tag{1.6.37}$$ [141]

将式(1.6.37)应用到下凸函数 $f(x)=-\log x$，经过一些计算得到式(1.6.37)的最大值在 $p_0=(b-L(a,b))/(b-a)$ 处取得，其中 $p_0a+(1-p_0)b=L(a,b)$，并且

$$0\leqslant \log\frac{\mathscr{A}(\boldsymbol{p},\boldsymbol{x})}{\mathscr{G}(\boldsymbol{q},\boldsymbol{x})}\leqslant \log\left(\frac{b-a}{\log(b/a)}\right)-\log(ab)=\frac{\log(b^b/a^a)}{b-a}-1$$

它等价于式(1.6.36)。最后，我们建立式(1.6.37)。对于某个 $\lambda_i\in[0,1]$，有 $x_i=\lambda_ia+(1-\lambda_i)b$。然后通过凸函数性质

$$\begin{aligned}0&\leqslant \sum p_if(x_i)-f(\sum p_ix_i)\\&\leqslant \sum p_i(\lambda_if(a)+(1-\lambda_i)f(b))-f(a\sum p_i\lambda_i+b\sum p_i(1-\lambda_i))\end{aligned}$$

记 $\sum p_i\lambda_i=p$ 和 $1-\sum p_i\lambda_i=q$，并对 p 求最大化，得到式(1.6.37)。 □

问题 1.36 令 f 是在直线 $\mathbb{R}$ 上严格正的概率密度函数(PDF)，定义 Kullback-Leibler 散度 $D(g\|f)$，证明 $D(g\|f)$ 下 $\geqslant 0$。

接下来，假设 $\int e^xf(x)\mathrm{d}x<\infty$ 和 $\int|x|e^xf(x)\mathrm{d}x<\infty$。证明表达式

$$-\int xg(x)\mathrm{d}x+D(g\|f) \tag{1.6.38}$$

在满足 $\int|x|g(x)\mathrm{d}x<\infty$ 的 PDF g 上在唯一的 PDF $g^*\propto e^xf(x)$ 处取得最小值，并计算这个最小值。

解答 Kullback-Leibler 散度 $D(g\|f)$ 定义为

$$D(g\|f)=\int g(x)\ln\frac{g(x)}{f(x)}\mathrm{d}x,\quad \int g(x)\left|\ln\frac{g(x)}{f(x)}\right|\mathrm{d}x<\infty$$

以及

$$D(g \| f) = \infty, \quad \int g(x) \left| \ln \frac{g(x)}{f(x)} \right| \mathrm{d}x = \infty$$

即 $D(g \| f) \geq 0$ 是 Gibbs 不等式。

现在，取 PDF $g^*(x) = \mathrm{e}^x f(x)/Z$，其中，$Z = \int \mathrm{e}^x f(x)\mathrm{d}x$。设 $W = \int x\mathrm{e}^x f(x)\mathrm{d}x$，则 $\frac{W}{Z} = \int x g^*(x)\mathrm{d}x$。接着，有

$$\begin{aligned} D(g^* \| f) &= \frac{1}{Z}\int \mathrm{e}^x f(x) \ln \frac{\mathrm{e}^x}{Z}\mathrm{d}x \\ &= \frac{1}{Z}\int \mathrm{e}^x f(x)(x - \ln Z)\mathrm{d}x = \frac{1}{Z}(W - Z\ln Z) = \frac{W}{Z} - \ln Z \end{aligned}$$

143

并得到

$$-\int x g^*(x)\mathrm{d}x + D(g^* \| f) = -\ln Z$$

这是问题最后部分涉及的最小值。

事实上，对于任何使得 $\int |x| g(x)\mathrm{d}x < \infty$ 的 PDF g，设 $q(x) = g(x)/f(x)$，有

$$\begin{aligned} D(g \| g^*) &= \int g(x) \ln \frac{g(x)}{g^*(x)}\mathrm{d}x = \int q(x) \ln[q(x)\mathrm{e}^{-x}Z] f(x)\mathrm{d}x \\ &= -\int x f(x) q(x)\mathrm{d}x + \int f(x) q(x) \ln q(x)\mathrm{d}x + \ln Z \\ &= -\int x g(x)\mathrm{d}x - D(g \| f) + \ln Z \end{aligned}$$

这说明

$$-\int x g(x)\mathrm{d}x + D(g \| f) = -\int x g^*(x)\mathrm{d}x + D(g^* \| f) + D(g \| g^*)$$

由于 $D(g \| g^*) \geq 0$，除非 $g = g^*$（此时 $D(g \| g^*) = 0$），结论成立。 □

备注 1.6.1 式(1.6.38)的最小值性质在多个领域中有重要且广泛的应用，如统计物理、遗传理论和金融数学。我们建议读者参考文献[109]做更多的了解。

143

编码理论简介

2.1 Hamming 距离，码字的几何特征，码本规模的基本界

为了更好地理解，建议在第一次读本章时专注于二进制的情况，即信道中传送的符号为 0 或 1。

就像我们之前看到的，在 MBSC 情况中，行错误概率 $p\in(0,1/2)$，ML 译码器寻找的是与接收到的二进制码字 $\boldsymbol{y}^{(N)}$ 具有最大数目相同数位的码字 $\boldsymbol{x}^{(N)}$。事实上，如果接收到 $\boldsymbol{y}^{(N)}$，ML 译码器会比较不同二进制码字 $\boldsymbol{x}^{(N)}$ 的概率

$$\begin{aligned}\boldsymbol{P}(\boldsymbol{y}^{(N)}\mid\boldsymbol{x}^{(N)})&=p^{\delta(\boldsymbol{x}^{(N)},\boldsymbol{y}^{(N)})}(1-p)^{N-\delta(\boldsymbol{x}^{(N)},\boldsymbol{y}^{(N)})}\\&=(1-p)^N\left(\frac{p}{1-p}\right)^{\delta(\boldsymbol{x}^{(N)},\boldsymbol{y}^{(N)})}\end{aligned}$$

在这里，

$$\delta(\boldsymbol{x}^{(N)},\boldsymbol{y}^{(N)})=\text{满足 } x_i\neq y_i \text{ 的数字 } i \text{ 的数目} \tag{2.1.1a}$$

是所谓的码字 $\boldsymbol{x}^{(N)}=x_1\cdots x_N$ 和 $\boldsymbol{y}^{(N)}=y_1\cdots y_N$ 的 Hamming 距离。由于第一项 $(1-p)^N$ 不依赖于 $\boldsymbol{x}^{(N)}$，译码器寻求最大化第二项，也就是最小化 $\delta(\boldsymbol{x}^{(N)},\boldsymbol{y}^{(N)})$（当 $0<p/(1-p)<1$，$p\in(0,1/2)$）。

Hamming 距离的定义式(2.1.1a)可以被扩展到 q 进制序列。q 进制码字空间 $\mathscr{H}_{N,q}=\{0,1,\cdots,q-1\}^{\times N}$（集合 $J_q=\{0,1,\cdots,q-1\}$ 的 N 次 Cartesian 幂）连同式(2.1.1a)中定义的距离叫作长度为 N 的 q 进制 Hamming 空间。它包含 q^N 个元素。在二进制情况下，$\mathscr{H}_{N,2}=\{0,1\}^{\times N}$。 141

码字 $\boldsymbol{x}^{(N)}=x_1\cdots x_N$ 和 $\boldsymbol{0}^{(N)}=0\cdots0$ 的距离 $\delta(\boldsymbol{x}^{(N)},\boldsymbol{0}^{(N)})$ 在编码理论中起着十分重要的作用，定义它为码字 $\boldsymbol{x}^{(N)}$ 的重量，记作 $w(\boldsymbol{x}^{(N)})$：

$$w(\boldsymbol{x}^{(N)})=\text{满足 } x_i\neq 0 \text{ 的数字 } i \text{ 的数目} \tag{2.1.1b}$$

引理 2.1.1 数值 $\delta(\boldsymbol{x}^{(N)},\boldsymbol{y}^{(N)})$ 在 $\mathscr{H}_{N,q}$ 上定义了一个距离：

(i) $0\leqslant\delta(\boldsymbol{x}^{(N)},\boldsymbol{y}^{(N)})\leqslant N$ 和 $\delta(\boldsymbol{x}^{(N)},\boldsymbol{y}^{(N)})=0$ 当且仅当 $\boldsymbol{x}^{(N)}=\boldsymbol{y}^{(N)}$

(ii) $\delta(\boldsymbol{x}^{(N)},\boldsymbol{y}^{(N)})=\delta(\boldsymbol{y}^{(N)},\boldsymbol{x}^{(N)})$

(iii) $\delta(\boldsymbol{x}^{(N)},\boldsymbol{z}^{(N)})\leqslant\delta(\boldsymbol{x}^{(N)},\boldsymbol{y}^{(N)})+\delta(\boldsymbol{y}^{(N)},\boldsymbol{z}^{(N)})$（三角不等式）

证明 (i)和(ii)的证明很明显。为了验证(iii)，观察到对于任何使得 $z_i\neq x_i$ 的 i 存在两种可能：或者 $y_i\neq x_i$，那么距离算在 $\delta(\boldsymbol{x}^{(N)},\boldsymbol{y}^{(N)})$ 里；或者 $z_i\neq y_i$，那么距离算在 $\delta(\boldsymbol{y}^{(N)},\boldsymbol{z}^{(N)})$ 里。 □

几何上，二进制 Hamming 空间 $\mathscr{H}_{N,2}$ 可以被认为是在 N 维空间中单位立方体的顶点集合，Hamming 距离等于从一个顶点到另一个顶点需要经过边的最小数目。在 N 较小的情况下通过图示辅助阐述是很好的方式，见图 2-1。

接下来将讨论 Hamming 距离的几何和代数性质，二者在编码理论中均有重要作用。在任意度量空间中，可以考虑一个给定码字 $\boldsymbol{x}^{(N)}$ 周围半径为 R 的球：

$$\mathscr{B}_{N,q}(\boldsymbol{x}^{(N)},R)=\{\boldsymbol{y}^{(N)}\in\mathscr{H}_{N,q};\delta(\boldsymbol{x}^{(N)},\boldsymbol{y}^{(N)})\leqslant R\} \tag{2.1.2}$$

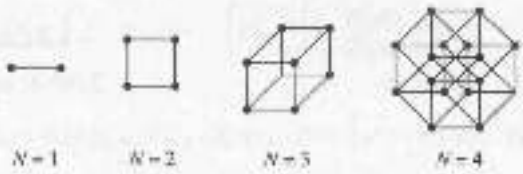


图 2-1

一个重要(且困难)的问题是计算在一个给定的 Hamming 空间里能装下的给定半径且不重叠的球的最大数目。

观察到码字服从 mod q 加运算：

145

$$x^{(N)}+y^{(N)}=(x_1+y_1)\bmod q\cdots(x_N+y_N)\bmod q \tag{2.1.3a}$$

这使得 Hamming 空间$\mathscr{H}_{N,q}$成为一个交换群，其中零码字$\mathbf{0}^{(N)}=0\cdots0$作为群中的零(码字可以通过乘积发挥强大的功能，参见下面)。

对于 $q=2$，我们有一个包含两点的码字符号集$\{0,1\}$，即两元域$\mathbb{F}_2$，其中存在下列运算：$0+0=1+1=0\cdot1=1\cdot0=0$，$0+1=1+0=1\cdot1=1$。(已知域是包含两个可交换运算(加法和乘法)的集合，满足结合律和分配律的标准公理。)因此，在二进制 Hamming 空间$\mathscr{H}_{N,2}$中的每个点都是自反的：$x^{(N)}+x^{(N)}=\mathbf{0}^{(N)}$当且仅当$x^{(N)}=x^{(N)}$。事实上，$\mathscr{H}_{N,2}$是域$\mathbb{F}_2$上的线性空间，包含$1\cdot x^{(N)}=x^{(N)}$，$0\cdot x^{(N)}=\mathbf{0}^{(N)}$。

所以，所有 q 进制的加法都可以理解为针对每个数位并且是 mod q 的。

引理 2.1.2 $\mathscr{H}_{N,q}$上的 Hamming 距离在群变换下不变：

$$\delta(x^{(N)}+z^{(N)},y^{(N)}+z^{(N)})=\delta(x^{(N)},y^{(N)}) \tag{2.1.3b}$$

证明 对于所有 $i=1,\cdots,N$ 和 $x_i,y_i,z_i\in\{0,1,\cdots,q-1\}$，数位 $x_i+z_i \bmod q$ 和 $y_i+z_i \bmod q$ 与数位 x_i 和 y_i 具有相同的关系(=或≠)。 □

一个编码被归为一个码字的集合$\mathscr{X}_N\subset\mathscr{H}_{N,q}$，这意味着我们不考虑任何特别的码字分布(与源信息是等概率分布的假设一致)。这里有一个假设，即发送方和接收方都知道编码方案。Shannon 编码定理保证在某种情况下，存在渐近最优的码字达到信源的信息速率和信道容量的极限。此外，Shannon SCT 证明几乎所有码字都是渐近最优的。然而，在实际情况中，这些事实作用有限：我们想要具有明晰形式的最优编码。此外，可以快速编码和译码并能最大化信息传输速率的编码也是我们所追求的。

所以，假设信源发送二进制序列$u^{(k)}=u_1\cdots u_k$，$u_i=0,1$。为了让系统整体的错误概率在 $n\to\infty$时趋于 0，我们必须以更长的码字$x^{(N)}\in\mathscr{H}_{N,2}$编码 $u^{(k)}$，其中 $N=R^{-1}n$，$0<R<1$。然后，信源码字$x^{(N)}$发送到信道，被转化成另一个码字$y^{(N)}\in\mathscr{H}_{N,2}$。以式(2.1.3a)的形式表示两个码字因数位内容不一致而造成的错误是十分方便的：$e^{(N)}=y^{(N)}-x^{(N)}=x^{(N)}+y^{(N)}$，或者写成$y^{(N)}=x^{(N)}+e^{(N)}$。所以，误差码字$e^{(N)}$有越多的数位为字 1，表示越多的符号被信道失真。随后，ML 译码器产生一个猜想码字$x_*^{(N)}$，它与$x^{(N)}$相同或者不相同，继而重建序列$u_*^{(k)}$。在一对一的编码准则下，译码最后的步骤(理论上)是明显的：我

146

们仅需将映射$u^{(k)}\to x^{(N)}$取逆映射。直观地说，一个好的码在码字$e^{(N)}$不包含"太多"的非零数位时，需能使接收方"纠正"错误序列$e^{(N)}$。

回到有行错误概率 $p<1/2$ 的 MBSC：ML 译码器选择一个码字$x_*^{(N)}$使得码字$e^{(N)}$有最少的含有 1 的数位。在几何上：

$$x_*^{(N)}\in\mathscr{X}_N\text{ 是在 Hamming 距离 }\delta\text{ 上与 }y^{(N)}\text{ 距离最近的码字} \tag{2.1.4}$$

相同的准则可以在 q 进制的情况下使用：我们寻找与接收序列最接近的码字。这个准则的

一个缺陷是，如果几个码字与接收的码字都有相同的最小距离，我们就会迷惑了。在这种情况下，我们或者从中任意选择一个码字(可以随机选取也可以依据消息内容来选取，这与所谓的列表译码相关)，或者当需要高质量传输时，我们拒绝译码接收到的码字并要求发送方重新传输。

定义 2.1.3 我们称 N 为二进制码$\mathscr{C}_N$的长度，M：$=\#\mathscr{C}_N$是码大小而ρ：$=\dfrac{\log_2 M}{N}$是码的信息速率。如果码本中任何合法码字有多达 D 位的改变都不会变成另一个码字，则码$\mathscr{C}_N$被叫作 D-错误可检测，而 E-错误可纠正要求，如果任意码字$x^{(N)}$中有多达 E 位改变，产生的码字依旧比其他码字(严格地)更接近$x^{(N)}$(也就是$x^{(N)}$在准则(2.1.4)下依旧可以从被污染的码字猜出来)。一个码的最小距离(简称距离)d 定义如下

$$d=\min[\partial(x^{(N)},x'^{(N)}):x^{(N)},x'^{(N)}\in\mathscr{C}_N,x^{(N)}\neq x'^{(N)}] \tag{2.1.5}$$

一个码$\mathscr{C}_N$的最小距离和信息速率有时被分别记为$d(\mathscr{C}_N)$和$\rho(\mathscr{C}_N)$。

对于信息速率$\rho=\dfrac{\log_q M}{N}$的 q 进制码$\mathscr{C}_N\subset\mathscr{H}_{N,q}$，上述定义几乎可以完全照字面重复。也就是说，一个码$\mathscr{C}_N$，如果对于所有 $r=1,\cdots,E$，$x^{(N)}\in\mathscr{C}_N$和$y^{(N)}\in\mathscr{H}_{N,q}$，$\partial(x^{(N)},y^{(N)})=r$，所有使得$x'^{(N)}\neq x^{(N)}$的$x'^{(N)}\in\mathscr{C}_N$距离$\partial(y^{(N)},x'^{(N)})>r$，则码$\mathscr{C}_N$叫作 E-错误可纠码。也就是说，一个码字中即使有多达 E 位错误却依旧比其他码字更接近原码字。几何上，这一性质意味着码字周围半径为 E 的球互不相交：

$$\mathscr{B}_{N,q}(x^{(N)},E)\cap\mathscr{B}_{N,q}(x'^{(N)},E)=\varnothing,\quad 对于所有的x^{(N)},x'^{(N)}\in\mathscr{C}_N$$

接下来，一个码$\mathscr{C}_N$叫作 D-错误可检码，如果一个码字周围半径为 D 的球不包含另一个码字。等价地说，交集$\mathscr{B}_{N,q}(x^{(N)},D)\cap\mathscr{C}_N$退化成一个点$x^{(N)}$。 147

长度为 N、大小为 M、最小距离为 d 的码被记为$[N,M,D]$码。说到$[N,M]$或者$[N,d]$码，我们指的是长度为 N 大小为 M 或者最小距离为 d 的码。

为了理解这个定义，我们来证明之前提到的 E-错误可纠码的等价定义。首先，假设多个半径 E 的球不相交。然后，对一个码字中有多达 E 位变化而产生的新码字依旧在原码字对应的球内，所以该新码字将离其他码字更远。反过来说，假设我们的码有这样的性质：在一个码字中改变 E 个数位不会产生跟另一个码字距离相同或更近(对比起原码字)的码字。在一个码字中正好改变 E 个数位，所产生的码字除了能落入原始码字周围的球，不能落在任何半径 E 的其他球内。如果我们做更少的数位改变，同样我们得到的码字不会落在其他任何球内，因为如果是的话，那么往第二个球中心移动新码字迟早会产生与原始码字距离为 E，距离第二个码字距离$<E$ 的码字，而这是不可能的。

对于一个 D-错误可测码字，距离 $d\geqslant D+1$。此外，距离为 d 的码字能检测 $d-1$ 个错误并能纠正$(d-1)/2$ 个错误。

备注 2.1.4 定义 2.1.3 意味着一个码字检测出至少 D 个错误并纠正至少 E 个错误。一些学者对这个事实总结了一点，将 D 和 E 确定为具有以上性质的最大值。我们遵从传统的做法将码字检测和纠正的能力以不等式而非等式来定义，尽管在接下来许多例子中声明一个码字检测 D 个或纠正 E 个错误就意味着 D 或 E 而不是更多。比如参见定义 2.1.7。

定义 2.1.5 在 2.3 节我们系统地学习了所谓的线性码。线性结构建立在空间$\mathscr{H}_{N,q}$中，其中的字母表大小 q 具有 p^s 的形式，其中 p 为奇数，s 为正整数；在这个情况下字母表$\{0,1,\cdots,q-1\}$形成一个域 $\mathbb{F}_q$，其中引入了两个合适的运算：加法和乘法。参见 3.1 节。当 $s=1$，即 q 是奇数时，两个运算都可以理解为标准的 mod q 运算。当 $\mathbb{F}_q$是一个有着加法+和乘

法·的域时，集合$\mathscr{H}_{N,q}=\mathbb{F}_q^{(N)}$变成了$\mathbb{F}_q$上的线性空间，其中的运算包含按元素逐个对应相加和与标量相乘，这些运算都可以与$\mathbb{F}_q$中的运算相对应。也就是说，对于$x^{(N)}=x_1\cdots x_N$和$y^{(N)}=y_1\cdots y_N$和$\gamma\in\mathbb{F}_q$，

$$x^{(N)}+y^{(N)}=(x_1+y_1)\cdots(x_N+y_N),\gamma\cdot x^{(N)}=(\gamma\cdot x_1)\cdots(\gamma\cdot x_N) \tag{2.1.6a}$$

当$q=p^r$，q进制$[N,M,d]$码$\mathscr{X}_N$被称为线性的如果它是$\mathscr{H}_{N,q}$的线性子空间。也就是说，$\mathscr{X}_N$有这样的性质：如果$x^{(N)}$，$y^{(N)}\in\mathscr{X}_N$，那么$x^{(N)}+y^{(N)}\in\mathscr{X}_N$和$\gamma\cdot x^{(N)}\in\mathscr{X}_N$对于所有$\gamma\in\mathbb{F}_q$成立。对于一个线性码字$\mathscr{X}$，码本大小$M$为$M=q^k$，其中$k$从1，…，$N$中取值并给出了码字的维度，即线性独立码字的最大数目。对应地，我们可以写出$k=\dim\mathscr{X}$。在常规几何中，如果$k=\dim\mathscr{X}$，那么在$\mathscr{X}$中存在k维的基底，即一个线性独立的码字集合$x^{(1)}$，…，$x^{(k)}$使得任何码字$x\in\mathscr{X}$均可以被写成线性组合$\sum_{1\le j\le k}a_j x^{(j)}$的形式，其中$a_j\in\mathbb{F}_q$。（事实上，如果$k=\dim\mathscr{X}$，那么任何$k$个线性独立的码字集合都是$\mathscr{X}$的基底。）在线性的情况中，我们主要讨论$[N,k,d]$或者$[N,k]$码。

[148]

从定义得知，一个线性的$[N,k,d]$码$\mathscr{X}_N$总是包含零序列$\mathbf{0}^{(N)}=0,\cdots,0$。此外，由于性质(2.1.3b)，线性码$X$的最小距离$d(\mathscr{X}_N)$等于非零码字$x^{(N)}\in\mathscr{X}_N$的最小重量$w(x_N)$，参见式(2.1.b)。

最后，我们定义所谓的码字x和y的外积，为$w=x\wedge y$，其元素为

$$w_i=\min[x_i,y_i],\quad i=1,\cdots,N \tag{2.1.6b}$$

线性码的许多性质已经在这一节中提及，证明中的一些细节将在后面提到。

线性码的一个简单例子是重复码$\mathscr{R}_N\subset\mathscr{H}_{N,q}$，形式为

$$\mathscr{R}_N=\{x^{(N)}=x\cdots x;x=0,1,\cdots,q-1\}$$

它能检测$N-1$个错误和纠正$\left\lfloor\frac{N-1}{2}\right\rfloor$个错误。一个线性的奇偶校验码

$$\mathscr{P}_N=\{x^{(N)}=x_1\cdots x_N;x_1+\cdots+x_N=0\}$$

只能检测一个错误并且不能纠正它。

观察到在中心为$z^{(N)}$的Hamming空间$\mathscr{H}_{N,q}$的球的体积为

$$v_{N,q}(R)=\#\mathscr{B}_{N,q}(z^{(N)},R)=\sum_{0\le k\le R}\begin{bmatrix}N\\k\end{bmatrix}(q-1)^k \tag{2.1.7}$$

它不依赖于中心$z^{(N)}\in\mathscr{H}_{N,q}$的选择。

有趣的是考虑N取大值（理论上，$N\to\infty$）并分析码$\mathscr{X}_N$的核心参数，比如信息速率$\rho(\mathscr{X})=\frac{\log\#\mathscr{X}}{N}$和每个数位的距离$d(\mathscr{X})=\frac{d(\mathscr{X})}{N}$。我们的目标集中在好的码，即有很多码字（这意味着可以增加信息速率）和大的码字间距离（这意味着可以增加检测和纠正的能力），从这一点来看，理解码字的基本界十分重要。

[149]

长度为N距离为d的q进制编码所含码字的最大数目通常由上界$M_q^*(N,d)$给出。我们先从最基本的事实说起：$M_q^*(N,1)=q^N$，$M_q^*(N,N)=q$，$M_q^*(N,d)\le qM_q^*(N-1,d)$。在二进制情况下，$M_2^*(N,2s)=M_2^*(N-1,2s-1)$（简单推演可得）。

事实上，如果我们希望保持很好的错误检测和错误纠正的性能，码字的数目不能过多。码的参数有多种界，最简单的界是在19世纪40年代晚期由Hamming发现的。

定理2.1.5 （Hamming界）(i) 如果q进制码$\mathscr{X}_N$纠正E个错误，那么它的大小$M=\#\mathscr{X}_N$遵从

$$M\le q^N/v_{N,q}(E) \tag{2.1.8a}$$

对于线性[N，k]码上式可写成

$$N-k \geqslant \log_q(v_{N,q}(E))$$

(ii) 相应地，$E=\lfloor (d-1)/2 \rfloor$，

$$M_q^*(N,d) \leqslant q^N / v_{N,q}(E) \tag{2.1.8b}$$

证明 (i)码字$x^{(N)} \in \mathscr{X}_N$周围的$E$-球必须是非交的。所以，覆盖的点的总数目等于乘积$v_{N,q}(E)M$，它不能超过$q^N$即 Hamming 空间$\mathscr{H}_{N,q}$的基数。

(ii) 类似地，如果$\mathscr{X}_N$是一个[N，M，d]码，那么如前所述，对于$E=\lfloor (d-1)/2 \rfloor$，球$\mathscr{B}_{N,q}(x^{(N)}, E)$，$x^{(N)} \in \mathscr{X}_N$不相交。体积$\#\mathscr{B}_{N,q}(x^{(N)}, E)$为

$$v_{N,q}(E) = \sum_{0 \leqslant k \leqslant E} \begin{bmatrix} N \\ k \end{bmatrix} (q-1)^k$$

而所有球的并集写为

$$\bigcup_{x^{(N)} \in \mathscr{X}_N} \mathscr{B}_{N,q}(x^{(N)}, E)$$

必须在$\mathscr{H}_{N,q}$中，基数也为$\#\mathscr{H}_{N,q}=q^N$。 □

我们看到寻找好的码实际上变成了一个几何问题，因为一个能纠正E个错误的好的码$\mathscr{X}_N$必须以半径为E的球对 Hamming 空间进行紧密填充。一个能够真正紧密填充的码$\mathscr{X}_N$ 150
有另外一个优势：码不但能纠正错误，而且不会导致拒绝译码。更确切地说法如下。

定义 2.1.7 一个大小为$\#\mathscr{X}_N=M$的E-纠错码$\mathscr{X}_N$，如果其能让 Hamming 界以等号成立：

$$M = q^N / v_{N,q}(E)$$

则称这个码是完美的。如果一个码字$\mathscr{X}_N$是完美的，每个码字$y^N \in \mathscr{H}_{N,q}$属于一个(唯一的)球$\#\mathscr{B}_E(x^{(N)})$。也就是说，我们总能通过一个码字来译码$y^N$：当错误数目$\leqslant E$时能得到正确的答案，而$>E$时就会得到错误的答案，但我们永远不会在译码的时候出现问题。

寻找完美二进制码问题在 20 年前就解决了。这些码只存在于：

(a) $E=1$：在这里$N=2^l-1$，$M=2^{2^l-l-1}$。这些码对应于所谓的 Hamming 码。

(b) $E=3$：在这里$N=23$，$M=2^{12}$，它们对应于所谓的(二进制)Golay 码。

Hamming 码和 Golay 码都将在这讨论。Golay 码(连同一些改动)被用在美国的航空项目上：早在 20 世纪 70 年代，经由这种码编码的照片就从火星和金星传出，照片的质量很不错，不需要任何进一步提升画质的处理。在苏联的(和早期美国的)宇宙飞船中其他码也同样被使用(我们以后也要讨论)，用它们编码的照片质量一般不高，往往需要根据照片的统计特征进一步地操作处理。

如果我们考虑非二元码，那么对于三元码进制存在不止一个完美码(同样以 Golay 命名)。

我们现在描述一系列直接从已知码中构造新码的构造方法。

例子 2.1.8 新码的构造包括：

(i)扩展：遵从一个公共的准则对码$\mathscr{X}_N$中的每个码字$x^{(N)}=x_1 \cdots x_N$增加一个数位x_{N+1}，即所谓的奇偶校验扩展需要在符号集域E_q中有$x_{N+1}+\sum_{1 \leqslant j \leqslant N} x_j = 0$。显然，扩展码$\mathscr{X}_{N+1}^{+}$与原始码$\mathscr{X}_N$有相同的大小，距离$d(\mathscr{X}_{N+1}^{+})$等于$d(\mathscr{X}_N)$或$d(\mathscr{X}_N)+1$。 151

(ii) 截断：从码字$x \in \mathscr{X}(=\mathscr{X}_N)$中删除一个数字。删除后得到一个$N-1$长的码字$\mathscr{X}_{N-1}^{-}$，当距离$d(\mathscr{X}_N) \geqslant 2$，$\mathscr{X}_{N-1}^{-}$和$\mathscr{X}_N$的大小相同，且$d(\mathscr{X}_{N-1}^{-}) \geqslant d(\mathscr{X}_N)-1$。

(iii) 清除：简单地删除一些码字$x \in \mathscr{X}_N$。例如，在二元码情况下，从一个线性码中删除所有包含奇数个非零数位的码字，将可以得到一个线性子码；在这种情况下，如果原

始的码字距离是奇数，则经删除后得到的码字会有更大的距离。

(iv) 添加：与删除相反，比方说，将每一个码字的补码(即 N 长码字中的 1 由 0 替换，反之亦然)加入到二进制码字$\mathscr{X}_N$ 中。记扩张后的码字为$\overline{\mathscr{X}_N}$，可以得到 $d\,\overline{\mathscr{X}_N}=\min[d(\mathscr{X}_N),\ N-d(\mathscr{X}_N)]$，其中

$$\overline{d}(\mathscr{X}_N)=\max[\partial(x^{(N)},x'^{(N)});x^{(N)},x'^{(N)}\in\mathscr{X}_N]$$

(v) 缩短：取出所有在第 i 位为零的码字$x^{(N)}\in\mathscr{X}_N$，然后将这个数位删除(也就是在 $x_i=0$ 位缩短)。这样一来，原始的二元线性$[N,\ M,\ d]$ 码$\mathscr{X}_N$ 就被缩减为二元线性码 $\mathscr{X}_{N-1}^{*,0}(i)$，其长度为 $N-1$，大小为 $M/2$ 或 M，距离大于 d 或在一般的情况下大于 0。

(vi) 重复：以固定次数重复码字 $x(=x^{(N)})\in\mathscr{X}_N$，比如说重复 m 次，则产生了一个级联的(Nm) 字 $xx\cdots x$。重复的结果就是长为 Nm 的码字$\mathscr{X}_{Nm}^{*}$，其距离为 $d(\mathscr{X}_{Nm}^{*})=md(\mathscr{X}_N)$。

(vii) 直和：给定两个码$\mathscr{X}_N$ 和$\mathscr{X}'_N$，形成一个新的码$\mathscr{X}\oplus\mathscr{X}'=\{xx'_{\pm}\ x\in\mathscr{X},\ x\in\mathscr{X}'\}$。值得指出的是重复和直和的构造都不是很有效率，两者中的任意一个在编码中都不是很流行(尽管我们将会在例子和问题中继续探讨这些构造方法)。事实上一个更有效率的构造是

(viii) 块积$(x|x+x')$：对于$[N,\ M,\ d]$ 和$[N,\ M',\ d']$ 码$\mathscr{X}_N$ 和$\mathscr{X}'_N$，如下定义一个 $2N$ 长的新码$\mathscr{X}_N|\mathscr{X}'_N$

$$\{x(x+x');x(=x^{(N)})\in\mathscr{X}_N,x'(=x'^{(N)})\in\mathscr{X}'_N\}$$

这就是说，每个$\mathscr{X}_N|\mathscr{X}'_N$ 中的码字是$\mathscr{X}_N$ 中的码字级联上其与$\mathscr{X}'_N$ 中码字相加的结果(通常情况下，在这种结构中$\mathscr{X}_N$或$\mathscr{X}'_N$中都不是线性的)。最终码字是由$\mathscr{X}_N|\mathscr{X}'_N$ 表示，大小为

$$\#(\mathscr{X}_N|\mathscr{X}'_N)=(\#\mathscr{X}_N(\#\mathscr{X}'_N))$$

为了便于理解，可以练习验证上述码的距离

152

$$d(\mathscr{X}_N|\mathscr{X}'_N)=\min[2d(\mathscr{X}_N),d(\mathscr{X}'_N)]$$

(ix) 对偶码：对偶性的概念是基于空间$\mathscr{H}_{N,q}$内部点积($q=p^r$)：对于 $x=x_1\cdots x_N$ 和 $y=y_1\cdots y_N$，点积运算如下

$$(x^{(N)}\cdot y^{(N)})=x_1\cdot y_1+\cdots+x_N\cdot y_N$$

运算结果将从 F_q 中产生。对于一个线性$[N,\ k]$ 码$\mathscr{X}_N$，它的对偶码$\mathscr{X}_N^{\perp}$ 是一个线性$[N,\ N-K]$ 码，定义如下

$$\mathscr{X}_N^{\perp}=\{y^{(N)}\in\mathscr{H}_{N,q};(x^{(N)}\cdot y^{(N)})=0,\ \ x\in\mathscr{X}_N\}\tag{2.1.9}$$

显然，$(\mathscr{X}_N^{\perp})^{\perp}=\mathscr{X}_N$。而且，$\dim\mathscr{X}_N^{\perp}+\dim\mathscr{X}_N=N$。如果$\mathscr{X}_N^{\perp}=\mathscr{X}_N$，这种码被称为自偶码。

举例 2.1.9 (a) 证明如果$[N,\ M,\ d]$ 码$\mathscr{X}_N$ 的距离是奇数，这个码字可以被扩展为码距为 $d+1$ 的$[N+1,\ M]$ 码$\mathscr{X}_N^{+}$。

(b) 证明 E-错误纠错码$\mathscr{X}_N$ 能被扩展为可检测 $2E+1$ 个错误的检错码$\mathscr{X}_N^{+}$。

(c) 证明一个完美的二进制码的距离是一个奇数。

解答 (a) 将数位 x_{N+1} 加到$[N,\ M]$ 码$\mathscr{X}_N$ 中的码字 $x=x_1\cdots x_N$ 中，得到$[N+1,\ M]$ 码字$\mathscr{X}_N^{+}$，其中 $x_{N+1}=\sum_{1\leqslant j\leqslant N}x_j$。如果$\mathscr{X}_N$ 的距离 d 是奇数，则$\mathscr{X}_N^{+}$ 的距离是 $d+1$。实际上，如果一对码字 $x,\ x'\in\mathscr{X}$，且 $\partial(x,\ x')>d$，则扩展的码字 $x_+,\ x'_+$ 有 $\partial(x_+,\ x'_+)\geqslant\partial(x,$

$x')>d$。否则，如果 $\partial(x, x')=d$，新码的距离就会增加：$\partial(x_+, x'_+)=d+1$。

(b) E-错误纠错码字的距离 d 严格大于 $2E$，因此，上述扩展得到的码距离严格大于 $2E+1$。

(c) 对于一个完美的 E-错误纠错码字，其距离最多为 $2E+1$，因此只能等于 $2E+1$。 □

举例 2.1.10 证明不存在定义在 $\mathbb{F}_2$ 上，码长为 90、码本大小为 2^{78} 的完美 2-错误纠错码。

解答 我们可能想知道是否存在长度 $N=90$，大小 $M=2^{78}$ 的完美 2-错误纠错码，因为

$$v_{N,2}(2)=1+90+\frac{90\cdot 89}{2}=4096=2^{12}$$

且

$$M\times v_{N,2}(2)=2^{78}\cdot 2^{12}=2^{90}=2^{N}$$

153

然而，这种码是不存在的。假定它是存在的，且全零字序列 $\mathbf{0}=0\cdots0$ 是一个码字。那么，这个码必须有 $d=5$。考虑 88 个仅有三个非零数位的字符，并且它们头两位都是 1：

$$1110\cdots00,\ 1101\cdots00,\ \cdots,\ 110\cdots01 \tag{2.1.10}$$

这些字符中的每一个到某个特定码字的距离不应该大于 2。就是说，1110…00 的码字必须包含 5 个非零数位(因为全零序列是一个码字，只有包含 5 个非零数位才可能满足 $d=5$)。假定该码字是

$$11111\cdots00$$

上面这个码字与(1110…00)随后的两个字符的距离都是 2

$$11010\cdots00 \quad 和 \quad 11001\cdots00$$

继续这种构建，可以看到表(2.1.10)中的任意字符都被有 5 个非零数位的码字吸引，并且(2.1.10)中该字符后面的另外两个字符也被同一码字吸引。但是 88 是不能被 3 整除的。□

下面我们讨论码的界。

定理 2.1.11 (Gilbert-Varshamov(GV)界)对于任意的 $q\geqslant2$，$d\geqslant2$，存在一个 q 进制的 $[N, M, d]$ 码 $\mathscr{C}_N$，也就是

$$M=\#\mathscr{C}_N\geqslant q^N/v_{N,q}(d-1) \tag{2.1.11}$$

证明 考虑这样一个在所有最小距离为 d 且长度为 N 的码中具有最大码本大小的码。那么，任意字 $\mathbf{y}^{(N)}\in\mathscr{H}_{N,q}$ 到某个码字的距离必须 $\leqslant d-1$。否则，可以将 $\mathbf{y}^{(N)}$ 加入码本中，并不改变码的最小距离。因此，码字周围半径为 $d-1$ 的球填充了整个 Hamming 空间 $\mathscr{H}_{N,q}$。就是说，对于最多码字数的码 $\mathscr{C}_N^{max}$ 有

$$(\#\mathscr{C}_N^{max})v_{N,q}(d-1)\geqslant q^N$$

正如此前提及，从一个码(或码的集合)生成另一个码有很多方法。通过运用截断且删除每个码字 $x^{(N)}$ 的最后一个数位 x_N，其中码字 $x^{(N)}$ 来自于最初码 $\mathscr{C}_N$。如果码 $\mathscr{C}_N$ 有最小距离 $d>1$，则新码 $\mathscr{C}_{N-1}$ 的最小距离 $\geqslant d-1$，且和 $\mathscr{C}_N$ 的大小相同。截断过程会导致有如下界。□

定理 2.1.12 (Singleton 界)对最小距离为 d 的任意 q 进制码 $\mathscr{C}_N$，可以得到如或于

$$M=\#\mathscr{C}_N\leqslant M_q^*(N,d)\leqslant q^{N-d+1} \tag{2.1.12}$$

154

证明 在 $[N, M, d]$ 码 $\mathscr{C}_N$ 中依然实施截断：对于每个码字 $x\in\mathscr{C}_N$，我们将删除最后一个数位。新码是 $d^-\geqslant d-1$ 的 $[N, M, d^-]$ 码。重复这个过程 $d-1$ 次，得到一个大小同样为 M，距离不小于 1 的 $(N-d+1)$ 码，这个码必须可以嵌入 Hamming 空间 $\mathscr{H}_{N-d+1,q}$，且 $\#\mathscr{H}_{N-d+1,q}=q^{N-d+1}$。因此可以得到结果。 □

与 Hamming 界一样，Singleton 界中等式的情况吸引了很多特殊的兴趣。

定义 2.1.13 如果在 Singleton 界中有如下等式，则称一个 q 进制的线性$[N, k, d]$码为最大距离分离(MDS)

$$d = N - k + 1 \tag{2.1.13}$$

我们将在后面看到，和完美码相似，MDS 码的家族中所包含的码也是相当少的。

推论 2.1.14 如果 $M_q^*(N, d)$ 是最小距离为 d 的码$\mathscr{X}_N$的最大尺寸，可得

$$\frac{q^N}{v_{N,q}(d-1)} \leqslant M_q^*(N,d) \leqslant \min\left(\frac{q^N}{v_{N,q}([(d-1)/2])}, q^{N-d+1}\right) \tag{2.1.14}$$

当不会产生误解时，我们会省略系数 N 和(N)。当 $d \sim N/2$ 时，式(2.1.14)中的上界将变得非常粗略。这就是说，在 $N=10$ 和 $d=5$ 的二元$[N, M, d]$-码的情况下，从表达式(2.1.14)得出了上界 $M_2^*(10, 5) \leqslant 18$，虽然实际上并没有 $M_2 \geqslant 13$ 的码，不过，确实存在一个 $M=12$ 的码。后者的码字如下所示：

0000000000, 1111100000, 1001011010, 0100110110,
1100001101, 0011010101, 0010011011, 1110010011,
1001100111, 1010111100, 0111001110, 0101111001

在这种情况下，该推理给出的下界值是 2(因为 $2^{10}/v_{10,2}(4)=2.6585$)。当然这个界精确程度远达不到令人满意的程度(在下面会给出一些更好的边界)。

定理 2.1.15 (Plotkin 界)对于长度为 N，距离为 d，且 $N<2d$ 的二元码$\mathscr{X}$，码本的大小 M 满足

155

$$M = \#\mathscr{X} \leqslant 2\left\lfloor \frac{d}{2d-N} \right\rfloor \tag{2.1.15}$$

证明 因为最小距离不能超过平均距离，即

$$M(M-1)d \leqslant \sum_{x\in\mathscr{X}}\sum_{x'\in\mathscr{X}} \delta(x,x')$$

另一方面，将码$\mathscr{X}$记为一个 $M\times N$ 的矩阵，行向量的集合被当作码字。假设矩阵的第 i 列包含 s_i 个 0 和 $M-s_i$ 个 1，那么

$$\sum_{x\in\mathscr{X}}\sum_{x'\in\mathscr{X}} \delta(x,x') \leqslant 2\sum_{1\leqslant i\leqslant N} S_i(M-s_i) \tag{2.1.16}$$

如果 M 是偶数，当 $s_i=M/2$ 时，可知式(2.1.16)的右半部分取得最大值。进而可以得到

$$M(M-1)d \leqslant \frac{1}{2}NM^2, \quad 或\ M \leqslant \frac{2d}{2d-N}$$

由于 M 是偶数，这就意味着

$$M \leqslant 2\left\lfloor \frac{d}{2d-N} \right\rfloor$$

当 M 是奇数时，式(2.1.16)的右端不大于 $N(M^2-1)/2$，基于此可得

$$M \leqslant \frac{N}{2d-N} = \frac{2d}{2d-N} - 1$$

这反过来说明

$$M \leqslant \left\lfloor \frac{2d}{2d-N} \right\rfloor - 1 \leqslant 2\left\lfloor \frac{d}{2d-N} \right\rfloor$$

因为，对于所有 $x>0$，都有$\lfloor 2x \rfloor \leqslant 2\lfloor x \rfloor + 1$。 □

定理 2.1.16 令 $M_q^*(N, d)$ 是二元$[N, d]$码的最大码本，那么对于任意 N 和 d，有

$$M_q^*(N, 2d-1) = M_q^*(N+1, 2d) \tag{2.1.17}$$

和

$$2M_2^*(N-1,d)=M_2^*(N,d) \tag{2.1.18}$$

证明　为了证明式(2.1.17)，取长度为 N，距离为 $2d-1$，码本为 $M_2^*(N,2d-1)$的码 $\mathscr{X}$。取它的奇偶校验扩展为$\mathscr{X}^+$。即在每个码字 $x=x_1\cdots x_N$后加上数位 x_{N+1}使得 $\sum_{i=1}^{N+1}x_i=0$，则$\mathscr{X}^+$是长度为 $N+1$，距离为 $2d$ 的码，且有相同码本 $M_2^*(N,2d-1)$，因此有

$$M_2^*(N,2d-1)\leqslant M_2^*(N+1,2d)$$

类似地，将最后一个数位删除将可得到相反的结果

$$M_2^*(N,2d-1)\geqslant M_2^*(N+1,2d)$$

[155]

现在证明(2.1.18)，给定一个$[N,d]$码，并将其码字分为两类：一类以 0 结尾，另一类以 1 结尾。其中一类至少包含一半码字，因此可以得到结果。 □

推论 2.1.17　如果 d 是偶数且 $2d>N$，那么

$$M_2^*(N,d)\leqslant 2\left\lfloor\frac{d}{2d-N}\right\rfloor \tag{2.1.19}$$

和

$$M_2^*(2d,d)\leqslant 4d \tag{2.1.20}$$

如果 d 是奇数且 $2d+1>N$，那么

$$M_2^*(N,d)\leqslant 2\left\lfloor\frac{d+1}{2d+1-N}\right\rfloor \tag{2.1.21}$$

和

$$M_2^*(2d+1,d)\leqslant 4d+4 \tag{2.1.22}$$

证明　不等式(2.1.19)由式(2.1.17)得到，而不等式(2.1.20)由式(2.1.18)和式(2.1.19)得到。如果 $d=2d'$，那么

$$M_2^*(4d',2d')=2M_2^*(4d'-1,2d')\leqslant 8d'=4d$$

更进一步，不等式(2.1.21)由式(2.1.17)得到

$$M_2^*(N,d)=M_2^*(N+1,d+1)\leqslant 2\left\lfloor\frac{d-1}{2d+1-N}\right\rfloor$$

最后，不等式(2.1.22)由式(2.1.17)和式(2.1.20)得到。 □

举例 2.1.18　证明 q-进制码的 Plotkin 界：

$$M_q^*(N,d)\leqslant\left\lfloor d\Big/\left(d-N\frac{q-1}{q}\right)\right\rfloor,\quad 如果\ d>N\frac{q-1}{q} \tag{2.1.23}$$

解答　给定一个 q-进制的$[N,M,d]$码$\mathscr{X}_N$，注意到其最小距离 d 是受到平均距离约束的

$$d\leqslant\frac{1}{M(M-1)}S,\quad 其中\ S=\sum_{x\in\mathscr{X}}\sum_{x'\in\mathscr{X}}\delta(x,x')$$

令 k_{ij}表示$\mathscr{X}$中所有码字的第 i 位置上字符 $j\in\{0,\cdots,q-1\}$出现的总次数，$i=1,\cdots,N$。显然 $\sum_{0\leqslant j\leqslant q-1}k_{ij}=M$，而第 i 位置对 S 的贡献是

$$\sum_{0\leqslant j\leqslant q-1}k_{ij}(M-k_{ij})=M^2-\sum_{0\leqslant j\leqslant q-1}k_{ij}^2\leqslant M^2-\frac{M^2}{q}$$

[157]

其中定义在$\{\boldsymbol{u}=u_1\cdots u_q:\ u_j\geqslant 0,\ \sum u_j=M\}$上形如$(u_1,\cdots u_q)\to\sum_{1\leqslant j\leqslant q}u_j^2$的二次函数在 $u_1=\cdots u_q=M/q$取得它的最小值。将所有 N 个数位求和并令 $\theta=(q-1)/q$，可得

$$M(M-1)d\leqslant\theta M^2N$$

这样会得到 M 的上界 $M\leqslant d(d-\theta N)^{-1}$，和二元码的情况类似，这里我们完成了对 q 进制码的证明。 □

目前已有很多关于 Plotkin 界（Hadamard 码）中等式成立的许多理论研究，但是本书将不对此做讨论。值得指出的是现有的编码界（Hamming，Singleton，GV，Plotkin）不仅仅只对线性码成立。对于 GV 界而言，它是可以由线性码达到的，详见定理 2.3.26.

举例 2.1.19 证明码长为 10，能纠正 2 个错误的二元纠错码，最多拥有 12 个码字。

解答 码的距离必须不小于 5。假定码包含 M 个码字，将它扩展为长度为 6 的[11，M]的码。Plotkin 界是以如下方式工作的：将扩展后的码列在一个 $M\times 11$ 矩阵的 M 行中。如果这个矩阵第 i 列包含了 s_i 个 0 和 $M-s_i$ 的 1，那么有

$$6(M-1)M\leqslant\sum_{x\in\mathscr{X}^+}\sum_{x'\in\mathscr{X}^-}\delta(x,x')\leqslant 2\sum_{i=1}^{11}s_i(M-s_i)$$

观察上式 RHS，当 M 是偶数时，RHS$\leqslant(1/2)\cdot 11M^2$。当 M 是奇数时，RHS$\leqslant(1/2)\cdot 11(M^2-1)$，因此，$M\leqslant 12$。

举例 2.1.20 （二元球尺寸的渐近性）令 $q=2$，$\tau\in(0,1/2)$，则有 $\eta(\tau)=-\tau\log_2\tau-(1-\tau)\log_2(1-\tau)$（参见式(1.2.2a)）

$$\lim_{N\to\infty}\frac{1}{N}\log v_{N,2}(\lfloor\tau N\rfloor)=\lim_{N\to\infty}\frac{1}{N}\log v_{N,2}(\lceil\tau N\rceil)=\eta(\tau) \tag{2.1.24}$$

解答 令 $R=\lceil\tau N\rceil$，下面和式的最后一项

$$v_{N,2}(R)=\sum_{i=0}^{R}\binom{N}{i},\quad R=\lceil\tau N\rceil$$

是最大的。实际上，两个连续项的比值是

158

$$\binom{N}{i+1}\Big/\binom{N}{i}=\frac{N-i}{i+1}$$

该比值在 $0\leqslant i\leqslant R$ 上不会小于 1。因此有

$$\binom{N}{R}\leqslant v_{N,2}(R)\leqslant(R+1)\binom{N}{R}$$

现在使用 Stirling 公式：$N!\sim N^{N+1/2}e^{-N}\sqrt{2\pi}$，则有

$$\log\binom{N}{R}=-(N-R)\log\frac{N-R}{N}-R\log\frac{R}{N}+O(\log N) \tag{2.1.25}$$

和

$$-\left(1-\frac{R}{N}\right)\log\left(1-\frac{R}{N}\right)-\frac{R}{N}\log\frac{R}{N}+\frac{O(\log N)}{N}$$
$$\leqslant\frac{\log v_{N,2}(R)}{N}\leqslant\frac{1}{N}\log(R+1)+\text{LHS}$$

取极限 $R/N\to\tau$ 就可证得结果。$R=\lfloor\tau N\rfloor$ 的情形可以用相似的方法讨论。 □

举例 2.1.20 在关于下式的渐近性研究中很有用

$$\alpha(N,\tau)=\frac{1}{N}\log M_2^*(N,\lceil\tau N\rceil) \tag{2.1.26}$$

上式给出了一个能纠正约 τN 个错误、检测约 $2\tau N$ 个错误（也就是数位 N 的所有数量的线性部分）的码的最大尺寸的信息速率。令

$$\underline{\alpha}(\tau):=\liminf_{N\to\infty}\alpha(N,\tau)\leqslant\limsup_{N\to\infty}\alpha(N,\tau)=:\overline{\alpha}(\tau) \tag{2.1.27}$$

对于这些极限我们有

定理 2.1.21 记 $\eta(\tau)=-\tau\log_2\tau-(1-\tau)\log_2(1-\tau)$，如下渐近界对二元码成立：

$$\overline{\alpha}(\tau)\leqslant 1-\eta(\tau/2),0\leqslant\tau\leqslant 1/2\quad(\text{Hamming}) \tag{2.1.28}$$

$$\overline{\alpha}(\tau)\leqslant 1-\tau,0\leqslant\tau\leqslant 1/2\quad(\text{Singleton}) \tag{2.1.29}$$

$$\underline{\alpha}(\tau)\geqslant 1-\eta(\tau),0\leqslant\tau\leqslant 1/2\quad(\text{GV}) \tag{2.1.30}$$

$$\overline{\alpha}(\tau)=0,1/2\leqslant\tau\leqslant 1\quad(\text{Plotkin}) \tag{2.1.31}$$

通过使用精细的界限(同样得益于 Plotkin 界)，我们将在问题 2.10 中证明

$$\overline{\alpha}(\tau)\leqslant 1-2\tau,\quad 0\leqslant\tau\leqslant 1/2 \tag{2.1.32}$$

定理 2.1.21 的证明正是基于对上面所提界的直接观察。在举例 2.1.22 中，将有对 Hamming 和 GV 界的应用。

159

图 2-2 展现了各类界的形状。码的“好”序列是指对于$(\tau,\alpha(N,\lceil\tau N\rceil))$是渐近局限于表示渐近边界曲线之间的域中。特别地，一个“好”码应该位于 GV 界的曲线之上。构造出这样的序列是一个困难的问题：第一个实现渐近 GV 界的例子出现在 1973 年(基于代数几何思想的 Goppa 码)。尽管本书讨论的所有码家族在特殊值 N，M 和 d 上表现了令人印象深刻的特性，但这些码产生的值都低于 GV 曲线(事实上，它们使 $\alpha(\tau)=0$)。

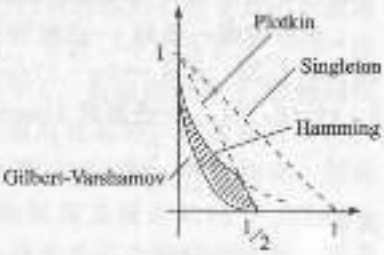


图 2-2

就上界而言，Hamming 和 Plotkin 不相上下，而 Singleton 界逐渐变得不重要(尽管对于特殊的 N，M 和 d 值它还是相当重要的)。大约存在有十几个其他上界，其中有些将会在本书的本节和后面章节讨论。

Gilbert-Varshamov 界本身不一定是最优的。在 1982 年之前，并不知道有其他更好的下界(在二元编码的情形下仍然不知道是否存在其他更好的下界)。然而，如果使用的字母表包含了 $q\geqslant 49$ 符号，其中 $q=p^{2m}$，$p\geqslant 7$，且 p 是一个素数，则存在一种同样基于几何代数的构造，它将产生一个不同的下界，当 $N\to\infty$ 时，给出(线性)码例子将能逐渐超过 GV 曲线[169]。此外，TVZ 构造会产生多项式的复杂度。随后，又提出另外两个下界：(a)Elkies界，其中 $q=p^{2m}-1$；(b)Xing 界，其中 $q=p^m$[43，175]。N. Elkies 的文章“Excellent codes from modular curves”指出通过使用不同的编码构造，同样可以改善其他字母表的 GV 界。 160

举例 2.1.22 证明式(2.1.28)和式(2.1.30)中给定的界(定理 2.1.21 中和渐近 Hamming 及 GV 界有关的部分)。

解答 考虑式(2.1.14)中 Hamming 和 GV 界的部分内容，有

$$2^N/v_{N,2}(d-1)\leqslant M_2^*(N,d)\leqslant 2^N/v_{N,2}(\lfloor(d-1)/2\rfloor) \tag{2.1.33}$$

Hamming 界中体积的下界是平凡的

$$v_{N,2}(\lfloor(d-1)/2\rfloor)\geqslant\begin{bmatrix}N\\ \lfloor(d-1)/2\rfloor\end{bmatrix}$$

对于上界而言，注意到当 $d/N\leqslant\tau<1/2$ 时

$$v_{N,2}(d-1)\leqslant\sum_{0\leqslant i\leqslant d-1}\left(\frac{d-1}{N-d+1}\right)^{d-1-i}\begin{bmatrix}N\\ i\end{bmatrix}$$

$$\leqslant \sum_{1\leqslant i\leqslant d-1}\left(\frac{\tau}{1-\tau}\right)^{d-1-i}\binom{N}{d-1}\leqslant \frac{1-\tau}{1-2\tau}\binom{N}{d-1}$$

然后，对于信息速率$(\log M_2^*(N,d))/N$，

$$1-\frac{1}{N}\log\left(\frac{1-\tau}{1-2\tau}\binom{N}{d-1}\right)$$

$$\leqslant \frac{1}{N}\log M_2^*(N,d)\leqslant 1-\frac{1}{N}\log\binom{N}{\lfloor (d-1)/2\rfloor}$$

由 Stirling 公式，当 $N\to\infty$ 时，上述不等式中的对数函数满足

$$\frac{1}{N}\log\binom{N}{\lfloor (d-1)/2\rfloor}\to \eta(\tau/2),\ \frac{1}{N}\log\binom{N}{d-1}\to \eta(\tau)$$

于是很容易得到式(2.1.28)和式(2.1.30)中的界。 □

现在考虑一般的 q-进制符号集的情况。

例子 2.1.23 令 $\theta:=(q-1)/q$。通过修改在举例 2.1.22 中的论证，证明对于任意的 $q\geqslant 2$，$\tau\in(0,\theta)$，q-进制的 Hamming 球的体积具有如下的对数渐近线形式：

181

$$\lim_{N\to\infty}\frac{1}{N}\log_q v_{N,q}(\lfloor \tau N\rfloor)=\lim_{N\to\infty}\frac{1}{N}\log_q v_{N,q}(\lceil \tau N\rceil)=\eta^{(q)}(\tau)+\tau\kappa \tag{2.1.34}$$

其中

$$\eta^{(q)}(\tau):=-\tau\log_q\tau-(1-\tau)\log_q(1-\tau),\quad \kappa:=\log_q(q-1) \tag{2.1.35}$$

下一步，同式(2.1.26)类似，引入

$$a^{(q)}(N,\tau):=\frac{1}{N}\log M_q^*(N,\lceil \tau N\rceil) \tag{2.1.36}$$

和极限

$$\underline{a}^{(q)}(\tau):=\liminf_{N\to\infty}a^{(q)}(N,\tau)\leqslant \limsup_{N\to\infty}a^{(q)}(N,\tau)=:\overline{a}^{(q)}(\tau) \tag{2.1.37}$$

定理 2.1.24 *对于所有 $0<\tau<\theta$，有*

$$\overline{a}^{(q)}(\tau)\leqslant 1-\eta^{(q)}(\tau/2)-\kappa\tau/2\quad(\text{Hamming}) \tag{2.1.38}$$

$$\overline{a}^{(q)}(\tau)\leqslant 1-\tau\quad(\text{Singleton}) \tag{2.1.39}$$

$$\underline{a}^{(q)}(\tau)\geqslant 1-\eta^{(q)}(\tau)-\kappa\tau\quad(\text{GV}) \tag{2.1.40}$$

$$\overline{a}^{(q)}(\tau)\leqslant \max[1-\tau/\theta,0]\quad(\text{Plotkin}) \tag{2.1.41}$$

当然，式(2.1.38)、(2.1.39)和(2.1.41)右侧的最小值提供了这三个上界的较优选择。我们在这省略了定理 2.1.24 的证明，把它留作一个练习，基本上是举例 2.1.22 证明的重复。

例子 2.1.25 通过修改举例 2.1.22 的解，证明式(2.1.38)和式(2.1.40)给出的界。

2.2 Shannon 第二编码定理的几何证明，码本规模的精细界

在这节中我们将利用 Hamming 空间集合给出 Shannon 第二编码定理(或 Shannon 噪声编码定理，SCT/NCT；参见定理 1.4.14 和 1.4.15)两个部分的另一种证明。我们采用了证明中的一些技巧来得到一些关于编码的高级界限，这些先进的界进一步确定了在定理 2.1.6 中给出的 Hamming 界，以及在定理 2.1.21 和 2.1.24 中相应的渐近结论。

SCT/NCT 的直接部分在下面的定理 2.2.1 中给出，与定理 1.4.14 和 1.4.15 相比这里给出的是一种修改过的形式。为了简洁，我们在这里只考虑在空间$\mathscr{H}_{N,2}=\{0,1\}^N$ 中

运作的(为了简洁下标 2 将会被省略)无记忆二进制对称信道(MBSC)。从 1.4 节中我们知道，SCT 的直接表述部分表明对于任意传输速率 $R<C$，存在： 162

(i) 一列码 f_n：$\mathscr{U}_n\to\mathscr{H}_N$，总共编码 $\#\mathscr{U}_n=2^n$ 个消息。

(ii) 一列译码规则 f_N^{-1}：$\mathscr{H}_N\to\mathscr{U}_n$，使得 $n\sim NR$ 且当 $n\to\infty$ 时，错误解码概率趋于零。

在这里 C 由式(1.4.11)和式(1.4.27)给出。为了方便，我们重新写出 C 的表达式：

$$C=1-\eta(p),\quad 其中\ \eta(p)=-p\log_2 p-(1-p)\log(1-p) \tag{2.2.1}$$

和信道矩阵是

$$\Pi=\begin{bmatrix}1-p & p\\ p & 1-p\end{bmatrix} \tag{2.2.2}$$

我们假定信道以概率 $1-p$ 正确传输一个字符，错误的概率为 p。上述概率对于不同字符是相互独立的。

定理 2.2.1 中，指出存在一个一对一的码映射 f_n 序列可以使得译码的任务简化为猜测码字 $f_n(u)\in\mathscr{H}_N$。也就是说，定理保证了对于所有的 $R<C$，存在一个包含 $\#\mathscr{X}_N\sim 2^{NR}$ 序列的子集 $\mathscr{X}_N\subset\mathscr{H}_N$，在其中进行猜测(译码)的错误概率趋于 0，而编码映射 f_n 的确切性质不是很重要。然而，让映射 f_n 能确切地展示在眼前是方便而有益的，因为它可以由一个概率化的构造(随机编码)得到，而其中某些编码映射不必是一对一的。不仅如此，解码的准则也是几何的：当接收到的码字 $u^{(N)}\in\mathscr{H}_N$，我们寻找距离其最近的码字 $f_n(u)\in\mathscr{X}_N$。因此，每个错误都会被声明，也就是码字是非唯一的或者是多个编码的结果，或者是单纯产生了错误信息。像我们早前看到的，当概率 $p\in(0,1/2)$ 时，几何译码规则与 ML 检测器是相对应的。这种检测器可以使我们利用几何的观点来构成证明的核心。

也是在 1.4 节中，SCT/NCT 的直接部分的新证明只能保证好码(甚至它们的衍生)的存在性，但是没有给出提示如何去构造这些码(除了再一次使用随机编码方案，且挑选它的典型的实现)。

在下面给出的关于 SCT/NCT 的阐述中，我们将处理最大差错概率(2.2.4)，而非平均差错概率。然而，证明的大部分仍然是基于对码字平均差错概率的直接分析。

定理 2.2.1 (SCT/NCT，直接部分)考虑一个 MBSC，其信道矩阵 Π 如式(2.2.2)所示，其中 $0\leqslant p\leqslant 1/2$，令 C 如式(2.2.1)中所示。于是，对任意 $R\in(0,C)$，存在一个一一对应的编码序列 f_n：$\mathscr{U}_n\to\mathscr{H}_N$，使得 163

(i)

$$n=\lfloor NR\rfloor,\quad \#\mathscr{U}_n=2^n \tag{2.2.3}$$

(ii) 当 $n\to\infty$，在几何解码法则下，最大错误概率趋于 0：

$$e^{\max}(f_n)=\max[\boldsymbol{P}_n(在几何编码规则下发生错误\ |f_n(u)\ 发送);u\in\mathscr{U}_n]\to 0 \tag{2.2.4}$$

其中，$\boldsymbol{P}_n(\cdot\,|f_n(u)发送)$表示，在已发送码字 $f_n(u)$的条件下，信道产生的$\mathscr{H}_N$ 中的接收码元的概率分布。其中 $u\in\mathscr{U}_n$ 为信源发出的初始信息。

为了进一步解释该结论，考虑下面这个例子。

例子 2.2.2 假设要发送一个信息 $u\in\mathscr{A}^n$，其中符号集$\mathscr{A}$的大小为 K，该信息通过一个信道矩阵为 $\begin{bmatrix}0.85 & 0.15\\ 0.15 & 0.85\end{bmatrix}$ 的 MBSC。在任意小的差错概率要求下，能达到的传输速率为多少？

这里，$C=1-\eta(0.15)=0.577291$。因此，根据定理 2.2.1，当 n 充分大时，可达到

任意小于0.577291的传输速率，且差错概率可以任意小。例如，当想要实现$0.5<R<0.577291$的传输速率且$e_{max}<0.015$，则当n充分大即$n>n_0$时，存在符合要求的码字f_n：$\mathscr{A}^n\to\{0,1\}^{\lceil nR\rceil}$。

假设我们知道这样一种码字f_n，如何对信息m进行编码？首先，将m划分为长度为L的块，其中

$$L=\left\lceil\frac{0.577291N}{\log K}\right\rceil,\quad 所以|\mathscr{A}^L|=K^L\leqslant 2^{0.577291N}$$

然后我们可以将来自$\mathscr{A}^L$的块嵌入到符号集$\mathscr{A}^n$中，从而对该块编码。对应的传输速率为$\log|\mathscr{A}^L|/\lceil n/R\rceil\sim 0.577291$。正如我们已知的，SCT告诉我们存在这样的码字，但是却没有说明如何找到(构造)这样的码字，而这正是困难之处。

在进行定理2.2.1的证明之前，先研究一下Hamming空间$\mathscr{H}_N$的几何特征和信道随机性之间的关联。正如1.4节所示，我们用$P(\cdot\mid f_n(u))$表示$P_n(\cdot\mid f_n(u)发送)$的缩写，
156 在该分布下的均值和方差分别为$E(\cdot\mid f_n(u))$和$\mathrm{Var}(\cdot\mid f_n(u))$。

注意到，在分布$P(\cdot\mid f_n(u))$下，在(随机)接收码字$Y^{(N)}$的失真位数可以表示为

$$\sum_{j=1}^{N}\mathbf{1}(位数\ j\ 在Y^{(N)}\neq 位数\ j\ 在\ f_n(u))$$

这是一个服从二项分布$\mathrm{Bin}(N,p)$的随机变量，其中均值为

$$E\left[\sum_{j=1}^{N}\mathbf{1}(位数\ j\ 在Y^{(N)}\neq 位数\ j\ 在\ f_n(u))\mid f_n(u)\right]$$

$$=\sum_{j=1}^{N}E[\mathbf{1}(位数\ j\ 在Y^{(N)}\neq 位数\ j\ 在\ f_n(u))\mid f_n(u)]=Np$$

方差为

$$\mathrm{Var}\left[\sum_{j=1}^{N}\mathbf{1}(位数\ j\ 在Y^{(N)}\neq 位数\ j\ 在\ f_n(u))\mid f_n(u)\right]$$

$$=\sum_{j=1}^{N}\mathrm{Var}[\mathbf{1}(位数\ j\ 在Y^{(N)}\neq 位数\ j\ 在\ f_n(u))\mid f_n(u)]=Np(1-p)$$

根据Chebyshev不等式，对于任意的$\varepsilon\in(0,1-p)$和正整数$N>1/\varepsilon$，发送码字$f_n(u)$后，至少有$\lceil N(p+\varepsilon)\rceil$个失真位数的概率为

$$\leqslant P(\geqslant N(p+\varepsilon)\ 个失真\mid f_n(u))\leqslant\frac{p(1-p)}{N(\varepsilon-1/N)^2}\tag{2.2.5}$$

定理2.2.1的证明　在整个证明过程中，我们假设和式(2.2.3)一样的条件，省略下标n和N，即

$$2^{nR}=M$$

我们还将假设ML/几何解码。和1.4节类似，我们用Hamming空间$\mathscr{H}_N$来确定信源信息集合$\mathscr{U}_n$。正如Shannon提出的，我们再次使用随机编码。更准确地说，一条信息$u\in\mathscr{H}_n$被映射为一个随机码字$F_n(u)\in\mathscr{H}_N$，其中每个数字均为独立同分布且取值为0和1概率分别为1/2。此外，我们令码字$F_n(u)$独立于不同的信息$u\in\mathscr{H}_n$；将来自$\mathscr{H}_n$的字符串标记为$u(1),\cdots,u(M)$(无特殊排序)，从而得到一族取自$\mathscr{H}_n$的IID随机字符串$F_n(u(1)),\cdots,F_n(u(M))$。最后，我们令码字独立于信道。再次由1.4节的分析得到，我们可以认为这些随机码元为$\mathscr{H}_{NM}=\{0,1\}^{NM}$中的随机超字符串或码符号集，其中每个数
165 字均为独立同分布且取值为0和1概率分别为1/2。每个给定的随机码本的采样$f(=f_n)$

(即任给取自$\mathscr{H}_{NM}$的超字符串)对应信息$u(1)$，…，$u(M)$的确定编码$f_n(u(1))$，…，$f_n(u(M))$，即编码f，参考图2-3。

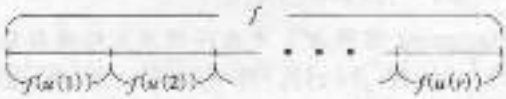


图 2-3

正如1.4节所示，我们将随机码元的概率分布用$\mathscr{P}_n$表示，即

$$\mathscr{P}_n(F_n = f) = \frac{1}{2^{NM}}, \quad 对于所有可抽样\ f \tag{2.2.6}$$

且其相对于$\mathscr{P}_n$的期望为$\mathscr{E}_n$。

接下来的证明步骤如下。首先，我们将证明(通过重复1.4节的部分讨论)，当传输速率$R\in(0, C)$时，随着$n\to\infty$，对以上随机编码的平均概率的期望为

$$\lim_{n\to\infty}\mathscr{E}_n[e^{m}(F_n)] = 0 \tag{2.2.7}$$

这里，$e^{m}(F_n)$是$e^{m}(F_n(u(1))$，…，$F_n(u(M)))$的缩写，且为一个取值为(0，1)的随机变量，表示前面提到的随机编码的平均错误概率。更准确地说，正如1.4节所示，对于所有给定的码字采样集合$f(u(1))$，…，$f(u(M))\in\mathscr{H}_N$(即对所有取自$\mathscr{H}_{NM}$的超字符串f)，我们定义

$$e^{m}(f_n) = \frac{1}{M}\sum_{1\le i\le M}\boldsymbol{P}(当用码体\ f|f(u(i))\ 时产生错误) \tag{2.2.8}$$

于是，平均错误概率的期望为

$$\mathscr{E}_n[e^{m}(F_n)] = \frac{1}{2^{NM}}\sum_{f(u(1)),\cdots,f(u(M))\in\mathscr{H}_N} e^{m}(f) \tag{2.2.9}$$

式(2.2.9)说明了(用和1.4节类似的方式)存在一个确定码元序列f^n，其平均错误概率$e^{m}(f_n)=e^{m}(f_n(u(1))$，…，$f_n(u(2^n)))$服从

$$\lim_{n\to\infty}e^{m}(f_n) = 0 \tag{2.2.10}$$

最后，我们将从式(2.2.10)式推出式(2.2.4)，参考引理2.2.6。

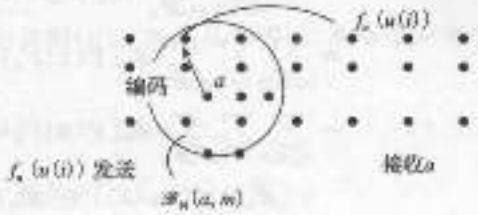


图 2-4

备注2.2.3 由于认为码字$f(u(1))$，…，$f(u(M))$来自随机码本的采样，因此，必须允许码字重复的情况($f(u(i))=f(u(j))$当$i\neq j$)，在这种情形下，ML译码器将会出错，当我们考虑式(2.2.8)的右式的概率时，必须要包含这种情形，因此，对$i=1$，…，M，我们定义

$$\begin{aligned}&\boldsymbol{P}(当用码体\ f|f(u(i))\ 时产生错误)\\ &=\begin{cases}1, & 若对某些\ i'\neq i\ 有\ f(u(i))=f(u(i'))\\ \boldsymbol{P}(\delta(\boldsymbol{Y}^{(N)}, f(u(j)))\le\delta(\boldsymbol{Y}^{(N)}, f(u(i))) & 对某些\ j\neq i|f(u(i)),\\ \quad 若对所有\ i'\neq i\ 有\ f(u(i))\neq f(u(i')) & \end{cases}\end{aligned} \tag{2.2.11}$$

现在我们将深入讨论细节。第一步如下。

引理2.2.4 考虑信道矩阵Π(参考式(2.2.2))其中$0\le p<1/2$，假设传输速率$R<C-1-\eta(p)$，令$N>1/\varepsilon$。于是，对任意$\varepsilon\in(0, 1/2-p)$，由式(2.2.8)和式(2.2.9)定义的平均

错误概率的期望 $\mathscr{E}_r[e^{ML}(F_r)]$ 服从

$$\mathscr{E}_r[e^{ML}(F_r)] \leqslant \frac{p(1-p)}{N(\varepsilon-1/N)^2}+\frac{M-1}{2^N}v_N(\lceil N(p+\varepsilon)\rceil) \tag{2.2.12}$$

其中 $v_N(b)$ 表示二进制 Hamming 空间 $\mathscr{H}_N$ 中在半径为 b 的球形区域中的点数。

[167] 证明 令 $m(=m_N(p,\varepsilon))=\lceil N(p+\varepsilon)\rceil$。当 $f_r(u(i))$ 为在以接收码字 $y(=y^{(N)})\in\mathscr{H}_N$ 为球心的 Hamming 球 $\mathscr{B}_N(y,m)$ 中唯一的码字时，ML 译码器必然返回码字 $f_r(u(i))$（参考图 2.4）。在任意其他的情形中（对某些 $k\neq i$，$f_r(u(i))\notin\mathscr{B}_N(y,m)$ 或 $f_r(u(k))\in\mathscr{B}_N(y,m)$），存在错误概率。

因此

$$\begin{aligned}
&\boldsymbol{P}(\text{当用码本 } f \mid f_r(u(i)) \text{ 时发生错误})\\
&\leqslant \sum_{y\in\mathscr{H}_N}\boldsymbol{P}(y\mid f_r(u(i)))\mathbf{1}(f_r(u(i))\notin\mathscr{B}_N(y,m)) \qquad (2.2.13)\\
&\quad+\sum_{x\in\mathscr{H}_N}\boldsymbol{P}(x\mid f_r(u(i)))\sum_{k\neq i}\mathbf{1}(f_r(u(k))\in\mathscr{B}_N(x,m))
\end{aligned}$$

借助(2.2.5)式，右端的求和可以简单地估计为

$$\begin{aligned}
&\sum_{y\in\mathscr{H}_N}\boldsymbol{P}(y\mid f_r(u(i)))\mathbf{1}(f_r(u(i))\notin\mathscr{B}_N(y,m))\\
&=\sum_{y\in\mathscr{H}_N}\boldsymbol{P}(y\mid f_r(u(i)))\mathbf{1}(\text{距离 } \partial(y,f_r(u(i)))\geqslant m) \qquad (2.2.14)\\
&=\boldsymbol{P}(\geqslant m \text{ 位失真} \mid f(u(i)))\leqslant\frac{p(1-p)}{N(\varepsilon-1/N)^2}
\end{aligned}$$

注意到由于式(2.2.14)的右式并不依赖于采样码字 f 的选取，当我们先取平均 $\frac{1}{M}\sum_{1\leqslant i\leqslant M}$ 再取期望就可以得到式(2.2.14)的界。 □

式(2.2.13)右端的第二个求和更为巧妙，它需要取平均再取期望。这里，由于随机码字 $F_r(u(1)),\cdots,F_r(u(M))$ 是相互独立的，我们有

$$\begin{aligned}
&\mathscr{E}_r\Big[\sum_{1\leqslant i\leqslant M}\sum_{x\in\mathscr{H}_N}\boldsymbol{P}(x\mid F_r(u(i)))\sum_{k\neq i}\mathbf{1}(F_r(u(k))\in\mathscr{B}_N(x,m))\Big]\\
&=\sum_{1\leqslant i\leqslant M}\sum_{k\neq i}\sum_{x\in\mathscr{H}_N}\mathscr{E}_r[\boldsymbol{P}(x\mid F_r(u(i)))\mathbf{1}(F_r(u(k))\in\mathscr{B}_N(x,m))]\\
&=\sum_{1\leqslant i\leqslant M}\sum_{k\neq i}\sum_{x\in\mathscr{H}_N}\mathscr{E}_r[\boldsymbol{P}(x\mid F_r(u(i)))]\\
&\quad\times\mathscr{E}_r[\mathbf{1}(F_r(u(k))\in\mathscr{B}_N(x,m))] \qquad (2.2.15)
\end{aligned}$$

接下来，由于这些码字的分布均为 $\mathscr{H}_N$ 中的相同分布，期望 $\mathscr{E}_r[\boldsymbol{P}(x\mid F_r(u(i)))]$ 和 $\mathscr{E}_r[\mathbf{1}(F_r(u(k))\in\mathscr{B}_N(x,m))]$ 可以计算如下

$$\mathscr{E}_r[\boldsymbol{P}(x\mid F_r(u(i)))]=\frac{1}{2^N}\sum_{x\in\mathscr{H}_N}\boldsymbol{P}(x\mid x) \tag{2.2.16a}$$

[168]

$$\mathscr{E}_r[\mathbf{1}(F_r(u(k))\in\mathscr{B}_N(x,m))]=\frac{v_N(m)}{2^N} \tag{2.2.16b}$$

接着在 x 域求和得到

$$\sum_{x\in\mathscr{H}_N}\sum_{x\in\mathscr{H}_N}\boldsymbol{P}(x\mid x)=\sum_{x\in\mathscr{H}_N}\sum_{x\in\mathscr{H}_N}\boldsymbol{P}(x\mid x)=2^N \tag{2.2.17}$$

最后在 $k\neq i$ 条件下求和得到

$$\text{式(2.2.15)的 RHS}=\frac{1}{M}\sum_{1\le i\le M}\sum_{k\ne i}\frac{v_N(m)}{2^N}$$

$$=\frac{v_N(m)M(M-1)}{2^N M}\le\frac{(M-1)v_N(m)}{2^N}\tag{2.2.18}$$

联立式(2.2.12)到式(2.2.18)，我们可以推出$\mathscr{E}_N[e^{\text{av}}(F_\varepsilon)]$不会超过式(2.2.12)的右式。

接下来，我们从熵$h(p+\varepsilon)$的角度估计$v_N(m)$，其中$m=\lceil N(p+\varepsilon)\rceil$。这里的讨论类似于1.4节且基于下面的结论。

引理 2.2.5 假设$0<p<1/2$，$\varepsilon>0$且正整数N满足$p+\varepsilon+1/N<1/2$。于是下述边界成立：

$$v_N(\lceil N(p+\varepsilon)\rceil)\le 2^{N\eta(p+\varepsilon)}\tag{2.2.19}$$

引理2.2.5的证明将在例2.2.7之后给出。现在我们先证明定理2.2.1。即，我们想要证明(2.2.7)。事实上，如果$p<1/2$且$R<C=1-\eta(p)$，则我们取$\zeta=C-R>0$且$\varepsilon>0$。于是得到(i)$p+\varepsilon<1/2$；(ii)$R+\zeta/2<1-\eta(p+\varepsilon)$。当$N$取值非常大时，(iii)$N>2/\varepsilon$。在这样的$\varepsilon$和$N$的选取下，我们得到

$$\varepsilon-\frac{1}{N}>\frac{\varepsilon}{2}\quad\text{和}\quad R-1+\eta(p+\varepsilon)<-\frac{\zeta}{2}\tag{2.2.20}$$

于是，从式(2.2.12)开始，我们得到

$$\mathscr{E}_N[e(F_\varepsilon)]\le\frac{4p(1-p)}{N\varepsilon^2}+\frac{2^{NR}}{2^N}2^{N\eta(p+\varepsilon)}$$

$$<\frac{4}{N\varepsilon^2}p(1-p)+2^{-N\zeta/2}\tag{2.2.21}$$

这证明了式(2.2.7)。因此，存在一个编码序列$f_\varepsilon:\mathscr{M}_n\to\mathscr{M}_N$满足式(2.2.10)。

为了完成定理2.2.1的证明，我们根据引理2.2.6，从式(2.2.7)推出式(2.2.4)。

引理 2.2.6 考虑一个二元信道(并不一定要无记忆)，令$C>0$为给定常数。$0<R<C$且$n=\lfloor NR\rfloor$，定义码字$f_\varepsilon:\mathscr{M}_n\to\mathscr{M}_N$和$\tilde{f}_\varepsilon:\mathscr{M}_n\to\mathscr{M}_N$的数量$e_{\max}(f_\varepsilon)$和$e^{\text{av}}(\tilde{f}_\varepsilon)$如式(2.2.4)、(2.2.8)和(2.2.10)中所示。于是，下述声明等价：

(i) 对所有的$R\in(0,C)$，存在码字f_ε且$\lim_{\varepsilon\to\infty}e_{\max}(f_\varepsilon)=0$。

(ii) 对所有的$R\in(0,C)$，存在码字$\tilde{f}_\varepsilon$且$\lim_{\varepsilon\to\infty}e_{\max}(\tilde{f}_\varepsilon)=0$。

169

引理 2.2.6 的证明 显而易见，声明(i)可以导出声明(ii)。为了从声明(ii)推导出声明(i)，取$R<C$并且让N足够大

$$R'=R+\frac{1}{N}<C,n'=\lfloor NR'\rfloor,M'=2^{n'}\tag{2.2.22}$$

我们知道，存在一个$\mathscr{M}_n\to\mathscr{M}_N$的码字序列$\tilde{f}_\varepsilon$，满足$e^{\text{av}}(\tilde{f}_\varepsilon)\to 0$。即

$$e^{\text{av}}(\tilde{f}_\varepsilon)\frac{1}{M'}\sum_{1\le i\le M'}\boldsymbol{P}(\text{采用码本}\tilde{f}_\varepsilon\mid\tilde{f}_\varepsilon(u(i))\text{时发生错误})\tag{2.2.23}$$

在接下来的证明中，设$M'=2^{\lfloor NR'\rfloor}$且$\tilde{f}_\varepsilon(u(1))$，…，$\tilde{f}_\varepsilon(u(M))$为对应信源消息$u(1)$，…，$u(M)\in\mathscr{M}_{n'}$的码字。

方便起见，我们用$\boldsymbol{P}(f_\varepsilon\text{-error}\mid\tilde{f}(u(i)))$代替$\boldsymbol{P}(\text{加入码体}f_\varepsilon\mid\tilde{f}(u(i))\text{时发生错误})$。现在，式(2.2.23)的右端至少一半的加数$\boldsymbol{P}(\tilde{f}_\varepsilon\text{-error}\mid\tilde{f}(u(i)))$小于$2e^{\text{av}}(\tilde{f}_\varepsilon)$。注意到，由式(2.2.22)有

$$M'/2\ge 2^{\lfloor NR\rfloor-1}\tag{2.2.24}$$

因此，我们有至少有$2^{\lfloor NR\rfloor-1}$个码字$f(u(i))$满足

$$\boldsymbol{P}(\text{error}\mid\tilde{f}_\varepsilon(u(i)))<2e^{\text{av}}(\tilde{f}_\varepsilon)$$

将这些码字重新列为一个新的二元码，其长度为 N 且信息速率为 $(\log M'/2)/N$。将此新的编码表示为 f_n，我们得到

$$e^{(n)}(f_n) \leqslant 2e^{(n)}(\tilde{f}_n)$$

因此，随着 $n\to\infty$，$e^{(n)}(f_n)\to 0$ 且 $(\log M'/2)/N\to R$。此即为声明(i)，到此完成了引理 2.2.6 的证明。 □

于是，在给出引理 2.2.5 的证明后，定理 2.2.1 的证明到此完成。

举例 2.2.7 （参考举例 2.1.20）证明对正整数 N 和 m，且 $m<N/2$，$\beta=m/N$，

170

$$2^{N\eta(\beta)}/(N+1) < v_N(m) < 2^{N\eta(\beta)} \tag{2.2.25}$$

证明 设

$$v_N(m) = \#\{\text{点之间的距离} \leqslant m \text{ 在 } \mathscr{H}_N \text{ 上与 } \mathbf{0} \text{ 的距离}\} = \sum_{0\leqslant k\leqslant m}\binom{N}{k}$$

且 $\beta=m/N<1/2$，我们得到 $\beta/(1-\beta)<1$，于是

$$\left(\frac{\beta}{1-\beta}\right)^m < \left(\frac{\beta}{1-\beta}\right)^k, \quad 对\ 0\leqslant k<m$$

对 $0\leqslant k<m$，乘积变为

$$\begin{aligned}\beta^k(1-\beta)^{N-k} &= \left(\frac{\beta}{1-\beta}\right)^k(1-\beta)^N \\ &> \left(\frac{\beta}{1-\beta}\right)^m(1-\beta)^N = \beta^m(1-\beta)^{N-m}\end{aligned}$$

因此

$$\begin{aligned}1 &= \sum_{0\leqslant k\leqslant N}\binom{N}{k}\beta^k(1-\beta)^{N-k} > \sum_{0\leqslant k\leqslant m}\binom{N}{k}\beta^k(1-\beta)^{N-k} \\ &> \beta^m(1-\beta)^{N-m}\sum_{0\leqslant k\leqslant m}\binom{N}{k} = v_N(m)\beta^m(1-\beta)^{N-m} \\ &= v_N(m)2^{N(-m/N)\log\beta+(1-m/N)\log(1-\beta)}\end{aligned}$$

该式说明了 $v_N(m)<2^{N\eta(\beta)}$，为了得到式(2.2.25)中左式的边界，有

$$v_N(m) > \binom{N}{m}$$

于是我们的目标变为检查右式不小于 $2^{N\eta(\beta)}/(N+1)$。考虑一个二项分布随机变量 $Y\sim \mathrm{Bin}(N,\beta)$，且

$$p_k = \mathbb{P}(Y=k) = \binom{N}{k}\beta^k(1-\beta)^{N-k}, k=0,\cdots,N$$

于是，可以证明当 $k=m$ 时，p_k 取得其最大值。

$$p_m = \binom{N}{m}\beta^m(1-\beta)^{N-m} \geqslant \frac{1}{N+1}, \quad 其中\ \beta^m(1-\beta)^{N-m} = 2^{-N\eta(\beta)}$$

这时，假设 $k\leqslant m$，有

$$\begin{aligned}\frac{p_k}{p_m} &= \frac{m!(N-m)!(N-m)^{m-k}}{k!(N-k)!m^{m-k}} \\ &= \frac{(k+1)\cdots m}{m^{m-k}}\cdot\frac{(N-m)^{m-k}}{(N-m+1)\cdots(N-k)}\end{aligned}$$

171

由于等式的右端是 $2(m-k)$ 个小于等于 1 的因式相乘，故其不大于 1。类似地，如果 $k\geq m$，我们可以得到乘积

$$\frac{m^{k-m}}{(m+1)\cdots k}\cdot\frac{(N-k+1)\cdots(N-m)}{(N-m)^{k-m}}$$

基于和上述相同的理由，其仍然小于等于 1。因此，比值 $p_k/p_m\leq 1$ 从而得到设定的边界。 □

接下来我们证明引理 2.2.5。

引理 2.2.5 的证明 首先，由 $p+\varepsilon<1/2$ 可以推出 $m=\lfloor N(p-\varepsilon)\rfloor<N/2$ 且

$$\beta_\varepsilon=\frac{m}{N}=\frac{\lfloor N(p-\varepsilon)\rfloor}{N}<p+\varepsilon$$

由于 $x\mapsto\eta(x)$ 在区间 $(0,1/2)$ 内是一个严格递增的函数，于是这反过来说明 $\eta(\beta)<\eta(p-\varepsilon)$。这导出了引理 2.2.5 的结论。 □

SCT/NCT 直接部分的几何证明澄清了（MBSC）容量的概念。具体而言，在式(1.4.11)、式(1.4.27)和式(2.2.1)中对 MBSC 的 $C=\eta(p)$ 的表达中，正项 1 说明了随机码在码字之间产生空隙时所对应的速率，而负项 $-\eta(p)$ 说明了当码字渐近地填充了该空间时所对应的速率。我们接下来展示一个例子。 □

举例 2.2.8 引用信道容量估计的相关理论，推导二元对称无记忆信道的信道容量表达式。特别地，估计(i)对称无记忆信道的容量和(ii)一个输入符号集为{0，1}且输入不会出现连续的 0 的完美信道的信道容量。

解答 信道容量的定义为，能正确解码接收码字的速率 R 的上确界，且正确解码的概率随着信息长度趋于无穷而趋于 1。对于给定的输入码字 $x^{(N)}=x_1\cdots x_N$，无记忆信道的概率的经典形式为

$$P^{(N)}(y^{(N)}\text{ 接收}\mid x^{(N)}\text{ 发送})=\prod_{1\leq i\leq N}P(y_i\mid x_i)$$

换言之，噪声独立作用于输入字符中 x 的每一个符号 x_i 上，$P(y\mid x)$ 表示在给定输入 x 的条件下，输出为 y 的概率。 172

符号 x 取自一个大小为 q 的输入符号集 $\mathscr{A}_{in}$，y 取自大小为 r 的输出符号集 $\mathscr{A}_{out}$。概率 $P(y\mid x)$ 形成了一个 $q\times r$ 的随机矩阵（信道矩阵）。对一个无记忆信道，若该矩阵的每一行都是其他行的排列，例如每一行都包含相同的概率集合 $p_1,\cdots,p_r$，则称该信道为对称的。同样地，对无记忆信道，如果该矩阵的每一列也是其他列的排列，则称该信道为双重对称的。如果 $m=n=2$（典型地，$\mathscr{A}_{in}=\mathscr{A}_{out}=\{0,1\}$），则该无记忆信道为二元的。对一个无记忆二元对称信道，信道矩阵元素 $P(y\mid x)$ 为 $P(0\mid 0)=P(1\mid 1)=1-p$，$P(1\mid 0)=P(0\mid 1)=p$，$p\in(0,1)$ 为单个二元符号翻转的概率，$1-p$ 为其正确传输的概率。

用信道容量 $C\geq 0$ 来表示信道的特性：

(a) 对所有的 $R<C$，R 为可靠传输的速率。

(b) 对所有的 $R>C$，R 为不可靠传输的速率。

R 为可靠传输的速率，如果存在一个码元序列 $f_n:\mathscr{H}_n\to\mathscr{H}_N$ 和解码规则 $\hat{f}_N:\mathscr{H}_N\to\mathscr{H}_n$，$n\sim NR$ 且误差概率

$$e(f_n,\hat{f}_N)\to 0,\quad\text{当 } N\to\infty \text{ 时}$$

换言之，

$$C=\lim_{N\to\infty}\frac{1}{N}\log M_N$$

其中 M_N 为错误解码的概率趋于 0 时码字 $x\in\mathscr{M}_N$ 的最大数量。

SCT 说明，对无记忆信道，

$$C=\max_{p_X} I(X;Y)$$

其中 $I(X;Y)$ 为随机输入符号 X 和对应的输出符号 Y 的互信息，且最大值取自 X 的所有可能的分布 p_X。

现在，对无记忆对称信道(MSC)，上述的最大化过程只用于输出符号

$$C=(\max_{p_X} h(Y))+\sum_{y\in\mathscr{M}_r} p_y\log p_y$$

求和 $-\sum_y p_y\log p_y$ 为信道矩阵 $(P(y|x))$ 每一行的熵。对双重对称信道，C 的表达式可以进一步化简为

[173]

$$C=\log M-h(p_1,\cdots,p_r)$$

其中 $h(Y)$ 在 p_X 等概率分布时达到，即 $p_X(x)=1/q$(且 $p_Y(y)=1/r$)。在 MBSC 中，我们得到

$$C=1-\eta(p)$$

从而完成了(i)部分的解答。

接下来，尽管(ii)部分对应的信道不是无记忆的，但通用的定义和一些相对应的结论仍可以使用。令 $n(j,t)$ 表示以 j，$j=0,1$ 结尾的长度为 t 的字符串数目，于是有

$$n(0,t)=n(1,t-1),$$
$$n(1,t)=n(0,t-1)+n(1,t-1)$$

因此

$$n(1,t)=n(1,t-1)+n(1,t-2)$$

将其写为递归的形式

$$\begin{bmatrix} n(1,t) \\ n(1,t-1) \end{bmatrix}=A\begin{bmatrix} n(1,t-1) \\ n(1,t-2) \end{bmatrix}$$

且递归矩阵为

$$A=\begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix}$$

通解为

$$n(1,t)=c_1\lambda_1^t+c_2\lambda_2^t$$

其中 λ_1 和 λ_2 为矩阵 A 的特征值，即特征等式的根

$$\det(A-\lambda I)=(1-\lambda)(-\lambda)-1=\lambda^2-\lambda-1=0$$

于是，$\lambda_\pm=(1\pm\sqrt{5})/2$ 且

$$\frac{1}{t}\log n(1,t)\to\log\left(\frac{\sqrt{5}-1}{2}\right)$$

从而，该信道的容量为

$$C=\lim_{t\to\infty}\frac{1}{t}\log\ \#\text{允许输入的字符串长度 } t$$

[174]

$$=\lim_{t\to\infty}\frac{1}{t}\log[n(1,t)+n(0,t)]=\log\left(\frac{\sqrt{5}+1}{2}\right)$$ □

备注 2.2.9 我们可以修改最后的问题，考虑 MBSC，其信道矩阵为 $\Pi=\begin{bmatrix} 1-p & p \\ p & 1-p \end{bmatrix}$，输入

约束为不能出现连续的 0。这种信道可以看作是两个连续信道的组合(参考举例 1.4.29a)，从而导出信道容量

$$C=\min\left(\log\left(\frac{\sqrt{5}+1}{2}\right),1-\eta(p)\right)$$

接下来，我们说明在 MBSC 中，Shannon 的 SCT 的强对称部分(参考定理 1.4.14)，我们再次使用 Hamming 空间的几何性证明它。术语“强”说明对任意传输速率 $R>C$，最大错误概率实际上任意接近于 1。为了简单起见，我们证明 MBSC 中的情形。

定理 2.2.10 (SCT/NCT，强对称部分) MBSC 的信道矩阵为 $\begin{bmatrix}1-p & p\\ p & 1-p\end{bmatrix}$，其中 $0<p<1/2$，信道容量为 C，且 $R>C$。令 $n=\lfloor NR\rfloor$。对所有的码元 $f_n:\mathscr{H}_n\to\mathscr{H}_N$ 和解码规则 $\hat{f}_N:\mathscr{H}_N\to\mathscr{H}_n$，最大错误概率为

$$\varepsilon^{\max}(f_n,\hat{f}_N):=\max[P(\text{在解码规则 }\hat{f}_N\mid f_n(\boldsymbol{u})\text{ 下发生错误}):\boldsymbol{u}\in\mathscr{H}_n] \tag{2.2.26a}$$

且满足

$$\limsup_{N\to\infty}\varepsilon^{\max}(f_n,\hat{f}_N)=1 \tag{2.2.26b}$$

证明 正如 1.4 节所示，我们可以假设码元 f_n 为一一映射且服从 $\hat{f}_N(f_n(\boldsymbol{u}))=\boldsymbol{u}$，对所有的 $\boldsymbol{u}\in\mathscr{H}_n$(否则，错误解码的概率将非常大)。和式(2.2.26b)的假设相反，

$$\varepsilon^{\max}(f_n,\hat{f}_N)\leqslant c,\quad \text{对某些 } c<1,\text{对所有足够大的 } N \tag{2.2.27}$$

我们的目标在于从式(2.2.27)推出 $R\leqslant C$。正如之前所示，令 $n=\lfloor NR\rfloor$，$f_n(\boldsymbol{u}(i))$ 为字符串 $\boldsymbol{u}(i)\in\mathscr{H}_n$，$i=1,\cdots,2^n$ 对应的码字。令 $\mathscr{D}_i\subset\mathscr{H}_N$ 为二元字符串的集合，且当且仅当 $\boldsymbol{a}\in\mathscr{D}_i$，$\hat{f}_N$ 返回 $f_n(\boldsymbol{u}(i))$：$\hat{f}_N(\boldsymbol{a})=f_n(\boldsymbol{u}(i))$。于是 $f(\boldsymbol{u}(i))\in\mathscr{D}_i$。集合 $\mathscr{D}_i$ 为成对分块的。如果 $\bigcup_i\mathscr{D}_i\neq\mathscr{H}_N$，则在 $\mathscr{H}_N\setminus\bigcup_i\mathscr{D}_i$ 部分，信道产生了错误。令 $s_i=|\mathscr{D}_i|$ 表示 $\mathscr{D}_i$ 的大小。

我们的第一步是“完善”解码规则，使其更“接近”ML 准则。换言之，我们将把每一个 $\mathscr{D}_i$ 用新的集合 $\mathscr{L}_i\in\mathscr{H}_N$ 代替且 $\#\mathscr{L}_i=s_i$，$\mathscr{L}_i$ 的边界更为“紧致”(即更贴近 Hamming 球 $B(f(\boldsymbol{u}(i)),b_i)$)。我们寻找两两非交的集合 $\mathscr{C}_i$($\mathscr{C}_i$ 的势为 $\#\mathscr{C}_i=s_i$)。对于一些稍后将详细给定的半径为 $b_i\geqslant 0$ 的球，$\mathscr{C}_i$ 满足 175

$$\mathscr{B}_N(f(\boldsymbol{u}(i)),b_i)\subseteq\mathscr{C}_i\subset\mathscr{B}_N(f(\boldsymbol{u}(i)),b_i+1),1\leqslant i\leqslant 2^n \tag{2.2.28}$$

我们认为通过一系列“不相交的交换”，$\mathscr{C}_i$ 可以从 $\mathscr{D}_i$ 中得到，并在其中添加字符串 $\boldsymbol{b}$，其参考如下的 Hamming 距离

$$\delta(\boldsymbol{b},f_n(\boldsymbol{u}(i)))\leqslant\delta(\boldsymbol{a},f_n(\boldsymbol{u}(i))) \tag{2.2.29}$$

我们去掉一个字符串 $\boldsymbol{a}$ 并添加了一个字符串 $\boldsymbol{b}$。

定义 $\hat{g}_N$ 为一个新的译码规则。因为翻转概率 $p<1/2$，所以式(2.2.29)中的关系意味着

$$\begin{aligned}&P(\hat{f}_N\text{ 返回 }f_n(\boldsymbol{u}(i))\mid f_n(\boldsymbol{u}(i)))=P(\mathscr{D}_i\mid f_n(\boldsymbol{u}(i)))\\&\leqslant P(\mathscr{C}_i\mid f_n(\boldsymbol{u}(i)))=P(\hat{g}_N\text{ 返回 }f_n(\boldsymbol{u}(i))\mid f_n(\boldsymbol{u}(i)))\end{aligned}$$

这等价于

$$P(\text{使用 }\hat{g}_N\text{ 时候的错误}\mid f_n(\boldsymbol{u}(i)))\leqslant P(\text{使用 }\hat{f}_N\text{ 时候的错误}\mid f_n(\boldsymbol{u}(i))) \tag{2.2.30}$$

然后，可以清楚看出

$$\varepsilon^{\max}(f_n,\hat{g}_N)\leqslant\varepsilon^{\max}(f_n,\hat{f}_N)\leqslant c \tag{2.2.31}$$

接下来，对于任意足够大的 N，假定存在 $p'<p$ 使得

$$b_i+1 \leqslant \lceil Np' \rceil, \quad 对于一些 1 \leqslant i \leqslant 2^n \tag{2.2.32}$$

然后，注意到$\mathscr{C}_i$代表$\mathscr{H}_N \setminus \mathscr{C}_i$的补集，根据式(2.2.28)和式(2.2.31)可得

$$\begin{aligned}
&\boldsymbol{P}(至少\ Np'\ 个数字失真 \mid f(\boldsymbol{u}(i))) \\
&\leqslant \boldsymbol{P}(至少\ b_i+1\ 个数字失真 \mid f(\boldsymbol{u}(i))) \\
&\leqslant \boldsymbol{P}(\mathscr{C}_i \mid f(\boldsymbol{u}(i))) \leqslant \varepsilon^{(N)}(f_n, \hat{g}_N) \leqslant \varepsilon
\end{aligned}$$

上式将导致矛盾，因为根据大数定理，随着$N \to \infty$，概率

$$\boldsymbol{P}(至少\ Np'\ 个数字失真 \mid 发送\ \boldsymbol{x}) \to 1$$

在所选择的输入码字上$\boldsymbol{x} \in \mathscr{H}_N$是均匀分布的(事实上，这个概率并不依赖于其体的$\boldsymbol{x} \in \mathscr{H}_N$)。

所以，对于足够大的N，我们不可能有$p' \in (0, p)$使得式(2.2.32)成立。这就是说，与之相反的事实是对的：对于任何给定的$p' \in (0, p)$，我们可以找到一个任意大的N使得

176

$$b_i > Np', \quad 对所有\ i = 1, \cdots, 2^n \tag{2.2.33}$$

(由于我们声明式(2.2.33)对于所有$p' \in (0, p)$成立，所以我们在式(2.2.33)中的左侧式子内代入b_i或者b_i-1都不影响结论。)

接下来，我们再次使用关于Hamming球体积的精确表达

$$\begin{aligned}
s_i = \# \mathscr{D}_i = \# \mathscr{C}_i &\geqslant v_N(b_i) = \sum_{0 \leqslant j \leqslant b_i} \binom{N}{j} \geqslant \binom{N}{b_i} \\
&\geqslant \binom{N}{\lfloor Np' \rfloor}, 条件是\ b_i \geqslant Np'
\end{aligned} \tag{2.2.34}$$

举例2.2.7给出了一个实用的界(参见式(2.2.25))

$$v_N(R) \geqslant \frac{1}{N+1} 2^{Nh_2(\frac{R}{N})} \tag{2.2.35}$$

我们现在继续完成定理2.2.10的证明。基于式(2.2.35)，对于所有$p' \in (0, p)$我们可以找到任意大的N使得对于所有的$1 \leqslant i \leqslant 2^n$都有$s_i \geqslant 2^{N(\eta(p')-\varepsilon_N)}$并且$\lim_{N \to \infty} \varepsilon_N = 0$。因为原始集合$\mathscr{D}_1, \cdots, \mathscr{D}_{2^n}$是不相交的，可知

177

$$s_1 + \cdots + s_{2^n} \leqslant 2^N, 这意味着\ 2^{N(\eta(p')-\varepsilon_N)} \times 2^{\lfloor NR \rfloor} \leqslant 2^N$$

或者

$$\eta(p') - \varepsilon_N + \frac{\lfloor NR \rfloor}{N} \leqslant 1, 这意味着\ R \leqslant 1 - \eta(p') + \varepsilon_N + \frac{1}{N}$$

当$N \to \infty$时上式右侧趋近于$1-\eta(p)$。那么，对给定任意$p' \in (0, p)$，都有$R \leqslant 1-\eta(p')$。注意到该结论对所有的$p' < p$都成立，所以$R \leqslant 1-\eta(p)-C$。这就完成了对定理2.2.10的证明。 □

我们已经看到，通过分析Hamming空间$\mathscr{H}_N$内给定集合$\mathscr{X}$与一系列球$\mathscr{B}_N(y, s)$之间交集的特性，能够揭示许多关于$\mathscr{X}$自身的特性。在本节剩下的部分，我们将采用上述方法导出q进制编码上一些更好的界：Elias界和Johnson界。这些界是编码领域最知名的普适界，其精确性难分伯仲。

Elias界的证明与Plotkin界的证明相似：参见定理2.1.15和举例2.1.18。我们在一个q进制$[N, M, d]$码$\mathscr{X}$中统计码字(这个q进制$[N, M, d]$码$\mathscr{X}$在关于码字$\boldsymbol{y} \in \mathscr{H}_N$半径为s的球$\mathscr{B}_{N,q}(\boldsymbol{y}, s)$中)。更准确地说，我们计算$(\boldsymbol{x}, \mathscr{B}_{N,q}(\boldsymbol{y}, s))$对，其中$\boldsymbol{x} \in \mathscr{X} \cap \mathscr{B}_{N,q}(\boldsymbol{y}, s)$。如果球$\mathscr{B}_{N,q}(\boldsymbol{y}, s)$包含$K_s$码字，那么

$$\sum_{\boldsymbol{x} \in \mathscr{H}_N} K_s = M v_{N,q}(s) \tag{2.2.36}$$

因为每一个码字 x 都落在球 $\mathscr{B}_{N,q}(y,s)$ 的 $v_{N,q}(s)$ 内。

引理 2.2.11 如果 $\mathscr{C}$ 是一个 q 进制 $[N,M]$ 码，那么对于所有的 $s=1,\cdots,N$，都存在一个关于 N-码字 $y\in\mathscr{H}_{N,q}$ 的球 $\mathscr{B}_{N,q}(y,s)$，其中包含 $K_y=\#(\mathscr{C}\cap\mathscr{B}_{N,q}(y,s))$ 个码字且 K_y 遵循

$$K_y\geqslant Mv_{N,q}(s)/q^N \tag{2.2.37}$$

证明 将式(2.2.36)两边同时除以 q^N。$\frac{1}{q^N}\sum_y K_y$ 给出了球 $\mathscr{B}_{N,q}(y,s)$ 中的码字个数的平均值。但是注意到一定有一个球至少包含和该平均值一样多的码字。 □

一个满足式(2.2.37)特性的球 $\mathscr{B}_{N,q}(y,s)$ 被称为临界的(对于码 $\mathscr{C}$)。

定理 2.2.12 (Elias 界)令 $\theta=(q-1)/q$，对于所有满足 $s<\theta N$ 和 $s^2-2\theta Ns+\theta Nd>0$ 的整数 $s\geqslant 1$，一个 N 长 q 进制编码的最大容量 $M_q^*(N,d)$ 和码距离 d 满足

$$M_q^*(N,d)\leqslant\frac{\theta Nd}{s^2-2\theta Ns+\theta Nd}\cdot\frac{q^N}{v_{N,q}(s)} \tag{2.2.38}$$

证明 固定一个临界球 $\mathscr{B}_{N,q}(y,s)$ 并考虑通过从码字 $\mathscr{C}_y$：$\mathscr{C}'(x-y;\ x\in\mathscr{C})$ 中减去码 y 得到码 $\mathscr{C}'$。那么 $\mathscr{C}'$ 同样是一个 $[N,M,d]$ 码。所以，我们可以假定 $y=\mathbf{0}$ 和 $\mathscr{B}_{N,q}(\mathbf{0},s)$ 是一个临界球。

继续取 $\mathscr{C}_1=\mathscr{C}\cap\mathscr{B}_{N,q}(\mathbf{0},s)=\{x\in\mathscr{C};\ w(x)\leqslant s\}$，并注意到码 $\mathscr{C}_1$ 的参数是 $[N,K,e]$，其中 $e\geqslant d$ 并且 $K(=K_y)\geqslant Mv_{N,q}(s)/q^N$。和对 Plotkin 界的证明相似，考虑 $\mathscr{C}_1$ 中的码字间距总和为

$$S_1=\sum_{x\in\mathscr{C}_1}\sum_{x'\in\mathscr{C}_1}\delta(x,x')$$

我们同样有 $S_1\geqslant K(K-1)e$。另一方面，如果 k_{ij} 是字母 $j\in J_q=\{0,\cdots,q-1\}$ 在所有码字 $x\in\mathscr{C}_1$ 第 i 位置上的总数，那么

$$S_1\sum_{1\leqslant i\leqslant N}\sum_{0\leqslant j\leqslant q-1}k_{ij}(K-k_{ij})$$

注意到总和 $\sum_{0\leqslant j\leqslant q-1}k_{ij}=K$。同时，由于 $w(x)\leqslant s$，在每一个码字 $x\in\mathscr{C}_1$ 中 0 的总数都满足 $\geqslant N-s$。0 在所有码字中的总数为 $\sum_{1\leqslant i\leqslant N}k_{i0}\geqslant K(N-s)$。记

$$S=NK^2-\sum_{1\leqslant i\leqslant N}(k_{i0}^2+\sum_{1\leqslant j\leqslant q-1}k_{ij}^2)$$

178

使用 Cauchy-Schwarz 不等式来估计

$$\sum_{1\leqslant j\leqslant q-1}k_{ij}^2\geqslant\frac{1}{q-1}(\sum_{1\leqslant j\leqslant q-1}k_{ij})^2=\frac{1}{q-1}(K-k_{i0})^2$$

然后可得

$$\begin{aligned}
S&\leqslant NK^2-\sum_{1\leqslant i\leqslant N}\left(k_{i0}^2+\frac{1}{q-1}(K-k_{i0})^2\right)\\
&=NK^2-\frac{1}{q-1}\sum_{1\leqslant i\leqslant N}[(q-1)k_{i0}^2+K^2-2Kk_{i0}+k_{i0}^2]\\
&=NK^2-\frac{1}{q-1}\sum_{1\leqslant i\leqslant N}(qk_{i0}^2+K^2-2Kk_{i0})\\
&=NK^2-\frac{N}{q-1}K^2-\frac{q}{q-1}\sum_{1\leqslant i\leqslant N}k_{i0}^2-\frac{2}{q-1}K\sum_{1\leqslant i\leqslant N}k_{i0}\\
&=\frac{q-2}{q-1}NK^2-\frac{q}{q-1}\sum_{1\leqslant i\leqslant N}k_{i0}^2+\frac{2}{q-1}KL
\end{aligned}$$

其中 $L=\sum_{1\le i\le N} k_i$。再次使用 Cauchy-Schwarz 不等式：

$$\sum_{1\le i\le N} k_i^2 \ge \frac{1}{N}\left(\sum_{1\le i\le N} k_i\right)^2 = \frac{1}{N}L^2$$

可得

$$\begin{aligned} S &\le \frac{q-2}{q-1}NK^2 - \frac{q}{q-1}\frac{1}{N}L^2 + \frac{2}{q-1}KL \\ &= \frac{1}{q-1}\left((q-2)NK^2 - \frac{q}{N}L^2 + 2KL\right) \end{aligned}$$

方括号中的二次函数的最大值出现在 $L=NK/q$，并且回想上文中提及的 $L\ge K(N-s)$。所以，选择 $K(N-s)\ge NK/q$，即 $s\le N(q-1)/q$，我们可以估计

$$\begin{aligned} S &\le \frac{1}{q-1}\left((q-2)NK^2 - \frac{q}{N}K^2(N-s)^2 + 2K^2(N-s)\right) \\ &= \frac{1}{q-1}K^2 s\left(2(q-1) - \frac{qs}{N}\right) \end{aligned}$$

这就得到了不等式 $K(K-1)e \le \frac{1}{q-1}K^2 s\left(2(q-1)-\frac{qs}{N}\right)$，进而对 K 求解可得

179

$$K \le \frac{\theta Ne}{s^2 - 2\theta Ns + \theta Ne}$$

如果 $s\le N\theta$，$s^2-2\theta Ns+\theta Ne>0$ 上式成立。最后，回忆 $\mathscr{X}^{(1)}$ 取自一个 $[N, M, d]$ 码 $\mathscr{X}$，其中 $K\ge Mv(s)/q^N$ 并且 $e\ge d$。最后，我们得到了

$$\frac{Mv_{H,q}(s)}{q^N} \le \frac{\theta Nd}{s^2 - 2\theta Ns + \theta Nd}$$

这就导出了 Elias 界(2.2.38)。 □

用于 Elias 界证明的思路(更早的用在 Plotkin 界中)对推导 $W_2^*(N, d, \ell)$ 的界有很大帮助。这里 $W_2^*(N, d, \ell)$ 表示一个 N 长二元(非线性)码 $\mathscr{X}\in\mathscr{H}_{N,2}$ 的最大容量，其中码距离 $d(\mathscr{X})\ge d$，码字重量 $w(x)=\ell$，$x\in\mathscr{X}$。首先，给出三个明显的性质

(i) $W_2^*(N, 2k, k)=\left\lfloor\frac{N}{k}\right\rfloor$

(ii) $W_2^*(N, 2k, \ell)=W_2^*(N, 2k, N-\ell)$

(iii) $W_2^*(N, 2k-1, \ell)=W_2^*(N, 2k, \ell)$，$\ell/2\le k\le \ell$

(建议读者自行证明作为练习。)

举例 2.2.13 证明对于所有的正整数 $N\ge 1$，$k\le\frac{N}{2}$ 和 $\ell<\frac{N}{2}-\sqrt{\frac{N^2}{4}-kN}$，有

$$W_2^*(N,2k,\ell) \le \left\lfloor \frac{kN}{\ell^2-\ell N+kN} \right\rfloor \tag{2.2.39}$$

解答 取一个 $[N, M, 2k]$ 码 $\mathscr{X}$ 使得 $w(x)=\ell$，$x\in\mathscr{X}$。和之前一样，令 k_i 为 1 在所有码字中第 i 个位置的总数。考虑点积和 $D=\sum_{x,x'\in\mathscr{X}} \mathbf{1}(x\ne x')(x\cdot x')$，我们得到

$$\begin{aligned} (x\cdot x') = w(x\wedge x') &= \frac{1}{2}(w(x)+w(x')-\delta(x,x')) \\ &\le \frac{1}{2}(2\ell-2k) = \ell-k \end{aligned}$$

所以

$$D \leqslant (\ell - k)M(M-1)$$

另一方面，i 位置对 D 贡献等于 $k_{i1}(k_{i1}-1)$，即

$$D = \sum_{1\leqslant i\leqslant N} k_{i1}(k_{i1}-1) = \sum_{1\leqslant i\leqslant N}(k_{i1}^2 - k_{i1}) = \sum_{1\leqslant i\leqslant N} k_{i1}^2 - \ell M$$ [180]

同样，最后的和项在 $k_{i1}=\ell M/N$ 处取得最小，即

$$\frac{\ell^2 M^2}{N} - \ell M \leqslant D \leqslant (\ell - k)M(M-1)$$

这就得出了式(2.2.39)。 □

接下来给出另一个有用的界。

举例 2.2.14 证明对于所有正整数 $N\geqslant 1$，$k\leqslant \frac{N}{2}$ 和 $2k<\ell<4k$，下面式子成立

$$W_2^*(N,2k,\ell) \leqslant \left\lfloor \frac{N}{\ell} W_2^*(N-1,2k,\ell-1) \right\rfloor \tag{2.2.40}$$

解答 同样取码 $\mathscr{X}$ 使得所有 $x\in\mathscr{X}$ 都有 $w(x)=\ell$。考虑 $\mathscr{X}$ 对 $x_1=1$ 的截断码(参考举例 2.1.8(v))：$[N, M, 2k]$ 码给出了一个码长为 $(N-1)$、距离 $\geqslant 2k$ 且重量为常数 $(\ell-1)$ 的码。所以，这部分码字的数量 $\leqslant W_2^*(N-1, 2k, \ell-1)$。因此，在码字 $\mathscr{X}$ 中第一数位上字符 1 的总数不超过 $W_2^*(N-1, 2k, \ell-1)$。重复这个论断，我们得到了字符 1 在所有位置上的总数 $\leqslant NW_2^*(N-1, 2k, \ell-1)$。但是这个数字事实上等于 ℓM，也就是说 $\ell M \leqslant NW_2^*(N-1, 2k, \ell-1)$。然后就可以得到式(2.2.40)中的界。 □

推论 2.2.15 对于所有正整数 $N\geqslant 1$，$k\leqslant \frac{N}{2}$ 和 $2k<\ell<4k-2$，有

$$\begin{aligned} W_2^*(N,2k-1,\ell) &= W_2^*(N,2k,\ell) \\ &\leqslant \left\lfloor \frac{N}{\ell}\left\lfloor \frac{N-1}{\ell-1}\left\lfloor \cdots \left\lfloor \frac{N-\ell+k}{k}\right\rfloor \cdots \right\rfloor\right\rfloor\right\rfloor \end{aligned} \tag{2.2.41}$$

2.2 节剩下的部分重点放在 Johnson 界。这个界旨在改进二元 Hamming 界(参考式(2.1.8b)，$q=2$)

$$M_2^*(N,2E+1) \leqslant 2^N/v_N(E) \quad 或 \quad v_N(E)M_2^*(N,2E+1) \leqslant 2^N \tag{2.2.42}$$

也就是说，Johnson 界表明

$$M_2^*(N,2E+1) \leqslant 2^N/v_N^*(E) \quad 或 \quad v_N^*(E)M_2^*(N,2E+1) \leqslant 2^N \tag{2.2.43}$$

其中

$$\begin{aligned} v_N^*(E) = v_N(E) + \frac{1}{\lceil N/(E+1)\rceil}\Bigg[&\binom{N}{E+1} \\ &- W_2^*(N,2E+1,2E+1)\binom{2E+1}{E}\Bigg] \end{aligned} \tag{2.2.44}$$ [181]

回忆一下，$v_N(E) = \sum_{0\leqslant s\leqslant E}\binom{N}{s}$ 代表了半径为 E 的二元 Hamming 球的体积。我们将从下面引理开始导出式(2.2.43)中的界。

引理 2.2.16 如果 x，y 是二元码，$\delta(x, y)=2\ell+1$，那么，存在 $\binom{2\ell+1}{\ell}$ 个二元码字 z，使得 $\delta(x, z)=\ell+1$ 并且 $\delta(y, z)=\ell$。

证明 留作练习。 □

考虑一个集合 $\mathscr{F}(=\mathscr{F}_{N,E+1})$，其中包含所有与 $\mathscr{X}$ 中码字距离为 $E+1$ 的二元 N 长

字符：

$$\mathscr{J}=\{z\in\mathscr{H}:\delta(z,x)=E+1,\quad 对某些\ x\in\mathscr{X}$$
$$和\ \delta(z,y)\leqslant E-1,\quad 对于所有的\ y\in\mathscr{X}\} \tag{2.2.45}$$

那么我们可以得到

$$Mv_N(E)+\#\mathscr{J}\leqslant 2^N \tag{2.2.46}$$

因为 $z\in\mathscr{J}$ 中没有任何字符落在码字 $y\in\mathscr{X}$ 周围半径为 E 的球中。当我们解决下一个举例的时候，式(2.2.43)中的界可被推导出。

举例 2.2.17 证明 $\#\mathscr{J}$ 是不小于式(2.2.44)右边的第二项

$$\frac{M}{[N/E+1]}\left[\binom{N}{E+1}-W_2^*(N,2E+1,2E+1)\binom{2E+1}{E}\right] \tag{2.2.47}$$

解答 我们希望寻找 $\#\mathscr{J}$ 的下界，考虑如下定义的与 N-码"匹配"的字符集合 $\mathscr{W}(=\mathscr{W}_{N,E+1})$

$$\begin{aligned}\mathscr{W}&=\{(x,z):x\in\mathscr{X},z\in\mathscr{J}_{E+1},\delta(x,z)=E+1\}\\&=\{(x,z):x\in\mathscr{X},z\in\mathscr{H}_N:\delta(x,z)=E+1,\\&\qquad\delta(y,z)\geqslant E+1,\quad 对所有的\ y\in\mathscr{X}\}\end{aligned} \tag{2.2.48}$$

给定一个 $x\in\mathscr{X}$，$\mathscr{X}$ 切片 $\mathscr{W}^x$ 定义为

$$\begin{aligned}\mathscr{W}^x&=\{z\in\mathscr{H}_N:(x,z)\in\mathscr{W}\}\\&=\{z:\delta(x,z)=E+1,\delta(y,z)\geqslant E+1,\quad 对于所有的\ y\in\mathscr{X}\}\end{aligned} \tag{2.2.49}$$

观察到如果 $\delta(x,y)=E+1$，那么对于所有的 $y\in\mathscr{X}\setminus\{x\}$ 都有 $\delta(y,z)\geqslant E$，否则应有 $\delta(x,y)<2E+1$。因此

182

$$\mathscr{W}^x=\{z:\delta(x,z)=E+1,\delta(y,z)\neq E,\quad 对于所有的\ y\in\mathscr{X}\} \tag{2.2.50}$$

为了计算 $\#\mathscr{W}^x$，我们需关注距离码字 x 为 $E+1$ 的字符集合，从该集合的字符总数 $\binom{N}{E+1}$ 中减去该集合中距离另一码字 $y\in\mathscr{X}$ 为 E 的字符数可得 $\#\mathscr{W}^x$。但是如果 $\delta(x,z)=E+1$ 而 $\delta(y,z)=E$，那么有 $\delta(x,y)=2E+1$。同样，我们注意到不可能有两个不同码字到同一个 N 字符 z 的距离都为 E。因此，基于之前的解释，可得

$$\#\mathscr{W}^x=\binom{N}{E+1}-\binom{2E+1}{E}\times\#\{y\in\mathscr{X}:\delta(x,y)=2E+1\}$$

进一步，如果我们以码字 $y\in\mathscr{X}$ 为参照从原码本中删除每一个满足 $\delta(x,y)=E+1$ 的 x，码，这将生成一个 N 长码，其码字为 z、码字重量为 $w(z)=2E+1$。所以，至多有 $W_2^*(N,2E+1,2E+1)$ 个码字 $y\in\mathscr{X}$，满足 $\delta(x,y)=2E+1$，进而可得，

$$\#\mathscr{W}^x\geqslant\binom{N}{E+1}-W^*(N,2E+1,2E+1)\binom{2E+1}{E} \tag{2.2.51}$$

并且

$$\#\mathscr{W}\geqslant M\times 式(2.2.51)的右端 \tag{2.2.52}$$

现在固定 $v\in\mathscr{J}$ 并且考虑 v-切片

$$\mathscr{W}^v=\{y\in\mathscr{X}:(y,v)\in\mathscr{W}\}=\{y\in\mathscr{X}:\delta(y,v)=E+1\} \tag{2.2.53}$$

如果 $y,z\in\mathscr{W}^v$，那么 $\delta(y,u)=\delta(z,u)=E+1$。所以

$$w(y-u)=w(z-u)=E+1$$

并且

$$\begin{aligned}2E+1 \leqslant \delta(y,z) &= \delta(y-v,z-v)\\ &= w(y-v)+w(z-v)-2w((y-v)\wedge(z-v))\\ &= 2E+2-2w((y-v)\wedge(z-v))\end{aligned}$$

这意味着

$$w((y-v)\wedge(z-v))=0 \quad 和 \quad \delta(y,z)=2E+2$$

所以 $y-v$ 和 $z-v$ 没有公共的数位 1。因此，至多存在 $\left\lceil\frac{N}{E+1}\right\rceil$ 个形如 $y-v$ 的字符，其中 $y\in\mathscr{W}^*$，即 $\mathscr{W}^*$ 中至多有 $\left\lceil\frac{N}{E+1}\right\rceil$ 个字符。因此

$$\#\mathscr{W} \leqslant \left\lceil\frac{N}{E+1}\right\rceil \#\mathscr{S} \tag{2.2.54}$$

综合式(2.2.51)式(2.2.52)和式(2.2.54)得到不等式(2.2.47)。 □ 183

推论 2.2.18 借鉴推论 2.2.15，下面的界成立

$$M^*(N,2E+1) \leqslant 2^N\left[v_N(E)-\frac{1}{\lfloor N/(E+1)\rfloor}\binom{N}{E}\left(\frac{N-E}{E+1}-\left\lfloor\frac{N-E}{E+1}\right\rfloor\right)\right]^{-1} \tag{2.2.55}$$

例子 2.2.19 定义 $N=13$，$E=2$，即 $d=5$，不等式(2.2.41)意味着 $W^*(13,5,5)\leqslant\left\lfloor\frac{13}{5}\left\lfloor\frac{12}{4}\left\lfloor\frac{11}{3}\right\rfloor\right\rfloor\right\rfloor=23$，而基于式(2.2.43)中的 Johnson 界可得

$$M^*(13,5) \leqslant \left\lfloor\frac{2^{13}}{1+13+78+(286-10\times 23)/4}\right\rfloor=77$$

相比起 Hamming 界给出的 $M^*(13,5)\leqslant 89$，Johnson 界要精确很多。事实上，众所周知 $M^*(13,5)=64$，参见 3.4 节。

2.3 线性码：基本构造

这部分，我们更深入地探讨线性码。首先我们考虑符号取 0 和 1 的二元制码，相应地，$\mathscr{H}_N$ 将代表长度为 N 的二元 Hamming 空间；$\mathscr{H}_N$ 中的码字 $x^{(N)}=x_1\cdots x_N$ 将被称为(行)向量。所有在二元数位上的操作都是二进制运算(即 mod 2)。在不致困惑情况下，我们忽略下标 N 和上标(N)。让我们重复一下线性码的定义(参考定义 2.1.5)。

定义 2.3.1 如果一个字符 $x+x'$ 的二元码 $\mathscr{X}\subseteq\mathscr{H}_N$ 联合一对向量 $x=x_1\cdots x_N$ 和 $x'=x'_1\cdots x'_N$，码字 $\mathscr{X}$ 包含 $x+x'$ 这个和，则该二元码被称为线性的。换句话说，一个线性码是域 $F_2=\{0,1\}$ 上 $\mathscr{H}_N$ 的线性子空间。因此，一个线性码总含有一个零行向量 $\mathbf{0}=0\cdots 0$。线性码 $\mathscr{X}$ 的一个基是 $\mathscr{X}$ 中具有最多独立线性独立码字的集合；一个线性码可通过它的基生成，也就是说，每一个向量 $x\in\mathscr{X}$ 可以被(唯一地)表示为(一些)基向量的线性加和。对于一个给定的线性码 $\mathscr{X}$，其所有可能的基包含相同数量的向量；基向量的数目被称作 $\mathscr{X}$ 的秩或维数。长度为 N 而秩为 k 的线性码被称作$[N,k]$码，当考虑距离为 d 时可以进一步称作$[N,k,d]$码。

现实中实用的码都是线性的，因为它们简单易用。比方说，为了识别一个线性码，我们只需要确定它的基就可以了，这就会有很多好处，我们可以在后面的内容中看到这些好处。

引理 2.3.2 任意一个秩为 k 的线性码包含 2^k 个向量，即大小为 $M=2^k$。 184

证明 码的基包含 k 个线性独立向量，这个码由它的基生成，因此它由这些基的线性组合

(加和)构成。更为精确地说一共有 2^k 个可能的加和($1, \cdots, k$)子集的数目显示了被加数的量)，并且它们能给出不同的向量。 □

因此，一个秩为 k 的二元线性码 $\mathscr{X}$ 可以编码长度为 k 的所有字符串。$[N, k]$码的信息速率是 k/N。所以，表明 $k \leqslant N$ 个线性独立的码字 $x \in \mathscr{H}_N$ 可以确定一个(唯一的)秩为 k 的线性码 $\mathscr{X} \subset \mathscr{H}_N$。换言之，一个秩为 k 的线性二元码可以由一个 0，1 元素构成的 $k \times N$ 矩阵表示

$$G = \begin{pmatrix} g_{11} & \cdots & g_{1N} \\ g_{21} & & g_{2N} \\ \vdots & & \vdots \\ g_{k1} & \cdots & g_{kN} \end{pmatrix}$$

该矩阵中的列向量是线性独立的。

也就是说，我们采用该矩阵中的行向量 $g(i) = g_{i1} \cdots g_{iN}$，$1 \leqslant i \leqslant k$ 作为一个线性码的基向量。

定义 2.3.3 矩阵 G 被称为线性码的生成矩阵，而且生成矩阵不是唯一的。

等价地说，一个线性$[N, k]$码 $\mathscr{X}$ 可以被描述成某个 $(N-k) \times N$ 二元矩阵 $\boldsymbol{H}$ 的核，该矩阵的元素为 0 和 1：$\mathscr{X} = \ker \boldsymbol{H}$，其中

$$\boldsymbol{H} = \begin{pmatrix} h_{11} & h_{12} & \cdots & h_{1N} \\ h_{21} & h_{22} & \cdots & h_{2N} \\ \vdots & \vdots & & \vdots \\ h_{(N-k)1} & h_{(N-k)2} & \cdots & h_{(N-k)N} \end{pmatrix}$$

并且

$$\ker \boldsymbol{H} = \{x = x_1 \cdots x_N : x\boldsymbol{H}^{\mathrm{T}} = 0^{(N-k)}\} \tag{2.3.1}$$

很明显矩阵 $\boldsymbol{H}$ 的行向量 $h(j)$，$1 \leqslant j \leqslant N-k$，与 $\mathscr{X}$ 正交，从点积的角度解释为

$$(x \cdot h(j)) = 0, \quad \text{对所有的 } x \in \mathscr{X} \text{ 和 } 1 \leqslant j \leqslant N-k$$

这里，对于 $x, y \in \mathscr{H}_N$

$$(x \cdot y) = (y \cdot x) = \sum_{i=1}^{N} x_i y_i, \quad \text{其中 } x = x_1 \cdots x_N, y = y_1 \cdots y_N \tag{2.3.2}$$

185 参考举例 2.1.8(ix)。

式(2.3.2)中的内积拥有 $\mathbb{R}^N$ 中欧式标量积的全部特性。但是，它不是正定的(所以它没有定义一个范数)，也就是，存在非零向量 $x \in \mathscr{H}_N$ 满足 $(x \cdot x) = 0$。幸运的是，我们并不需要正定性。

然而，这里的点积满足秩-零特性：如果 $\mathscr{L}$ 是 $\mathscr{H}_N$ 上秩为 k 的线性子空间，那么它的正交补 $\mathscr{L}^{\perp}$(即于任意 $x \in \mathscr{L}$ 满足 $(x \cdot z) = 0$ 的所有向量 $z \in \mathscr{H}_N$ 的集合)是一个秩为 $N-k$ 的线性子空间。所以，$\boldsymbol{H}$ 的 $(N-k)$ 行可视为 $\mathscr{X}^{\perp}$ 中的基，$\mathscr{X}^{\perp}$ 是 $\mathscr{X}$ 的正交补。

满足特性 $\mathscr{X} = \ker \mathscr{H}$ 或者 $(x \cdot h(j)) = 0$(参见式(2.3.1))的矩阵 $\boldsymbol{H}$(或者有时它的转置 $\boldsymbol{H}^{\mathrm{T}}$)被称作 $\mathscr{X}$ 的奇偶校验(或者简单叫作校验)矩阵。在很多情况下，通过校验矩阵来描述一个码比生成矩阵更方便。

校验矩阵也不是唯一的，因为在 $\mathscr{X}^{\perp}$ 中的基可以有多种选择。另外，对于一组长度为 N 的不同的码，我们更习惯辨别行数大于 $N-k$ 的校验矩阵(但是一些行将是线性相关的)；第 3 章有这样的例子。然而，我们将暂时把 $\boldsymbol{H}$ 视为一个行线性独立的 $(N-k) \times N$ 矩阵。

举例 2.3.4 $\mathscr{X}$是信息速率为$\rho=k/N$的线性二元$[N, k, d]$码。$\boldsymbol{G}$和$\boldsymbol{H}$分别为$\mathscr{X}$的生成矩阵和奇偶校验矩阵。在这个例子中我们参考举例 2.1.8 的结构，给出以下问题

(a) $\mathscr{X}$的奇偶校验扩展是一个通过相加得到的长度为$N+1$的二元码$\mathscr{X}^+$，即给每个码字$x\in\mathscr{X}$加上符号$x_{N+1}=\sum_{1\le i\le N}x_i$，使得$\sum_{1\le i\le N+1}x_i$为 0。证明$\mathscr{X}^+$是一个线性码，得到它的秩和最小距离。$\mathscr{X}$和$\mathscr{X}^+$的信息率、生成矩阵及奇偶校验矩阵是怎样联系的？

(b) $\mathscr{X}$的截断$\mathscr{X}^-$被定义为截断每个$x\in\mathscr{X}$码字最后符号后得到的$N-1$长线性编码。假定$\mathscr{X}$的码距离$d\ge 2$。证明$\mathscr{X}^-$是线性的并给出它的秩、生成矩阵和奇偶校验矩阵。证明$\mathscr{X}^-$的最小距离至少为$d-1$。

(c) $\mathscr{X}$的m重复是通过将码$x\in\mathscr{X}$重复m次得到的Nm长码$\mathscr{X}^m(m)$。证明$\mathscr{X}^m(m)$是一个线性码，给出它的秩和最小距离，讨论$\mathscr{X}^m(m)$的信息率和生成矩阵以及奇偶校验矩阵是怎样与ρ、$\boldsymbol{G}$和$\boldsymbol{H}$联系的？ 186

解答 (a) 生成矩阵和校验矩阵是

$$\boldsymbol{G}^+=\left[\begin{array}{c|c}\boldsymbol{G} & \begin{matrix}\sum_{1\le i\le N} g_i\\ \vdots\\ \vdots\\ \sum_{1\le i\le N} g_k\end{matrix}\end{array}\right],\quad \boldsymbol{H}^+=\left[\begin{array}{c|c}\boldsymbol{H} & \begin{matrix}1\\1\\\cdot\\\cdot\\1\end{matrix}\\ \hline 0\ \cdots\ 0 & 1\end{array}\right]$$

$\mathscr{X}^+$的秩等于$\mathscr{X}$的秩，$\mathscr{X}=k$。如果$\mathscr{X}$的最小距离是偶数，则不变；如果为奇，则增加 1。信息速率为$\rho^+=(N-1)\rho/N$。

(b) 生成矩阵

$$\boldsymbol{G}^-=\begin{bmatrix} g_{11} & \cdots & g_{1(N-1)}\\ & \vdots & \\ & \vdots & \\ & \vdots & \\ g_{k1} & \cdots & g_{k(N-1)}\end{bmatrix}$$

$\mathscr{X}$的奇偶校验矩阵$\boldsymbol{H}$，经过适当地列变换之后，可以写成：

$$\boldsymbol{H}=\left[\begin{array}{c|c}\boldsymbol{H}^- & \begin{matrix}\cdot\\\cdot\\\cdot\\\cdot\\\cdot\end{matrix}\\ \hline 0\ \cdots\ 0 & *\end{array}\right]$$

$\mathscr{X}^-$的奇偶校验矩阵是通过$\boldsymbol{H}^-$给定的。它的秩不变；距离可能最大减 1。信息率为$\rho^-=N\rho/(N-1)$。

(c) 生成矩阵和奇偶校验矩阵分别是

$$\boldsymbol{G}^m(m)=(\boldsymbol{G}\cdots\boldsymbol{G})(m\text{ 次})$$

利

$$H''(m)=\begin{bmatrix} H & 0 & 0 & \cdots & 0 \\ I & I & 0 & \cdots & 0 \\ I & 0 & I & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ I & 0 & 0 & \cdots & I \end{bmatrix}$$

[187]

其中，I 是一个 $N\times N$ 的单位矩阵，0 表示的是一个零矩阵（因此维度为 $(N-k)\times N$ 和 $N\times N$），第一列中的单位阵的数目等于 $m-1$。（这不是 $H''(m)$ 的唯一形式。）$H''(m)$ 是 $(Nm-k)\times Nm$ 维的，

这里线不变，$\mathscr{X}''(m)$ 的最小距离为 md，信息速率为 a/m。 □

举例 2.3.5 一个线性二进制码 $[N, k]$ 的对偶码 $\mathscr{X}$ 定义为集合 $\mathscr{X}^{\perp}$，包含 $y=y_1\cdots y_N$，满足以下点积关系

$$(y\cdot x)=\sum_{1\le i\le N} y_i\cdot x_i=0,\quad 对于每一个属于\mathscr{X}的\ x=x_1\cdots x_N$$

类比例子 2.1.5(ix)。证明当且仅当 H 是对偶码的生成矩阵时，$(N-k)\times N$ 的矩阵 H 是 $\mathscr{X}$ 码的奇偶校验矩阵。因此，验证 G 和 H 矩阵分别是一个线性码的生成矩阵和奇偶校验矩阵，当且仅当：

(i) 矩阵 G 的行是线性无关的。

(ii) 矩阵 H 的列是线性无关的。

(iii) 矩阵 G 的行数加上矩阵 H 的列数等于矩阵 G 的列数和矩阵 H 的行数。

(iv) $GH^{T}=0$。

解答 矩阵 H 的行 $h(j)$，$j=1,\cdots,N-k$ 符合 $(x\cdot h(j))=0$，$x\in\mathscr{X}$。而且，如果一个向量 y 满足 $(x\cdot y)=0$，$x\in\mathscr{X}$，那么 y 是 $y(j)$ 的线性组合。因此，H 是矩阵 $\mathscr{X}^{\perp}$ 的生成矩阵，另一方面，$\mathscr{X}^{\perp}$ 的任何生成矩阵都是一个 $\mathscr{X}$ 的奇偶校验矩阵。

因此，对于任何一对线性码的生成矩阵和奇偶校验矩阵 G，H，(i)(ii)(iv) 可通过定义证得，(iii) 通过秩的公式得到

$$N=\dim(行一致\ G)+\dim(行一致\ H)$$

遵从于 (iv) 和 G、H 的极大性。

另一方面，任何遵循 (i)—(iv) 的矩阵的 G，H 对具有最大化的性质（根据 (i)—(iii)）和正交性 (iv)。因此，它们是 $\mathscr{X}=$ 行 $-G$ 的秩的生成矩阵和奇偶校验矩阵。 □

举例 2.3.6 一个线性二元 $[N, k]$ 码的码字个数是多少？其中不同的基数量是多少？计算 $k=4$ 的基的数目，列出所有 $k=2$ 和 $k=3$ 的基。

[188]

试证明一个线性二元码中由所有偶数重量的码字组成的一个子集是可构成线性码，证明对于偶数 d，如果线性 $[N, k, d]$ 码存在的话，那么其码字重量可以是偶数。

解答 二元线性 $[N, k]$ 码中的码字数量为 2^k，不同的基的数目为 $\frac{1}{k!}\prod_{i=1}^{k-1}(2^k-2^i)$。事实上，如果第一个基底向量被选择了，则在下一步它所有的线性组合应该被排除。对于 $k=4$ 是 840，对于 $k=3$ 是 28。

最终，对于偶数 d，我们可以截短原始码字然后使用奇偶校验扩展获得一个线性 $[N, k, d]$ 码。 □

例子 2.3.7 二元 Hamming $[7, 4]$ 码可由一个 3×7 的奇偶校验矩阵确定。该校验矩阵的

列是长度为 3 的非零码字。将这些码字按照词典编纂顺序排列我们可以得到

$$H_{\text{ex}}^{\text{Ham}}=\begin{bmatrix}1&0&1&0&1&0&1\\0&1&1&0&0&1&1\\0&0&0&1&1&1&1\end{bmatrix}$$

相应的生成矩阵为

$$H_{\text{ex}}^{\text{Ham}}=\begin{bmatrix}0&0&1&1&0&0&1\\0&1&0&0&1&0&1\\0&0&1&0&1&1&0\\1&1&1&0&0&0&0\end{bmatrix}\tag{2.3.3}$$

在很多情况下将一个线性[N，k]码的校验矩阵以范式(或者标准形式)表述是很方便的

$$H_{\text{ex}}=(I_{N-k}H')\tag{2.3.4a}$$

Hamming[7，4]码的校验矩阵如下

$$H_{\text{ex}}^{\text{Ham}}=\begin{bmatrix}1&0&0&1&1&0&1\\0&1&0&1&0&1&1\\0&0&1&0&1&1&1\end{bmatrix}$$

其生成矩阵的范式如下

$$G_{\text{ex}}=(G'I_k)\tag{2.3.4b}$$

也就是

$$G_{\text{ex}}^{\text{Ham}}=\begin{bmatrix}1&1&0&1&0&0&0\\1&0&1&0&1&0&0\\1&1&1&0&0&1&0\\1&1&1&0&0&0&1\end{bmatrix}$$

189

形式上，G_{ex}和G_{ex}确定不同的码字，然而这些码字是等价的。

定义 2.3.8 如果两种码字只是在数位排列上有区别，那么称这些码字是等价的。对于线性码，等价是指它们的生成矩阵可以通过列变换或者包含加或者标量乘等运算行变换相互转化。显然，等价的码拥有相同的参数(长度，秩，距离)。

在接下来的部分中，除非特别说明，我们不区分彼此等价的线性码。

备注 2.3.9 将 G 以范式表示的好处是对信源序列$u^{(k)}\in\mathscr{M}_k$的编码可以表示为一个 N 维向量$u^{(k)}G^{\text{ex}}$；根据式(2.3.4b)，$u^{(k)}G^{\text{ex}}$的最后 k 个数位(被称为信息数位)形式上依旧是字符$u^{(k)}$，而前面的 $N-k$ 个被用于奇偶校验(被称为奇偶校验数位)。形象地来说，奇偶校验数位携带着译码器可以用来检测错误的冗余信息。

就像通过解剖生物来研究生命一样，
你发现它的时候也就是你失去它的时候。

Alexander Pope(1668—1744)，英国诗人

定义 2.3.10 一个二元码 $x=x_1\cdots x_N$ 的码字重量 $w(x)$是其中的非零数字的个数

$$w(x)=\#\{i;1\leqslant i\leqslant N,x_i\neq 0\}\tag{2.3.5}$$

定理 2.3.11

(i) 线性二元码的距离等于其非零码字的最小码字重量。

(ii) 线性二元码的距离等于校验矩阵中线性相关的列的数量。

证明 (i) 因为码$\mathscr{X}$是线性的，对于每一对码字 x，$y\in\mathscr{X}$ 都有 $x+y\in\mathscr{X}$。由于 Hamming 距离的平移不变性(见引理 2.1.1)，对于任意一对码字，$\delta(x, y)=\delta(0, x+y)=w(x+y)$，因此$\mathscr{X}$的最小距离等于码字 $\mathbf{0}$ 和其余的码字之间的最小距离，即$\mathscr{X}$中非零码字的最小重量。

(ii) 记线性码$\mathscr{X}$的奇偶校验矩阵为 $\boldsymbol{H}$，其最小距离为 d。那么，存在一个码字 $x\in\mathscr{X}$ 恰好有 d 个非零数位。因为 $x\boldsymbol{H}^{\mathrm{T}}=0$，我们可以得出矩阵 $\boldsymbol{H}$ 中有 d 列是线性相关的(它们对应着 x 中的非零数位)。另一方面，如果 $\boldsymbol{H}$ 中存在$(d-1)$列线性独立的列向量，那么它们的和是零。但是，这也意味着存在一个码字 y 的重量 $w(y)=d-1$，所以 $y\boldsymbol{H}^{\mathrm{T}}=0$。然而 y 不可能一定属于$\mathscr{X}$，因为 $\min\{w(x): x\in\mathscr{X}, x\neq\mathbf{0}\}=d$。 □

[190]

定理 2.3.12 Hamming[7, 4]码的最小距离为 3，即能够检测 2 个错并纠正 1 个错，而且它是一个能纠正单个错误的完美码字。

证明 对于任意一对列向量，奇偶校验矩阵 $\boldsymbol{H}^{\mathrm{H}}$ 包含它们的和，进而可以得到线性相关的三个列向量(比如这里的第 1，6 和 7 列)。其中，任意两列线性无关，因为它们是不同的($x+y=\mathbf{0}$ 意味着 $x=y$)。此外，$\mathscr{B}(1)$的体积等于 $1+7=2^3$ 并且码长是 2^4，$2^4\times 2^3=2^7$，因此该码是完美的。 □

Hamming[7, 4]码的结构可以轻松扩展到任意长度为 $N=2^l-1$ 的码设计；比方说，考虑一个$(2^l-1)\times l$ 的矩阵 $\boldsymbol{H}^{\mathrm{Ham}}$，该矩阵的列表示所有的可能存在的长度为 l 的非零二元向量

$$\boldsymbol{H}^{\mathrm{Ham}}=\begin{bmatrix}1 & 0 & \cdots & 0 & 1 & \cdots & 1\\ 0 & 1 & \cdots & 0 & 1 & \cdots & 1\\ 0 & 0 & \cdots & 0 & 0 & \cdots & 1\\ \vdots & \vdots & & \vdots & \vdots & & 1\\ 0 & 0 & \cdots & 1 & 0 & \cdots & 1\end{bmatrix} \tag{2.3.6}$$

矩阵 H^{H} 的行是线性无关的，因此 $\boldsymbol{H}^{\mathrm{Ham}}$ 可以被视为一个长度为 $N=2^l-1$，秩为 $N-l=2^l-1-l$ 的线性码的校验矩阵。$\boldsymbol{H}^{\mathrm{Ham}}$ 的任何两列是线性无关的，但是存在线性相关的三个列向量，例如，x，y 和 $x+y$。因此，校验矩阵为 $\boldsymbol{H}^{\mathrm{Ham}}$ 的码$\mathscr{X}^{\mathrm{Ham}}$ 的最小距离为 3，即它可以检测 2 个错，纠正 1 个错。

这种码也被叫作 Hamming$[2^l-1, 2^l-1-l]$码，它是完美的单错误纠错码：1-球的列 $v_{2^l-1}(1)$等于 $1+2^l-1=2^l$，又大小×列$=2^{2^l-l-1}\times 2^l=2^{2^l-1}=2^N$。随着 $l\to\infty$，信息速率 $\frac{2^l-l-1}{2^l-1}\to 1$。这证明了如下定理。

定理 2.3.13 上述的结构定义了一系列$[2^l-1, 2^l-1-l, 3]$线性二元码$\mathscr{X}^{\mathrm{Ham}}_{2^l-1}$，$l=1, 2, \cdots$，这些码是完美的单一错误纠错码。

[191]

例子 2.3.14 假设任意数位的错误概率 $p\ll 1$，并且与其他的数位错误无关。那么发送一个没有编码的具有$(4N)$-数位消息，出现错误的概率为

$$1-(1-p)^{4N}\simeq 4Np$$

但是，如果我们使用[7, 4]码，我们需要传输 $7N$ 个数字。在传输中发生一次错误至少需要两个数位出现错误，概率

$$\approx 1-\left[1-\binom{7}{2}p^2\right]^N\simeq 21Np^2\ll 4Np$$

我们可以看出在 Hamming 码中额外使用 3 个校验数位置是有效果的。

线性码的标准译码过程是基于陪集和伴随式的概念。回想 ML 准则对向量 $y=y_1\cdots y_N$ 进行译码的结果是最接近 y 的码字 $x\in\mathscr{X}$。

定义 2.3.15 $\mathscr{X}$是 N 长二元线性码，$w=w_1\cdots w_N$ 是$\mathscr{H}_N$ 中的一个码字。一个由 w 决定的$\mathscr{X}$的陪集是形如$w+x$ 的码字集合，这里 $x\in\mathscr{X}$。我们用 $w+\mathscr{X}$来表示该陪集。

在线性代数中一个简单(而有用的)练习如下：

例子 2.3.16 $\mathscr{X}$是一个线性码，w，v 是长度为 N 的码字。那么：

(1) 如果 w 在陪集$v+\mathscr{X}$中，那么 v 在陪集$w+\mathscr{X}$之中；换句话说，陪集中每一个码字都可以决定这个陪集。

(2) $w\in w+\mathscr{X}$。

(3) 当且仅当 $w+v\in\mathscr{X}$时，w 和 v 在同样的陪集中。

(4) 每个长度为 N 的字符属于并且只属于一个陪集。也就是说，陪集划分了整个 Hamming 空间$\mathscr{H}_N$。

(5) 所有的陪集包含相同数量的字符，并且等于$\#\mathscr{X}$。如果$\mathscr{X}$的秩是 k，那么存在 2^{N-k}个不同的陪集，每个包含 2^k 个码字。码$\mathscr{X}$是任意码字的陪集。

(6) $w+v$ 所决定的陪集与形为 $x+y$ 的元素组成的集合相一致，这里 $x\in w+\mathscr{X}$，$y=\mathscr{X}+v$。

现在得到线性码的解码规则：事先知道码$\mathscr{X}$，因此你可以计算出所有的陪集。当接收到一个码字 y 之后，可以找出它的陪集 $y+\mathscr{X}$并且找到一个码字 $w\in y+\mathscr{X}$使它的重量最小，这个码字被称为陪集 $y+\mathscr{X}$的陪集首。陪集首可能不是唯一的：这样的话你需要在陪集首列表里面做出选择或者拒绝译码并且请求重传。假设你已选择了一个陪集首 w，那么对于码字 y 的解调结果是

$$x_e=y+w \tag{2.3.7}$$

192

举例 2.3.17 试证明码字 x_e 总能最小化 y 和$\mathscr{X}$中码字的距离。

解答 因为 y 和 w 属于同一个陪集，于是有 $y+w\in\mathscr{X}$(见例 2.3.16(3))。$\mathscr{X}$中的所有码字可视为 $y+v$ 的和，其中 v 取自陪集 $y+\mathscr{X}$。因此，对于任意 $x\in\mathscr{X}$

$$\delta(y,x)=w(y+x)\geqslant\min_{v\in y+\mathscr{X}}w(v)=w(w)=d(y,x_e)$$ □

用奇偶校验矩阵来描述陪集 $y+\mathscr{X}$是很方便的。

定理 2.3.18 陪集 $w+\mathscr{X}$与形式为 $\boldsymbol{y}\boldsymbol{H}^{\mathrm{T}}$ 的向量集合一一对应：当且仅当 $\boldsymbol{y}\boldsymbol{H}^{\mathrm{T}}=\boldsymbol{y}'\boldsymbol{H}^{\mathrm{T}}$ 时两个向量 $\boldsymbol{y}$ 和 $\boldsymbol{y}'$在相同的陪集之内。换句话说，陪集是用奇偶校验矩阵的列空间(值域)来确定的。

证明 当且仅当 $y+y'\in\mathscr{X}$时，两个向量 y 和 y'在相同的陪集之内，即

$$(y+y')\boldsymbol{H}^{\mathrm{T}}=\boldsymbol{y}\boldsymbol{H}^{\mathrm{T}}+\boldsymbol{y}'\boldsymbol{H}^{\mathrm{T}}=0,\quad 即\ \boldsymbol{y}\boldsymbol{H}^{\mathrm{T}}=\boldsymbol{y}'\boldsymbol{H}^{\mathrm{T}}$$ □

在实际中，解码规则的实施如下。形为 $\boldsymbol{y}\boldsymbol{H}^{\mathrm{T}}$ 的向量称为伴随式：对于一个线性(N,k)码存在 2^{N-k}个伴随式。它们被列在伴随式列表中，对于每一个伴随式计算一个相应的陪集首。当接收到一个字符 y 之后，计算伴随式 $\boldsymbol{y}\boldsymbol{H}^{\mathrm{T}}$，然后在伴随式列表中找到相应的陪集首 w，然后根据式(2.3.7)：用$x_e=y+w$ 来译码 y。

这个过程被称为伴随式译码：虽然它相对来说简单，但是需要写出一个较长的陪集首列表。而且最好将整个解码算法设计得与具体码字(即它的生成矩阵)彼此独立。这个目标可以在 Hamming 码的情况下实现：

定理 2.3.19 对于 Hamming 码而言，与每伴随式对应的陪集首 w 是唯一的，并且该陪集首最多包含一个的非零数位。更精确地说，如果伴随式 $y(\boldsymbol{H}^{(\mathrm{Ham})})^{\mathrm{T}}=y$ 给出校验矩阵 $\boldsymbol{H}^{(\mathrm{Ham})}$ 的列 i，那么相应陪集的头只有一个非零数字 i。

证明 陪集首最小化了接收符号和码之间的距离，Hamming码是完美的单符号纠错码。因此，每个字符要么是个码字要么与一个唯一的码字只有1的距离。因此，陪集首是唯一的并且包含最多一个非零数位。如果伴随式 $yH^{\mathrm{T}}=s$ 占据了奇偶校验矩阵中第 i 列的位置，那么对于只有一位非零数字 i 的码 $e_i=0\cdots10\cdots0$，我们有

[193]

$$(y+e_i)H^{\mathrm{T}}=s+s=0$$

也就是说，$(y+e_i)\in\mathscr{X}$，$e_i\in y-\mathscr{X}$。显然，e_i 是陪集首。 □

二元Hamming码的对偶 $\mathscr{X}^{\mathrm{Ham}\perp}$ 形成一个类特别的码，叫作单纯形码。如果 $\mathscr{X}^{\mathrm{Ham}}$ 是 $[2^l-1,\ 2^l-1-l]$ 码，它的对偶 $(\mathscr{X}^{\mathrm{Ham}})^{\perp}$ 是 $[2^l-1,\ l]$ 码，原始奇偶校验矩阵 H^{Ham} 是 $\mathscr{X}^{\mathrm{Ham}\perp}$ 的生成矩阵。

举例 2.3.20 证明一个二元单纯形码 $\mathscr{X}^{\mathrm{Ham}\perp}$ 中的每个非零的码字重量为 2^{l-1}，并且任何两个码字之间的距离等于 2^{l-1}，因此符合单纯形码的定义。

解答 如果 $\mathscr{X}=\mathscr{X}^{\mathrm{Ham}}$ 是二元Hamming $[2^l-1,\ 2^l-1-l]$ 码，那么它的对偶 $\mathscr{X}^{\perp}$ 是 $[2^l-1,\ l]$ 码，它的 $l\times(2^l-1)$ 生成矩阵为 H^{Ham}。H^{Ham} 中任意一行的重量等于 2^{l-1}（因此 $d(\mathscr{X}^{\perp})=2^{l-1}$）。事实上，$H^{\mathrm{Ham}}$ 中第 j 行的重量等于在 j 位置具有长度1的非零向量的数量。这样可以计算出重量 2^{l-1}，因为 $\mathscr{H}_l$ 的所有 2^l 个向量中在任意指定位置为1的数量为 2^l 的一半。

现在考虑 $\mathscr{X}^{\mathrm{Ham}\perp}$ 中的一个一般码字，它可以通过 H^{Ham} 的行 j_1，…，j_s 的和来表示，这里 $s\leqslant l$，$1\leqslant j_1<\cdots<j_s\leqslant l$。这个码字的重量为 2^{l-1}；这给出了非零码字 $v=v_1\cdots v_{2^l-1}\in\mathscr{H}_{1,2}$ 的数量，其中加和项 $v_{j_1}+\cdots+v_{j_s}=1$。而且，2^{l-1} 给出 $\mathscr{H}_{1,2}$ 中一半的向量的重量。事实上，我们需要让 $v_{j_1}+\cdots v_{j_s}=1$，这样对于包含的 s 个数位就产生了 2^{s-1} 种可能性。接下来，我们对剩下的可能性为 2^{l-s} 的 $l-s$ 个数位不加以限制，那么 $2^{l-s}\times2^{s-1}=2^{l-1}$。所以，对于所有非零码 $x\in\mathscr{X}^{\perp}$ 都有重量 $w(x)=2^{l-1}$。最后，对于任意 x，$x'\in\mathscr{X}$，$x\neq x'$，距离 $\delta(x,\ x')=\delta(0,\ x+x')=w(x+x')$ 总是等于 2^{l-1}。所以，码字 $x\in\mathscr{X}^{\perp}$ 形成一个具有 2^l 个"顶点"的"单纯形"几何图样。 □

接下来我们简单总结关于定义在符号集 $\mathbb{F}_q=\{0,\ 1,\ \cdots,\ q-1\}$ 上的线性码基本性质，这里 $q=p^r$。现在我们针对Hamming空间 $\mathscr{H}_{q,N}$ 使用符号 $\mathbb{F}_q^{\times N}$。

定义 2.3.21 如果 q 元码 $\mathscr{X}\subseteq\mathbb{F}^{\times N}$ 中任意一对向量 $x=x_1\cdots x_N$，$x'=x'_1\cdots x'_N$ 的线性组合 $\gamma\cdot x+\gamma'\cdot x'$（任意 $\gamma\cdot\gamma'\in\mathbb{F}_q$ 并且加法定义为数位相加 $\gamma\cdot x_j+\gamma'\cdot x'_j$）依然属于 $\mathscr{X}$，那么 $\mathscr{X}$ 是线性的。也就是说 $\mathscr{X}$ 是 $\mathbb{F}^{\times N}$ 的一个线性子空间。因此，就像在二进制的情况下，一个线性码总是包含向量 $\mathbf{0}=0\cdots0$。线性码的基再次被定义为 $\mathscr{X}$ 中包含最大数量线性无关码字的集合。线性码通过它的基生成，也就是说每个码向量（唯一地）表示为基码向量的线性组合。基向量的数量叫作码的维度或者秩；因为一个给定线性码的所有基都包含相同数量的向量，这个对象（码的维度）可谓被正确地定义了。就像二进制的情况一样，长度为 N 秩为 k 的线性码被称为 $[N,\ k]$ 码，或者当它的距离等于 d 时是可进一步定义 $[N,\ k,\ d]$ 码。

[194]

就像二进制的情况一样，线性码 $\mathscr{X}$ 的最小距离等于最小非零码字的重量

$$d(\mathscr{X})=\min[w(x):x\in\mathscr{X},x\neq\mathbf{0}]$$

其中

$$w(x)=\#\{j:1\leqslant j\leqslant N,x_j\text{ 在 }\mathbb{F}_q\text{ 上不为 }0\}\tag{2.3.8}$$

$$x=x_1\cdots x_N\in\mathbb{F}_q^{\times N}$$

线性码 $\mathscr{X}$ 可通过一个生成矩阵 G 或者奇偶校验矩阵 H 定义。线性 $[N,\ k]$ 码的生成矩阵是一个 $k\times N$ 的矩阵 G，其元素取自 $\mathbb{F}_q$，这里 $\mathbb{F}_q$ 的行 $g(i)=g_{i1}\cdots g_{iN}$，$1\leqslant i\leqslant k$ 构成 $\mathscr{X}$ 的基。

奇偶校验矩阵是一个$(N-k)\times N$矩阵$\boldsymbol{H}$，它的元素取自$\mathbb{F}_q$，而$\mathbb{F}_q$的行$\boldsymbol{h}(j)=h_{j1}\cdots h_{jN}$，$1\leqslant j\leqslant N-k$线性独立并且和$\mathscr{X}$是点-正交的：对于所有$j=1,\cdots N-k$和$\mathscr{X}$中的码字$\boldsymbol{x}=x_1\cdots x_N$，均有

$$(\boldsymbol{x}\cdot\boldsymbol{h}(j))=\sum_{1\leqslant s\leqslant N}x_s h_{js}=0$$

换句话说，$\mathscr{X}$的所有q^k个码字由矩阵$\boldsymbol{G}$行的线性组合得到的。子空间$\mathscr{X}$可被视为矩阵$\boldsymbol{G}$对Hamming空间$\mathbb{F}_q^{\times N}$的一个操作：具体而言，$\mathscr{X}=\mathbb{F}_q^{\times k}\boldsymbol{G}$。这说明了码$\mathscr{X}$是如何编码$q^k$个长度为$k$"消息"的(并且解释了信息速率$\rho(\mathscr{X})=k/N$)。另一方面，$\mathscr{X}$是$\boldsymbol{H}^{\top}$的核(零空间)：$\mathscr{X}\boldsymbol{H}^{\top}=0$。这里一个有用的联系是检验：对于$\mathscr{X}$的对偶码$\mathscr{X}^{\perp}$，上述情况是相反的，即$\boldsymbol{H}$是生成矩阵，$\boldsymbol{G}$是奇偶校验矩阵。与举例2.3.5比较。

当然，一个给定码的生成矩阵和奇偶校验矩阵不是唯一的，例如，我们可以交换$\boldsymbol{G}$中的行$\boldsymbol{g}(j)$或者进行行间运算，用行的线性变换取代其中的一行(但是要保证原始行的系数不能为零)。交换$\boldsymbol{G}$中的行得到一个不同但是等价的码，它的基础几何参数和$\mathscr{X}$的是一样的。

引理2.3.22 *对于任意$[N,k]$码，存在生成矩阵为$\boldsymbol{G}$且满足范式$\boldsymbol{G}=(\boldsymbol{G}'\ \boldsymbol{I}_k)$的等价码，这里$\boldsymbol{I}_k$是一个标识$k\times k$的单位矩阵，$\boldsymbol{G}'$是一个$k\times(N-k)$矩阵。类似地，奇偶校验矩阵$\boldsymbol{H}$也可有一个标准形式$(\boldsymbol{I}_{N-k}\ \boldsymbol{H}')$。* 196

现在我们讨论一般秩为k的线性码$\mathscr{X}$的译码过程。正如我们之前提到的，它可能被用来编码长度为k的源信息(字符序列)$\boldsymbol{u}=u_1\cdots u_k$。当生成矩阵和奇偶校验矩阵是典型(或者标准)形式时，源信息编码$\boldsymbol{u}\in\mathbb{F}_q^k\to\mathscr{X}$将变得特别简单。

定理2.3.23 *对于任何线性码$\mathscr{X}$，存在与之等价的码$\mathscr{X}'$，$\mathscr{X}'$的生成矩阵$\boldsymbol{G}'^{\text{std}}$和校验矩阵$\boldsymbol{H}'^{\text{std}}$符合标准形式(2.3.4a)和(2.3.4b)，并且$\boldsymbol{G}'=-(\boldsymbol{H}')^{\top}$。*

证明 假设码$\mathscr{X}$是非平凡的(即不会变成字符$\boldsymbol{0}$)。写出$\mathscr{X}$的一个基然，后得到相应的生成矩阵$\boldsymbol{G}$。通过行变换(交换行i和j或者用行i加j来替换行i)我们可以改变基但不改变码。矩阵$\boldsymbol{G}$包含一个非零列，l_1：通过行变换来让g_{1l_1}成为该列中的唯一非零元素。通过变换数字(列)，将列l_1放到位置$N-k$。放弃行1和列$N-k$(即老的列l_1)并且对剩下的进行相同处理，结果得到列l_2中的唯一的非零元素g_{2l_2}。将列l_2放置在位置$N-k+1$处，继续直到出现一个上三角$k\times k$子矩阵。下一步的计算仅仅关于这个矩阵。如果这个矩阵是一个单位矩阵，那么运算结束。如果没有，那么选取第一列不多于一个非零元素。将相对应的行从底加上去来"消灭"多余的非零元素。重复这一步骤直到出现一个单位矩阵。现在生成矩阵为标准形式，并且新的码与初始码等价。 □

为了完善证明，观察出现在式(2.3.4a)和式(2.3.4b)中的矩阵$\boldsymbol{G}'^{\text{std}}$和$\boldsymbol{H}'^{\text{std}}$有$\boldsymbol{G}'=-(\boldsymbol{H}')^{\top}$，两个矩阵分别有$k$个独立的行和$N-k$个独立的列，除此之外，$k\times(N-k)$维矩阵$\boldsymbol{G}'^{\text{std}}(\boldsymbol{H}'^{\text{std}})^{\top}$消失了。事实上

$$(\boldsymbol{G}'^{\text{std}}(\boldsymbol{H}'^{\text{std}})^{\top})_{ij}=(\boldsymbol{G}'\text{ 的行 }i\cdot(\boldsymbol{H}')^{\top}\text{ 的列 }j)=g'_{ij}-g'_{ij}=0$$

因此，$\boldsymbol{H}'^{\text{std}}$是$\boldsymbol{G}'^{\text{std}}$的校验矩阵。

回到源编码，选择满足范式的生成矩阵$\boldsymbol{G}^{\text{std}}$，然后给出一个字符序列$\boldsymbol{u}=u_1\cdots u_k$，令$\boldsymbol{x}=\sum_{i=1}^{k}u_i\boldsymbol{g}^{\text{std}}(i)$，这里$\boldsymbol{g}^{\text{std}}(i)$表示$\boldsymbol{G}^{\text{std}}$的行$i$。$\boldsymbol{x}$中的最后$k$数位给出了字符序列$\boldsymbol{u}$，它们被称为信息数位。前面的$N-k$位确保$\boldsymbol{x}\in\mathscr{X}$，它们被称为奇偶校验数位。

标准形式非常方便，因为在上面的表达式$\mathscr{X}=\mathbb{F}^{\times k}\boldsymbol{G}$中每个码字前面的$(N-k)$长序列 196

被用于编码(用于检测和纠错)，后面的 k 长序列表示选自域 $\mathbb{F}_q^{\times k}$ 的消息。在二进制的情况下，奇偶校验矩阵 $\boldsymbol{H}$ 满足定理 2.3.11。特别地，一个码的最小距离等于其奇偶校验矩阵 $\boldsymbol{H}$ 中线性相关的列的数量。

定义 2.3.24 给出一个[N，k]线性 q 元码 $\mathscr{X}$，奇偶校验矩阵为 $\boldsymbol{H}$，N 维向量的伴随式 $\boldsymbol{y}\in\mathbb{F}_q^{\times N}$ 是 k 维向量 $\boldsymbol{y}\boldsymbol{H}^{\mathrm{T}}\in\mathbb{F}_q^{\times k}$，伴随式子空间是 $\mathbb{F}_q^{\times k}\boldsymbol{H}^{\mathrm{T}}$ 的值域(列空间)。根据向量 $\boldsymbol{w}\in\mathbb{F}_q^{\times N}$ 决定的 $\mathscr{X}$ 的一个陪集表示为 $\boldsymbol{w}+\mathscr{X}$ 并且形式如同 $\boldsymbol{w}+\boldsymbol{x}$ 的码字，这里 $\boldsymbol{x}\in\mathscr{X}$。所有的陪集拥有相同数量的元素，等同于 q^k，并且将整个 Hamming 空间 $\mathbb{F}_q^{\times N}$ 分成 q^{N-k} 个不相交的子集：码 $\mathscr{X}$ 是其中一个，陪集和相应的伴随式 $\boldsymbol{y}\boldsymbol{H}^{\mathrm{T}}$ 一一对应。伴随式编码过程在二进制情况下表示为：一个接收向量 $\boldsymbol{y}$ 表示为 $\boldsymbol{x}$，$=\boldsymbol{y}+\boldsymbol{w}$，这里 $\boldsymbol{w}$ 是陪集 $\boldsymbol{y}+\mathscr{X}$ 的头(即 $\boldsymbol{y}+\mathscr{X}$ 中权重最小的码字)。

所有在二进制情况下存在的缺点在 q 元的情况下依然存在(事实上更加严重)：陪集表非常庞大，陪集首可能不是独一无二的。然而，对于 q 元 Hamming 码，伴随式译码性能良好，我们将在 2.4 节中看到。

在线性码的情况下，一些界还可以改进(或者可以产生新的界限)。

举例 2.3.25 $\mathscr{X}$ 是一个二进制线性[N，k，d]码。

(a) 选取一个有 d 个非零数位的码字 $\boldsymbol{x}\in\mathscr{X}$。证明对 $\mathscr{X}$ 中码字 $\boldsymbol{x}$ 在非零数位截断可产生一个长度为 $N-d$，秩为 $k-1$ 且距离 $d'\geqslant\lceil d/2\rceil$ 的新码 $\mathscr{X}'_{N-d}$。

(b) 推导 Griesmer 界进而提高 Singleton 界(2.1.12)

$$N\geqslant d+\sum_{1\leqslant i\leqslant k-1}\left\lceil\frac{d}{2^i}\right\rceil \tag{2.3.9}$$

解答 (a) 不失一般性，假设 $\boldsymbol{x}$ 中的非零数字是 $x_1=\cdots=x_d=1$。在数字 1，…，d 上进行截断会产生码 $\mathscr{X}'_{N-d}$，秩减 1。事实上，假设 $k-1$ 个向量的一个线性组合在位置 $d+1$，…，N 上趋于 0。那么在位置 1，…，d 所有的值等于 0 或者 1 因为 d 是最小距离。但是除非向量是线性相关的，否则第一种情况不可能发生。第二种情况同样会引出矛盾，比如通过增加序列 $\boldsymbol{x}$ 可获得码 $\mathscr{X}$ 中 k 个线性相关的向量。接下来，假设 $\mathscr{X}$ 的距离 $d'<\lceil d/2\rceil$ 并且有

[197] $\boldsymbol{y}'\in\mathscr{X}$ 和 $w(\boldsymbol{y}')=\sum_{j=d+1}^{N}y'_j=d'$。

令 $\boldsymbol{y}\in\mathscr{X}$ 为 $\boldsymbol{y}'$ 在截断下的翻转镜像。根据式(2.1.6b)，我们可以写出二进制码外积的性质

$$w(\boldsymbol{y})=w(\boldsymbol{x}\wedge\boldsymbol{y})+w(\boldsymbol{y}+(\boldsymbol{x}\wedge\boldsymbol{y}))\geqslant d$$

因此，我们必须有 $w(\boldsymbol{x}\wedge\boldsymbol{y})>d-\lceil d/2\rceil$，见图 2-5。

那么

$$w(\boldsymbol{x})=w(\boldsymbol{x}\wedge\boldsymbol{y})+w(\boldsymbol{x}+(\boldsymbol{x}\wedge\boldsymbol{y}))=d$$

表示 $w(\boldsymbol{x}-(\boldsymbol{x}\wedge\boldsymbol{y}))<\lceil d/2\rceil$，但是这是一个矛盾，因为

$$w(\boldsymbol{x}-\boldsymbol{y})=w(\boldsymbol{x}+(\boldsymbol{x}\wedge\boldsymbol{y}))+w(\boldsymbol{y}+(\boldsymbol{x}\wedge\boldsymbol{y}))<d$$

我们可以总结出 $d'\geqslant\lceil d/2\rceil$。

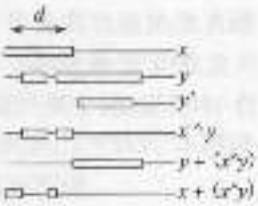


图 2-6

(b) 迭代(a)中的论证得到

$$N\geqslant d+d_1+\cdots+d_{k-1}$$

其中 $d_i\geqslant\left\lceil\frac{d_{i-1}}{2}\right\rceil$。当 $\left\lceil\frac{d/2}{2}\right\rceil\geqslant\left\lceil\frac{d}{4}\right\rceil$ 时，我们有

$$N\geqslant d+\sum_{1\leqslant i\leqslant k-1}\left\lceil\frac{d}{2^i}\right\rceil$$

总结这一部分，我们对线性码的 GV 界进行了详尽描述。

定理 2.3.26 (Gilbert 界)如果 $q=p^s$ 是一个素数幂，那么对于所有的整数 N 和 d 有 $2\leqslant d\leqslant N/2$，当满足下面条件时，

$$q^d \geqslant q^N/v_{N,q}(d-1) \tag{2.3.10}$$

存在一个最小距离 $\geqslant d$ 的 q 元线性$[N, k, d]$码。 [198]

证明 令 $\mathscr{Z}$ 为一个线性码，这个码的最大秩和距离都至少为 d。如果不等式(2.3.10)不成立，那么所有围绕码字半径为 $d-1$ 的 Hamming 球所组成的集合不能覆盖整个 Hamming 空间。所以必须存在一个点 $\boldsymbol{y}$ 属于任意码字周围的 Hamming 球中。那么，对于任意码字 $\boldsymbol{x}$ 和任意标量 $b\in\mathbb{F}_q$，向量 $\boldsymbol{y}$ 和 $\boldsymbol{y}+b\cdot\boldsymbol{x}$ 在 $\mathscr{Z}$ 有相同的陪集。同样地 $\boldsymbol{y}+b\cdot\boldsymbol{x}$ 不能在半径为 $d-1$ 的任意 Hamming 球中。这同样适用于 $\boldsymbol{x}+b\cdot\boldsymbol{y}$，因为如果 $\boldsymbol{x}+b\cdot\boldsymbol{y}$ 在某个 Hamming 球中的话，$\boldsymbol{y}$ 将会在其他一个码字周围的 Hamming 球中。这里我们用到了 $\mathbb{F}_q$ 是一个域这个事实。$\mathscr{Z}$ 和 $\boldsymbol{y}$ 张成的向量子空间是一个比 $\mathscr{Z}$ 还要大的码，其最小距离至少为 d。显然，这是一个矛盾，因此得证。 □

举例，令 $q=2$，$N=10$，那么 $2^5<v_{10,2}(2)=56<2^6$。当 $d=3$ 时，Gilbert 界保证了二元$[10, 5]$码在 $d\geqslant 3$ 情况下存在。

2.4 Hamming 码，Golay 码，Reed-Muller 码

在这部分我们会系统地学习在一个一般有限字母表 $\mathbb{F}_q$ 中具有 q 个元素的码，这里假设 $\mathbb{F}_q$ 是一个域。我们再一次申明 q 需要是一个形式为 p^s 的数，其中 p 是一个素数，s 是一个自然数；元素符加(+)和乘(·)也需要被特殊定义。(如前文所说，如果 $q=p$ 并且都是素数，我们可以认为 $\mathbb{F}_q=\{0, 1, \cdots, q-1\}$，在 $\mathbb{F}_q$ 中的加和乘也是标准的，即模 q 运算。)见 3.1 节。相应地，长度为 N 的 Hamming 空间 $\mathscr{H}_{N,q}$ 像前面一样被确定为 Cartesian 积 $\mathbb{F}_q^{\times N}$，其中的数字来自于 $\mathbb{F}_q$，并且 Hamming 空间还通过标量继承分量形式加和乘。

定义 2.4.1 给定正整数 q，$\ell\geqslant 2$，设 $N=\dfrac{q^\ell-1}{q-1}$，$k=N-\ell$，根据符号集 $\mathbb{F}_q$ 构建 q 元$[N, k, 3]$Hamming 码 $\mathscr{Z}_{N,q}^{\mathrm{Ham}}$ 如下。(a)选取任意非零 q 元 ℓ 字符 $\boldsymbol{h}^{(1)}\in\mathscr{H}_{\ell,q}$。(b)选取任意非零 q 元 ℓ 字符 $\boldsymbol{h}^{(2)}\in\mathscr{H}_{\ell,q}$，但不能是 $\boldsymbol{h}^{(1)}$ 的标量乘积。(c)继续：如果 $\boldsymbol{h}^{(1)}, \cdots, \boldsymbol{h}^{(s)}$ 包含当前已经选取的 q 元 ℓ 字符，选取任意非零向量 $\boldsymbol{h}^{(s+1)}\in\mathscr{H}_{\ell,q}$，它不能是 $\boldsymbol{h}^{(1)}, \cdots, \boldsymbol{h}^{(s)}$ 的标量积，其中 $1\leqslant s\leqslant N-1$。(d)这个过程终在选择 N 个向量 $\boldsymbol{h}^{(1)}, \cdots, \boldsymbol{h}^{(N)}$ 之后终止；最终形成了一个列为 $\boldsymbol{h}^{(1)\mathrm{T}}, \cdots, \boldsymbol{h}^{(N)\mathrm{T}}$ 的 $\ell\times N$ 矩阵 $\boldsymbol{H}^{\mathrm{Ham}}$。码 $\mathscr{Z}_{N,q}^{\mathrm{Ham}}\subset\mathbb{F}_q^{\times N}$ 通过奇偶校验矩阵 $\boldsymbol{H}^{\mathrm{Ham}}$ 定义。(事实上，我们在这里处理一族等效码，对字 $\boldsymbol{h}^{(j)}$，$1\leqslant j\leqslant N$ 的选择取模)。

为了简洁，我们将会使用 $\mathscr{Z}^{\mathrm{H}}$ 和 $\boldsymbol{H}^{\mathrm{H}}$(或者如果可能的话使用更简单的 $\boldsymbol{H}$)而不是 $\mathscr{Z}_{N,q}^{\mathrm{Ham}}$ 和 $\boldsymbol{H}^{\mathrm{Ham}}$。在二进制的情况下($q=2$)，矩阵 $\boldsymbol{H}^{\mathrm{H}}$ 由长度为 ℓ 的非零二元列向量组成。对于一般情况我们会排除列标量乘积所获得的冗余列。对于一般的 q，需要去除互为倍数的列。最后，我们能选择所有非零 ℓ 码字作为列，这些码字在它们的最顶层非零元素中含有 1。这样的列是线性独立的，它们的总和等于 $\dfrac{q^\ell-1}{q-1}$。接着，在二进制的情形下，可以按照字典顺序把码字和 $\mathbb{F}_q$ 中的数字编写在一起，通过构造，$\boldsymbol{H}^{\mathrm{H}}$ 中任何两列都是线性独立的，但是存在三个线性相关的列。所以，$d(\mathscr{Z}^{\mathrm{H}})=3$，$\mathscr{Z}^{\mathrm{H}}$ 能检查两个错误，纠正一个。此外，$\mathscr{Z}^{\mathrm{H}}$ 是纠正单个错误的完美码字，因为 [199]

$$M(1+(q-1)N)=q^k\left(1+(q-1)\frac{q^\ell-1}{q-1}\right)=q^{k+\ell}=q^N$$

正如在二进制情况下，一般的 Hamming 码只承认一个有效率（和简洁的）的译码过程。假设如上构造了一个奇偶检验矩阵 $\boldsymbol{H}(=\boldsymbol{H}^{n})$，根据接收的码字 $\boldsymbol{y}\in\mathbb{F}_q^N$，我们计算出校验子 $\boldsymbol{y}\boldsymbol{H}^{\mathrm{T}}\in\mathbb{F}_q^{\ell}$。如果 $\boldsymbol{y}\boldsymbol{H}^{\mathrm{T}}=0$，那么 $\boldsymbol{y}$ 是一个码字，另外，对于一些 $j=1,\cdots,N$，$a\in\mathbb{F}_q\setminus\{0\}$，列向量 $\boldsymbol{H}^{\mathrm{T}}\boldsymbol{y}$ 是 $\boldsymbol{H}$：$\boldsymbol{H}^{\mathrm{T}}\boldsymbol{y}=a\cdot\boldsymbol{h}^{(j)}$ 的列 $\boldsymbol{h}^{(j)}$ 的标量倍。换句话说，$\boldsymbol{y}\boldsymbol{H}^{\mathrm{T}}=a\cdot\boldsymbol{e}(j)\boldsymbol{H}^{\mathrm{T}}$，其中码字 $\boldsymbol{e}(j)=0\cdots1\cdots0\in\mathscr{H}_{N,q}$（第 j 位是 1，其他是 0）。那么可以用 $\boldsymbol{x}_{\ell}=\boldsymbol{y}-a\cdot\boldsymbol{e}(j)$ 来解码 $\boldsymbol{y}$，即把 $\boldsymbol{y}$ 中的数位 y_j 简单修改为 y_j-a。

我们总结如下。

定理 2.4.2 q 进制的 Hamming 码对于 $N=\dfrac{q^{\ell}-1}{q-1}$，$\ell=1,2,\cdots$，扭成一个族

$$\left[\frac{q^{\ell}-1}{q-1},\frac{q^{\ell}-1}{q-1}-\ell,3\right]$$

完美码字 $\mathscr{H}_{N,q}^{H}$ 根据译码规则可以纠正一个错误，这个译码规则是在一个接收码字 $\boldsymbol{y}=y_1\cdots y_N\in\mathbb{F}_q^{(N)}$ 中把 $\boldsymbol{y}$ 中的数位 y_j 修改为 y_j-a，其中 $1\leqslant j\leqslant N$，$a\in\mathbb{F}_q\setminus\{0\}$ 由条件 $\boldsymbol{H}^{\mathrm{T}}\boldsymbol{y}=a\cdot\boldsymbol{h}^{(j)}$，即奇偶校验矩阵 $\boldsymbol{H}$ 的列 j 的 a 待确定。

Hamming 码在 20 世纪 40 年代的后期由 R. Hamming 和 M. Golay 发现。Hamming 工作在 Los Alamos，那时他从一个电子工程师转型为一个计算机科学家（他自称是当地核物理学家们的“聪明的看门人”）。这个发现重新定义了 20 多年来的编码理论：人们努力地把 Hamming 码的性质扩展到更宽的编码领域（取得了很多进展）。本书中讨论的大多数关于编码的话题都以某种方式和 Hamming 码相联系。Richard Hamming 不仅是出色的科学家，他也有巨大的人格魅力，他的文章既有趣又引人深思。

200

直到 20 世纪 50 年代的后期，Hamming 码仍然是唯一一个存在于维度 $N\to\infty$ 的拥有“正则”性质的编码族，人们随后发现这个编码有很深的代数背景，基于这些观察得到的代数理论的发展仍然是现代编码理论的主要主题。

另一个重要的例子是四个 Golay 编码（两个二进制，两个三进制）。Marcel Golay (1902—1989) 是瑞典的一个电子工程师，他长期工作生活在美国。他有惊人的能力去“洞察” Hamming 空间的离散几何学，而不纠结于证明而“猜想”到多种码字的构造。

二进制 Golay 编码是一个 $[24,12]$ 编码，它的生成矩阵是 $\boldsymbol{G}(\boldsymbol{I}_{12}\ \boldsymbol{G}')$，其中 $\boldsymbol{I}_{12}$ 是一个 12×12 的特征矩阵，$\boldsymbol{G}'(=\boldsymbol{G}'_{12})$ 有如下的形式

$$\boldsymbol{G}'=\begin{vmatrix}
0&1&1&1&1&1&1&1&1&1&1&1\\
1&1&1&0&1&1&1&0&0&0&1&0\\
1&1&0&1&1&1&0&0&0&1&0&1\\
1&0&1&1&1&0&0&0&1&0&1&1\\
1&1&1&1&0&0&0&1&0&1&1&0\\
1&1&1&0&0&0&1&0&1&1&0&1\\
1&1&0&0&0&1&0&1&1&0&1&1\\
1&0&0&0&1&0&1&1&0&1&1&1\\
1&0&0&1&0&1&1&0&1&1&1&0\\
1&0&1&0&1&1&0&1&1&1&0&0\\
1&1&0&1&1&0&1&1&1&0&0&0\\
1&0&1&1&0&1&1&1&0&0&0&1
\end{vmatrix}\tag{2.4.1}$$

构造矩阵 $\boldsymbol{G}'$ 的规则是专门设计的（这是由 M. Golay 在 1949 年定义的）。在分析 Golay 编码

的时候，将会有专门的一些论据。

备注 2.4.3 有趣的是，存在一个构造所有的$\mathscr{X}_{24}^{\mathrm{Gol}}$码字(或者等价于它)的系统方法，这个方法把两种版本的 Hamming[7，4]编码$\mathscr{X}_7^H$结合在一起。首先，观察到颠倒一个 Hamming 编码$\mathscr{X}_7^H$所有位的顺序产生一个等价的编码，用$\mathscr{X}_7^K$表示。然后对$\mathscr{X}_7^H$和$\mathscr{X}_7^K$添加一个奇偶校验，产生编码$\mathscr{X}_8^{H,+}$和$\mathscr{X}_8^{K,+}$。最后，选择两个不同的编码 a，$b\in\mathscr{X}_8^{H,+}$ 和一个编码 $x\in\mathscr{X}_8^{K,+}$。那么所有的长度是 24 的$\mathscr{X}_{24}^{\mathrm{Gol}}$的 2^{12}个码字可以写成级联$(a+x)(b+x)(a+b+x)$。这可以用生成矩阵来检验。

引理 2.4.4 二进制编码$\mathscr{X}_{24}^{\mathrm{Gol}}$是自对偶的，满足$\mathscr{X}_{24}^{\mathrm{Gol}\perp}=\mathscr{X}_{24}^{\mathrm{Gol}}$。编码$\mathscr{X}_{24}^{\mathrm{Gol}}$也可以由矩阵 $\tilde{G}(G\mid I_{12})$生成。 [201]

证明 一个直接的计算表明，矩阵 G 的任意两行是点正交的，所以$\mathscr{X}_{24}^{\mathrm{Gol}}\subset\mathscr{X}_{24}^{\mathrm{Gol}\perp}$。但是$\mathscr{X}_{24}^{\mathrm{Gol}}$和$\mathscr{X}_{24}^{\mathrm{Gol}\perp}$的维数是相同的。所以，$\mathscr{X}_{24}^{\mathrm{Gol}\perp}=\mathscr{X}_{24}^{\mathrm{Gol}}$。引理的最后一个准则是根据性质$(G')^{\mathrm{T}}=G'$。 □

举例 2.4.5 证明距离 $d(\mathscr{X}_{24}^{\mathrm{Gol}})=8$。

解答 首先，对于所有的 $x\in\mathscr{X}_{24}^{\mathrm{Gol}}$，权 $w(x)$能被 4 整除。对于矩阵 $G(I_{12}\mid G')$的每一列这个结论都成立：1 的数量是 12 或者 8 中的一个。接着，对于所有二进制 N 长度的码子 x，x'

$$\mathrm{w}(x+x')=\mathrm{w}(x)+\mathrm{w}(x')-2w(x\wedge x')$$

其中$(x\wedge y)$是楔积，满足数位$(x\wedge y)_j=\min(x_j,y_j)$，$1\leqslant j\leqslant N$(参考式(2.1.6b))。但是对于任何矩阵 G 的行对 $g(j)$，$g(j')$，$w(g(j)\wedge g(j'))=0\mathrm{mod}2$。所以，对于所有 $x\in\mathscr{X}_{24}^{\mathrm{Gol}}$，4 可以整除 $w(x)$。

另一方面，$\mathscr{X}_{24}^{\mathrm{Gol}}$没有权重是 4 的码字，为了证明这一点，比较两个生成矩阵，(I_{12}/G')和$(G')^{\mathrm{T}}\mid I_{12}$。如果 $x\in\mathscr{X}_{24}^{\mathrm{Gol}}$中有 $w(x)=4$，把 x 写为一个级联$x_L\,x_R$。$(I_{12}\mid G')$行的任何非平凡和有至少是 1 的半-L 权，所以 $w(x_L)\geqslant 1$。相似地，$w(x_R)\geqslant 1$。但是如果 $w(x_L)=1$，那么 x 必须是$(I_{12}\mid G)$中的一行，加权不可能是 3。所以，$w(x_L)\geqslant 2$。相似地，$w(x_R)\geqslant 2$。但是，唯一的可能性 $w(x_L)=w(x_R)=2$，这是不可以直接校验的。所以，$w(x)\geqslant 8$。但是，$(I_{12}\mid G')$含有加权是 8 的行。所以，$d(\mathscr{X}_{24}^{\mathrm{Gol}})=8$。 □

当我们在任意数位把$\mathscr{X}_{24}^{\mathrm{Gol}}$截短，得到$\mathscr{X}_{23}^{\mathrm{Gol}}$，一个[23，12，7]编码。这种编码是完美的 3 纠错码。我们添加一个奇偶校验就可以从$\mathscr{X}_{23}^{\mathrm{Gol}}$中恢复$\mathscr{X}_{24}^{\mathrm{Gol}}$。

Hamming 码$[2^\ell-1,\ 2^\ell-1-\ell,\ 3]$和 Golay 码[23，12，7]是唯一可能完美的二进制线性编码。

长度为 12 的三进制 Golay 编码$\mathscr{X}_{12,3}^{\mathrm{Gol}}$的生成矩阵是$(I_6\mid G'_{(3)})$，其中

$$G'_{(3)}=\begin{pmatrix}0&1&1&1&1&1\\1&0&1&2&2&1\\1&1&0&1&2&2\\1&2&1&0&1&2\\1&2&2&1&0&1\\1&1&2&2&1&0\end{pmatrix},\quad 满足(G'_{(3)})^{\mathrm{T}}=G'_{(3)} \tag{2.4.2}$$

三进制 Golay 编码$\mathscr{X}_{11,3}^{\mathrm{Gol}}$是$\mathscr{X}_{12,3}^{\mathrm{Gol}}$在最后一位的一个截短。 [202]

定理 2.4.6 三进制 Golay 码$\mathscr{X}_{12,3}^{\mathrm{Gol}\perp}=\mathscr{X}_{12,3}^{\mathrm{Gol}}$是[12，6，6]，编码$\mathscr{X}_{11,3}^{\mathrm{Gol}}$是[11，6，5]，所以

是完美的。

证明　编码[11，6，5]是完美的，因为 $v_{11,3}(2)=1+11\times 2+\frac{11\times 10}{2}\times 2^2=3^5$。余下的证明留作一个练习。 □

Hamming 码[$\frac{3^l-1}{2}$，3^l-1-l，3]和 Golay 码[11，6，5]是唯一可能完美的三进制线性编码。此外，Hamming 码和 Golay 码是唯一完美线性编码，它存在于任何$\mathbb{F}_q$中，其中 $q=p^l$ 是一个素数幂。所以，这些编码是唯一可能的完美线性编码。甚至非线性的完美编码都没有带来任何本质上新的特性：它们和 Hamming 码和 Golay 码都有相同的参数(长度、大小和距离)。Golay 码在 20 世纪 80 年代用于美国的航海者宇宙飞船项目，用于传输木星和土星的特写照片。

下一个广泛运用的例子是 Reed-Muller 编码。对于 $N=2^m$，考虑二进制 Hamming 空间$\mathscr{X}_{m,2}$和$\mathscr{X}_{N,2}$。令 $\boldsymbol{M}(=\boldsymbol{M}_m)$是一个 $m\times N$ 的矩阵，其中列是整数 $j=0,1,\cdots,N-1$ 的二进制表示，最低有效位在首位置：

$$j=j_1\cdot 2^0+j_2\cdot 2^1+\cdots+j_m 2^{m-1} \tag{2.4.3}$$

所以

$$\boldsymbol{M}=\begin{array}{c} \begin{matrix} 0 & 1 & 2 & \cdots & 2^m-1 \end{matrix} \\ \begin{bmatrix} 0 & 1 & 0 & \cdots & 1 \\ 0 & 0 & 1 & \cdots & 1 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \\ 0 & 0 & 0 & \cdots & 1 \end{bmatrix} \end{array} \begin{array}{l} \\ \boldsymbol{v}^{(1)} \\ \boldsymbol{v}^{(2)} \\ \vdots \\ \boldsymbol{v}^{(m-1)} \\ \boldsymbol{v}^{(m)} \end{array} \tag{2.4.4}$$

$\boldsymbol{M}$ 的列给出了所有$\mathscr{X}_{m,2}$中的向量，行是所有$\mathscr{X}_{N,2}$中的向量，用$\boldsymbol{v}^{(1)}$，…，$\boldsymbol{v}^{(m)}$表示。特别地，$\boldsymbol{v}^{(m)}$前 2^{m-1}个元素是 0，最后 2^{m-1}是 1。为了从$\boldsymbol{M}_m$得到$\boldsymbol{M}_{m-1}$，我们必须丢弃最后一行，从剩下的$(m-1)\times N$ 矩阵中取两个相同一半中的一个。相反地，为了从 $\boldsymbol{M}_{m-1}$得到 $\boldsymbol{M}_m$，需要级联两个 $\boldsymbol{M}_{m-1}$的拷贝，并且添加行 $\boldsymbol{v}^{(m)}$：

203

$$\boldsymbol{M}_m=\begin{bmatrix} \boldsymbol{M}_{m-1} & \boldsymbol{M}_{m-1} \\ 0\cdots 0 & 1\cdots 1 \end{bmatrix} \tag{2.4.5}$$

考虑 $\boldsymbol{M}_m$ 的列$\boldsymbol{w}^{(1)}$，…，$\boldsymbol{w}^{(m)}$，它们对应数字 1，2，4，…，2^{m-1}。这些列组成了$\mathscr{X}_{m,2}$中的标准基：

$$\begin{matrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{matrix}$$

那么在位置 $j=\sum_{1\le i\le m} j_i 2^{i-1}$ 上的列是 $\sum_{1\le i\le m} j_i \boldsymbol{w}^{(i)}$。

向量$\boldsymbol{v}^{(i)}$，$i=1,\cdots,m$，可以解释为集合$\mathscr{A}_i\subset\mathscr{H}_{m,2}$的示性函数，其中第 i 个位置是 1：

$$\mathscr{A}_i=\{j\in\mathscr{H}_{m,2}:j_i=1\} \tag{2.4.6}$$

根据楔形乘积(参考式(2.1.6b))，$\boldsymbol{v}^{(i_1)}\wedge\boldsymbol{v}^{(i_2)}\wedge\cdots\boldsymbol{v}^{(i_k)}$是交集$\mathscr{A}_{i_1}\cap\cdots\cap\mathscr{A}_{i_k}$的示性函数。如果对于所有的 i_1，…i_k 是不同的，集的势$\#(\bigcap_{1\le j\le k}\mathscr{A}_{i_j})=2^{m-k}$。换句话说，我们有如下结论。

引理 2.4.7 设 $w(\wedge_{1\leqslant j\leqslant k} v^{(i_j)})=2^{m-k}$。

一个重要的事实如下。

定理 2.4.8 向量 $v^{(0)}=11\cdots1$ 和 $\wedge_{1\leqslant j\leqslant k} v^{(i_j)}$，$1\leqslant i_1<\cdots<i_k\leqslant m$，$k=1,\cdots,m$ 构成了 $\mathscr{H}_{N,2}$ 的基。

证明 我们能够验证 N 长码字的标准基 $e(j)=0\cdots1\cdots0$（1 在 j 位置处，其他是 0）可以表示成上述向量的线性组合。但是

$$e(j)=\wedge_{1\leqslant i\leqslant m}(v^{(0)}+(1+v_i^{(0)})\,v^{(i)}),0\leqslant j\leqslant N-1 \tag{2.4.7}$$

（在 j 位置的所有因子等于 1，至少有一个因子在 $i\neq j$ 的位置处等于 0。） [204]

例子 2.4.9 对于 $m=4$，$N=16$

$$\begin{aligned}
v^{(0)} &= 1111111111111111\\
v^{(1)} &= 0101010101010101\\
v^{(2)} &= 0011001100110011\\
v^{(3)} &= 0000111100001111\\
v^{(4)} &= 0000000011111111\\
v^{(1)}\wedge v^{(2)} &= 0001000100010001\\
v^{(1)}\wedge v^{(3)} &= 0000010100000101\\
v^{(1)}\wedge v^{(4)} &= 0000000001010101\\
v^{(2)}\wedge v^{(3)} &= 0000001100000011\\
v^{(2)}\wedge v^{(4)} &= 0000000000110011\\
v^{(3)}\wedge v^{(4)} &= 0000000000001111\\
v^{(1)}\wedge v^{(2)}\wedge v^{(3)} &= 0000000100000001\\
v^{(1)}\wedge v^{(2)}\wedge v^{(4)} &= 0000000000010001\\
v^{(1)}\wedge v^{(3)}\wedge v^{(4)} &= 0000000000000101\\
v^{(2)}\wedge v^{(3)}\wedge v^{(4)} &= 0000000000000011\\
v^{(1)}\wedge v^{(2)}\wedge v^{(3)}\wedge v^{(4)} &= 0000000000000001
\end{aligned}$$

□

定义 2.4.10 已知 $0\leqslant r\leqslant m$，阶数为 r 的 Reed-Muller(RM)编码 $\mathscr{Y}^{RM}(r,m)$ 是长度 $N=2^m$ 的二进制编码，它由所有的楔形乘积 $\wedge_{1\leqslant j\leqslant k} v^{(i_j)}$ 和 $v^{(0)}$ 张成，其中 $1\leqslant k\leqslant r$，$1\leqslant i_1<\cdots<i_k\leqslant m$。$\mathscr{Y}^{RM}(r,m)$ 的秩等于 $1+\begin{pmatrix}m\\1\end{pmatrix}+\cdots+\begin{pmatrix}m\\r\end{pmatrix}$。

所以 $\mathscr{Y}^{RM}(0,m)\subset\mathscr{Y}^{RM}(1,m)\subset\cdots\mathscr{Y}^{RM}(m-1,m)\subset\mathscr{Y}^{RM}(m,m)$。这里 $\mathscr{Y}^{RM}(m,m)=\mathscr{V}_{N,2}$ 是整个 Hamming 空间，并且 $\mathscr{Y}^{RM}(0,m)=\{00\cdots00,11\cdots1\}$ 是重复编码。接着，$\mathscr{Y}^{RM}(m-1,m)$ 由所有偶数加权的 $x\in\mathscr{H}_{N,2}$ 码子组成（简称为：偶码字）。实际上，根据引理 2.4.7，任何基向量都是偶的。进一步可推出，如果 x，x' 是偶的，那么

$$w(x+x')=w(x)+w(x')-2w(x\wedge x')$$

也是偶的。所以，所有码字 $\mathscr{Y}^{RM}(m-1,m)$ 是偶的。最后可知，$\mathscr{Y}^{RM}(m-1,m)$ 的维数与偶码字子空间的维数相等。得证。当 $\mathscr{Y}^{RM}(r,m)\subset\mathscr{Y}^{RM}(m-1,m)$，$r\leqslant m-1$，任何 RM 编码由偶码字组成。 [205]

对偶码是 $\mathscr{Y}^{RM}(r,m)^{\perp}=\mathscr{Y}^{RM}(m-r-1,m)$，的确，如果 $a\in\mathscr{Y}^{RM}(r,m)$，$b\in\mathscr{Y}^{RM}(m-r-1,m)$，那么楔形积 $a\wedge b$ 是一个偶码字，所以点乘 $(a\cdot b)=0$。但是

$$\dim(\mathscr{Y}^{RM}(r,m))+\dim(\mathscr{Y}^{RM}(m-r-1,m))=N$$

因此得证。作为推论，编码$\mathscr{X}^{RM}(m-2, m)$是 Hamming 编码的奇偶校验扩展。

根据定义，码字$x\in\mathscr{X}^{RM}(r, m)$与楔形乘积多项式有联系，其中的幂等变量$v^{(1)}, \cdots v^{(m)}$的阶数小于等于$r$，系数是0和1。(这里，多项式的阶数由在和项单项式中的变量$v^{(1)}, \cdots v^{(m)}$的最大数目计算。)在这个多项式中的0阶单项式与$v^{(0)}$成比例。

把这种一致性写为

$$x\in\mathscr{X}^{RM}(r,m)\leftrightarrow p_x(v^{(1)},\cdots,v^{(m)}),\quad 直径\ p_x\leqslant r \tag{2.4.8}$$

每一个这种多项式可以写成下面的形式

$$p_x(v^{(1)},\cdots,v^{(m)}) = v^{(m)}\wedge q(v^{(1)},\cdots,v^{(m-1)})-l(v^{(1)},\cdots,v^{(m-1)})$$

满足直径$q\leqslant r-1$，$\deg l\leqslant r$。码字$v^{(m)}\wedge q(v^{(1)}, \cdots, v^{(m-1)})$在前$2^{m-1}$个位置为0。

利用如上相同的表征

$$\begin{aligned} q(v^{(1)},\cdots,v^{(m-1)})&\leftrightarrow b\in\mathscr{X}^{RM}(r-1,m-1),\\ l(v^{(1)},\cdots,v^{(m-1)})&\leftrightarrow a\in\mathscr{X}^{RM}(r,m-1)\end{aligned} \tag{2.4.9}$$

进一步，2^m-码字x可以写为级联2^{m-1}码字的和：

$$x=(a\ a)-(\mathbf{0}|b)=(a|a-b) \tag{2.4.10}$$

这意味着 Reed-Muller 编码通过块积结构发生联系(参考例2.1.8(viii))：

$$\mathscr{X}^{RM}(r,m)=\mathscr{X}^{RM}(r,m-1)|\mathscr{X}^{RM}(r-1,m-1) \tag{2.4.11}$$

所以，可以归纳为

$$d(\mathscr{X}^{RM}(r,m))=2^{m-r} \tag{2.4.12}$$

实际上，对于$m=r=0$，$d(\mathscr{X}^{RM}(0, 0))=2^0$，对于所有$m$，$d(\mathscr{X}^{RM}(m, m))=1=2^0$。对于所有$m\geqslant r-1$，假设$d(\mathscr{X}^{RM}(r-1, m))=2^{m-r+1}$，$d(\mathscr{X}^{RM}(r-1, m-1))=2^{m-r}$。那么(参见式(2.4.14))

$$\begin{aligned} d(\mathscr{X}^{RM}(r,m))&=\min[2d(\mathscr{X}^{RM}(r,m-1)),d(\mathscr{X}^{RM}(r-1,m-1))]\\ &=\min[2\cdot 2^{m-1-r},2^{m-1-r+1}]=2^{m-r}\end{aligned} \tag{2.4.13}$$

208

总结一下可得到以下定理。

定理2.4.11 RM 编码$\mathscr{X}^{RM}(r, m)$，$(0\leqslant r\leqslant m)$是长度为$N=2^m$，秩$k=\sum_{0\leqslant l\leqslant r}\binom{N}{l}$，距离$d=2^{m-r}$的二进制编码，并且

(1) $\mathscr{X}^{RM}(0, m)=\{0\cdots0, 1\cdots1\}\subset\mathscr{X}^{RM}(1, m)\subset\cdots\subset\mathscr{X}^{RM}(m-1, m)\subset\mathscr{X}^{RM}(m, m)=\mathscr{X}_{N,2}$；$\mathscr{X}^{RM}(m-1, m)$是所有码长为偶数$N$的集合，$\mathscr{X}^{RM}(m-2, m)$是 Hamming 码$[2^m-1, 2^m-1-m]$的奇偶校验扩展。

(2) $\mathscr{X}^{RM}(r, m)=\mathscr{X}^{RM}(r, m-1)|\mathscr{X}^{RM}(r-1, m-1)$，$1\leqslant r\leqslant m-1$，

(3) $\mathscr{X}^{RM}(r, m)=\mathscr{X}^{RM}(m-r-1, m)$，$0\leqslant r\leqslant m-1$。

举例2.4.12 定义二进制线性编码$\mathscr{X}_1$和$\mathscr{X}_2$的块积为$\mathscr{X}_1|\mathscr{X}_2$，其中$\mathscr{X}_2$是$\mathscr{X}_1$的子码。$\mathscr{X}_1|\mathscr{X}_2$的权和最小距离与$\mathscr{X}_1$和$\mathscr{X}_2$相关。证明如果$\mathscr{X}^\perp$表示$\mathscr{X}$的对偶码，那么

$$(\mathscr{X}_1|\mathscr{X}_2)^\perp=\mathscr{X}_1^\perp|\mathscr{X}_2^\perp$$

利用块积的构造，反之亦成立。定义 Reed-Muller 编码$\mathscr{X}^{RM}(r, m)(0\leqslant r\leqslant m)$。证明如果$0\leqslant r\leqslant m-1$，那么$\mathscr{X}^{RM}(r, m)$的对偶也是一个 Reed-Muller 编码。

解答 两个线性编码$\mathscr{X}_1\subseteq\mathscr{X}_2\subseteq\mathbb{F}_2^N$的块积定义如下

$$\mathscr{X}_1|\mathscr{X}_2=\{(x|x+y):x\in\mathscr{X}_1,y\in\mathscr{X}_2\}$$

这是一个长度为$2N$的线性编码。如果$\mathscr{X}_1$的基是$x_1, \cdots, x_k$，$\mathscr{X}_2$的基是$y_1, \cdots, y_l$，

那么$\mathscr{X}_1|\mathscr{X}_2$的基是

$$(x_1|x_1),\cdots,(x_k|x_k),(0,y_1),\cdots,(0|y_l)$$

并且$\mathscr{X}_1|\mathscr{X}_2$的秩等于$\mathscr{X}_1$与$\mathscr{X}_2$的和。

接下来，我们计算最小距离

$$d(\mathscr{X}_1|\mathscr{X}_2)=\min[2d(\mathscr{X}_1),d(\mathscr{X}_2)] \tag{2.4.14}$$

令$\mathbf{0}\neq(x|x+y)\in\mathscr{X}_1|\mathscr{X}_2$。如果$y\neq\mathbf{0}$，那么权$w(x|x+y)\geqslant w(y)\geqslant d(\mathscr{X}_2)$。如果$y=0$，那么$w(x|x+y)=2w(x)\geqslant 2d(\mathscr{X}_1)$。这表明

$$d(\mathscr{X}_1|\mathscr{X}_2)\geqslant\min[2d(\mathscr{X}_1),d(\mathscr{X}_2)] \tag{2.4.15}$$

另一方面，如果$x\in\mathscr{X}_1$具有$w(x)=d(\mathscr{X}_1)$，那么$d(\mathscr{X}_1|\mathscr{X}_2)\leqslant w(x|x)=2d(\mathscr{X}_1)$。最后，如果$y\in\mathscr{X}_2$具有$w(y)=d(\mathscr{X}_2)$，那么$d(\mathscr{X}_1|\mathscr{X}_2)\leqslant w(0|y)=d(\mathscr{X}_2)$。可以总结为

$$d(\mathscr{X}_1|\mathscr{X}_2)\leqslant\min[2d(\mathscr{X}_1),d(\mathscr{X}_2)] \tag{2.4.16}$$

证明了式(2.4.14)。 [207]

现在，我们验证

$$(\mathscr{X}_2^{\perp}|\mathscr{X}_1^{\perp})\subseteq(\mathscr{X}_1|\mathscr{X}_2)^{\perp}$$

确实我们可以令$(u|u+v)\in\mathscr{X}_2^{\perp}|\mathscr{X}_1^{\perp}$和$(x|x+y)\in\mathscr{X}_1|\mathscr{X}_2$，点乘

$$\begin{aligned}\langle(u|u+v)\cdot(x|x+y)\rangle&=u\cdot x+(u+v)\cdot(x+y)\\&=u\cdot y+v\cdot(x+y)=0\end{aligned}$$

因为$u\in\mathscr{X}_2^{\perp}$，$y\in\mathscr{X}_1$，$v\in\mathscr{X}_1^{\perp}$且$(x+y)\in\mathscr{X}_2$。另外，已知

$$\begin{aligned}\operatorname{rank}(\mathscr{X}_2^{\perp}|\mathscr{X}_1^{\perp})&=N-\operatorname{rank}(\mathscr{X}_2)+N-\operatorname{rank}(\mathscr{X}_1)\\&=2N-\operatorname{rank}(\mathscr{X}_1|\mathscr{X}_2)=\operatorname{rank}(\mathscr{X}_1|\mathscr{X}_2)^{\perp}\end{aligned}$$

这实际上表明

$$(\mathscr{X}_2^{\perp}|\mathscr{X}_1^{\perp})=(\mathscr{X}_1|\mathscr{X}_2)^{\perp} \tag{2.4.17}$$

回到RM编码，它们由下面来确定：

$\mathscr{X}^{\mathrm{RM}}(0,m)$表示长度为$N=2^m$的二进制重复码，$\mathscr{X}^{\mathrm{RM}}(m,m)$表示长度为$N=2^m$的整个空间$\mathscr{V}_{N,2}$，对于$0<r<m$，$\mathscr{X}^{\mathrm{RM}}(r,m)$被$\mathscr{X}^{\mathrm{RM}}(r,m)=\mathscr{X}^{\mathrm{RM}}(r,m-1)|\mathscr{X}^{\mathrm{RM}}(r-1,m-1)$递归定义。

通过构造，$\mathscr{X}^{\mathrm{RM}}(r,m)$具有秩$\sum_{j=0}^{r}\binom{m}{j}$和最小距离$2^{m-r}$。特别地，$\mathscr{X}^{\mathrm{RM}}(m-1,m)$是奇偶校验码，所以$\mathscr{X}^{\mathrm{RM}}_{m-1,m}$是$\mathscr{X}(0,m)$的对偶。我们将证明，通常对于$0\leqslant r\leqslant m-1$，有

$$\mathscr{X}^{\mathrm{RM}}(r,m)^{\perp}=\mathscr{X}^{\mathrm{RM}}(m-r-1,m)$$

$m\geqslant 3$可以用归纳法证明。基于以上，我们可以假设当$0\leqslant r\leqslant m-1$时，$\mathscr{X}^{\mathrm{RM}}(r,m-1)^{\perp}=\mathscr{X}^{\mathrm{RM}}(m-r-2,m-1)$成立。那么对于$0\leqslant r<m$，存在

$$\begin{aligned}\mathscr{X}^{\mathrm{RM}}(r,m)^{\perp}&=(\mathscr{X}^{\mathrm{RM}}(r,m-1)|\mathscr{X}^{\mathrm{RM}}(r-1,m-1))^{\perp}\\&=\mathscr{X}^{\mathrm{RM}}(r-1,m-1)^{\perp}|\mathscr{X}^{\mathrm{RM}}(r,m-1)^{\perp}\\&=\mathscr{X}^{\mathrm{RM}}(m-r-1,m-1)|\mathscr{X}^{\mathrm{RM}}(m-r-2,m-1)\\&=\mathscr{X}^{\mathrm{RM}}(m-r-1,m)\end{aligned}$$

□

对RM码的编译码基于以下的观察。利用式(2.4.5)，对于所有的$i\notin\{i_1,\cdots,i_k\}$，当且仅当$v_j^{(i)}=0$时，乘积$v^{(i_1)}\wedge\cdots\wedge v^{(i_k)}$存在于$e(j)\in\mathscr{H}_{n,2}$的扩展中。 [208]

定义2.4.13 对于$1\leqslant i_1<\cdots<i_k\leqslant m$，定义

$$\begin{aligned}&C(i_1,\cdots,i_k)\text{表示所有整数}j=\sum_{1\leqslant i\leqslant m}j_i2^{i-1}\text{的集合}\\&\text{对于}i\notin\{i_1,\cdots,i_k\},\quad\text{满足}j_i=0\end{aligned} \tag{2.4.18}$$

对于一个空集($k=0$)，$C(\varnothing)=\{1,\cdots,2^n-1\}$。此外，集合

$$C(i_1,\cdots,i_k)-z=\{j+z:j\in C(i_1,\cdots,i_k)\} \tag{2.4.19}$$

那么，再次利用式(2.4.5)，对于所有的 $y=y_0\cdots y_{N-1}\in\mathscr{H}_{n,r}$

$$y=\sum_{0\leqslant k\leqslant n}\sum_{1\leqslant i_1<\cdots<i_k\leqslant n}\Big(\sum_{j\in C(i_1,\cdots,i_k)}y_j\Big)v^{(i_1)}\wedge\cdots\wedge v^{(i_k)} \tag{2.4.20}$$

(对于 $k=0$，取 $v^{(i_1)}\cdots$)

对 $\mathscr{H}_{n,r}$ 中的一个信息符号序列 $a=a_0\cdots a_{N-1}$ 进行编码，$k=1+\binom{m}{1}+\cdots+\binom{m}{r}$，关于 $\mathscr{X}_{n,m}^{\mathrm{RM}}$，重写为 $(a_{i_1,\cdots,i_l})$：这里，$i_1,\cdots,i_l$ 是 1 的依次的位置。那么构造一个码字为 $x=(x_0,\cdots,x_{N-1})\in\mathscr{X}_{n,m}^{\mathrm{RM}}$，其中

$$x=\sum_{1\leqslant l\leqslant r}\sum_{1\leqslant i_1<\cdots<i_l\leqslant m}a_{i_1,\cdots,i_l}v^{(i_1)}\wedge\cdots\wedge v^{(i_l)} \tag{2.4.21}$$

我们可以看到，通过识别元素 $a_j\sim a_{i_1,\cdots,i_l}$(其中 $j=j_0 2^0+j_1 2^1+\cdots+j_{m-1}2^{m-1}$，$i_1,\cdots,i_l$ 是 $j_1,\cdots,j_m$，$1\leqslant l\leqslant r$ 中 1 的依次的位置)，"信息空间" $\mathscr{H}_{n,r}$ 被嵌入到 $\mathscr{H}_{n,m}$ 中。利用这个辨识，可以得到下面的引理。

引理 2.4.14 对于所有 $0\leqslant l\leqslant m$ 和 $1\leqslant i_1<\cdots<i_l\leqslant m$，

$$\begin{aligned}\sum_{j\in C(i_1,\cdots,i_l)}x_j&=a_{i_1,\cdots,i_l},\quad 如果\ l\leqslant r\\&=0,\quad 如果\ l>r\end{aligned} \tag{2.4.22}$$

证明 可以参考式(2.4.20)。 □

引理 2.4.15 对于所有 $1\leqslant i_1<\cdots<i_r\leqslant m$，且对于任何 $1\leqslant t\leqslant m$，使得 $t\notin\{i_1,\cdots,i_r\}$

209

$$a_{i_1,\cdots,i_r}=\sum_{j\in C(i_1,\cdots,i_r)-2^{t-1}}x_j \tag{2.4.23}$$

证明 证明可由以下这个事实推导出，即 $C(i_1,\cdots,i_r,t)$ 是不相交并集 $C(i_1,\cdots,i_r)\cup C(i_1,\cdots,i_r,2^{t-1})$ 和等式 $\sum_{j\in C(i_1,\cdots,i_r,t)}x_j=0$(见式(2.4.19))。 □

而且有如下定理。

定理 2.4.16 对于与 $v^{(i_1)\cdots(i_r)}$ 相对应的任意信息符号 $a_{i_1,\cdots,i_r}$，我们能够将集合 $\{0,\cdots,N-1\}$ 分割成 2^{m-r} 个不相交的子集 S，每一个集合都包含 2^r 个元素，使得对于所有的 S，$a_{i_1,\cdots,i_r}=\sum_{j\in S}x_j$。

证明 集合 S 的清单开始于 $C(i_1,\cdots,i_r)$，接下来是 $(m-r)$ 个不相交集合 $C(i_1,\cdots,i_r)+2^{t-1}$，其中 $1\leqslant t\leqslant m$，$t\notin\{i_1,\cdots,i_r\}$。接下来，我们取任意一对值 $1\leqslant t_1\leqslant t_2\leqslant m$ 使得 $\{t_1,t_2\}\cap\{i_1,\cdots,i_r\}=\varnothing$。那么 $C(i_1,\cdots,i_r,t_1,t_2)$ 包含不相交的集合 $C(i_1,\cdots,i_r)$，$C(i_1,\cdots,i_r)+2^{t_1-1}$ 以及 $C(i_1,\cdots,i_r)+2^{t_2-1}$，对于它们中的任何一个，$a_{i_1,\cdots,i_r}=\sum_{j\in C(i_1,\cdots,i_r)+2^{t_k-1}}x_j$，$k=1,2$，那么对于剩下的集合

$$\begin{aligned}&C(i_1,\cdots,i_r)+2^{t_1-1}+2^{t_2-1}=C(i_1,\cdots,i_r,t_1,t_2)\setminus\\&[C(i_1,\cdots,i_r)\cup(C(i_1,\cdots,i_r)+2^{t_1-1})\cup(C(i_1,\cdots,i_r)+2^{t_2-1})]\end{aligned} \tag{2.4.24}$$

同样适用。

有 $\binom{m-r}{2}$ 个组合，并且它们仍然互相以及与以前的集合不相交。集合(2.4.24)进一

少组成了一群集合 S。 □

同理，集合 S 的一般形式是

$$C(i_1,\cdots,i_r)+2^{i_1-1}+\cdots+2^{i_r-1}$$

它与集合理论上的差分相同

$$C(i_1,\cdots,i_r,t_1,\cdots,t_s)$$
$$\setminus\Big[\bigcup_{\{j_1,\cdots,j_s\}\subset\{t_1,\cdots,t_s\}}(C(i_1,\cdots,i_r)+2^{j_1-1}+\cdots+2^{j_s-1})\Big] \tag{2.4.25}$$

这里每一个这样的集合都被集合 $C(t_1,\cdots,t_s)$ 标注，其中 $0\leqslant s\leqslant m-r$，$t_1<\cdots<t_s$ 以及 $\{t_1,\cdots,t_s\}\cap\{i_1,\cdots,i_r\}=\varnothing$。[在(2.4.25)中的并集 $\bigcup_{\{j_1,\cdots,j_s\}\subset\{t_1,\cdots,t_s\}}$ 是针对集合 $\{t_1,\cdots,t_s\}$ 中(严格的)子集 $\{t'_1,\cdots,t'_s\}$，其中 $t'_1<\cdots<t'_s$ 以及 $s'=0,\cdots,s-1$($s'=0$ 给出了一个空子集)]。集合 $C(i_1,\cdots,i_r)$ 的数量等于 2^{m-r}，并且它们中的每一个通过构造以后都有 2^r 个元素。

定理 2.4.16 为 Reed Muller 编码所谓的大数判决译码提供了一个基本原理。即依靠一个自码字 $x^h\in\mathscr{R}^{RM}_{r,m}$ 的接收编码 $y=(y_1,\cdots,y_{N-1})$，我们取任意的 $1\leqslant i_1<\cdots<i_r\leqslant m$ 并考虑在上述的 2^{m-r} 个集合 S 上计算和 $\sum_{j\in S}y_j$。如果 $y\in\mathscr{R}^{RM}_{r,m}$，所有这些和都是相同的并得到 $a_{i_1,\cdots,i_r}$。如果 y(即 Hamming 距离 $\delta(x^h,y)$ 中的错误数 $<2^{m-r-1}-d(\mathscr{R}^{RM}_{r,m})/2$，那么大多数的和将仍然给出一个正确的 $a_{i_1,\cdots,i_r}$，最糟糕的情况是每一个集合 S 不包含或者只包含一个错误)。通过改变 $\{i_1,\cdots,i_r\}$，我们将确定只包含 r 阶单项式的码字 $x^{(1)}\in\mathscr{R}^{RM}_{r,m}$。注意 $x^h-x^{(1)}$ 将成为 $\mathscr{R}^{RM}_{r-1,m}$ 中的一个码字。 210

接着 y 会被减少为 $y-x^{(1)}$，和 $x^h-x^{(1)}$ 相比，减少的码字 $y-x^{(1)}$ 将有 $\delta(x^h-x^{(1)},y^h-x^{(1)})=\delta(x^h,y)$ 个错误，它们 $<2^{m-r-1}-d(\mathscr{R}^{RM}_{r-1,m})/2$。我们重复上边的步骤并获得对于任意 $1\leqslant i_1<\cdots<i_{r-1}\leqslant m$ 来说正确的 $a_{i_1,\cdots,i_{r-1}}$。最后，我们恢复整个信息符号 $a_{i_1,\cdots,i_0}$。

因此，对于任何距离为 $\delta(y,\mathscr{R}^{RM}_{r,m})<d(\mathscr{R}^{RM}_{r,m})/2$ 的码字 $y\in\mathscr{H}^n_{X,2}$ 都能被唯一译码。

……正确，插入，重定义，
变大，变小，在字行间插入。

Jonathan Swift(1667—1745)，Anglo-Irish 作家

Reed-Muller 编码在 20 世纪 50 年代由 David Muller(1924—2008)首次研究得出；Irwin Reed(1923—2012)提出了上面的解码步骤。在 20 世纪 70 年代早期，RM 编码被太空飞船用来传递来自太空的图片。传输的质量在那个时候被认为是非常好的，然而，接下来，NASA 工程师决定在拍摄木星和水星时用 Golay 编码。

举例 2.4.17 最大距离可分(MDS)编码在早前定义为 q 进制线性[N，k，d]码，其中 $d=N-k+1$(Singleton 界的等式，见定义 2.1.13)。

(a) 证明 $\mathscr{C}$ 是 MDS，当且仅当

(i) 它的奇偶校验矩阵 $\boldsymbol{H}$ 的任意 $N-k$ 列都是线性独立的。

(ii) $\boldsymbol{H}$ 中存在 $N-k+1$ 列是线性相关的。

(b) 证明一个 MDS 的对偶码是 MDS，并推出 $\mathscr{C}$ 是 MDS 当且仅当它的生成矩阵 $\boldsymbol{G}$ 的任意 k 列都是线性无关的，并且 k 是最大这个数。

(c) 因此证明当 $\boldsymbol{G}$ 被写入标准形式($\boldsymbol{I}_k|\boldsymbol{G}'$)中时，那么 $\mathscr{C}$ 是 MDS 当且仅当 $\boldsymbol{G}'$ 的任意正方形子矩阵都是非奇异的。

(d) 最后，验证[N，k，d]编码 $\mathscr{C}$ 是 MDS 当且仅当对于任意 d 个位置 $1\leqslant i_1<\cdots<$

211 $i_d \leqslant N$，存在一个在数字 $i_1, \cdots i_d$ 中是非零数的权为 d 的码字。

解答 (a) MDS$[N, k, d]$编码有 $d=N-k+1$。如果线性码$\mathscr{X}$有 $d(\mathscr{X})=d$，它的奇偶校验矩阵 $\boldsymbol{H}$ 的任意$(d-1)$列都是线性独立的，$(d-1)$是具有这个性质的最大数，反之亦然。所以，任意$(N-k)$列是行线性独立的，$(N-k)$是最大的这种数，反之亦然，相当于 $\boldsymbol{H}$ 的任意$(N-k)\times(N-k)$子矩阵是可逆的。

(b) 令$\mathscr{X}$是关于奇偶校验矩阵 $\boldsymbol{H}$ 的$[N, k, d]$MDS 编码，那么 $\boldsymbol{H}$ 是$\mathscr{X}^{\perp}$的生成矩阵。$\boldsymbol{H}$ 的任意$(N-k)\times(N-k)$子矩阵是可逆的。那么，$\boldsymbol{H}^{\mathrm{T}}$ 行的任意非平凡组合有$\leqslant N-k-1$个非零元素，即权$\geqslant k+1$。最小权重等于 $k+1$。所以 $d(\mathscr{X}^{\perp})=k+1=N-(N-k)+1$。由于$\mathscr{X}^{\perp}$是$[N, N-k]$码，所以它就是 MDS。

那么，很明显，$[N, k]$码$\mathscr{X}$是 MDS，当且仅当 k 是最大数 l 使得它的生成矩阵 $\boldsymbol{G}$ 的任意 l 列都是线性独立的。等效于$\mathscr{X}$在任意 k 个位置上是系统的。

(c) 再一次，令$\mathscr{X}$为$[N, k, d]$MDS 码，记为 $\boldsymbol{G}=(\boldsymbol{I}_k|\boldsymbol{G}')$。取 $\boldsymbol{G}'$的一个$(\mu\times\mu)$子矩阵 $\bar{\boldsymbol{G}}_{\mu}$。通过使用行和列的排列组合，我们假设$\bar{\boldsymbol{G}}_{\mu}$ 占据了$\boldsymbol{G}'$的最左上角。那么考虑 $\boldsymbol{I}_k$ 的后$(k-u)$列和包含$\bar{\boldsymbol{G}}_{\mu}$ 的$\boldsymbol{G}'$的 u 列；相对应的 $k\times k$ 矩阵是非奇异的，并且组成了一个$k\times k$的子矩阵 $\boldsymbol{G}_k$。

$$\boldsymbol{G}_k=\begin{bmatrix} 0 & \bar{\boldsymbol{G}}_{\mu} \\ \boldsymbol{I}_{k-u} & * \end{bmatrix}$$

并满足

$$\det\boldsymbol{G}_k=\pm\det\bar{\boldsymbol{G}}_{\mu}\det\boldsymbol{I}_{k-u}=\pm\det\bar{\boldsymbol{G}}_{\mu}\neq 0, \quad 由(b)$$

所以，$\bar{\boldsymbol{G}}_{\mu}$ 是非奇异的。逆问题的证明是相似的。

(d) 最后，选择 $d=N-k+1$ 个数字，即 $i_1, \cdots, i_d$。考虑 i_1 和剩下的数字 $j_1, \cdots, j_{k-1}$。那么，$i_1, j_1, \cdots, j_{k-1}$是信息符号。所以存在有非零数字 i_1 和 $j_1, \cdots, j_{k-1}$ 为零的码字 $\boldsymbol{x}$。那么，$\boldsymbol{x}$ 一定有着非零数 $i_1, \cdots, i_d$。

逆问题：考虑一个$(N-d+1)\times N$ 的矩阵

$$\bar{\boldsymbol{G}}=[\boldsymbol{I}_{N-d+1}|\boldsymbol{E}_{(N-d+1)\times(d-1)}]$$

其中 I_{N-d+1}是一个单位矩阵，$\boldsymbol{E}$ 是一个具有全 1 元素的$(N-d+1)\times(d-1)$矩阵(单位是 $\mathbb{F}_2$)。$\bar{\boldsymbol{G}}$的行是线性独立的，权重为 d。对于任意一行，在同样的位置，存在一个有着非零数的码字$\boldsymbol{x}^{(i)}\in\mathscr{X}$(并且，可能，在别处)。那么，码字的秩 $k\geqslant N-d+1$，因此，$k=N-$
212 $d+1$。 □

举例 2.4.18 MDS 编码$[N, N, 1]$，$[N, 1, N]$以及$[N, N-1, 2]$总是存在并称为是平凡的。任何 $2\leqslant k\leqslant N-2$ 的$[N, k]$MDS 码被称为是非平凡的。证明不存在满足 $q\leqslant k\leqslant N-q$ 的$\mathbb{F}_q$上的非平凡 MDS 码。特别是，不存在非平凡二进制 MDS 码(这造成了人们对二进制 MDS 码缺乏明显的兴趣)。

解答 事实上，$[N, N, 1]$，$[N, N-1, 2]$和$[N, 1, N]$码都是 MDS。取 $q\leqslant k\leqslant N-q$，假设$\mathscr{X}$是一个 q 进制的 MDS。在标准形式$(\boldsymbol{I}_k|\boldsymbol{G}')$中取生成矩阵 $\boldsymbol{G}$，其中 $\boldsymbol{G}'$是 $k\times(N-k)$，$N-k\geqslant q$。

如果 $\boldsymbol{G}'$中某一列的一些元素是零，而且这列是 $\boldsymbol{I}_{k-1}$的 $k-1$ 列的线性组合，这在以前的例子(b)中是不可能的；因此 $\boldsymbol{G}'$有非 0 元素。接下来假设 $\boldsymbol{G}'$的第一行是 $1\cdots1$；否则我们能执行列的标量乘，同时保持这些编码的等价性。

现在取出 $\boldsymbol{G}'$的第二行：它的长度是 $N-k\geqslant q$，没有 0 元素。那么重复此过程，即，

$$G=\left[\begin{array}{c|ccccccc} & 1 & \cdots & 1 & \cdots & 1 & \cdots & 1 \\ I_k & \cdots & \cdots & a & \cdots & a & \cdots & \cdots \\ & & & \cdots & & \cdots & & \end{array}\right], a\neq 0$$

然后选取码字

$$x=\text{行}\ 1-a^{-1}(\text{行}\ 2)$$

它有 $w(x)\leqslant N-k-2+2=N-k$，并且$\mathscr{Y}$不可能是 MDS。

通过使用对偶编码，得出存在满足 $k\geqslant q$ 的非平凡 q 进制 MDS 码。因此，非平凡 MDS 码可能只有

$$N-q+1\leqslant k \text{ 或 } k\leqslant q-1$$

即不存在非平凡二进制 MDS 码，但存在一个非平凡的[3，2，2]三元 MDS 码。 □

备注 2.4.19 给定 k 和 q，对于一个 q 进制的 MDS[N，k]码，探究 N 的最大值是一件有趣的工作。我们证明，N 必须$\leqslant qk+q-1$，但计算的证据显示这个值是 $q+1$。

2.5 循环码和代数多项式，BCH 码简介

一类有用的线性码由所谓的循环码组成(特别是 Hamming 码，Golay 码和 Reed-Muller 码等都是循环码)。循环码是由 Eugene Prange 在 1957 年提出的，它的重要性立刻得到认识，而且产出了大量的文献。但更重要的是，循环码的思想和在 20 世纪 50 年代后期形成的一些其他的敏锐观察一起，特别是 BCH 码的发明，开启了从线性编码理论(那时是初始阶段)到代数的连接，特别是有限域理论。这创造了代数编码理论，一个在现代线性编码理论中非常有前景的方向。 [213]

我们首先从二进制循环码开始。对于长度为 N 的二进制循环码的编解码步骤是基于与二进制系数多项式有关的代数：

$$a(X)=a_0+a_1X+\cdots+a_{N-1}X^{N-1},$$
$$\text{其中 } a_k\in \mathbb{F}_2, k=0,\cdots,N-1 \tag{2.5.1}$$

这种多项式除了 $X^N+X^N=0$ 之外能以通常方式进行相加和相乘。这定义了一个二进制多项式代数$\mathbb{F}_2[X]$；这种在二进制多项式进行的运算适用于这种代数。多项式 $a(X)$的次数 $\deg a(X)$等于它非零系数的最大标号。零多项式的次数被设定为 0。因此，式(2.5.1)的表示包含次数$<N$ 的多项式。

定理 2.5.1 (a) $(1+X)^{2^l}=1+X^{2^l}$ (Freshman's Dream)。

(b) (分割算法)令 $f(X)$和 $h(X)$是两个二进制多项式，并且 $h(X)\not\equiv 0$ 的。那么存在唯一的多项式 $g(X)$和 $r(X)$，使得

$$f(X)=g(X)h(X)+r(X),\quad \text{其中}\quad \deg r(X)<\deg h(X) \tag{2.5.2}$$

多项式 $g(X)$被称为比率或者商，$r(X)$是余数。

证明 (a) 证明过程服从二项式分解，其中所有的交叉项消失。

(b) 如果 $\deg h(X)>\deg f(X)$，我们简单设置

$$f(X)=0\cdot h(X)+f(X)$$

如果 $\deg h(X)\leqslant\deg f(X)$，我们可以执行长除法的“标准”步骤，满足二进制加法和乘法的规则。 □

例子 2.5.2 对于二进制多项式：

(a) $(1+X+X^2+X^3)(X+X^2+X^3)=X+X^7$

(b) $1+X^N=(1+X)(1+X+\cdots X^{N-1})$

(c) 商$(X+X^2+X^3+X^4+X^5)/(1+X+X^2+X^3)=X^2+X^4=X^2+X^4$，余数是$X+X^2+X^3$。

214

定义 2.5.3 两个多项式，$f_1(X)$和$f_2(X)$，如果它们的余数在被$h(X)$除后相同，则称为等数模$h(X)$或者$f_1(X)=f_2(X)$模$h(X)$，即

$$f_i(X)=g_i(X)h(X)+r(X),\quad i=1,2$$

并且$\deg r(X)<\deg h(X)$。

定理 2.5.4 多项式的加法和乘法遵守等效性，即如果

$$f_1(X)=f_2(X) \bmod h(X) \qquad p_1(X)=p_2(X) \bmod h(X) \tag{2.5.3}$$

那么

$$\begin{cases} f_1(X)+p_1(X)=f_2(X)+p_2(X) \bmod h(X) \\ f_1(X)p_1(X)=f_2(X)p_2(X) \bmod h(X) \end{cases} \tag{2.5.4}$$

证明 对于$i=1,2$，我们有

$$f_i(X)=g_i(X)h(X)+r(X),\ p_i(X)=q_i(X)h(X)+s(X)$$

满足

$$\deg r(X),\ \deg s(X)<\deg h(X)$$

因此

$$f_i(X)+p_i(X)=(g_i(X)+q_i(X))h(X)+(r(X)+s(X))$$

满足

$$\deg(r(X)+s(X))\leqslant \max[r(X),s(X)]<\deg h(X)$$

因此

$$f_1(X)+p_1(X)=f_2(X)+p_2(X) \bmod h(X)$$

另外，对于$i=1,2$，乘积$f_i(X)p_i(X)$表示为

$$(g_i(X)q_i(X)h(x)+r(X)q_i(X)+s(X)g_i(X))h(X)+r(X)s(X)$$

因此，两个多项式$f_1(X)p_1(X)$和$f_2(X)p_2(X)$的余数可能仅仅来自于$r(X)s(X)$，因此，对于它们都是相同的。□

注意到每个线性二进制编码$\mathscr{X}_N$对应一个多项式集合，它们的系数是0，1，次数是$N-1$，对于模2加运算是封闭的，

$$a(X)=a_0+a_1X+\cdots+a_{N-1}X^{N-1}\leftrightarrow a^{(N)}=a_0\cdots a_{N-1}$$

$$b(X)=b_0+b_1X+\cdots+b_{N-1}X^{N-1}\leftrightarrow b^{(N)}=b_0\cdots b_{N-1}$$

215

$$a(X)+b(x)\leftrightarrow a^{(N)}+b^{(N)}=(a_0+b_0)\cdots(a_{N-1}+b_{N-1}) \tag{2.5.5}$$

(在一个长度为N的码字中，数字的编号用的是0，…，$N-1$而不是1，…，N，这样会更方便一些。)

当码字$a^{(N)}=a_0\cdots a_{N-1}$时，我们有意写为$a(X)\in\mathscr{X}$，它表示属于码字$\mathscr{X}$的多项式$a(X)$。

定义 2.5.5 给定一个二进制码字$a=a_0a_1\cdots a_{N-1}$，我们定义循环移位πa是一个码字$a_{N-1}a_0\cdots a_{N-2}$。如果每一个码字循环移位后仍然是同一个码字，那么这个二进制码$\mathscr{X}$就被称为循环码。

一个“直接”形成循环码的方法如下：取一个码字a，那么它的后向循环移位是πa，$\pi^2 a$等。最后获得这些向量的所有和。这样的一个构造允许我们从单独的一个码字中构建一个码字，最终这个码字的整个特性可以从码字a的特性中推断出来。结果显示，每一个循环码都可能以这种方式获得，相应的码字被称为循环码的生成子。

引理 2.5.6 一个二进制线性码$\mathscr{C}$是循环码，当且仅当对任何一个来自$\mathscr{C}$的基的向量u，有$\pi u\in\mathscr{C}$。

证明 在$\mathscr{C}$中的每一个码字都是基向量的和，但是$\pi(u+v)=\pi u+\pi v$，因此得证。 □

循环移位的一个有用特性在下面确立。

引理 2.5.7 如果码字a对应一个多项式$a(X)$，那么码字πa对应$Xa(X)$模$(1+X^N)$。

证明 关系

$$\begin{aligned}Xa(X)&=a_0X+a_1X^2+\cdots+a_{N-2}X^{N-1}+a_{N-1}X^N\\&=a_{N-1}+a_0X+a_1X^2+\cdots+a_{N-2}X^{N-1}\bmod(1+X^N)\end{aligned}$$

意味着对应于πa的多项式

$$a_{N-1}+a_0X+\cdots+a_{N-2}X^{N-1}$$

等于$Xa(X)\bmod(1+X^N)$。 □

相似的证明意味着码字π^2a对应$X^2a(X)\bmod(1+X^N)$，等等。更一般地，我们有以下结论。

例子 2.5.8 逆循环移位π^{-1}：$a_0\cdots a_{N-2}a_{N-1}\in\{0,1\}^N\mapsto a_1a_2\cdots a_{N-1}a_0$作用于次数至少为$N-1$的多项式$a(X)$上可表达为

$$\pi^{-1}a(X)=\frac{1}{X}[a(X)+a_0]+a_0X^{N-1}$$

316

定理 2.5.9 对于每一对多项式$a(X)$和$b(X)$，一个二进制循环码包含$a(X)+b(X)$和任何一个多项式$v(X)a(X)\bmod(1+X^N)$。

证明 根据线性特性，和$a(X)+b(x)\in\mathscr{C}$。如果$v(X)=v_0+v_1X+\cdots+v_{N-1}X^{N-1}$，则每一个多项式$X^ia(X)\bmod(1+X^N)$对应于$\pi^ia$，因此属于$\mathscr{C}$。因为

$$v(X)a(X)\bmod(1+X^N)=\sum_{i=0}^{N-1}v_iX^ia(X)\bmod(1+X^N)$$

所以LHS属于$\mathscr{C}$。 □

换句话说，次数至少为$N-1$并满足$*$-乘法的二进制多项式被定义为

$$a*b(X)=a(X)b(X)\bmod(1+X^N) \tag{2.5.6}$$

和通常的$\mathbf{F}_2$ $\mathscr{C}$-加法，构成了交换环，用$\mathbf{F}_2[X]/(1-X^N)$表示。二进制循环码恰恰是这种环的理想状态。

定理 2.5.10 令$g(X)=\sum_{i=0}^{N-k}g_iX^i$是一个在二进制循环码$\mathscr{C}$中最小次数的非零多项式。那么：

(i) $g(X)$是唯一最小次数的多项式。

(ii) 码$\mathscr{C}$的秩是k。

(iii) 对应$g(X)$，$Xg(x)$，…，$X^{k-1}g(X)$的码字，构成了在$\mathscr{C}$中的基；它们是码字$g=g_0\cdots g_{N-k}0\cdots0$的循环移位。

(iv) $a(X)\in\mathscr{C}$当且仅当对于一些次数$<k$的多项式$v(X)$有$a(X)=v(X)g(X)$（即$g(X)$是来自$\mathscr{C}$的每一个多项式的除数）。

证明 (i) 假设$c(X)=\sum_{i=0}^{N-k}c_iX^i$是在$\mathscr{C}$中最小次数为$N-k$的另一个多项式。那么$g_{N-k}=c_{N-k}=1$，因而，$\deg(c(X)+g(X))<N-k$。但是因为$N-k$是最小次数，$c(X)+g(X)$应该等于0，当且仅当$g(X)=c(X)$，这种情况才会发生。因此$g(X)$是唯一的。

(ii) 可以由(iii)推出。

(iii) 假设性质(iv)成立。那么每一个多项式 $a(X)\in\mathscr{X}$ 有如下的形式

217

$$g(X)v(X)=\sum_{i=1}^{r}v_iX^ig(X),\quad r<k$$

因此，每个多项式 $a(X)\in\mathscr{X}$ 是多项式 $g(X)$，$Xg(X)$，…，$X^{k-1}g(X)$（全部属于 $\mathscr{X}$）的一个线性组合。另一方面，多项式 $g(x)$，$Xg(X)$，…，$X^{k-1}g(X)$ 是不同次数的，因此是线性无关的。所以对应 $g(X)$，$Xg(X)$，…，$X^{k-1}g(X)$ 的码字 $\mathbf{g}$，$\pi\mathbf{g}$，…，$\pi^{k-1}\mathbf{g}$ 形成了 $\mathscr{X}$ 的基。

(iv) 我们知道每个多项式 $a(X)\in\mathscr{X}$ 的次数 $\geq\deg g(X)$。根据除法，

$$a(X)=v(X)g(X)+r(X)$$

在这里，我们一定有

$$\deg v(X)<k\quad 且\quad \deg r(X)<\deg g(X)=N-k$$

但是，由于定理 2.5.9（因为 $v(X)g(X)$ 的次数 $\leq N-1$，它与 $v(X)g(X)\bmod(1+X^N)$ 相同），因此 $v(X)g(X)$ 属于 $\mathscr{X}$。所以，通过线性特性可得出

$$r(X)=a(X)+v(X)g(X)\in\mathscr{X}$$

由于 $g(X)$ 是 $\mathscr{X}$ 中最小次数的唯一的多项式，所以 $r(X)=0$。 □

推论 2.5.11 通过循环移位和线性组合，每一个二进制循环码从对应于最小次数的多项式的码字中获得。

定义 2.5.12 在 $\mathscr{X}$ 中最小次数多项式 $g(X)$ 被称为二进制代码最小次数生成子，或者简称为 $\mathscr{X}$ 生成子。

备注 2.5.13 根据推论 2.5.11 可知，可能还有其他生成 $\mathscr{X}$ 的多项式。但是最小次数多项式是唯一的。

定理 2.5.14 一个次数小于等于 $N-1$ 的多项式 $g(X)$ 是一个长度为 N 的二进制循环码的生成子，当且仅当 $g(X)$ 可整除 $1+X^N$。也就是：对某些多项式 $h(X)$ 而言（次数 $N-\deg g(X)$）

$$1+X^N=h(X)g(X)\tag{2.5.7}$$

证明 （仅当且部分）根据除法

$$1+X^N=h(X)g(X)+r(X)$$

其中

$$\deg r(X)<\deg g(X)$$

即

218

$$r(X)=h(X)g(X)+1+X^N,\quad 即\ r(X)=h(X)g(X)\bmod(1+X^N)$$

根据定理 2.5.10，$r(X)$ 属于由 $g(X)$ 生成的循环码 $\mathscr{X}$。但是 $g(X)$ 一定是 $\mathscr{X}$ 中最小次数的唯一的多项式。所以，$r(X)=0$ 且 $1+X^N=h(X)g(X)$。 □

（当且部分）假设 $1+X^N=h(X)g(X)$，$\deg h(X)=N-\deg g(X)$。考虑集合 $\{a(X):a(X)=u(X)g(X)\bmod(1+X^N)\}$，即对应 $g(X)$ 的 * 乘法多项式环中的主理想。这个集合形成了一个线性编码；它包含 $g(X)$，$Xg(X)$，…，$X^{k-1}g(X)$，其中 $k=\deg h(X)$。我们足以证明 $X^kg(X)$ 同样也属于这个集合。但是 $X^kg(X)=1+X^N+\sum_{i=0}^{k-1}h_iX^ig(X)$，即 $X^kg(X)$ 与 $g(X)$，$Xg(X)$，…，$X^{k-1}g(X)$ 的线性组合是等价的。

推论 2.5.15 所有长度为 N 的二进制循环码与多项式 $1+X^N$ 的除数是一一对应的。

因此，循环码用多项式 $1+X^N$ 的因式分解来描述它。更准确地来说，我们对 $1+X^N$ 分解为不可约因子更感兴趣；将这些因子组合成乘积，从而产生长度为 N 的所有可能的循环码。

定义 2.5.16 如果 $a(X)$ 不能被写作两个多项式 $b(X)$ 和 $b'(X)$ 的乘积，则多项式 $a(X)=a_0+a_1X+\cdots+a_{N-1}X^{N-1}$ 被称为不可约，其中 $\min[\deg(b)(X),\ \deg b(X)]\geqslant 1$。

描述循环码的不可约多项式的重要性(方便性)是显而易见的：长度为 N 的每一个循环码生成多项式都是不可约因子 $1+X^N$ 的乘积。

例子 2.5.17 (a) 多项式 $1+X^N$ 有两个"标准"的除数：

$$1+X^N=(1+X)(1+X+\cdots+X^{N-1})$$

第一个因子 $1+X$ 生成二进制奇偶校验码 $\mathscr{P}_N=\{x=x_0\cdots x_{N-1}:\ \sum_i x_i=0\}$。然而多项式 $1+X+\cdots X^{N-1}$(它可能是可约的)生成重复编码 $\mathscr{R}_N=\{00,\ \cdots 0,\ 11\cdots 1\}$。

(b) 在字典形式中，选择 Hamming[7，4]码的生成和校验矩阵。如果我们重新排序数字 $x_1x_7x_6x_5x_4x_3x_2$(这会产生一个等价编码)，那么生成矩阵的行变成其他行的后向循环移位。

$$G^{\mathscr{V}}_{4\times 7}=\begin{pmatrix}1101000\\0110100\\0011010\\0001101\end{pmatrix}$$

219

最后一行的循环移位也在码中：

$$\begin{aligned}\pi(0001101)&=(1000110)\\&=(1101000)+(0110100)+(0011010)\end{aligned}$$

根据引理 2.5.6，码是循环的。根据定理 2.5.10(iii)可知，$g(X)$ 的生成多项式对应于矩阵 $G^{\mathscr{V}}_{4\times 7}$ 的下划线部分：

$$\underline{1101}\sim g(X)=1+X+X^3=\text{生成的多项式}$$

但我们可以用一个类似的命题去说明一个等效的循环码可以从码字 $1011\sim 1+X^2+X^3$ 获得。这并不矛盾，它并不说明一个循环码的多项式理想是一个唯一元素的主理想。

如果我们改变奇偶校验矩阵中列的顺序，这个码等效于原来的码；也就是说，生成矩阵为 $1+X^2+X^3$ 的码也是 Hamming[7，4]码。

在问题 2.3 中我们会验证 Golay[23，7]码由多项式 $g(X)=1+X+X^5+X^6+X^7+X^9+X^{11}$ 生成。

举例 2.5.18 利用在 $\mathbb{F}_2[X]$ 上的因式分解

$$X^7+1=(X+1)(X^3+X+1)(X^3+X^2+1) \tag{2.5.8}$$

找到所有长度为 7 的循环二进制码，找出其中的 Hamming 码和它们的对偶。

解答 见下表。

码字	$\mathscr{C}$的生成多项式	$\mathscr{C}^\perp$的生成多项式
$\{0,\ 1\}^7$	1	$1-X^7$
Parity check	$1+X$	$\sum_{0\leqslant i\leqslant 6}X^i$
Hamming	$1+X+X^3$	$1-X^2+X^3+X^4$
Hamming	$1+X^2+X^3$	$1+X+X^2+X^4$
Dual Hamming	$1+X^2+X^3+X^4$	$1+X+X^3$
Dual Hamming	$1+X+X^2+X^4$	$1+X^2+X^3$
Repetition	$\sum_{0\leqslant i\leqslant 6}X^i$	$1+X$
zero	$1-X^7$	1

220

很容易验证式(2.5.8)中的所有因子都是不可约的。任何不可约的因子都有可能包含或

者不包含在生成多项式的分解中。这一命题证明了在$\mathscr{H}_{7,2}$中正好存在如表中所示的8个二进制码。

例子 2.5.19 (a) 一次多项式 $1+X$ 和 X 是不可约的(但 X 并不出现在 $1+X^N$ 的分解中)。存在一个不可约的二次二进制多项式：$1+X+X^2$，两个三次式：$1+X+X^3$ 和 $1+X^2+X^3$，和三个四次式：

$$1+X+X^4,1+X^3+X^4 \quad 和 \quad 1+X+X^2+X^3+X^4 \tag{2.5.9}$$

它们每个都出现在对于不同 N 值的 $1+X^N$ 的分解中(见下面)。更大的区别是多项式 $1+X+X^3$ 和 $1+X^2+X^3$ 是"本原"的，但 $1+X^2+X^3+X^4$ 不是；见下面的例子 2.5.34 和 3.1～3.3 节。另一方面，多项式

$$1+X^5,1+X^4+X^5+X^6+X^7 \quad 和 \quad 1+X^2+X^3+X^5 \tag{2.5.10}$$

是可约的。多项式 $1+X^N$ 总是可约：

$$1+X^N=(1+X)(1+X+\cdots+X^{N-1})$$

(b) 一般来说，将多项式 $1+X^N$ 因式分解为不可约因子并不是很容易。在 N 的头 13 个奇数值中，可以简单分解成两个不可约因子的多项式 $1+X^N$ 如下所示：

$$1+X,1-X^3,1-X^5,1-X^{11},1-X^{13}$$

此外，多项式 $1+X^{15}$ 可以简单地分解为 $(1+X)(1-X+\cdots+X^{14})$，其余的有如下因子(忽略共同因子 $(1+X)$)：

$$\begin{aligned}
&1+X^7:(1+X+X^3)(1+X^2+X^3),\\
&1+X^9:(1+X+X^3)(1+X^2+X^3),\\
&1+X^{21}:(1+X+X^3)(1+X+X^4)\\
&\qquad\times(1+X^3+X^4)(1+X+X^2+X^3+X^4),\\
&1+X^{17}:(1+X^3+X^4+X^5+X^8)\\
&\qquad\times(1+X+X^2+X^3+X^5+X^7+X^8),\\
&1+X^{33}:(1+X+X^2)(1+X+X^5)(1+X^2+X^5)\\
&\qquad\times(1+X+X^2+X^3+X^6)(1+X^2+X^4+X^5+X^6),\\
&1+X^{23}:(1+X+X^2+X^4+X^7+X^9+X^{11})\\
&\qquad\times(1+X^3+X^4+X^5+X^9+X^{10}+X^{11})\\
&1+X^{35}:(1+X+X^2+X^3+X^4)(1-X^5+X^{10}+X^{15}+X^{20})
\end{aligned}$$

221

对于 N 是偶数的情况，$1+X^N$ 可能有多个根(见例子 2.5.35(c))。

例子 2.5.20 在域$\mathbb{F}_2$上(也就是来自$\mathbb{F}_2[X]$)的二次和三次不可约多项式如下。有三个在$\mathbb{F}_2$上的二次不可约多项式：X^2+1，X^2+X+2 和 X^2+2X+2。存在八个在$\mathbb{F}_2$上的三次不可约多项式：X^3+2X+2，X^3+X^2+2，X^3+X^2+X+2，X^3+2X^2+2X+2，X^3+2X+1，X^3+X^2+2X+2，X^3+X^2+1 和 X^3+2X^2+X+1。

循环码允许以多项式的方式进行编码和解码。这样很方便就能得到一个循环码$\mathscr{C}$的生成矩阵，形式与 Hamming[7，4]码的 $\boldsymbol{G}_{7,4}$ 类似(见上面)。也就是说，我们想要在$\mathscr{C}$找到能给形式如下的对应生成矩阵的基：

$$\boldsymbol{G}_{7,4}=\begin{bmatrix} \boxed{} & & & \\ & \boxed{} & & \mathbf{0} \\ & & \boxed{} & \\ & \mathbf{0} & & \boxed{} \end{bmatrix} \tag{2.5.11}$$

定理 2.5.10(iii)提供了这样一个基：选取生成多项式 $g(X)$ 和它的乘积：

$$g(X),\ Xg(X),\cdots,X^{k-1}g(X),\quad \deg g(X)=N-k$$

用符号写成

$$G_{\text{cycl}}=\begin{bmatrix} g(X) \\ Xg(X) \\ \vdots \\ X^{k-1}g(X) \end{bmatrix} \tag{2.5.12}$$

这个码的秩为 k，可以像下面这样解码长度为 k 的码字。给定一个码字 $a=a_0\cdots a_{k-1}$，产生多项式 $a(X)=\sum_{0\le i\le k-1} a_iX^i$ 并取乘积 $a(X)g(X)$。根据定理 2.5.9 可知它属于 $\mathscr{X}$，所以定义了一个码字。因此我们只需要存储多项式 $g(X)$：编码对应于多项式相乘。如果编码由乘积给出，那么解码一定与除法相关。回想在几何译码器的情况下，我们通过 Hamming 距离最近的码字来解码收到的码字。这样一个码字与对应陪集的引导段有关：我们已经看到陪集与形如 yH^T 的伴随码字有一一对应的关系。在循环码的情况中，伴随码字可以很直接地被计算。回想一下，如果 $g(X)$ 是一个循环码 $\mathscr{X}$ 的生成多项式，并且 $\deg g(x)=N-k$，那么 $\mathscr{X}$ 的秩为 k，必然有 2^{N-k} 个不同的陪集(见定理 2.5.10(v))。 222

定理 2.5.21 *陪集 $y+\mathscr{X}$ 与余子式 $y(X)=u(X)\bmod g(X)$ 一一对应。即两个码字 y，y' 属于同一个陪集，当且仅当在除法表示中，*

$$y(X)=a(X)g(X)+u(X),\ y'(X)=a'(X)g(X)+u'(X),\quad 且\ u(X)=u'(X)$$

证明 y 和 y' 属于同一陪集合当且仅当 $y+y'\in\mathscr{X}$。这等效于 $u(X)+u'(x)=0$，即由定理 2.5.14 可知 $u(X)=u'(X)$。 □

因此，陪集可以用 $\deg u(X)<\deg g(X)=N-k$ 的多项式 $u(X)$ 标记：这样就恰好就有 2^{N-k} 个这样的多项式。为了确定陪集 $y+\mathscr{X}$，只需计算余子式 $u(X)=y(X)\bmod g(X)$。遗憾的是，我们还有一个任务，就是在每个情况下找到一个引导段，对于一般的循环码还没有简单的算法用来找到引导段。然而，有一些已知的特定类型的循环码，它们采用一种相对简单的解码，第一种这样的类型发现于 1959 年，由 BCH 码组成(见 2.6 节)

像之前观察的那样，一个循环码不仅可以通过它的最小次数的多项式产生。为了某些目的，拥有这种性质的其他多项式可能也会有用。不管怎样，它们都是 $1+X^N$ 的除式。

定理 2.5.22 *令 $\mathscr{X}$ 是长度为 N 的二进制循环码，那么任何使得 $\mathscr{X}$ 是 $\tilde{g}(X)$ 的主理想的多项式 $\tilde{g}(X)$ 是 $1+X^N$ 的除式。*

证明 代数中的一个练习。 □

我们看到循环码自然而然被它们的生成多项式标记。

定义 2.5.23 *令 $\mathscr{X}$ 是由 $g(X)$ 产生的长度为 N 的二进制循环码。$\mathscr{X}$ 的校验多项式 $h(X)$ 定义为比率 $(1+X^N)/g(X)$，即 $h(X)$ 是 $h(X)g(X)=1-X^N$ 的唯一多项式。*

我们用标准记法 $\gcd(f(X),\ g(X))$ 来表示 $f(X)$ 和 $g(X)$ 的最大公倍数，$\operatorname{lcm}(f(X),\ g(X))$ 表示它们的最小公约数。记 $\mathscr{X}_1+\mathscr{X}_2$ 为两个线性码 $\mathscr{X}_1$，$\mathscr{X}_2\subset\mathscr{H}_{N,2}$ 的直和。即 $\mathscr{X}_1+\mathscr{X}_2$ 由线性组合 $\alpha_1a^{(1)}+\alpha_2a^{(2)}$ 构成，其中 α_i，$\alpha_i=0,1$，$a^{(i)}\in\mathscr{X}_i$，$i=1,2$。比较例子 2.1.8(vii)。 223

举例 2.5.24 *令 $\mathscr{X}_1$ 和 $\mathscr{X}_2$ 是两个长度为 N 的二进制循环码，生成多项式分别为 $g_1(X)$ 和 $g_2(X)$。证明：*

(a) *$\mathscr{X}_1\subset\mathscr{X}_2$ 当且仅当 $g_2(X)$ 整除 $g_1(X)$。*

(b) 交集$\mathscr{X}_1\cap\mathscr{X}_2$产生一个由$\mathrm{lcm}(g_1(X), g_2(X))$生成的循环码。

(c) 直和$\mathscr{X}_1+\mathscr{X}_2$是由$\gcd(g_1(X), g_2(X))$生成的一个循环码。

解答 (a)我们知道$a(X)\in\mathscr{X}_i$当且仅当在环$\mathbb{F}_2[X]/(1+X^N)$中，多项式$a(X)=f_i*g_i(X)$，$i=1, 2$。假设$g_2(X)$能整除$g_1(X)$并写成$g_1(X)=r(X)g_2(X)$。那么形如$f_1*g_1(X)$的每一个多项式$a(X)$具有$f_1*r*g_2(X)$的形式，也就是说，如果$a(X)\in\mathscr{X}_1$，那么$a(X)\in\mathscr{X}_2$。所以$\mathscr{X}_1\subset\mathscr{X}_2$。

相反地，假设$\mathscr{X}_1\subset\mathscr{X}_2$。令$d_i$为$g_i(X)$的次数，$1\leqslant d_i\leqslant N$，$i=1, 2$，并写成

$$g_1(X)=f(X)g_2(X)+r(X),\quad 其中\quad \deg r(X)<d_2$$

我们有每个在$\mathbb{F}_2[X]/(1+X^N)$中能被$g_1(X)*$整除的多项式同样也能被$g_2(X)*$整除。特别地，基本多项式$X^ig_1(X)$，$0\leqslant i\leqslant N-d_1-1$都能被$g_2(X)*$整除，即形式为

$$X^ig_1(X)=h^{(i)}(X)g_2(X)+a_i(X^N-1),\quad 其中\ a_i=0\ 或\ 1$$

如果对于某个i，系数$a_i=0$，那么我们比较两个恒等式

$$X^ig_1(X)=X^if(X)g_2(X)+X^ir(X)\quad 和\quad X^ig_1(X)=h^{(i)}(X)g_2(X)$$

并得出结论$X^ir(X)=0$。这意味着$r(X)=0$，因此$g_2(X)$能整除$g_1(X)$。

剩下的情况就是所有的系数$a_i=1$。于是我们比较

$$Xg_1(X)=Xh^{(0)}(X)g_2(X)+X+X^{d_1+1}$$

和

$$Xg_1(X)=h^{(1)}(X)g_2(X)+1+X^N$$

发现这种情况不可能。

(b) 这一部分很简单：交集$\mathscr{X}_1\cap\mathscr{X}_2$是$\mathscr{X}_1$和$\mathscr{X}_2$的子码，很显然它是个循环码；因此，由(a)可知，它的生成多项式$g(X)$能被$g_1(X)$和$g_2(X)$整除。所以它能被$\mathrm{lcm}(g_1(X), g_2(X))$整除。我们必须排除经过这次相除$g(X)$产生非平凡比率的情况。但是$\mathrm{lcm}(g_1(X), g_2(X))$本身就是一个包含在$\mathscr{X}_1$和$\mathscr{X}_2$中的循环码(跟原始码长度一样)的生成子。所以，当$g(X)\neq\mathrm{lcm}(g_1(X), g_2(X))$时，由$\mathrm{lcm}(g_1(X), g_2(X))$生成的码必然严格大于$\mathscr{X}_1\cap\mathscr{X}_2$。这与$\mathscr{X}_1\cap\mathscr{X}_2$的定义矛盾。

224

(c) 类似地，$\mathscr{X}_1+\mathscr{X}_2$是包含在$\mathscr{X}_1$和$\mathscr{X}_2$中的最小线性码。所以，它的生成子能整除$g_1(X)$和$g_2(X)$，即为它们的公因式。如果它不等于$\gcd(g_1(X), g_2(X))$，那么它与上面最小化性质相矛盾。

举例 2.5.25 令$\mathscr{X}$为长度为N、生成多项式为$g(X)$、校验多项式为$h(X)$的二进制循环码。证明$a(X)\in\mathscr{X}$当且仅当多项式$(1+X^N)$能整除$a(X)h(X)$，即在$\mathbb{F}_2[x]/(1+X^N)$中，$a*h(X)=0$。

解答 如果$a(X)\in\mathscr{X}$，那么对于某个多项式$f(X)\in\mathbb{F}_2[x]/(1+X^N)$有$a(X)=f(X)g(X)$。那么

$$a(X)h(X)=f(X)g(X)h(X)=f(X)(1+X^N)$$

在$\mathbb{F}_2[X]/(1+X^N)$等于0。相反地，令$a(X)\in\mathbb{F}_2[x]/(1+X^N)$并且假设$a(X)h(X)=0\bmod(1+X^N)$。记$a(X)=f(X)g(X)+r(X)$，其中$\deg r(X)<\deg g(X)$。那么

$$a(X)h(X)=f(X)(1+X^N)+r(X)h(X)=r(X)h(X)\bmod(1+X^N)$$

因此，只有当$r(X)=0$时，$r(X)h(X)=0\bmod(1+X^N)$才有可能。所以$a(X)=f(X)g(X)$并且$a(X)\in\mathscr{X}$。

举例 2.5.26 证明一个循环码的对偶也是循环的，并求出它的生成矩阵。

解答 如果$y\in\mathscr{X}^\perp$，即为对偶码，那么对于所有$x\in\mathscr{X}$的点乘$(xx*y)=0$。但是$(xx*y)=$

$(\boldsymbol{x}+\pi\boldsymbol{y})$，即 $\pi\boldsymbol{y}\in\mathscr{C}^{\perp}$，意味着$\mathscr{C}^{\perp}$也是循环的。

令 $g(X)=g_0+g_1X+\cdots+g_{N-k}X^{N-k}$ 为$\mathscr{C}$的生成多项式，其中 $N-k=d$ 是 $g(X)$的次数，k 是$\mathscr{C}$的秩。我们知道$\mathscr{C}$的生成矩阵 $\boldsymbol{G}$ 可以写成

$$\boldsymbol{G}\sim\begin{pmatrix} g(X) \\ Xg(X) \\ \vdots \\ X^{k-1}g(X) \end{pmatrix}\sim\begin{pmatrix} \boxed{} & & & \\ & \boxed{} & & \mathbf{0} \\ & & \boxed{} & \\ & \mathbf{0} & & \ddots \\ & & & \boxed{} \end{pmatrix} \tag{2.5.13}$$

225

选取 $h(X)=(1+X^N)/g(X)$并记 $h(X)=\sum_{j=0}^{k}h_jX^j$ 和 $\boldsymbol{h}=h_0\cdots h_{N-k}$。那么

$$\sum_{j=0}^{i}g_jh_{i-j}\begin{cases}=1, & i=0,N,\\ =0, & 1\leqslant i<N.\end{cases}$$

实际上，对于 $i=0$，N，我们有 $h_0g_0=1$ 以及 $h_kg_{N-k}=1$。对于 $1\leqslant i<N$，我们得到点积

$$(x^jg\cdot x^{j'}\boldsymbol{h}^{\perp})=0,\quad j=0,1,\cdots,N-k-1,\quad j'=0,\cdots,k-1$$

其中$\boldsymbol{h}^{\perp}=h_kh_{k-1}\cdots h_0$。那么我们很容易看出$\boldsymbol{h}^{\perp}$给出了$\mathscr{C}^{\perp}$的生成多项式$h^{\perp}(X)$。

另一种解法是基于举例 2.5.25 的。我们知道当且仅当 $a*h(X)=0$ 时 $a(X)\in\mathscr{C}$。令 k 为$g(X)$的次数，那么 $h(X)$的次数等于 $N-k$。次数 $\deg[a(X)h(X)]<2N-k$，所以 $a(X)h(X)$中 X^{N-k+1}，…，X^{N-1}的系数都为零。也就是说：

$$\begin{aligned} &a_0h_{N-k}+a_1h_{N-k-1}+\cdots+a_{N-k}h_0=0,\\ &a_1h_{N-k}+a_2h_{N-k-1}+\cdots+a_{N-k+1}h_0=0,\\ &\quad\vdots\qquad\qquad\qquad\qquad\vdots\\ &a_{k-1}h_{N-k}+a_kh_{N-k-1}+\cdots+a_{N-1}h_0=0 \end{aligned}$$

换句话说，$\boldsymbol{a}\boldsymbol{H}^{\mathrm{T}}=0$，其中 $\boldsymbol{a}=a_0$，…，a_{N-1}是 $a(X)$二进制系数的码字，$\boldsymbol{H}$ 是一个$(N-k)\times N$ 的矩阵

$$\boldsymbol{H}\sim\begin{pmatrix} h^{\perp}(X) \\ Xh^{\perp}(X) \\ \vdots \\ X^{N-k-1}h^{\perp}(X) \end{pmatrix}\sim\begin{pmatrix} \boxed{} & & & \\ & \boxed{} & & \mathbf{0} \\ & & \boxed{} & \\ & \mathbf{0} & & \ddots \\ & & & \boxed{} \end{pmatrix} \tag{2.5.14}$$

其中 $h^{\perp}(X)=X^{N-k}h(N^{-1})$，系数序列$\boldsymbol{h}^{\perp}=h_kh_{k-1}\cdots h_0$。

我们得出结论：矩阵 $\boldsymbol{H}$ 生成了码$\mathscr{C}'\subset\mathscr{C}^{\perp}$。但由于 $h_{N-k}=1$，$\mathscr{C}'$的秩等于 $N-k$，所以$\mathscr{C}'=\mathscr{C}^{\perp}$。

剩下就是验证多项式 $h^{\perp}(X)$能整除 $1+X^N$。为此，我们从 $g(X)h(X)=1+X^N$ 推出 $h(X^{-1})g(X^{-1})=X^{-N}+1$。所以 $h^{\perp}(X)X^kg(X^{-1})=1+X^N$。并且由于 $X^kg(X^{-1})$等于多项式 $g_k+g_{k-1}X+\cdots+g_0X^k$，即得到需要的结论。 □

举例 2.5.27 令$\mathscr{C}$为长度为 N、生成多项式为 $g(X)$的二进制循环码。 225

(a) 证明权为偶数的码字集合 $a\in\mathscr{C}$是循环码，并求出它的生成多项式。

(b) 证明$\mathscr{C}$包含一个权为奇数的码字当且仅当 $g(1)\neq0$，即码字 $\mathbf{1}\in\mathscr{C}$。

解答 (a) 如果码$\mathscr{C}$是偶数(即只包含偶数权重的码字)，那么每个多项式 $a(X)\in\mathscr{C}$有 $a(1)=\sum_{0\leqslant i\leqslant N-1}a_i=0$。所以，$a(X)$包含因子$(X-1)$。所以，生成多项式 $g(X)$有一个因子

$(X+1)$。反之亦然：如果$(X+1)$能整除$g(X)$，即$g(1)=0$，那么每个码字$\underline{a}\in\mathscr{X}$具有偶权重。

现在假设$\mathscr{X}$包含权重为奇数的一个码字，即$g(1)=1$；也就是$(1+X)$不能整除$g(X)$。令$\mathscr{X}^{\infty}$是在$\mathscr{X}$中由偶码字形成的子码。循环移位不会改变权重，所以$\mathscr{X}^{\infty}$是一个循环码。对于相应的多项式$a(X)$，与之前一样，我们有$(1+X)$能整除$a(X)$。所以，$\mathscr{X}^{\infty}$的生成多项式$g^{\infty}(X)$能被$(1+X)$整除，所以$g^{\infty}(X)=g(X)(X+1)$。

(b) 接下来证明$g(1)=1$当且仅当$\underline{1}\in\mathscr{X}$。对应的多项式是$1+\cdots+X^{N-1}$，它是在因式分解$1+X^N=(1+X)(1+\cdots+X^{N-1})$中$(1+X)$的补因子。所以，如果$g(1)=1$，即$g(X)$不包含因子$(1+X)$，那么$g(X)$一定是$1+\cdots+X^{N-1}$的一个除数。这表明$\underline{1}\in\mathscr{X}$。类似地可证明逆命题。 □

举例 2.5.28 令$\mathscr{X}$为长度为N、生成多项式为$g(X)$、校验多项式为$h(X)$的二进制循环码。

(a) 证明$\mathscr{X}$是自正交的，当且仅当$h^*(X)$能整除$g(X)$；并且是自对偶的，当且仅当$h^*(X)=g(X)$，其中$h^*(X)=h_k+h_{k-1}X+\cdots+h_0X^{k-1}$，$h(X)=h_0+\cdots+h_{k-1}X^{k-1}+h_kX^k$是校验多项式，并有$g(X)h(X)=1+X^N$。

(b) 令r是N的除数：$r|N$。一个二进制码$\mathscr{X}$叫作r-退化，如果每个码字$a\in\mathscr{X}$是一个级联$c\cdots c$，其中c是长度为r的序列。证明$\mathscr{X}$是r-退化当且仅当$h(X)$能整除$(1-X^r)$。

解答 (a) 自正交性意味着$\mathscr{X}\subset\mathscr{X}^{\perp}$，即对所有的$\boldsymbol{a}$，$\boldsymbol{b}\in\mathscr{X}$都有$(\boldsymbol{a},\boldsymbol{b})=0$。从举例2.5.26，我们知道$h^{\perp}(X)$是$\mathscr{X}^{\perp}$的生成多项式。然后，根据例2.5.26可得，$\mathscr{X}\subset\mathscr{X}^{\perp}$当且仅当$h^{\perp}(X)$能整除$g(X)$。

自对偶性意味着$\mathscr{X}=\mathscr{X}^{\perp}$即$h^{\perp}(X)=g(X)$。

(b) 对于$N=rs$，我们有因式分解

227

$$1-X^N=(1+X^r)(1+X^r+\cdots+X^{(s-1)r})$$

现在假设长度为N、生成多项式为$g(X)$的循环码$\mathscr{X}$是r-退化的。那么对于某个长度为$r-1$的序列$\tilde{c}$(其中$c=1\ \tilde{c}$)，码字$\boldsymbol{g}$具有形如$1\ \tilde{c}\ 1\ \tilde{c}\cdots 1\ \tilde{c}$的形式。令$\tilde{c}(X)$为对应于$\tilde{c}$的多项式(次数$\leq r-2$)。那么$g(X)$为

$$\begin{aligned}&1+X\,\tilde{c}(X)+X^r+X^{r+1}\,\tilde{c}(X)+\cdots+X^{(s-1)r}+X^{(s-1)r+1}\,\tilde{c}(X)\\&=(1+X^r+\cdots+X^{(s-1)r})[1+X\,\tilde{c}(X)]\end{aligned}$$

对于校验多项式$h(X)$我们有

$$\begin{aligned}h(X)&=(1+X^N)/((1+X^r+\cdots+X^{(s-1)r})[1+X\,\tilde{c}(X)])\\&=(1+X^r)/[1+X\,\tilde{c}(X)]\end{aligned}$$

即$h(X)$是$(1+X)^r$的一个除数。

相反地，令$h(X)|(1+X^r)$，其中$g(X)h(x)=1+X^r$，其中$g(X)=\sum_{0\leq i\leq r-1}c_iX^i$，$c_0=1$。选取$c=c_0\cdots c_{r-1}$，以相反的顺序重复上面的论证，我们得出结论：码字$\boldsymbol{g}$是级联$c\cdots c$。那么循环移位$\pi\boldsymbol{g}$是级联$c^{(1)}\cdots c^{(1)}$，其中$c^{(1)}=c_{r-1}c_1\cdots c_{r-2}(=\pi c$，$c$在$\{0,1\}^r$上的循环移位)。接下来的循环移位迭代$\pi^2\boldsymbol{g}$也是类似的步骤。所以，$\mathscr{X}$的基向量是$r$-退化的，所以整个$\mathscr{X}$也是这样。 □

在"标准"运算中，一个给定次数d的(实或虚的)多项式$p(X)$可以方便地通过它的根(或零点)a_1，…，a_d(通常为复数)来识别，依靠单项式分解：$p(X)=p_d\prod_{1\leq i\leq d}(X-a_i)$。在二进制运算(更宽泛地说，$q$进制运算)中，多项式的根仍然是极其有用的概念。在这

里，根能够帮助我们利用重要的可预测性来构建二进制循环码的生成多项式 $g(X)=\sum_{0\le i\le d} g_i X^i$。假设此时 $g(X)$ 的根 α_1，…，α_d 已知，并且表达式

$$g(X)=\prod_{1\le i\le d}(X-\alpha_i)$$

具有固定的含义(在有限域的框架内被提供)。即使不了解形式理论，我们也可以得到一些有用的结论。

第一个结论是：α_i 是 N 次单位根，因为它们应该在多项式 $1+X^N$ 的零点之中。所以，它们可以乘积或转置，即形成一个大小为 N 的 Abel 乘积群，可能是循环的。第二，在二进制运算中，如果 α 是 $g(X)$ 的一个零点，那么 α^2 亦然，因为 $g(X)^2=g(X^2)$。那么由 α^2 是零点推出 α^4 也是零点，以此类推。我们得出序列 α，α^2，…呈现循环：$\alpha^{2^d}=\alpha$(或 $\alpha^{2^d-1}=1$)，其中 d 是 $g(X)$ 的次数。也就是说，所有 N 次单位根可分解成不相交的类，形如 $\mathscr{C}=\{\alpha,\alpha^2,\cdots,\alpha^{2^c-1}\}$，大小为 c，$c=c(\mathscr{C})$ 是一个正整数(满足 2^c-1 除以 N)。记号 $\mathscr{C}(\alpha)$ 是很有益的，其中 $c=c(\alpha)$。相同类的成员被叫作彼此共轭。如果我们想获得根为 α 的一个生成多项式，那么所有单位根 $\alpha^i\in\mathscr{C}(\alpha)$ 的共轭根也为 $g(X)$ 的单位根。 228

因此，构造一个生成多项式 $g(X)$，我们必须从类 $\mathscr{C}$ "借来" 根，并对每个借来的单位根，我们把它们类的所有成员都集中在一起。然后，由于任何来自由 $g(X)$ 产生的循环码的多项式 $a(X)$ 是 $g(X)$ 的倍数(见定理 2.5.10(iv))，$g(X)$ 的根也是 $a(X)$ 的根。相反地，如果 $a(X)$ 具有 $g(X)$ 的根 α_i，那么 $a(X)$ 也在此码中。我们看到循环码可以方便地用单位根进行描述。

例子 2.5.29 (Hamming[7，4]码)回忆二进制 Hamming[7，4]码 $\mathscr{H}^{(7)}$ 的奇偶校验矩阵 $\boldsymbol{H}$ 是 3×7 的矩阵；它的列为所有长度为 3 的非零二进制码字；这些行的不同排序确定了不同的码。在这节的后面我们将解释任何给定长度为 2^l-1，以某个顺序(或某些顺序)写出来的非零二进制码字序列可以表示成一个单独元素 ω 的幂序列：ω^1，ω，ω^2，…，ω^{2^l-2}。产生这些幂的乘法法则具有一种特殊的类型(多项式的乘积对某个次数为 l 的不可约的多项式取余)。为了强调这一事实，我们在这一节用记号 * 表示这一乘法法则，用 ω^{*i} 替代 ω^i。不管怎样，对于 $l=3$，二进制非零 3-码字的一种正确排序(取自两种可能的排序)为

$$\boldsymbol{H}=\begin{pmatrix}0&0&1&0&1&1&1\\0&1&0&1&1&1&0\\1&0&0&1&0&1&1\end{pmatrix}\sim(\omega^{*3}\omega\,\omega^{*2}\omega^{*4}\omega^{*5}\omega^{*6}\omega^{*4})$$

因此，关于这种表示，等式 $a\boldsymbol{H}^T=0$，它确定了码字 $\boldsymbol{a}=a_1\cdots a^6$(或它的多项式 $a(X)=\sum_{1\le i\le 7}a_iX^i$)在 $\mathscr{H}^{(7)}$ 中，可以被重新写为

$$\sum_{0\le i\le 6}a_i\omega^{*i}=0,\quad 或\ a(*\omega)=0$$

换句话说，$a(X)\in\mathscr{H}^{(7)}$，当且仅当 ω 在乘法法则 * 下是 $a(X)$ 的一个根(在这种情况为次数 $\le$ 2 的二进制多项式对多项式 $1+X+X^3$ 取余的乘积)。

最后的证明可以用这种方式转述：Hamming[7，4]码等同于关于包含 ω 根的生成矩阵 $g(X)$ 的循环码；在这种情况中生成多项式 $g(X)=1+X+X^3$，满足 $g(*\omega)=\omega^{*3}+\omega+\omega^{*3}=0$。$\boldsymbol{H}^{(7)}$ 行的另一种排序以相同的方式与多项式 $1+X^2+X^3$ 相关联。 229

假定我们通过幂运算定义了适当的操作方式，可以看出 Hamming[7，4]码是被单根 ω 确定的。由于这个原因，我们可以称 ω 为这个码的定义根(或定义零)。这就是为什么称元

素 α 为主元的原因；参考3.1~3.3节。

举例2.5.30 当 $a=a_0a_1\cdots a_{N-1}\in\mathscr{Z}$ 时意味着 $\overleftarrow{a}=a_{N-1}\cdots a_1a_0\in\mathscr{Z}$，则码 $\mathscr{Z}$ 被称作是可逆的。证明具有生成多项式 $g(X)$ 的循环码是可逆的当且仅当 $g(\alpha)=0$，意味着 $g(\alpha^{-1})=0$。

解答 对于生成多项式 $g(X)=\sum_{0\leqslant i\leqslant d}g_iX^i$，有 $\deg g(X)=d<N$ 和 $g_0=g_d=1$，逆多项式是 $g^m(X)=X^{N-1}g(X^{-1})$，因此如果循环码 $\mathscr{Z}$ 是可逆的，且 α 是 $g(X)$ 的根，那么 α 是 $g^m(X)$ 的根，当且仅当 $g(\alpha^{-1})=0$ 这才是可能的。

相反，令 $g(X)$ 满足性质 $g(\alpha)=0$ 意味着 $g(\alpha^{-1})=0$。上面的公式对于所有多项式的次数 $<N$ 的 $a(X)$ 都成立：$a^m(X)=X^{N-1}a(X^{-1})$。如果 $a(X)\in\mathscr{Z}$ 可以得到对于 $g(X)$ 的所有根 α 有 $a(\alpha)=a(\alpha^{-1})=0$。那么对于 $g(X)$ 的所有根 α 有 $a^m(\alpha)=a^m(\alpha^{-1})=0$。因此，$a^m(X)$ 是 $g(X)$ 的倍数，且 $a^m(\mathscr{Z})\in\mathscr{Z}$。 □

研究多项式根的自然架构是基于有限域理论或 Galois 理论的（我们已经看到过多项式环域是如何被使用的）。在这节的剩余部分我们给出一个简短的关于 Galois 理论的介绍，以更好理解前面介绍过的编码的例子。在第3章我们将会更深入地探讨 Galois 理论，以获得足够的知识来进行码字的构造。

备注2.5.31 域是一个交换环，其中每个非零元素都是可逆的。就是说，当乘法产生了一个群时环就是域。事实上，一个域的非零元素的乘法群是循环的。

定理2.5.32 令 $g(X)\in\mathbb{F}_2[X]$ 是一个次数为 d 的不可约的二元多项式，那么，模 $g(X)$ 乘在具有 2^d 个元素的域上产生一个次数 $\leqslant d-1$（也就是空间 $\mathbb{F}_2^{(d)}$）的二进制多项式集合。相反，如果模 $g(X)$ 乘产生一个域，则 $g(X)$ 是不可约的。

证明 唯一一个需验证的重要性质是逆元素的存在性。取非零多项式 $f(X)$，且有 $\deg f(X)\leqslant d-1$，考虑形如 $f(X)h(X)$（一般的乘法）的所有多项式，其中 $h(X)$ 遍历所有次数 $\leqslant d-1$ 的多项式集合。这些结果一定是不同的模 $g(X)$。实际上，如果

230

$$f(X)h_1(X)=f(X)h_2(X)\bmod g(X)$$

那么，对于一些次数 $\leqslant d-2$ 的多项式 $v(X)$，有

$$f(X)(h_1(X)-h_2(X))=v(X)g(X) \tag{2.5.15}$$

这就意味着或者是 $g(X)\mid f(X)$，或者是 $g(X)\mid h_1(X)-h_2(X)$。我们得出如果多项式 $g(X)$ 是不可约的，除非 $h_1(X)=h_2(X)$ 且 $v(X)=0$，否则式(2.5.15)是不可能的。对于一个且只有一个多项式 $h(X)$，我们可得

$$f(X)h(X)=1\bmod g(X)$$

$h(X)$ 表示 $f(X)$ 取模 $g(X)$ 乘的逆。我们记为 $h(X)=f(X)^{*-1}$。

另一方面，如果 $g(X)$ 是可约的，有 $g(X)=b(X)b'(X)$，其中 $b(X)$ 和 $b'(X)$ 都是非零的，并且次数 $<d$，即 $b(X)b'(X)=0\bmod g(X)$。如果模 g 乘生成一个域，$b(X)$ 和 $b'(X)$ 都会有逆，即 $b(X)^{*-1}$，$b'(X)^{*-1}$。然而

$$b(X)^{*-1}*b(X)*b'(X)=b'(X)=0$$

和相似的 $b(X)=0$。 □

通过以上构造方法得到的域被称为多项式域，通常表示为 $\mathscr{P}_2[X]/(g(X))$。它包含了 2^d 个元素，其中 $d=\deg g(X)$（代表次数 $<d$ 的多项式）。我们称 $g(X)$ 为域的核多项式。在这节的剩余部分，我们将在一个给定多项式域的乘法用 $*$ 号表示。零多项式和1多项式分别用 **0**，**1** 表示；它们明显是多项式域的零和单位1。如下的结果起到关键作用。

定理 2.5.33 (a) 在多项式域$\mathbb{F}_2[X]/\langle g(X)\rangle$的非零元素的乘法群对于大小为$2^d-1$的循环群$\mathbb{Z}_{2^d-1}$是同构的。

(b) 通过挑选次数为d的不同的不可约多项式来获得的多项式域都是同构的。

证明 在这里我们只证明结论(a)。结论(b)将会在3.1节中证明。从域中取出任意一个元素，$a(X)\in\mathbb{F}_2[X]/\langle g(X)\rangle$，可以观察到

$$a^{*i}(X)_i=\underbrace{a*\cdots*a}_{i次}(X)$$

(域中的乘法)取最多2^d-1值(域中的元素数少一个，因为**0**是被排除的)。因此存在一个正整数r使得$a^{*r}(X)=\mathbf{1}$；r的最小值被称为$a(X)$的阶数。 231

选择一个最大阶数为r的多项式$a(X)\in\mathbb{F}_2[X]/\langle g(X)\rangle$。那么我们称任何一个其他元素$b(X)$的阶数能除尽$r$。事实上，令$s$为$b(X)$的阶数。取$s$的素数因子$p$，然后写为

$$s=p^{c'}l' \quad 且 \quad r=p^{c}l$$

满足整数c'，$c\geq 0$，且l，$l'\geq 1$，其中l，l'是不可被p整除的。我们可以得到$c\geq c'$。事实上，元素$a^{*l'}(X)$的阶数是l，$b^{*l'}(X)$的阶数是$p^{c'}$，且乘积$a^{*l'}*b^{*l'}(X)$的阶数是$lp^{c'}$。因此，$c'\leq c$或者别的r都不是最大的。这对于任何的素数p都是正确的，因此，s能够整除r。 □

因此，对于r是最大的阶数，域中的每一个元素$b(X)$满足$b^{*r}(X)=\mathbf{1}$。通过使用pigeon-hole原理，我们可以得到$r=2^d-1$，这是域中非零元素的数目。因此，$a(X)$是阶数r的一个元素，幂级数$\mathbf{1}$，$a(X)$，…，$a^{*(2^d-1)}(X)$彻底讨论了域的乘法群。

在定理2.5.33的证明中，我们使用符号$\mathbb{F}_{2^d}$表示任意多项式域$\mathbb{F}_2[X]/\langle g(X)\rangle$，其中$g(X)$是一个次数为$d$的不可约二元多项式。更进一步，在$\mathbb{F}_{2^d}$中的非零元素乘法群用$\mathbb{F}_{2^d}^*$表示；它是循环的(根据定理2.5.33，有$\simeq\mathbb{Z}_{2^d-1}$)。任意群$\mathbb{F}_{2^d}^*$的生成子(其中$*$幂穷尽$\mathbb{F}_{2^d}^*$)被称为域$\mathbb{F}_{2^d}$的主元。

例子 2.5.34 我们能看到写出的全部不可约多项式列表的重要性。有6个次数为3的不可约二元多项式(其中每个都是主元多项式)：

$$\begin{aligned}&1+X^2+X^3,1+X^2+X^3,1+X+X^2+X^3+X^3\\&1+X+X^2+X^3+X^3,1+X+X^3+X^4+X^3\\&1+X^2+X^3+X^4+X^3\end{aligned}\tag{2.5.16}$$

并且有9个次数为4的(其中六个是主元多项式)：

$$\begin{aligned}&1+X+X^4,1+X+X^2+X^3+X^4,1+X^3+X^4\\&1+X+X^2+X^3+X^4,1+X^2+X^3+X^4+X^4,1+X+X^3+X^2+X^4\\&1+X+X^2+X^3+X^4,1+X^2+X^3+X^4+X^4,1+X^3+X^4\end{aligned}\tag{2.5.17}$$

不可约多项式的数目会随着次数的增加显著增加：次数为7时有18个，次数为8时有30个，等等。然而，在各种有限域中，存在有大量可用的不可约多项式。 232

例子 2.5.35 (a) 域$\mathbb{F}_2[X]/\langle 1+X+X^2\rangle$有四个元素：**0**，**1**，$X$，$1+X$，满足乘法表：

$$X*X=1+X,\quad 当\ X^2=1+X\bmod(1+X+X^2)$$

$$X*(1+X)=X+X*X=1$$

$$(1-X)*(1+X)=1-X+X+X*X=1+1-X=X$$

因为$X^{*2}(1+X)*X=\mathbf{1}$，群与$\mathbb{Z}_3$是同构的。对于这个域的一个备选符号是$\mathbb{F}_4$。

(b) 域$\mathbb{F}_2[X]/\langle 1+X+X^3\rangle$ $\mathbb{F}_2[X]/\langle 1+X^2+X^3\rangle$每个包含了8个元素，代表了次数$\leq 2$的所有多项式。每个这样的多项式$a_0+a_1X+a_2X^2$都是通过它的系数序列$a_0a_1a_2$(一个二

进制字)来定义。域表都是通过查看后面的幂 X^{i} 来得到，见图 2-6。

$1+X+X^3$			$1+X^2+X^3$		
X^{i}	多项式	码	X^{i}	多项式	码
—	0	000	—	0	000
$X^{i=1}$	1	100	$X^{i=0}$	1	100
X	X	010	X	X	010
$X^{i=2}$	X^2	001	$X^{i=2}$	X^2	001
$X^{i=3}$	$1+X$	110	$X^{i=3}$	$1+X^2$	101
$X^{i=4}$	$X+X^2$	011	$X^{i=4}$	$1+X+X^2$	111
$X^{i=5}$	$1+X+X^2$	111	$X^{i=5}$	$1+X$	110
$X^{i=6}$	$1+X^2$	101	$X^{i=6}$	$X+X^2$	011

图 2-6

在两种情况下，非零元素的乘法群是 $\mathbb{Z}_7$，这两个域明显是同构的，因为它们共享同样的乘法循环群形式。这些域的共同符号是 $\mathbb{F}_8$。注意的是两个域表对于 $0\leqslant i<3$ 的幂 X^{i} 是一致的；事实上，这是一个一般的模式，见 3.1～3.3 节。

此外，元素 $X=X^{i=1}\in\mathbb{F}_2[X]/\langle 1+X+X^3\rangle$ 可以被认为是核多项式 $1+X+X^3$ 的一个根，且元素 $X=X^{i=1}\in\mathbb{F}_2[X]/\langle 1+X^2+X^3\rangle$ 被认为是 $1+X^2+X^3$ 的一个根，因为这些多项式在它们各自的域里产生零。剩余的两个根是 $X^{i=2}$，$X^{i=4}$(同样在它们各自的域里被计算得出)。

将这个例子应用到 Hamming[7，4]码中(参考例子 2.5.29)，域 $\mathbb{F}_2[X]/\langle 1+X+X^3\rangle$ 产生生成多项式 $1+X+X^3$ 的根，域 $\mathbb{F}_2[X]/\langle 1+X^2+X^3\rangle$ 产生生成多项式 $1+X^2+X^3$ 的根。即，如果在两个同构域 $\mathbb{F}_2[X]/\langle 1+X+X^3\rangle$ 中的任意一个有定义根 $\omega=X$，Hamming[7，4]码等效于长度为 7 的循环码。不精确地说(该说法将会在 3.1 节中修正)，这个码是被它的根 ω 定义的，这个根是 $\mathbb{F}_8$ 的素数元素。

235

(c) 域 $\mathbb{F}_2[X]/\langle 1+X+X^4\rangle$ 包含了 16 个元素。域表如图 2-7 所示。在这种情况下，乘法群是 $\mathbb{Z}_{15}$。域可以被 $\mathbb{F}_{16}$ 表示。如上述，元素 $X\in\mathbb{F}_2[X]/\langle 1+X+X^4\rangle$ 产生了多项式 $1+X+X^4$ 的一个根；其他的根是 $X^{i=2}$，$X^{i=4}$，$X^{i=8}$。

X^{i}	多项式	系数字符串
—	0	0000
$X^{i=0}$	1	1000
X	X	0100
$X^{i=2}$	X^2	0010
$X^{i=3}$	X^3	0001
$X^{i=4}$	$1+X$	1100
$X^{i=5}$	$X+X^2$	0110
$X^{i=6}$	X^2+X^3	0011
$X^{i=7}$	$1+X+X^3$	1101
$X^{i=8}$	$1+X^2$	1010
$X^{i=9}$	$X+X^3$	0101
$X^{i=10}$	$1+X+X^2$	1110
$X^{i=11}$	$X+X^2+X^3$	0111
$X^{i=12}$	$1+X+X^2+X^3$	1111
$X^{i=13}$	$1+X^2+X^3$	1011
$X^{i=14}$	$1+X^3$	1001

图 2-7

这个例子可以被把 Hamming[15，11]码视为生成多项式 $g(X)=1+X+X^4$ 的循环码。我们现在可以说 Hamming[15，11]码是(模等价)长度为 15 的循环码，其中域 $\mathbb{F}_2[X]/\langle 1+X+X^4\rangle$ 中的定义根为 $\omega(=X)$。因为 X 是域中乘法群的生成子，我们同样可以说定义根 ω 是 $\mathbb{F}_{16}$ 的素数元素。

通常情况下，取域 $\mathbb{F}_2[X]/\langle g(X)\rangle$，其中 $g(X)=\sum_{0\leqslant i\leqslant m}g_iX^i$ 是一个次数为 m 的不可约二进制多项式，那么元素 X，$X^{i=2}$，$X^{i=4}$，…，$X^{i=2^{m-1}}$ 将满足等式

$$\sum_{0\leqslant i\leqslant m}g_i(X^{i=s})^{i}=0,\quad s=1,2,\cdots,2^{m-1}$$

也就是说，在不可约多项式 g 的域 $\mathbb{F}_2[X]/\langle g(X)\rangle$ 中，X，$X^{i=2}$，$X^{i=4}$，…，$X^{i=2^{m-1}}$ 严格为零。

234

从例子 2.5.35 中体现出的另一个特征是在所有部分(a)-(c)中，元素 X 代表核多项式 $g(X)$ 的根。然而，在通常意义上这是不正确的。只有当 $g(X)$ 为"主"二元多项式时才成

立；对于这个性质的详细讨论见3.1—3.3节。此外，对于一个主核多项式$g(X)$我们有对于$i<d-\deg g(X)$的幂X^i与X^{i+1}是相同的。而更多的幂X^{i+1}，$m\leqslant i\leqslant 2^\ell-1$，能相对容易地计算出来。继续这个思想，我们就可以讨论一般的二进制Hamming码。

例子2.5.36 令$\mathscr{H}^H$是二元$[2^\ell-1,\ 2^\ell-1-\ell]$Hamming码。我们知道它的奇偶校验矩阵$\boldsymbol{H}$由所有长度为$\ell$的非零列向量来表征。这些以特别的顺序写出的向量，列出了在域$\mathbb{F}_2[X]/(g(X))$中的连续幂ω^{*i}，$i=0,1,\cdots,2^\ell-2$。其中$\omega=X$且$g(X)=g_0+g_1X+\cdots+g_\ell X^\ell=X^\ell$是次数为$\ell$的本原多项式。因此

$$\boldsymbol{H}=\begin{pmatrix}1 & 0 & \cdots & 0 & g_0 & \cdots\\ 0 & 1 & \cdots & 0 & g_1 & \cdots\\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots\\ 0 & 0 & \cdots & 0 & g_{\ell-1} & \cdots\\ 0 & 0 & \cdots & 1 & 0 & \cdots\end{pmatrix} \tag{2.5.18}$$

或$\boldsymbol{H}=(1\ \omega\cdots\omega^{*(\ell-1)}\ \omega^{*\ell}\cdots\omega^{*(2^\ell-2)})$。

因此，如前所述，码字的等式$a\boldsymbol{H}^{\mathrm{T}}=\boldsymbol{0}$等效于对应的多项式$a(*\omega)=\boldsymbol{0}$。因此，我们可以说$a(X)\in\mathscr{H}^H$当且仅当$\omega$是$a(X)$中的一个根。

另一方面，通过构造，ω是$g(X)$的一个根；$g(*\omega)=0$。因此，我们把Hamming$[2^\ell-1,\ 2^\ell-1-\ell]$码与由生成多项式$g(X)$产生的长度为$2^\ell-1$的循环码视为等效，其中$g(X)$有定义根$\omega$。根$\omega$的作用可以由任意其他元素从$\{\omega,\ \omega^{*2},\ \cdots,\ \omega^{*(2^\ell-1)}\}$中体现出来。

上面的思想产生了一个直接的(且深远的)归纳。取$N=2^\ell-1$且令ω为域$\mathbb{F}_{2^\ell}\simeq\mathbb{F}_2[X]/(g(X))$的一个素数元素，其中$g(X)$是一个本原多项式。(在这章所有的例子和问题中，这个条件都是满足的。)考虑根的一个定义集合，以ω，ω^2，ω^3的形式开始，但是更普遍的是ω，ω^3，$\cdots$，$\omega^{(\delta-1)}$。(其中参数δ通常是个整数且大于3。)考虑具有这些根的循环码：我们有什么想法吗？长度为$N=2^\ell-1$，我们可以猜测它可以产生Hamming$[2^\ell-1,\ 2^\ell-1-\ell]$码的子码，可以纠正一个以上的错误。这就是所谓的(二元)BCH码构造的主旨(Bose-Choudhury，Hocquenguem，1959)。

在这节中我们只是对BCH码做了一个简短的介绍；这些码字更详细和更具普适性的介绍将在3.2节中讨论。对于$N=2^\ell-1$，域$\mathbb{F}_{2^\ell}\simeq\mathbb{F}_2[X]/(g(X))$具有如下性质：它的非零元素是单位根的第$N$个根(即多项式$1+X^N$的零点)。也就是说，多项式$1+X^N$可因式分解为线性因子$\prod_{1\leqslant i\leqslant N}(X-\omega_i)$的乘积，其中所有$\omega_i$列出了整个$\mathbb{F}_{2^\ell}^*$。(在3.1节中的说法中，$\mathbb{F}_{2^\ell}$是$1+X^N$在$\mathbb{F}_2$上的分裂域。)这里我们使用符号$\omega_1=X$表示乘法循环群$\mathbb{F}_{2^\ell}^*$的生成器。(事实上它可以是这个群中的任意生成器。) 235

因为$\omega^N=1$，且根据性质指数N是最小的，因此元素ω通常被称为第N个本原单位根。因此，当$0\leqslant k<N$时功率ω^k产生了这个域的不同元素。当我们推断出如下乘积结果时，会使用这个事实

$$\prod_{1\leqslant i<j\leqslant\delta-1}(\omega^{k_i}-\omega^{k_j})\neq 0$$

上述乘积满足每一个功率ω^{k_1}，$\cdots$，$\omega^{k_{\delta-1}}$的集合。(这样的集合是从定理2.5.39的证明中$(\delta-1)\times N$奇偶校验矩阵生成的$(\delta-1)\times(\delta-1)$子阵。)

定义2.5.37 给定$N=2^\ell-1$和$\delta=3,\cdots,N$，定义一个长度为N，设计距离为δ的狭义二元BCH码$\mathscr{H}_{N,\delta}^{\mathrm{BCH}}$，即由阶数小于$N$的二元多项式$a(X)$形成循环码，有

$$a(\omega)=a(\omega^2)=\cdots=a(\omega^{\delta-2})=0 \tag{2.5.19}$$

也就是说，$\mathscr{C}_{N,\delta}^{BCH}$是长度为$N$的循环码，它的生成器$g(X)$是包含根$\omega$，$\omega^2$，…，$\omega^{\delta-1}$的最小二元多项式：

$$\begin{aligned} g(X)&=\operatorname{lcm}\{(X-\omega),\cdots,(X-\omega^{\delta-1})\}\\ &=\operatorname{lcm}\{M_\omega(X),\cdots,M_{\omega^{\delta-1}}(X)\} \end{aligned} \tag{2.5.20}$$

在这里lcm代表着最小公倍数且$M_\alpha(X)$代表着根为α的最小二元多项式。为了简化我们将在这章中使用术语二元BCH码。(在3.2节中将介绍更具普适性的BCH码类。)

例子2.5.38 当$N=7$，恰当的多项式域是$\mathbb{F}_2[X]/\langle 1+X+X^3\rangle$或者$\mathbb{F}_2[X]/\langle 1+X^2+X^3\rangle$，即域$\mathbb{F}_8$两个实现中的一个，因为7是一个素数，这个域中的任意非零多项式都有乘法阶7，即非零多项式是$\mathbb{F}_2[X]/\langle 1+X^2+X^3\rangle$中乘法群的生成器。事实上，我们可以将多项式$1+X^7$分解为不可约因子乘积：

$$1+X^7=(1+X)(1+X+X^3)(1+X^2+X^3)$$

如果我们选择多项式域$\mathbb{F}_2[X]/\langle 1+X+X^3\rangle$，则$\omega=X$满足

236

$$\omega^3=1+\omega,(\omega^2)^3=1+\omega^2,(\omega^4)^3=1+\omega^4$$

也即共轭项ω，ω^2，ω^4是核心多项式$1+X+X^3$的根：

$$1+X+X^3=(X-\omega)(X-\omega^2)(X-\omega^4)$$

下一步有ω^3，ω^6，$\omega^{12}=\omega^5$是$1+X^2+X^3$的根：

$$1+X^2+X^3=(X-\omega^3)(X-\omega^5)(X-\omega^6)$$

因此，长为7，设计距离为3的二元BCH码是由阶数$\leqslant 6$的二元多项式$a(X)$生成的，即

$$a(\omega)=a(\omega^2)=0,\text{其中 } a(X)\text{ 是 } 1+X+X^3\text{ 的倍数}$$

这个码与Hamming[4，7]码是相等的，尤其是它的“真实”距离等于3。

长度为7，设计距离为4的二元BCH码是由阶数$\leqslant 6$的二元多项式$a(X)$生成的，即$a(\omega)=a(\omega^3)=a(\omega^5)=0$，其中$a(X)$就是下式的倍数

$$(1+X+X^3)(1+X^2+X^3)=1+X+X^2+X^3+X^4+X^5+X^6$$

这只是一个简单的重复码$\mathscr{R}_7$。

BCH码理论的主要部分是

定理2.5.39 (BCH界)设计距离为δ的二元BCH码的最小距离大于等于δ。

定理2.5.39(有时被称为BCH定理)的证明基于如下结果。

引理2.5.40 考虑$m\times m$的Vandermonde行列式，行列式由交换环中的项构成：

$$\det\begin{pmatrix} \alpha_1 & \alpha_2 & \cdots & \alpha_m\\ \alpha_1^2 & \alpha_2^2 & \cdots & \alpha_m^2\\ \vdots & \vdots & \ddots & \vdots\\ \alpha_1^m & \alpha_2^m & \cdots & \alpha_m^m \end{pmatrix}=\det\begin{pmatrix} \alpha_1 & \alpha_1^2 & \cdots & \alpha_1^m\\ \alpha_2 & \alpha_2^2 & \cdots & \alpha_2^m\\ \vdots & \vdots & \ddots & \vdots\\ \alpha_m & \alpha_m^2 & \cdots & \alpha_m^m \end{pmatrix} \tag{2.5.21}$$

这个行列式的值是

$$\prod_{1\leqslant i\leqslant m}\alpha_i\times\prod_{1\leqslant i<j\leqslant m}(\alpha_i-\alpha_j) \tag{2.5.22}$$

引理2.5.40的证明 式(2.5.21)的行列式都以α_1，…，α_m多项式的形式表达。当$i<j$如果有$\alpha=\alpha_j$，则行列式有重复的行(或列)，并且因此可以删去(正如在标准的算法中)。因

237

此，行列式分解为乘积$\prod_{1\leqslant i<j\leqslant m}(\alpha_i-\alpha_j)$。然后我们比较式(2.5.21)和式(2.5.22)中α_i的功率，这会立即得出引理2.5.40。 □

定理 2.5.39 的证明　令多项式 $a(X)\in\mathscr{Z}$，对于所有的 $j=1,\cdots,\delta-1$ 有 $a(*\omega^{*j})=0$，也即，

$$\begin{pmatrix} 1 & \omega & \omega^{*2} & \cdots & \omega^{*(N-1)} \\ 1 & \omega^{*2} & \omega^{*4} & \cdots & \omega^{*2(N-1)} \\ \vdots & & \vdots & \ddots & \vdots \\ 1 & \omega^{*(\delta-1)} & \omega^{*2(\delta-1)} & \cdots & \omega^{*(N-1)(\delta-1)} \end{pmatrix}\begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_{N-1} \end{pmatrix}=\mathbf{0}$$

根据引理 2.5.40，$((\delta-1)\times N)$ 矩阵的任意 $(\delta-1)$ 列都是线性独立的。因此，在 $a(X)$ 中至少有 δ 个非零系数。因此，$\mathscr{Z}$ 的距离 $\geqslant\delta$。 □

例子 2.5.41　(在文献[18]106 页中的一个错误已被纠正)考虑一个 $N=15$，$\delta=5$ 的 BCH 码。使用以下分解成不可约多项式：

$$X^{15}-1=(X+1)(X^2+X+1)(X^4+X+1)(X^4+X^3+1)$$
$$\times(X^4+X^3+X^2+X+1)$$

码的生成多项式是

$$g(x)=(X^4+X+1)(X^4+X^3+X^2+X+1)=X^8+X^7+X^6+X^4+1$$

事实上，$g(\omega^3)=g(\omega^5)=0$。$X^4+X^3+X^2+X+1$ 的零点集合是$(\omega^3,\omega^6,\omega^{12},\omega^9)$。$X^4+X+1$ 的零点集合是$(\omega,\omega^2,\omega^4,\omega^8)$。$X^4+X^3+1$ 的零点集合是$(\omega^7,\omega^{14},\omega^{13},\omega^{11})$。$X^2+X+1$ 的零点集合是(ω^5,ω^{10})。

(b) 令 $N=31$，ω 是F_{32}的本原元。根为 ω 的最小多项式是

$$M_\omega(X)=(X-\omega)(X-\omega^2)(X-\omega^4)(X-\omega^8)(X-\omega^{16})$$

我们同样可以得到根为 ω^3 的最小多项式：

$$M_{\omega^3}(X)=(X-\omega^3)(X-\omega^{30})(X-\omega^{27})(X-\omega^9)(X-\omega^{18})$$

利用定义可得长度为 31，设计距离为 $\delta=8$ 的 BCH 码的生成器是 $g(X)=\mathrm{lcm}(M_\omega(X),M_{\omega^3}(X),M_{\omega^5}(X),M_{\omega^7}(X))$。事实上，BCH 码(很显然长度至少为 9)的最小距离至少为 11。这可从定理 2.5.39 中得到，因为所有的功率 $\omega,\omega^2,\cdots,\omega^{10}$ 都可从 $g(X)$ 的根中得到。 238

对于 BCH 码来说有一个简单的解码实现：它可以归纳为 Hamming 码解码过程。根据定理 2.5.39 知，设计距离为 δ 的 BCH 码至少可以纠正 $t=\left\lfloor\frac{\delta-1}{2}\right\rfloor$ 个错误。假定发送一个码字 $c=c_0\cdots c_{N-1}$，接收到存在错误的码字为 $r=c+e$，其中 $e=e_0\cdots e_{N-1}$。假定 e 最多有 t 个非零项。引入所有阶数$<N$ 的相关多项式 $c(X)$，$r(X)$，$e(X)$。对于 $c(X)$ 有 $c(\omega)=c(\omega^2)=\cdots=c(\omega^{(\delta-1)})=0$。显然

$$r(\omega)=e(\omega),r(\omega^2)=e(\omega^2),\cdots,r(\omega^{(\delta-1)})=e(\omega^{(\delta-1)}) \tag{2.5.23}$$

因此，对于 $i=1,\cdots,\delta-1$ 我们可以计算相应的 $r(\omega^i)$。如果结果都是 0，则 $r(X)\in\mathscr{Z}$(则无错误或最少有 $t+1$ 个错误)。否则，令 $E=\{i:e_i=1\}$ 代表着错误的位数，且假定 $0<\#E\leqslant t$。引入错误位置多项式

$$\sigma(X)=\prod_{i\in E}(1-\omega^iX) \tag{2.5.24}$$

它有二进制系数，阶数为 $\#E$ 且最小系数为 1。如果知道 $\sigma(X)$，我们可以找到它的根为功率 ω^{-i}，因此能找到错误位数 $i\in E$。然后我们可以简单地改变这些位数就能纠正差错。

为了计算 $\sigma(X)$，考虑如下形式幂级数

$$\zeta(X)=\sum_{i\geqslant1}e(\omega^i)X^i$$

(观察到因为 $\omega^N=1$，故这个幂级数的系数是循环。)对于最初的 $(\delta-1)$ 个系数，借助式(2.5.23)特点，我们有等式

$$e(\omega^j)=r(\omega^j),j=1,\cdots,\delta-1$$

只有这些是我们需要的，然后依据接收的字 r 来计算它们。

现在令

$$\omega(X)=\sum_{i\in E}\omega^i X\prod_{j\in E,j\neq i}(1-\omega^j X) \tag{2.5.25}$$

然后将上面形式级数重写为

$$\zeta(X)=\sum_{j\geq 1}\sum_{i\in E}\omega^{ij}X^j=\sum_{i\in E}\sum_{j\geq 1}\omega^{ij}X^j=\sum_{i\in E}\frac{\omega^i X}{1-\omega^i X}=\frac{\omega(X)}{\sigma(X)} \tag{2.5.26}$$

观察到多项式 $\omega(X)$，$\sigma(X)$ 阶数都是 $\#E\leq t$。

根据下式以及系数

$$e(\omega^j)=r(\omega^j),j=1,\cdots,2t$$

[239]

从式(2.5.26)中得到的等式 $\zeta(X)\sigma(X)=\omega(X)$ 可以写为

$$\begin{aligned}&(\sigma_0+\sigma_1X+\cdots+\sigma_tX^t)\times(r(\omega)X+r(\omega^2)X^2+r(\omega^{(2t-1)})X^{2t-1}+\cdots)\\&=\omega_1-\omega_2X+\cdots-\omega_tX^t\end{aligned} \tag{2.5.27}$$

我们感兴趣的是 X^k 的系数，其中 $t<k\leq 2t$，它们满足

$$\sum_{0\leq j\leq t}\sigma_j r(\omega^{(k-j)})=0 \tag{2.5.28}$$

且没有涉及 $e(\omega^j)$ 中的任何项。我们得到下面的等式

$$\begin{pmatrix}r(\omega^{(t+1)}) & r(\omega^t) & \cdots & r(\omega)\\ r(\omega^{(t+2)}) & r(\omega^{(t-1)}) & \cdots & r(\omega^2)\\ \vdots & \vdots & \ddots & \vdots\\ r(\omega^{(2t)}) & r(\omega^{(2t-1)}) & \cdots & r(\omega^t)\end{pmatrix}\begin{pmatrix}\sigma_t\\ \sigma_t\\ \vdots\\ \sigma_t\end{pmatrix}=\mathbf{0}$$

上面的矩阵是一个 $t\times(t+1)$ 的矩阵，故其核中总存在一个非零向量。该向量标识了错误定位多项式 $\sigma(X)$。可以看出，上述的过程(叫作 Berlekamp-Massey 解码算法)确保我们能够明确指定集合 E 并纠正小于等于 t 个错误。

不幸的是，BCH 码是渐近“坏”的：对任意 BCH 码序列，随着其长度 $N\to\infty$，k/N 或 d/N 趋于 0。换言之，它们位于图 2-2 的底部。为了获得满足 Gilbert-Varshamov(GV)界的编码，需要使用基于代数几何的更有效的方式。这种编码早在 20 世纪 70 年代构造出来(Goppa 和 Justesen 编码)，但仍然存在着构造一个位于 Gilbert-Varshamov 曲线之上编码的问题。正如之前提到的，Tsfasman、Vladut 和 Zink 在 1982 年发明了一类新的编码，当码表中的符号数足够大时，这些编码位于 GV 曲线上。然而，对于二元编码，问题仍然存在。

举例 2.5.42 对于生成多项式 $g(X)=X^3+X+1$ 和奇偶校验多项式 $h(X)=X^4+X^2+X+1$，计算循环码的秩和最小距离。现在，令 ω 为域 $\mathbb{F}_8$ 中 $g(X)$ 的根。我们接收到码字 $r(X)=X^5+X^4+X(\mathrm{mod}\ X^7-1)$。证明 $r(\omega)=\omega^4$，并使用最小距离译码对 $r(X)$ 进行解码。

解答 长度为 N 的循环码的生成多项式为 $g(X)\in\mathbb{F}_2[X]$，奇偶校验多项式为 $h(X)\in\mathbb{F}_2[X]$ 且 $g(X)h(X)=1-X^N$。注意到如果 $g(X)$ 的次数为 k，即 $g(X)=a_0+a_1X+\cdots+a_kX^k$，其中 $a_k\neq 0$，那么，$g(X)$，$Xg(X)$，…，$X^{N-k-1}g(X)$ 形成了该循环码的一组基。特别地，该循环码的秩为 $N-k$。在该例中，$N=7$，$k=3$，秩为 4。

[240]

对于$\mathscr{C}$，如果$h(X)=b_0+b_1X+\cdots+b_{N-K}X^{N-K}$则奇偶校验矩阵$H$具有如下形式

$$\underbrace{\begin{bmatrix} b_{N-K} & b_{N-K+1} & \cdots & b_1 & b_0 & 0 & \cdots & 0 & 0 \\ 0 & b_{N-K} & b_{N-K+1} & \cdots & b_1 & b_0 & \cdots & 0 & 0 \\ 0 & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & & \\ 0 & 0 & \cdots & 0 & b_{N-K} & b_{N-K+1} & \cdots & b_1 & b_0 \end{bmatrix}}_{N}$$

$\mathscr{C}$的码字为$\boldsymbol{H}$的两两列的线性相关。在该例中，

$$\boldsymbol{H}=\begin{bmatrix} 1 & 0 & 1 & 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & 1 & 1 & 0 \\ 0 & 0 & 1 & 0 & 1 & 1 & 1 \end{bmatrix}$$

于是我们有如下关系：

没有零列 ⇒ 没有权重为1的码字

没有重复列 ⇒ 没有权重为2的码字

线性编码$\mathscr{C}$的最小距离$d(\mathscr{C})$是码字的最小非零码重。在该例中，$d(\mathscr{C})=3$。（实际上，该循环码等价于Hamming码[7，4]。）

由于$g(X)\in\mathbb{F}_2[X]$是不可约的，编码$\mathscr{C}\mathbb{F}_2[X]/(X^7-1)$是由$\omega$定义的循环码。在$\mathbb{F}_8$中非零元素的乘法循环群$\mathscr{G}$为

$$\omega^0=1,\omega,\omega^2,\omega^3=\omega+1,\omega^4=\omega^2+\omega$$
$$\omega^5=\omega^3+\omega^2=\omega^2+\omega+1,\omega^6=\omega^3+\omega^2+\omega=\omega^2+1$$
$$\omega^7=\omega^3+\omega=1$$

其次，正如所要求的，$r(\omega)$的值为

$$\begin{aligned} r(\omega)&=\omega-\omega^3+\omega^5 \\ &=\omega-(\omega+1)+(\omega^2+\omega+1) \\ &=\omega^2+\omega=\omega^4 \end{aligned}$$

[241]

令$c(X)=r(X)-X^4\bmod(X^7-1)$。那么$c(\omega)=0$，即$c(X)$是一个码字。由于$d(X)=3$，该编码可以纠正一个错误。我们刚好发现存在一个码字$c(X)$，它与$r(X)$的距离为1。于是根据最小距离译码，$r(X)=X-X^3-X^5$能被译码通过

$$c(X)=X+X^4+X^5-X^6\bmod(X^7-1)$$ □

最后我们用两个有用的结论来总结这一节。

举例2.5.43 （多项式的欧式算法）欧式算法是用于计算在相同域$\mathbb{F}$中的两个多项式$f(X)$和$g(X)$的最大公因式的一种方法。假设$\deg g(X)\leqslant\deg f(X)$，并且设$f(X)=r_{-1}(X)$，$g(X)=r_0(X)$，则

(1) 用$r_0(X)$除$r_{-1}(X)$：

$$r_{-1}(X)=q_1(X)r_0(X)+r_1(X)，其中\ \deg r_1(X)<\deg r_0(X)$$

(2) 用$r_1(X)$除$r_0(X)$：

$$r_0(X)=q_2(X)r_1(X)+r_2(X)，其中\ \deg r_2(X)<\deg r_1(X)$$

⋮

(k) 用$r_{k-1}(X)$除$r_k(X)$：

$$r_{k-2}(X)=q_k(X)r_{k-1}(X)+r_k(X)，其中\ \deg r_k(X)<\deg R_{k-1}(X)$$

⋮

该算法持续到余数为0。

(s)用 r_{r-2} 除 $r_{r-1}(X)$：

$$r_{r-2}(X)=q_r(X)r_{r-1}(X)$$

则

$$\gcd(f(X),g(X))=r_{r-1}(X) \tag{2.5.29}$$

在每个阶段，当前的余数 $r_k(X)$ 涉及两个之前的余数，因此，所有的余数，包括 $\gcd(f(X),g(X))$，都可以写成用 $f(X)$ 和 $g(X)$ 表达的形式。实际上，

引理 2.5.44 欧式算法中的余数 $r_k(X)$ 满足

242

$$r_k(X)=a_k(X)f(X)+b_k(X)g(X),k\leqslant r-1$$

其中

$$a_{-1}(X)=b_{-1}(X)=0,$$
$$a_0(X)=0,b_0(X)=1,$$
$$a_k(X)=-q_k(X)a_{k-1}(X)+a_{k-2}(X),k\geqslant 1,$$
$$b_k(X)=-q_k(X)b_{k-1}(X)+b_{k-2}(X),k\geqslant 1$$

特别地，存在多项式 $a(X)$ 和 $b(X)$，使得

$$\gcd(f(X),g(X))=a(X)f(X)+b(X)g(X)$$

此外：

(1) $\deg a_k(X)=\sum_{1\leqslant i\leqslant k}\deg q_i(X)$，$\deg b_k(X)=\sum_{1\leqslant i\leqslant k}\deg q_i(X)$，

(2) $\deg r_k(X)=\deg f(X)-\sum_{1\leqslant i\leqslant k+1}\deg q_i(X)$。

(3) $\deg b_k(X)=\deg f(X)-\deg r_{k-1}(X)$。

(4) $a_k(X)b_{k+1}(X)-a_{k+1}(X)b_k(X)=(-1)^{k+1}$。

(5) $a_k(X)$ 和 $b_k(X)$ 是互素的。

(6) $r_k(X)b_{k+1}(X)-r_{k+1}(X)b_k(X)=(-1)^{k+1}f(X)$。

(7) $r_{k+1}(X)a_k(X)-r_k(X)a_{k+1}(X)=(-1)^{k+1}g(X)$。

证明 该证明留作练习。 □

2.6 本章附加问题

问题 2.1 长度为 N 的二元循环码 $\mathscr{C}$ 的校验多项式 $h(X)$ 被定义为条件 $a(X)\in\mathscr{C}$ 当且仅当 $a(X)h(X)=0 \bmod (1+X^N)$。该校验多项式是如何与 $\mathscr{C}$ 的生成多项式联系的？给定 $h(X)$，构造奇偶校验矩阵并解释 $\mathscr{C}$ 的陪集 $\mathscr{C}+y$。

描述所有长度为 16 和 15 的循环码。找出重复码和奇偶校验码的生成多项式和校验多项式，找出长度为 7 的 Hamming 码的生成多项式和校验多项式。

解答 所有长度为 16 的循环码均是 $1+X^{16}=(1+X)^{16}$ 的因子，即由 $g(X)=(1+X)^k$ 生成，其中 $k=0,1,\cdots,16$。这里 $k=0$ 给出了全部的 $\{0,1\}^{16}$，$k=1$ 为奇偶校验码，$k=15$ 为重复码(00…0，11…1)，$k=16$ 为全零码。对 $N=15$，分解为不可约多项式的过程如下：

$$1+X^{15}=(1+X)(1+X+X^2)(1+X+X^4)(1+X^3+X^4)$$
$$\times(1+X+X^2+X^3+X^4)$$

243

其中所列不可约多项式的任意乘积都可产生一个循环码。

实际上，$1+X^N=(1+X)(1+X+\cdots+X^{N-1})$；$g(X)=1+X$ 生成了奇偶校验码，$g(X)=1+X+\cdots+X^{N-1}$ 生成了重复码。在 Hamming 码[7，4]中，经过检验，生成多项

式为$g(X)=1+X+X^3$。

校验多项式$h(X)$等于$(1+X^N)/g(X)$。实际上，对所有的$a(X)\in\mathscr{X}$，$a(X)h(X)=v(X)g(X)h(X)=v(X)(1+X^N)=0 \bmod (1+X^N)$。相反，如果$a(X)h(X)=v(X)(1+X^N)$，根据不可约分解的唯一性，$a(X)$必定具有$v(X)g(X)$的形式。

陪集$y+\mathscr{X}$和余数$y(X)=u(X)\bmod g(X)$是一一对应的。换言之，两个码字y^1和y^2属于相同的陪集当且仅当在除法表达式中，

$$y^{(i)}(X)=v^{(i)}(X)g(X)+u^{(i)}(X),i=1,2,\quad u^{(1)}(X)=u^{(2)}(X)$$

实际上，$y^{(1)}$和$y^{(2)}$属于相同的陪集当且仅当$y^{(1)}+y^{(2)}\in\mathscr{X}$。这等价于$u^{(1)}(X)-u^{(2)}(X)=0$，即$u^{(1)}(X)=u^{(2)}(X)$。

如果我们写出$h(X)=\sum_{j=1}^{N}h_jX^j$，则点积为

$$\sum_{j=0}^{N}g_jh_{i+j}=\begin{cases}1, & i=0,N\\ 0, & 1\leqslant i\leqslant N\end{cases}$$

于是，$(g(X)\cdot h^{\dagger}(X))=0$，其中$h^{\dagger}(X)=h_N-h_{N-1}X-\cdots+h_1X^{N-1}$。因此，对于$\mathscr{X}$的奇偶校验矩阵$\boldsymbol{H}$的行由$\boldsymbol{h}=h_0h_1\cdots h_N\cdots0$的循环移位而形成。重复码和奇偶校验码的校验多项式分别为$1+X$和$1-X+\cdots+X^{N-1}$，且它们互为对偶。经过检验，Hamming 码[7，4]的校验多项式等于$1+X+X^3+X^4$。 □

问题 2.2 (a) 根据半径为d的N维 Hamming 球的体积$v_N(d)$，证明二元$[N,d]$码大小的 Hamming 和 Gilbert-Varshamov 界。

假设对于一些固定的$\lambda\in(0,1/4)$的最小距离为$\lfloor\lambda N\rfloor$。令$\gamma(N,\lfloor\lambda N\rfloor)$为能纠正$\lfloor\lambda N\rfloor$个错误的任意二元码的最大信息速率。证明

$$1-\eta(\lambda)\leqslant\liminf_{N\to\infty}\alpha(N,\lfloor\lambda N\rfloor)\leqslant\limsup_{N\to\infty}\alpha(N,\lfloor\lambda N\rfloor)\leqslant1-\eta(\lambda/2)\qquad(2.6.1)$$

(b) 固定$R\in(0,1)$，并假设我们想要发送长度为N的信息集合U_N中的一个信息，其大小为$\#U_N=2^{NR}$。该信息通过一个错误概率$p<1/2$的 MBSC 信道。所以我们预计有pN个错误。根据(a)中的渐近边界，对于大的N值，p取何值时我们能纠正pN个错误？ 244

解答 (a) 如果对于所有的$x,y\in\mathscr{X}$且$x\neq y$，有$B(x,E)\cap B(y,E)=\varnothing$，$\mathscr{X}\subset\mathbb{F}_2^N$被称为能纠正$E$个错误的编码。对于大小为$M$，距离为$d$，纠正$E=\left\lfloor\frac{d-1}{2}\right\rfloor$个错误的编码，其 Hamming 界如下。每个码字的半径为E的球是互不相交的；它们的总体积等于$M\times v_N(E)$。但它们的并集位于$\mathbb{F}_2^N$内，因此$M\leqslant2^N/v_N(E)$。

在另一方面，取一个大小为$\#\mathscr{X}$的E纠错码$\mathscr{X}^*$，则不存在编码使得

$$y\in\mathbb{F}_2^N\setminus\bigcup_{x\in\mathscr{X}^*}B(x,2E+1)$$

或者我们可以给$\mathscr{X}^*$增加一个这样的码字，增加其大小同时保留纠错的性质。由于每个码字$y\in\mathbb{F}_2^N$与这个码字的距离小于$d-1$，我们可以将y加入此编码中。因此，半径为$d-1$的球包含了整个$\mathbb{F}_2^N$，即$M\times v_N(d-1)\geqslant2^N$，或者

$$M\geqslant2^N/v_N(d-1)\text{(Varshamov-Gilbert 界)}$$

对于$\alpha(N,E)=(\log\#\mathscr{X})/N$，结合这些界可以导出

$$1-\frac{\log v_N(2E+1)}{N}\leqslant\alpha(N,E)\leqslant1-\frac{\log v_N(E)}{N}$$

注意到对于任意$s<\kappa N$并且$0<\kappa<1/2$

$$\binom{N}{s-1}=\frac{s}{N-s+1}\binom{N}{s}<\frac{\kappa}{1-\kappa}\binom{N}{s}$$

因此，

$$\binom{N}{E}\leqslant v_N(E)\leqslant\binom{N}{E}\sum_{j=1}^{E}\left(\frac{\kappa}{1-\kappa}\right)^{j}$$

现在，根据 Stirling 公式，随着 N，$E\to\infty$ 且 $E/N\to\lambda\in(0,1/4)$

$$\frac{1}{N}\log\binom{N}{E}\to\eta(\lambda/2)$$

所以，我们证明了 $\lim\limits_{N\to\infty}\frac{1}{N}\log v_N([\lambda N])=\eta(\lambda)$，并且

245

$$1-\eta(\lambda)\leqslant\liminf_{N\to\infty}\frac{1}{N}\log M\leqslant\limsup_{N\to\infty}\frac{1}{N}\log M\leqslant 1-\eta\left(\frac{\lambda}{2}\right)$$

(b) 如果最小距离 d 满足 $\left[\frac{d-1}{2}\right]\geqslant pN$，即 $\lambda/2\geqslant p$，我们可以纠正 pN 个错误。利用渐近 Hamming 界我们得到 $R\leqslant 1-\eta(\lambda/2)\leqslant 1-\eta(p)$。所以，如果 $p\leqslant\eta^{-1}(1-R)$，可靠传输是可能实现的。

Shannon SCT 谈到：

$$\text{无记忆信道的容量 } C=\sup_{p_X} I(X;Y)$$

其中 $I(X;Y)=h(Y)-h(Y|X)$ 为信道单符号随机输入和输出的互信息，在输入字符 X 的整个分布中取最大值。对于一个错误概率为 p 的 $MBSC$，条件熵 $h(Y|X)$ 等于 $\eta(p)$。于是

$$C=\sup_{p_X} h(Y)-\eta(p)$$

通过使用等概率分布的输入 X(当然 Y 也是)可以使 $h(Y)$ 取到最大值。因此，对于 MBSC，$C=1-\eta(p)$。于是，借助于 MBSC，满足 $R\leqslant 1-\eta(p)$，即 $p\leqslant\eta^{-1}(1-R)$，可靠传输是可能的。这两个证明给出了相同的答案。 □

问题 2.3 证明由 $g(X)=1+X+X^2+X^4+X^7+X^8+X^{11}$ 生成的长度为 23 的二元码的最小距离为 7，并且是完美的。提示：注意到根据 BCH 界(定理 2.5.39)，如果一个循环码的生成多项式有根 $\{\omega,\omega^2,\cdots,\omega^{\delta-1}\}$，则码距 $\geqslant\delta$，检查 $X^{23}+1=(X+1)g(X)g^{\sim}(X)$ mod 2，其中 $g^{\sim}(X)=X^{11}g(1/X)$ 为 $g(X)$ 的逆。

解答 首先证明该码为码距等于 5 的 BCH 码。回忆如果 ω 为多项式 $p(X)\in\mathbb{F}_2[\mathscr{X}]$ 的根，则 $\omega^2,\omega^4,\omega^8,\omega^{16},\omega^{32},\omega^{64},\omega^{128},\omega^3,\omega^6,\omega^{12}$ 也是它的根。这导出了设计序列 $\{\omega,\omega^2,\omega^4,\omega^6\}$。根据 BCH 定理，码字 $\mathscr{X}=\langle g(X)\rangle$ 的码距 $\geqslant 5$。

接下来，奇偶校验扩展 $\mathscr{X}^{\sim}$ 是自正交的。为了验证这点，我们需要证明，$\mathscr{X}^{\sim}$ 的生成矩阵任意两列是正交的。它们被表示为

246

$$(X^i g(X)\,|\,1)\quad 和\quad (X^j g(X)\,|\,1)$$

并且它们的点积为

$$\begin{aligned}1+(X^i g(X))(X^j g(X))&=1+\sum_r g_{r-i}g_{r-j}=1+\sum_r g_{r-i}g^{\sim}_{23-r+j}\\&=1+\text{在}\underbrace{g(X)\times g^{\sim}(X)}_{1+\cdots+X^{22}}\text{上 }X^{i+j}\text{ 的系数}\\&=1+1=0\end{aligned}$$

所以

$$\mathscr{X}^{+} \text{ 中的任意两个码字是点正交的} \tag{2.6.2}$$

这说明$\mathscr{X}^{+}$中所有码字的码重可被 4 整除。实际上，根据检验，$\mathscr{X}^{+}$的生成矩阵所有行$(X^i g(X)\mid 1)$的码重为 8。那么，根据和式中包含的行数导出，如果$c\in\mathscr{X}^{+}$和$g^{(i)}\sim(X^i g(X)\mid 1)$为$\mathscr{X}^{+}$的生成矩阵的行，则

$$w(g^{(i)}+c)=w(g^{(i)})+w(c)-2w(g^{(i)}\wedge c)$$

其中$(g^{(i)}\wedge c)_l=\min[(g^{(i)})_l,\ c_l]$，$l=1,\cdots,24$。我们已知$8\mid w(g^{(i)})$和引入的假设，$4\mid w(c)$。接下来，$w(g^{(i)}\wedge c)$为偶数，于是$2w(g^{(i)}\wedge c)$能被 4 整除。于是，上式的左端，$w(g^{(i)}\wedge c)$能被 4 除。因此，$\mathscr{X}^{+}$的码距为 8，因为它$\geq 5$并能被 4 整除。(很明显，它不可能大于 8，否则它的值将是 12。)于是，原码$\mathscr{X}$的码距为 7。

最后，码$\mathscr{X}$是一个完美的纠 3 个错误的编码，由于F_2^{23}中 3 球的体积等于

$$\binom{23}{0}+\binom{23}{1}+\binom{23}{2}+\binom{23}{3}=1+23+253+1771=2048=2^{11}$$

并且$2^{12}\times 2^{11}=2^{23}$。因此，很明显，秩为 12，长度为 23。 □

问题 2.4 证明校验多项式为X^3+X^2+X+1的 Hamming 码为循环码，它的生成多项式是多少？Hamming 原码是否包含等于它对偶的子码？假设不可约单调多项式$M_j(X)$分解如下

$$X^N+1=\prod_{j=1}^{l}M_j(X)^{k_j} \tag{2.6.3}$$

证明循环码的码长$N=\prod_{j=1}^{l}(k_j+1)$。

解答 在F_2^3中我们有

$$X^7-1=(X^3+X+1)(X^3+X^2+X+1) \tag{2.47}$$

生成多项式为$g(X)=X^3+X+1$的循环码的校验多项式为$h(X)=X^4+X^2+X+1$。奇偶校验矩阵为

$$\begin{pmatrix}1&0&1&1&1&0&0\\0&1&0&1&1&1&0\\0&0&1&0&1&1&0\end{pmatrix} \tag{2.6.4}$$

该矩阵的列为F_2^3中的非零元素。所以，这等价于 Hamming 原始码[7，4]。

Hamming[7，4]码的对偶的生成多项式为$X^4+X^3+X^2+1$(即$h(X)$的逆)。由于$X^4+X^3+X^2+1=(X+1)g(X)$，它为 Hamming 码的子码。

最后，任意不可约多项式$M_j(X)$都可能包含在一个为循环码的生成多项式中，其能量为 0，…，k_j。所以，构造这种生成多项式的概率数等于$\prod_{j=1}^{l}(k_j+1)$。 □

问题 2.5 描述 Reed-Muller 码的构造，推出它的信息率和码距。

解答 F_2^m空间中有$N=2^m$个点。如果$A\subset F_2^m$，令$\mathbf{1}_A$为A的示性函数。考虑如下超平面的集合

$$\prod_j=\{p\in F_2^m: p_j=0\}$$

令$h^j=\mathbf{1}_{\prod_j}$，$j=1,\cdots,m$，且$h^0=\mathbf{1}_{F_2^m}=1$。定义函数集合$F_2^m\to F_2$：

$$\mathscr{A}_0=\{h^0\},$$

$$\mathscr{A}_1=\{h^j: j=1,2,\cdots,m\},$$

$$\mathscr{A}_2 = \{h^i \cdot h^j : i,j = 1,2,\cdots,m, i<j\},$$
$$\vdots$$
$$\mathscr{A}_{i+1} = \{a \cdot h^j : a \in \mathscr{A}_i, j = 1,2,\cdots,m, h^j \nmid a\},$$
$$\vdots$$
$$\mathscr{A}_m = \{h^1 \cdots h^m\}$$

这些集合的并集的势为 $N-2^m$（一共有 2^m 个函数）。所以，来自 $\bigcup_{i=0}^{r}\mathscr{A}_i$ 的函数可以被当成 $\mathbb{F}_2^N$ 的一组基。

然后，长度 $N=2^m$ 为 Reed-Muller 码 $\mathrm{RM}(r, m)=\mathscr{R}_{r,m}^{\mathrm{RM}}$ 可定义为 $\bigcup_{i=0}^{r}\mathscr{A}_i$ 的张成，秩为 $\sum_{i=0}^{r}\begin{bmatrix} m \\ i \end{bmatrix}$。其信息率为

$$\frac{1}{2^m}\sum_{i=0}^{r}\begin{bmatrix} m \\ i \end{bmatrix}$$

248

接下来，如果 $a \in \mathrm{RM}(r, m)$，则对于一些 $x \in \mathrm{RM}(m-1, r)$ 和 $y \in \mathrm{RM}(m-1, r-1)$，有

$$a = (y,y)h^1 + (x,x) = (x,x+y)$$

因此，$\mathrm{RM}(r, m)$ 与 Descartes 积 $(\mathrm{RM}(m-1, r) \mid \mathrm{RM}(m-1, r-1))$ 相同。根据 Descartes 积界

$$d[\mathrm{RM}(m,k)] \geq \min(2d[\mathrm{RM}(m-1,k)], d[\mathrm{RM}(m-1,k-1)])$$

经过归纳，推导出

$$d[\mathrm{RM}(r,m)] \geq 2^{m-r}$$

另一方面，向量 $h^1 \cdot h^2 \cdots h^m$ 距离 $\mathrm{RM}(m, r)$ 为 2^{m-r}。因此，

$$d[\mathrm{RM}(r,m)] = 2^{m-r}$$

□

问题 2.6 (a) 定义在域 $\mathbb{F}_2$ 中长度为 N 的奇偶校验码。证明编码是线性的当且仅当它是奇偶校验码。根据奇偶校验来定义原始 Hamming 码并找出它的生成矩阵。

(b) 令 $\mathscr{X}$ 为循环码。定义对偶码

$$\mathscr{X}^{\perp} = \{y = y_1 \cdots y_N : \sum_{i=1}^{N} x_i y_i = 0, \quad 对所有的\ x = x_1 \cdots x_N \in \mathscr{X}\}$$

证明 $\mathscr{X}^{\perp}$ 为循环码并说明 $\mathscr{X}$ 和 $\mathscr{X}^{\perp}$ 的生成多项式是如何相互关联的，说明重复码和奇偶校验码是循环的，并确定它们的生成多项式。

解答 (a)（非必要线性）编码 $\mathscr{X}$ 的奇偶校验码 $\mathscr{X}^{\mathrm{PC}}$ 为向量 $y = y_1 \cdots y_N \in \mathbb{F}_2^N$ 的集合使得点积

$$y \cdot x = \sum_{i=1}^{N} x_i y_i = 0（在\mathbb{F}_2上）, \quad 对所有的\ x = x_1 \cdots x_N \in \mathscr{X}$$

从定义来看，很明显 $\mathscr{X}^{\mathrm{PC}}$ 也是 $\overline{\mathscr{X}}$ 的奇偶校验码，它是由线性码 $\mathscr{X}$：$\mathscr{X}^{\mathrm{PC}} = \overline{\mathscr{X}}^{\mathrm{PC}}$ 张成的。实际上，如果 $y \cdot x = 0$ 且 $y \cdot x' = 0$ 则 $y \cdot (x - x') = 0$。因此，奇偶校验码 $\mathscr{X}^{\mathrm{PC}}$ 始终为线性的，且形成了一个对 $\mathscr{X}$ 点正交的子空间。因此，给定的码字 $\mathscr{X}$ 为线性的当且仅当它为奇偶校验码。线性码 $\mathscr{X}$ 和 $\mathscr{X}^{\mathrm{PC}}$ 形成了一个对偶对：$\mathscr{X}^{\mathrm{PC}}$ 是 $\mathscr{X}$ 的对偶，反之亦然。$\mathscr{X}^{\mathrm{PC}}$ 的生成矩阵 $\boldsymbol{H}$ 是 $\mathscr{X}$ 的奇偶校验矩阵，反之亦然。

249

长度为 $N=2^l-1$ 的 Hamming 码的校验矩阵是一个 $l \times N$ 的矩阵，所有非零列都来自 $\mathbb{F}_2^l$（在某些约定好的排序情况下）。所以，Hamming[7, 4]码对应 $l=3$；它的奇偶校验是

$$x_1 + x_2 + x_3 + x_5 = 0,$$
$$x_2 + x_3 + x_6 + x_7 = 0,$$

$$x_1+x_2+x_6+x_7=0$$

生成矩阵是

$$\begin{bmatrix}1&1&0&1&0&0&0\\0&1&1&0&1&0&0\\0&0&1&1&0&1&0\\0&0&0&1&1&0&1\end{bmatrix}$$

(b) 对偶码的生成多项式是 $g^{\perp}(X)=X^{N-1}g(X^{-1})$。重复码有 $g(X)=1+X+\cdots+X^{N-1}$，并且秩为 1。奇偶校验码 $g(X)=1+X$，并且秩为 $N-1$。 □

问题 2.7 (a) 当错误率 $p>1/2$ 时，编码定理如何使用？

(b) 给出一个不是线性码的例子。

(c) 给出一个不是循环码的线性码例子。

(d) 定义一个二进制 Hamming 码和它的对偶。证明 Hamming 码是完美的。解释为什么 Hamming 码不能纠正两个错误。

(e) 证明在对偶码中有：

(i) 任意非零码字的码重等于 $2^{\ell-1}$。

(ii) 任何一对码字的距离是 $2^{\ell-1}$。

解答 (a) 如果 $p>1/2$，我们保留输出结果使得 $p'=1-p$。

(b) 满足 $\mathscr{X}=\{11\}$ 的码 $\mathscr{X}\subset\mathbb{F}_2^2$ 不是线性的，因为 $00\notin\mathscr{X}$。

(c) 满足 $\mathscr{X}=\{00,10\}$ 的码 $\mathscr{X}\subset\mathbb{F}_2^2$ 是线性的，但是不是循环的，因为 $01\notin\mathscr{X}$。

(d) 初始 Hamming[7，4]码距离为 3，是完美的纠正 1 个错误的编码。所以，码字中有两个错误的会导致码字落在半径为 1 的球外，即，对于一个不同码字会落到半径为 1 的球上(最近的距离是 1，距初始码字的距离是 2)。所以，我们可以检测到两个错误但是却不能纠正它们。

(e) Hamming$[2^{\ell-1}, 2^{\ell}-\ell-1, 3]$码的对偶是线性的，长度是 $N=2^{\ell}-1$，秩为 ℓ。它的生成矩阵是 $\ell\times(2^{\ell}-1)$，其中列给出了的全部非零向量(初始码的奇偶校验矩阵)。矩阵的行是线性独立的；此外，任何行 $i=1,\cdots,\ell$ 都有 $2^{\ell}-1$ 个数字 1。这是因为每个这样的数字都来源于一列，也就是说，一个长度为 ℓ 的非零向量，其中数字 1 在位置 i 上；严格地说，有 $2^{\ell-1}$ 个这样的向量。虽然任何一对矩阵的列都是线性独立的，但是列的三元组是线性相关的(每对列通过它们的和来补充)。 250

每一个非零对偶码字 x 都是上面生成矩阵的行和。假定这些被加数是行 $i_1,\cdots,i_s$，其中 $1\leqslant i_1<\cdots<i_s\leqslant\ell$。那么，正如前文所示，和中数字 1 的数量等于这个矩阵的列数，因为数字 $i_1,\cdots,i_s$ 的和是 1。在剩余的 $\ell-s$ 个数字，我们没有限制，所以对于它们来说，还有 $2^{\ell-s}$ 种可能。对于数字 $i_1,\cdots,i_s$，我们有 2^{s-1} 种可能(总数 2^s 的一半)，所以，又有 $2^{\ell-s}\times2^{s-1}=2^{\ell-1}$。

我们证明了每一个非零对偶码字的码重等于 $2^{\ell-1}$，即，从零向量到任何对偶码字的距离是 $2^{\ell-1}$。因为对偶码是线性的，所以任何一对不同的对偶码字 x，x' 间的距离等于 $2^{\ell-1}$：

$$\delta(x,x')=\delta(\mathbf{0},x'-x)=w(x-x')=2^{\ell-1}$$

令 $J\subset\{1,\cdots,\ell\}$ 为起作用行的集合

$$x=\sum_{i\in J}g^{(i)}$$

那么 x 中非零数字的 $\delta(\mathbf{0}, x)=\#$ 可以被计算为

$$\underset{\substack{\uparrow \\ J\text{ 外到 }0\text{ 和 }1\text{ 的 }\#}}{2^{\ell-|J|}} \times \underset{\substack{\uparrow \\ \text{得到 }\sum_{j\in J} x_j=1\text{ 对 }2\text{ 求余数的 }\#}}{(\text{子集 } K\subset J \text{ 的 } \#(\text{其中 } |K| \text{ 为奇数}))}$$

其中 $x_j=0$ 或 1

由此推出 $2^{\ell-|J|}2^{|J|-1}=2^{\ell-1}$。换句话说，为了从一个数字 $x_j=\sum_{i\in J} g^{(i)}$ 中得到一个贡献，我们必须确定(i)0 和 1 在 $\{1, \cdots, \ell\}\setminus J$ 中的配置(因为它是一个长度为 N 的非零向量的一部分)，(ii)0 和 1 在 J 中的配置，其中 1 的个数是奇数。

为了验证 $d(\mathscr{Z}^{N\perp})=2^{\ell-1}$，它足以证实零码字和任何其他码字 $x\in\mathscr{Z}^{N\perp}$ 的距离等于 $2^{\ell-1}$。 □

问题 2.8 (a) 多项式 $g(X)$ 成为一个长度为 N 的循环码生成多项式的充要条件是什么？什么是 BCH 码？证明关于 $\{\omega, \omega^3\}$ 的 BCH 码是 Hamming 初始码，其中 ω 是在适当域中 X^3+X-1 的根。

251 (b) 定义和估算 Vandermonde 行列式。定义 BCH 码并获得它的最小距离的范围。

解答 (a) $g(X)$ 成为长度为 N 的循环码生成多项式的充要条件是 $g(X)\mid(X^N-1)$。生成多项式 $g(X)$ 可能是不可约的，也可能是可约的；在后一种情况中，通过它不可约因子的乘积 $g(X)=M_1(X)\cdots M_k(X)$ 来表示，其中 $k\leqslant d-\deg g$。令 s 是使 $N\mid 2^s-1$ 成立的最小数。那么 $g(X)$ 可以在域 $\mathbb{K}=\mathbb{F}_{2^s}\supset\mathbb{F}_2$ 中被因式分解为一阶单项式：$g(X)=\prod_{i=1}^{d}(X-\omega_i)$，其中 $\omega_1, \cdots, \omega_d\in\mathbb{K}$。[通常指最小域—关于 g 的分裂域，但是这不是必要的。]每一个元素 ω_i 都是 $g(X)$ 的一个根，并且也是它的不可约因子 $M_1(X), \cdots, M_k(X)$ 的至少一个根。[更准确地说，每一个 $M_i(X)$ 都是一个前面一阶单项式的子积。]

我们想选择一个根在 $\omega_1, \cdots, \omega_d\in\mathbb{K}$ 上的定义集 D；对于每一个因子 $M_i(X)$，它是一个至少包含一个根 ω_{j_i} 的集合。它自然地倾向于采用最小定义集，其中每个非可约因子用一个根来表示，但是这个集合也许不容易严格描述。很明显，定义集合的势 $|D|$ 是在 k 和 d 之间的。形成 D 的根都来源于域 $\mathbb{K}$，但事实上，也有一些来自于包含所有 ω_{j_i} 的子域 $\mathbb{K}'\subset\mathbb{K}$。[当然，$\mathbb{F}_2\subset\mathbb{K}'$。]那么，我们可以识别出 $g(X)$ 生成的循环码，其多项式集合

$$(f(X)\in\mathbb{F}_2[X]/(X^N-1): f(\omega)=0, \quad \text{对所有的 } \omega\in D)$$

$\mathscr{Z}$ 可以认为是一个关于根(或零)的定义集 D 的循环码。

(b) 长度为 N、设计距离为 δ 的二元 BCH 码是一个循环码，其典型集为 $\omega, \omega^2, \cdots, \omega^{\delta-1}$，其中 $\delta\leqslant N$ 并且 ω 是第 N 个主单位根，满足 $\omega^N=1$。值得注意的是，如果 ω 是多项式 $p(X)$ 的一个根，那么 $\omega^2, \cdots, \omega^{2^{\ell-1}}$ 也是。考虑到一个形如 $\{\omega, \omega^3, \cdots, \omega^{\delta-1}\}$ 的定义集，我们对上面的二进制序列进行"补空"，并生成一个理想多项式，它的特性可以被分析研究。

最简单的例子是当 $N=7$ 和 $D=\{\omega, \omega^3\}$ 时，其中 ω 是 X^3+X+1 的根。这里，$\omega^7=(\omega^3)^2\omega=(\omega+1)^2\omega=\omega^3+\omega=1$，所以 ω 是第七个单位根。[我们用到了特征值为 2 的结论。]事实上，它是一个本原根，所以，像刚才说的，ω^3 是 X^3+X+1 的一个根，X^3+X+1；$(\omega^3)^2+\omega^3+1=(\omega^2+\omega+1)^2=0$，还有 ω^6 也是。那么关于定义集 $\{\omega, \omega^3\}$ 的循环码生成多项式为 X^3+X+1，因为这个多项式所有的根都被使用了。我们知道它和 Hamming [7, 4]码相同。

Vandermonde 行列式是

$$\Delta = \det \begin{bmatrix} 1 & 1 & 1 & \cdots & 1 \\ x_1 & x_2 & x_3 & \cdots & x_n \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ x_1^{n-1} & x_2^{n-1} & x_3^{n-1} & \cdots & x_n^{n-1} \end{bmatrix}$$

252

可以观察到，如果 $x_i = x_j (i \neq j)$，行列式就等于 0(两行是一样的)。所以 $x_i - x_j$ 是 Δ 的一个因子，

$$\Delta = P(x) \prod_{i<j} (x_i - x_j)$$

其中 P 是变量 $x_1, \cdots, x_n$ 的一个多项式。现在考虑中具有形如 $a \prod_i x_i^{m(i)}$ 的和中关于展开式 Δ 的项，其中 $\sum m(i) = 0 + 1 + \cdots + (n-1) = n(n-1)/2$，但是 $\prod_{i<j} (x_i - x_j)$ 是 $a \prod_i x_i^{m(i)}$ 的和，其中 $\sum m(i) = n(n-1)/2$，所以 $P(x)$ 是常数，考虑到 $x_2 x_3^2 \cdots x_n^{n-1}$，我们可得常数为 1，所以

$$\Delta = \prod_{i<j} (x_i - x_j) \tag{2.6.5}$$

假定 N 是奇数，K 是一个包含 F_2 的域，其中 $X^N - 1$ 可以因式分解成线性因子。[这个域被选为 F_{2^r}，其中 $N \mid 2^r - 1$]由码字 $c = c_0 c_1 \cdots c_{N-1}$ 组成的循环码被称作设计距离 $\delta < N$ 的 BCH 码，其中对于所有的 $r = 1, 2, \cdots, \delta - 1$ 都有 $\sum_{j=0}^{N} c_j \omega^{rj} = 0$，$\omega$ 是一个第 N 个本原单位根。接下来，$\mathscr{X}^{BCH}$ 是一个在 F_2 上的向量空间，并且 $c \in \mathscr{X}^{BCH}$，当且仅当

$$\boldsymbol{c}\boldsymbol{H}^T = 0 \tag{2.6.6}$$

其中

$$\boldsymbol{H} = \begin{bmatrix} 1 & \omega & \omega^2 & \cdots & \omega^{N-1} \\ 1 & \omega^3 & \omega^4 & \cdots & \omega^{3N-3} \\ 1 & \omega^5 & \omega^8 & \cdots & \omega^{5N-5} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & \omega^{\delta-1} & \omega^{2\delta-2} & \cdots & \omega^{(N-1)(\delta-1)} \end{bmatrix} \tag{2.6.7}$$

现在 $\operatorname{rank} \boldsymbol{H} = \delta$。的确，根据(2.6.5)，对于任何 $\delta \times \delta$ 的子式 $\widetilde{\boldsymbol{H}}$

$$\det \widetilde{\boldsymbol{H}} = \prod_{i<j} (\omega^i - \omega^j) \neq 0$$

所以式(2.6.6)告诉我们

$$c \in \mathscr{X}, c \neq \mathbf{0} \Rightarrow \sum |c_i| \geqslant \delta$$

所以，$\mathscr{X}^{BCH}$ 中的最小距离大于等于 δ。 □

问题 2.9 如果一个 Hamming 空间 $\{0, 1\}^N$ 的子集 $\mathscr{X}$，势长 $\mathscr{X} = M$，并且最小 Hamming 距离是 $d = \min[\delta(x, x'): x, x' \in \mathscr{X}, x \neq x']$，那么这个子集被称为 $[N, M, d]$ 码(不一定线性)。如果一个 $[N, M, d]$ 码不被包含在任何 $[N, M+1, d]$ 码内，那么此被称为最大码。证明一个 $[N, M, d]$ 码当且仅当对于任意 $y \in \{0, 1\}^N$ 都存在 $x \in \mathscr{X}$ 使得 $\delta(x, y) < d$ 成立时是最大码。结论：如果在一个码字中发生 d 个或更多改变时，那么新码字比初始码字更接近于某些其他的码字。 253

假定一个最大 $[N, M, d]$ 码被用在通过二元无记忆信道传输信息，错误率为 p，并且接收端用最大似然译码器。证明错误译码的概率 π_{cr}^{ML} 服从下面的界

$$1 - b(N, d-1) \leqslant \pi_{cr}^{ML} \leqslant 1 - b(N, \lfloor (d-1)/2 \rfloor)$$

其中 $b(N, m)$ 是一个部分二项式和

$$b(N,m)=\sum_{0\leqslant k\leqslant m}\binom{N}{k}p^k(1-p)^{N-k}$$

解答 如果一个码是最大码，那么再添加一个码字将减小距离。因此，对于所有 y 都存在 $x\in\mathscr{X}$ 使得 $\delta(x, y)<d$。相反，如果这个性质成立，那么码只有在减小 d 时才可扩展。那么在码字中产生 d 个或更多个可以给出一个更加靠近不同码字的字。这肯定不会在 ML 译码下给出一个正确的推测，因为它选择了最近的码字。

因此，

$$\pi_{ML}\geqslant\sum_{0\leqslant k\leqslant N}\binom{N}{k}p^k(1-p)^{N-k}=1-b(N,d-1)$$

在另一方面，此码纠正了 $\lceil(d-1)/2\rceil$ 个错误。所以

$$\pi_{ML}\leqslant 1-b(N,\lfloor d-1/2\rfloor)$$ □

问题 2.10 对于 $[N, M, d]$ 二元码，Plotkin 界表明如果 $d>N/2$，那么有 $M\leqslant\frac{d}{d-N/2}$。一个码的长度是 N、距离是 d，令 $M_2^*(N, d)$ 为该码的最大数量，并令

$$\alpha(\lambda)=\lim_{N\to\infty}\frac{1}{N}\log M_2^*(N,\lfloor\lambda N\rfloor)$$

从 Plotkin 界推出，对于 $\lambda\geqslant\frac{1}{2}$ 有 $\alpha(\lambda)=0$。

假定上述的界成立，证明如果 $d\leqslant N/2$，那么

$$M\leqslant 2^{N-(2d-1)}\frac{d}{d-(2d-1)/2}=2d2^{N-(2d-1)}$$

254 推到渐近 Plotkin 界：$\alpha(\lambda)\leqslant 1-2\lambda$，$0\leqslant\lambda\leqslant\frac{1}{2}$。

解答 如果 $d>N/2$，应用 Plotkin 界并且得到 $\alpha(\lambda)=0$。如果 $d\leqslant N/2$，按照最后 $N-(2d-1)$ 个数字，考虑长度为 N、距离为 $d\leqslant N/2$ 的码 $\mathscr{X}$ 的分割，也就是说，将 $\mathscr{X}$ 分割为不相交的子集，最后的 $N-(2d-1)$ 个数字不变。作为这些子集之一的 $\mathscr{X}'$。其大小一定是 M' 使得 $M'2^{N-(2d-1)}\geqslant M$。

所以，$\mathscr{X}'$ 是一个长度为 $N'=(2d-1)$，距离为 $d'=d$ 的码，其中 $d'>N'/2$。将 Plotkin 界应用到 $\mathscr{X}'$ 中，得到

$$M'\leqslant\frac{d'}{d'-N/2}=\frac{d}{d-(2d-1)/2}=2d$$

因此

$$M\leqslant 2^{N-(2d-1)}2d$$

令 $d=\lfloor\lambda N\rfloor$，当 $N\to\infty$，可以得到 $\alpha(\lambda)\leqslant 1-2\lambda$，$0\leqslant\lambda\leqslant 1/2$。 □

问题 2.11 论述和证明 Hamming、Singleton 和 Gilbert-Varshamov 界。给出(a)一个可以到达 Hamming 限制的例子；(b)可以到达 Singleton 界的例子。

解答 Hamming 界给出了一个长度为 N、大小为 M 的 E-纠错编码，

$$M\leqslant\frac{2^N}{v_N(E)}$$

其中 $v_N(E)=\sum_{0\leqslant i\leqslant E}\binom{N}{i}$ 是 E-球在 Hamming 空间 $\{0, 1\}^N$ 中的体积。它遵循这样的事实：

关于码字 $x\in\mathscr{X}$ 的 E-球一定是不相交的：

$$M\times v_N(E)=\#\ M\,E\text{-球覆盖的点}$$
$$\leqslant 2^N=\#\ \text{在}\{0,1\}^N\text{上的点}$$

Singleton 界是长度为 N 的码 $\mathscr{X}$ 的大小 M 和距离 d 满足

$$M\leqslant 2^{N-d+1}$$

上面的结论可由下面的观察得出：截断 $\mathscr{X}$（即，从码字 $x\in\mathscr{X}$ 中删除一个数字）$d-1$ 次仍然无法合并码字（即保留 M）而最后的码字适合 $\{0,1\}^{N-d+1}$。

Gilbert-Varshamov 界是指二元 $[N,d]$ 码的最大尺寸 $M^*=M_2^*(N,d)$ 满足

$$M^*\geqslant\frac{2^N}{v_N(d-1)}$$ 255

这个界服从如下的观察：对于最大尺寸编码 $\mathscr{X}^*$，任何码字 $y\in\{0,1\}^N$ 一定距离 $\leqslant d-1$。所以

$$M^*\times v_N(d-1)\geqslant\#\ \text{在距离}\ d-1=2^N\ \text{中的点}$$

达到 Hamming 界的码被称为完美码，比如，Hamming$[2^\ell-1,2^\ell-1-\ell,3]$码。这里，$E=1$，$v_N(1)=1+2^\ell-1=2^\ell$，$M=2^{2^\ell-\ell-1}$。除了这些码，只有一个（二元）完美码的例子，Golay$[23,12,7]$码。

达到 Singleton 界的码被称为最大距离可分离（MDS）码：它们的校验矩阵的任意 $N-M$ 行线性独立。这种码的例子是（i）整个 $\{0,1\}^N$，（ii）重复码 $\{0\cdots0,1\cdots1\}$ 以及所有偶码的重码字 $x\in\{0,1\}^N$ 的集合。事实上，这就是二元 MDS 码的所有例子。更多有趣的例子将由非二进制的 Reed-Solomon 码给出：见 3.2 节。达到 Gilbert-Varshamov 界的，具有一般 N 和 d 的二元码到目前为止还没有被构造出来（虽然针对非二元字母表已经构造出来了）。□

问题 2.12 （a）请用 Hamming 初始码作为你的例子，解释纠错码对于电脑工程师的存在性和重要性。

（b）在 Hamming 码中码重分别为 1、2、3、4、5 的有多少个码字？

解答 （a）考虑一个形如（2.6.4）的矩阵 $\boldsymbol{H}$ 给出的线性映射 $\mathbb{F}_2^7\to\mathbb{F}_2^3$。Hamming 码 $\mathscr{X}$ 是核 $\ker\boldsymbol{H}$，即满足 $x\boldsymbol{H}^{\mathrm{T}}=\boldsymbol{0}$ 的码字 $x=x_1x_2x_3x_4x_5x_6x_7\in\{0,1\}^7$ 的集合。这里，我们可以选择四个数字 x_4，x_5，x_6，x_7，从 $\{0,1\}$ 中任选；那么 x_1，x_2，x_3 可以通过下式得到：

$$x_1=x_4+x_5+x_7$$
$$x_2=x_4+x_6+x_7$$
$$x_3=x_5+x_6+x_7$$

它意味着码 $\mathscr{X}$ 可以被用来编码长度为 4 的 16 个二进制“信息”。如果 $y=y_1y_2y_3y_4y_5y_6y_7$ 在一个位置上不同于码字 $x\in\mathscr{X}$，也就是 $y=x+e_k$，那么等式 $y\boldsymbol{H}^{\mathrm{T}}=e_k\boldsymbol{H}^{\mathrm{T}}$ 给出了一个数字 k 的二元分解，它用来解码 x。因此，码 $\mathscr{X}$ 允许出现一个可以被纠正的错误。

假定任何数字的错误概率是 $p\ll1$，独立于其他符号。那么在传送一个非编码（$4N$）-符号信息时的错误概率是

$$1-(1-p)^{4N}\simeq 4Np$$

但是使用 Hamming 码时我们需要发送 $7N$ 个符号。一个错误传输需要至少有两个错误数字，它的概率是

$$\approx 1-\left(1-\binom{7}{2}p^2\right)^N\simeq 21Np^2\ll 4Np$$ 256

所以，在 Hamming 码中额外使用 3 个校验数字是合理的。

(b) 由二进制码字 $\boldsymbol{x}=x_1\cdots x_N$ 组成的，长度为 $N=2^\ell-1(\ell\geqslant 3)$ 的码 $\mathscr{X}_{H,\ell}$ 满足 $\boldsymbol{x}\boldsymbol{H}^\top=0$，其中 $\boldsymbol{H}$ 是一个 $\ell\times N$ 矩阵，它的列 $h^{(1)}$，…，$h^{(N)}$ 全是长度为 ℓ 的非零二进制向量。所以，码重为 $w(\boldsymbol{x})=\sum_{j=1}^{N}x_j=s$ 的码字数量等于 s 个二元，非零，逐点不等且和为 0 的 ℓ-向量的集合数。事实上，如果 $\boldsymbol{x}\boldsymbol{H}^\top=0$，$w(\boldsymbol{x})=s$ 和 $x_{j_1}=x_{j_2}=\cdots=x_{j_s}=1$，那么行向量的和 $h^{(j_1)}+\cdots+h^{(j_s)}=\boldsymbol{0}$。

因此，码重是 0 的有一个码字，没有码重是 1 或 2 的码字，码重是 3 的有 $N(N-1)/3$ 个，(即，对于 $\ell=3$ 和 $\ell=4$，码重是 3 的分别有 7 和 35 个码字)。进一步我们可以找到码重是 4 的有 $[N(N-1)(N-2)-N(N-1)]/4!=N(N-1)(N-3)/4!$ 个(即，对于 $\ell=3$ 和 $\ell=4$，码重是 4 的分别有 7 和 105 个码字)。最后，我们有码重是 5 的有 $N(N-1)(N-2)(N-7)/5!$ 个(即，对于 $\ell=3$ 和 $\ell=4$，码重是 5 的分别有 0 和 168 个)。每当我们增加一个因子的时候，我们应该避免 ℓ-向量等于先前所选向量的线性组合。在问题 3.9 中，我们将计算对于 $N=15$ 时的枚举多项式：

$$1+35X^3+105X^4+168X^5+280X^6+435X^7+435X^8$$
$$+280X^9+168X^{10}+105X^{11}+35X^{12}+X^{15}$$ □

问题 2.13 (a) 二元 Hamming 空间向量 $\mathscr{H}_N$ 中向量 $\boldsymbol{x}$，$\boldsymbol{y}$ 的点积被定义为 $\boldsymbol{x}\cdot\boldsymbol{y}=\sum_{i=1}^{N}x_iy_i(\mathrm{mod}\ 2)$，并且如果 $\boldsymbol{x}\cdot\boldsymbol{y}=0$，我们称 $\boldsymbol{x}$ 和 $\boldsymbol{y}$ 是正交的。那么说 $\mathscr{X}\subseteq\mathscr{H}_N$ 是一个满足生成矩阵为 $\boldsymbol{G}$、奇偶校验矩阵为 $\boldsymbol{H}$ 的线性 $[N,k]$ 码意味着什么？证明

$$\mathscr{X}^\perp=\{\boldsymbol{x}\in\mathscr{H}_N:\boldsymbol{x}\cdot\boldsymbol{y}=0,\quad 对所有的\ \boldsymbol{y}\in\mathscr{X}\}$$

是一个线性 $[N,N-k]$ 码，并且找出它的生成多项式和奇偶校验矩阵。

(b) 如果线性码 $\mathscr{X}$ 满足 $\mathscr{X}=\mathscr{X}^\perp$，那么称 $\mathscr{X}$ 是自正交的。证明如果 $\boldsymbol{G}$ 的行是自正交和两两正交的，那么 $\mathscr{X}$ 是自正交的。如果 $\mathscr{X}=\mathscr{X}^\perp$，那么称线性码是自对偶的。证明自对偶码一定是 $[N,N/2]$ 码(因此 N 一定为偶数)。相反，证明一个自正交 $[N,N/2]$ 码当 N 为偶数时是自对偶的。请给出对于任意偶数 N 时这种码的例子，并证明一个自对偶码总是包含字符 $1\cdots1$。

257

(c) 现在考虑一个 Hamming $[2^\ell-1,2^\ell-\ell-1]$ 码 $\mathscr{X}_{H,\ell}$，描述 $\mathscr{X}_{H,\ell}^\perp$ 的生成矩阵。证明任意在 $\mathscr{X}_{H,\ell}^\perp$ 上的两个码字距离等于 $2^{\ell-1}$。

解答 通过定义可知，$\mathscr{X}^\perp$ 在线性变化下保持不变；所以 $\mathscr{X}^\perp$ 是一个线性码。从代数的角度考虑，$\dim\mathscr{X}^\perp=N-k$。$\mathscr{X}^\perp$ 的生成矩阵 $\boldsymbol{G}^\perp$ 和 $\boldsymbol{H}$ 相同，奇偶校验矩阵 $\boldsymbol{H}^\perp$ 和 $\boldsymbol{G}$ 相同。

如果 $\mathscr{X}\subseteq\mathscr{X}^\perp$，那么矩阵 $\boldsymbol{G}$ 的行 $g^{(1)}$，…，$g^{(k)}$ 是自正交的，同时也是两两正交的。反之亦然。从前面的观察中可以看出，如果 $\mathscr{X}$ 是自对偶的，那么 $k=N-k$，即 $k=N/2$，并且 N 应该是偶数。相似地，如果 $\mathscr{X}$ 是自正交的并且 $k=N/2$，那么 $\mathscr{X}$ 是自对偶的。

令 $\boldsymbol{1}=1\cdots1$。如果 $\mathscr{X}=\mathscr{X}^\perp$，那么 $\boldsymbol{1}\cdot g^{(i)}=g^{(i)}\cdot g^{(i)}=0$，所以有 $\boldsymbol{1}\subseteq\mathscr{X}^\perp$ 并且 $\boldsymbol{1}\subseteq\mathscr{X}$。有这样一个编码的例子，其生成矩阵为

$$\boldsymbol{G}=\begin{array}{c}\left.\begin{pmatrix}1&1&1&\cdots&1&1&\cdots&1\\1&1&0&\cdots&0&1&\cdots&0\\1&0&1&\cdots&0&1&\cdots&0\\\vdots&\vdots&\vdots&\ddots&\vdots&\vdots&\ddots&\vdots\\1&0&0&\cdots&1&1&\cdots&1\end{pmatrix}\right\}N/2\\ \leftarrow N/2\rightarrow\qquad\qquad\leftarrow N/2\rightarrow\end{array}$$

Hamming 码$\mathscr{H}_H$的对偶$\mathscr{H}_H^{\perp}$被称为极长码。通过上面的论述，我们可以知道它的码长为$2^\ell-1$、秩为ℓ，并且它的生成矩阵G_H大小是$\ell\times(2^\ell-1)$，其中列给出了长度是ℓ的所有非零向量。为了检验$\operatorname{dist}\mathscr{H}_H^{\perp}=2^{\ell-1}$可通过证明一个的非零码字$x=\mathscr{H}^{\perp}$的码重等于$2^{\ell-1}$。但是非零码字$x\in\mathscr{H}^{\perp}$是$G_H^t$行向量的一个非零线性组合。令$J\subset\{1,\cdots,\ell\}$是贡献行的集合：

$$x=\sum_{i\in J}g(i)$$

很清楚，$w(g^{(i)})=2^{\ell-1}$刚好是所有2^ℓ个向量的一半，在任何给定的位置上都是1。该证明将在关于$\#J$的推导中完成。

一个简单而优雅的方法是用 MacWilliams 等式(参考引理3.4.4)，立即可以得出

$$W_{\mathscr{H}^{\perp}}(s)=1+(2^\ell-1)s^{2^{\ell-1}} \tag{2.6.8}$$

给出这个推导是很有意义的。我们将在问题3.9中，针对一个 Hamming 码的加权枚举多项式，证明这个公式。然后将这个表达式代入 MacWilliams 等式中可以得到 [258]

$$\begin{aligned}W_{\mathscr{H}^{\perp}}(s)&=\left(\frac{1}{2^{N-\ell}}\frac{1}{N+1}\left(1+\frac{1-s}{1+s}\right)^N\right.\\&\quad\left.+\frac{N}{N+1}\left(1+\frac{1-s}{1+s}\right)^{(N-1)/2}\left(1-\frac{1-s}{1+s}\right)^{(N+1)/2}\right)(1+s)^N\\&=2^\ell\left(\frac{1}{2^\ell}+\frac{2^\ell-1}{2^\ell}s^{2^{\ell-1}}\right)\end{aligned}$$

它等价于(2.6.8)。

问题2.14 简要描述 Hamming$[2^\ell-1,2^\ell-1-\ell]$码的解码过程。

对于形如式(2.3.4a)的码字字典奇偶校验矩阵，Hamming[7，4]码可以用于编码16符号、字母表的前15个字符和间隔符号*。编码规则是

A 0011001	E 0111100	I 1010101	M 1111111
B 0100101	F 0001111	J 1100110	N 1000011
C 0010110	G 1101001	K 0101010	O 0000000
D 1110000	H 0110001	L 1001100	* 1011010

你已经收到一个105位的数字信息

1000110 0000000 0110001 1000011 1000011 1110101
0111100 0011010 0100101 0111100 1011000 1101001
0000000 0010000 1010000

其中一些码字是被损坏的，解码接收到的信息。

解答 Hamming$[2^\ell-1,2^\ell-1-\ell]$码($\ell=2,3,\cdots$)可以作为一个长度为$N=2^\ell-1$二进制"字符串"$x=x_1\cdots x_N$集合使得$xH^t=0$来获得。矩阵H是一个$(\ell\times2^\ell-1)$的矩阵，其中列是由$2^\ell-1$个非零二进制字符串组成：就是，

$$H=\begin{pmatrix}1&0&\cdots&0&1&\cdots&1\\0&1&\cdots&0&0&\cdots&1\\\cdots&\cdots&\cdots&\cdots&\cdots&\cdots&\cdots\\0&0&\cdots&1&1&\cdots&1\end{pmatrix}$$

这里列被约定按字典排序。通过排列组合上面矩阵的行，可以获得不同矩阵，这些矩阵定义了不同但等价的码：它们被称为 Hamming 码。

为了解码，我们必须确定一个矩阵 $\boldsymbol{H}$（校验矩阵）并让收发两端都知道这个矩阵。依靠
259 收到的字符（串）$\boldsymbol{y}=y_1\cdots y_N$，我们构造一个伴随向量 $\boldsymbol{yH}^{\mathrm{T}}$。如果 $\boldsymbol{yH}^{\mathrm{T}}=0$，我们可以通过它自身来解码 $\boldsymbol{y}$。（我们无法确定初始码字是否被信道损坏。）

如果 $\boldsymbol{yH}^{\mathrm{T}}\neq 0$，那么 $\boldsymbol{yH}^{\mathrm{T}}$ 与 $\boldsymbol{H}$ 中的列相同。假设 $\boldsymbol{yH}^{\mathrm{T}}$ 给出 $\boldsymbol{H}$ 中的列 j，那么我们通过如下式子解码 $\boldsymbol{y}$

$$\boldsymbol{x}^{*}=\boldsymbol{y}+\boldsymbol{e}_j,\quad \text{其中}\boldsymbol{e}_j=0\cdots1\cdots0\text{（在位 }j\text{ 上的 1）}$$

换句话说，我们改变 $\boldsymbol{y}$ 中的数字 j，然后判定它是通过信道发送的哪个码字。当信道传输错误很少时我们判断得很准确。

如果 $\ell=3$ 的一个 Hamming[7，4]码包含 $2^4=16$ 个码字。当 $\boldsymbol{H}$ 被固定时这些码字是固定的：在例子中被用来编码 15 个字母 A 到 O 和空字符 *。依靠接收到信息我们将其分为长度为 7 的码字：在例子中这里一共有 15 个码字。执行解码过程得到

JOHNNIE * BE * GOOD □

问题 2.15 一个长度为 $N=2^\ell-1$ 的（二进制）Hamming 码，当 $\ell\geq 2$ 时，Hamming 码被定义为一个线性二元码，它的奇偶校验矩阵 $\boldsymbol{H}$ 的列由长度为 ℓ 的非零二进制向量组成。求这样一个码的秩（也就是相应的线性子空间的维度）和码字的数目。求码的最小距离并且证明它是纠单错误的纠错码。证明码字是完美的（即围绕码字的单位球的并集覆盖了所有码字空间）。

给出一个 $\ell=3$ 的 Hamming 码的奇偶校验矩阵和生成矩阵。这个码的信息速率是多少？为什么当 $\ell=2$ 时不合算？

解答 Hamming 码的奇偶校验矩阵 $\boldsymbol{H}$ 是 $\ell\times 2^\ell-1$，它由非零的长度为 ℓ 的列组成；特别地，它包含所有的权重为 1 的 ℓ 列。后者是线性独立的；因此 $\boldsymbol{H}$ 中的 ℓ 列是线性独立的。因为 $\mathscr{X}_{\text{Ham}}=\ker\boldsymbol{H}$，我们有 $\dim\mathscr{X}=2^\ell-1-\ell=\operatorname{rank}\mathscr{X}$，那么码字的数目等于 $2^{2^\ell-\ell-1}$。

因为 $\boldsymbol{H}$ 的所有列是不同的，任意两列都是线性独立的。所以，$\mathscr{X}$ 的最小距离 >2。但是 $\mathscr{X}$ 包含有线性相关的三列，例如

$$100\cdots0^{\mathrm{T}},\ 010\cdots0^{\mathrm{T}}\ \text{和}\ 110\cdots0^{\mathrm{T}}$$

因此，最小距离等于 3，所以如果发生了单个错误，也就是接收到的码字与这个码字的距
260 离是 1，那么这个码字唯一可译。因此 Hamming 码是纠单错误的纠错码。

为了证明它是完美的，我们必须检查

$$\text{码字的}\#\times 1\text{-球的列}=\text{码字的总}\#$$

事实上，令 $2^\ell-1=N$，我们有

$$\text{码字的}\#=2^{2^\ell-\ell-1}=2^{N-\ell}$$

$$1\text{-球的列}=\binom{N}{0}+\binom{N}{1}=1+N$$

$$\text{字的总}\#=2^N$$

和

$$(1+N)2^{N-\ell}=2^\ell 2^{N-\ell}=2^N$$

此码的信息速率等于

$$\text{秩}/\text{长度}=\frac{2^\ell-\ell-1}{2^\ell-1}$$

$\ell=3$ 的码具有形如(2,6,4)的 3×7 奇偶校验矩阵；通过行的任意排列组合可以得到等价的码。生成矩阵是 4×7 的：

$$\begin{bmatrix} 1 & 0 & 0 & 0 & 1 & 1 & 1 \\ 0 & 1 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 & 1 & 1 \end{bmatrix}$$

信息速率为 4/7。$t=2$ 的 Hamming 码很简单，它只有一个唯一的码字 1 1 1。 □

问题 2.16 定义一个在域$\mathbb{F}_q$上的长度为 N 的 BCH 码，设计距离为 δ。证明这样一个码的最小码重至少为 δ。

考虑一个在域$\mathbb{F}_2$上的长度为 31 的 BCH 码，设计距离为 8。证明最小距离至少为 11。

解答 一个在域$\mathbb{F}_q$上的长度为 N 的 BCH 码定义为一个循环码$\mathscr{C}$，它的最小次数的生成多项式 $g(X)\in\mathbb{F}_q[X]$，关于 $g(X)\mid(X^N-1)$(因此有 $\deg g(X)\leqslant N$)，包含了它的根中的幂级数 ω，ω^2，$\cdots\omega^{\delta-1}$，其中 $\omega\in\mathbb{F}_{q^l}$ 是第 N 个本源单位根。(这个根 ω 位于一个扩展域$\mathbb{F}_{q^l}$-X^N-1在$\mathbb{F}_q$上的分裂域，也就是 $N\mid(q^l-1)$)，那么 δ 被称为$\mathscr{C}$的设计距离；实际距离(一般情况下可能很难计算)$\geqslant\delta$。

如果我们考虑长度为 31 的二进制 BCH 码，ω 应该是一个次数为 31 的本源单位根，其中 $\omega^{31}=1$(根 ω 位于扩展域$\mathbb{F}_{32}$上)。我们知道在二进制代数中，如果阶数为 s 的多项式 $f(X)\in\mathbb{F}_2[X]$有一个根 ω，那么它就有根 ω^2，ω^4，$\cdots$，$\omega^{2^{s-1}}$，即 [261]

$$(X-\omega^{2^r})\mid f(X),r=0,\cdots,s-1$$

因此，假如$\mathscr{C}$的生成子 $g(X)$有根 ω，ω^2，ω^3，ω^4，ω^5，ω^6，ω^7，那么它还有根

$$\omega^8=(\omega^4)^2,\omega^9=(\omega^3)^3,\omega^{10}=(\omega^5)^2$$

也就是说定义集合可以扩展为

$$\omega,\omega^2,\omega^3,\omega^4,\omega^5,\omega^6,\omega^7,\omega^8,\omega^9,\omega^{10}$$

(所有的这些元素是不同的，因为 ω 是第 31 个本源单位根。)事实上，码$\mathscr{C}$的设计距离$\geqslant$11，因此，$\mathscr{C}$的最小距离$\geqslant$11。 □

问题 2.17 令$\mathscr{C}$是二进制域$\mathbb{F}_2$上的线性$[N,k,d]$码，G是$\mathscr{C}$的生成矩阵，有 k 行 N 列，使得第一行元素的 d 刚好是 1。令 G_1 是 $k-1$ 行 $N-d$ 列的矩阵，通过删除 G 的第一行和第一行中不为 0 元素的列构成。证明由 G_1 生成的线性码$\mathscr{C}_1$的最小距离 $d'\geqslant\lceil d/2\rceil$。这里对于一个实数 x，$\lceil x\rceil$满足 $x\leqslant\lceil x\rceil\leqslant x+1$ 的整数。

证明$\mathscr{C}_1$的维为 $k-1$。推导

$$N\geqslant\sum_{0\leqslant i\leqslant k-1}\lceil d/2^i\rceil$$

解答 令 x 是$\mathscr{C}$中由 G 的第一行表示的码字，挑选其他两行，叫作 y 和 z，删除第一行之后得到 y'和 z'。两个码重 $w(y')$和 $w(z')$一定$\geqslant\lceil d/2\rceil$，否则初始码 y 和 z 中至少其中一个，不妨假设为 y，在被删掉的 d 个数字中应该有最少的$\lceil d/2\rceil$个数字 1(因为根据条件 $w(y)\geqslant d$)。但是

$$w(x+y)=w(y')+d-\lceil d/2\rceil<d$$

这与$\mathscr{C}$的距离为 d 相矛盾。 [262]

我们要检验码重 $w(y'+z')\geqslant\lceil d/2\rceil$。假设相反的情况

$$w(y'-z')=m<\lceil d/2\rceil$$

那么 $m'=w(y^0+z^0)$一定$\geqslant d-m\geqslant\lceil d/2\rceil$，其中 y^0 是 y 的删除部分，长度为 d，z^0 是 z 的删除部分，长度也为 d。事实上，如前所述，如果 $m'<d-m$，那么 $w(y+z)<d$ 是不可能的。但是如果 $m'>d-m$，那么

$$w(x+y+z)=d+m'+m<d$$

也是不可能的。因此 G_1 中任意两行之和的码重 $\geq\lceil d/2\rceil$。

对于 G_1 中任意行数之和(不超过 $k-1$)这个证明可以重复。事实上，这种加和 $x+y+\cdots+z$ 情况下，我们可以过渡到新的矩阵，$\bar{G}$ 和 $\bar{G}_1$，涉及行之间的这个和。我们得出结论：$\mathscr{X}_1$ 的最小距离 $d'\geq\lceil d/2\rceil$。$\mathscr{X}_1$ 的秩是 $k-1$，因为 G_1 的任意 $k-1$ 行都是线性独立的(上面的和不可能为0)。

现在，删除过程可以被应用到 $\mathscr{X}_1$(删除 G_1 中的 d' 列产生 G_2 中行的数字1恰好是 d' 个数字1)。继续，直到将初始秩从 k 降到了1。这会导致需要的界

$$N\geq d+\lceil d/2\rceil+\lceil d/2^2\rceil+\cdots+\lceil d/2^{k-1}\rceil \qquad \square$$

问题2.18 定义一个循环线性码 $\mathscr{X}$，然后证明在正规化的条件下，它有唯一的最小长度的码字，多项式 $g(X)$ 的系数是这个码字符号(最小次数)的生成多项式；证明所有这个码的码字通过某种特别的方式与 $g(X)$ 相关。

进一步证明当且仅当 $g(X)$ 满足一个特定条件(需要被证明)，它可能是一个长度为 N 的循环码的生成多项式。

至少有三种方式根据已知的生成多项式确定这个码的奇偶校验矩阵。解释其中一种。

解答 令 $\mathscr{X}$ 是长度为 N，次数为 d 的生成多项式 $g(X)=\sum_{0\leq i\leq d}g_iX^i$ 的循环码，不失一般性，假设码是非平凡的，有 $1<d<N-1$。令 $\boldsymbol{g}$ 表示相应的码字 $g_d\cdots g_d0\cdots0$(这里有 $d+1$ 系数 g_i 填补了 $N-d-1$ 个0)，那么

(a) 对于次数为 $k=N-d$ 的某些多项式 $h(X)=\sum_{0\leq i\leq k}h_iX^i$，$g(X)\mid(X^N-1)$ 也就是 $g(X)h(X)=X^N-1$；

(b) 字符串 $\boldsymbol{a}=a_0\cdots a_{N-1}\in\mathscr{X}$ 当且仅当多项式 $a(X)=\sum_{0\leq i\leq N-1}a_iX^i$ 有形如 $a(X)=f(X)g(X)\bmod(X^N-1)$；

(c) 字符串 $\boldsymbol{g}$ 和它的循环移位 $\pi\boldsymbol{g}$，…，$\pi^{k-1}\boldsymbol{g}$(对应多项式 $g(X)$，$Xg(X)$，…，$X^{k-1}g(X)$)形成 $\mathscr{X}$ 的一个基础。

根据(a)，$g_0=h_0=1$，并且对于所有的 $l=1$，…，$N-1$，表示 $g(X)h(X)$ 的第 l 项系数的和 $\sum_{0\leq i\leq l}g_ih_{l-i}$ 等于0。根据(c)，$\mathscr{X}$ 的秩等于 k。

263 一种确定奇偶校验矩阵的方法是取比值 $(X^N-1)/g(X)=h(X)=h_0+h_1X+\cdots+h_kX^k$，那么可以形成 $N\times(N-k)$ 矩阵

$$\boldsymbol{H}=\begin{bmatrix} h_k & h_{k-1} & \cdots & 0 & \cdots & 0 & 0 \\ 0 & h_k & h_{k-1} & \cdots & h_1 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & h_k & \cdots & h_1 & h_0 \end{bmatrix} \tag{2.5.9}$$

$\boldsymbol{H}$ 中的行是字符串 $\boldsymbol{h}^T=h_k\cdots h_10\cdots0$ 的循环移位 $(\pi^j\boldsymbol{h}^T)$，$0\leq j\leq d-1=N-k-1$。

我们声明对于所有的 $\boldsymbol{a}\in\mathscr{X}$，$\boldsymbol{a}\boldsymbol{H}^T=0$。事实上，它足以检验对于基本码字 $\pi^j\boldsymbol{g}$，$\pi^j\boldsymbol{g}H^T=0$，$j=0$，…，$k-1$。即点积为

$$\pi^{j_1}\boldsymbol{g}\cdot\pi^{j_2}\boldsymbol{h}^T=0,0\leq j_1<k,0\leq j_2<N-k-1 \tag{2.5.10}$$

但是对于 $j=k-1$ 和 $j_2=0$，我们有

$$\pi^{k-1}\boldsymbol{g}\cdot\boldsymbol{h}^T=g_kh_k+g_{k-1}h_{k-1}=0$$

因为它给出乘积 $g(X)h(X)=X^N-1$ 的第一个系数(在单项式 X)。同样，对于 $j_1=k-2$ 和 $j_2=0$，$\pi^{k-2}\boldsymbol{g}\cdot\boldsymbol{h}^{\dagger}$ 给出 $g(X)h(X)$ 的第二个系数(在单项式 X^2)并且也等于 0。并且，对于 $j_1=j_2=0$，$\boldsymbol{g}\cdot\boldsymbol{h}^{\dagger}=0$ 是 $g(X)h(X)$ 的第 k 次的系数。

继续，$\boldsymbol{g}\cdot\pi\boldsymbol{h}^{\dagger}$ 等于 $g(X)h(X)$，$\boldsymbol{g}\cdot\pi^2\boldsymbol{h}^{\dagger}$ 的第 $(k+1)$ 次的系数，$\boldsymbol{g}\cdot\pi^7\boldsymbol{h}^{\dagger}$ 第 $(k+2)$ 次，等等；$\boldsymbol{g}\cdot\pi^{N-k-1}\boldsymbol{h}^{\dagger}=g_{k-1}h_k+g_kh_{k-1}$，第 $(N-1)$ 次。和前面的一样，它们全都为零。

当我们同时循环移动两个码字(如果可能的话)，同样正确，它将会导出(2.6.10)。

相反，假设对于某些码字 $\boldsymbol{a}=a_0\cdots a_{N-1}$ 有 $\boldsymbol{aH}^{\mathrm{T}}=0$。写出对应的多项式 $a(X)=f(X)g(X)+r(X)$，其中比值是 $f(X)=\sum_{0\le i\le k-1}f_iX^i$，$r(X)$是余项。那么或者 $r(X)=0$ 或者 $1\le\deg r(X)=d'<d$(并且对于 $d'<l\le n-1$，有 $r_d'=1$，和 $r_l=0$)。然后集合 $r=r_0\cdots r_d'$。

假设 $r(X)\neq0$，根据上面的讨论

(i) $\boldsymbol{aH}^{\mathrm{T}}=\boldsymbol{rH}^{\mathrm{T}}$，因此 $\boldsymbol{rH}^{\mathrm{T}}=0$。

(ii) 向量 $\boldsymbol{rH}^{\mathrm{T}}$ 的元素与乘积 $r(X)h(X)$ 的系数相同，从 $r_0h_k+\cdots+r_dh_{k-d}$ 开始到 r_dh_k 结束。所以，这些系数一定是 0。但是等式 $r_dh_k=0$ 是不可能的，因为 $r_d=h_k=1$。因此，$r(X)=0$ 并且 $a(X)=f(X)g(X)$，即 $\boldsymbol{a}\in\mathscr{C}$。我们得出结论：$\boldsymbol{H}$ 是$\mathscr{C}$的奇偶校验矩阵。

同样，$\boldsymbol{H}$ 是由码字对应的多项式 $X^{k-1}h^{\dagger}(X)$形成的矩阵，其中

$$h^{\dagger}(X)=\sum_{0\le i\le k}h_iX^{k-i}$$

264

作为另一种选择，令 $h(X)$是长度为 N、生成多项式为 $g(X)$的循环码$\mathscr{C}$的校验多项式，使得 $g(X)h(X)=X^N-1$。那么：

(a) $\mathscr{C}^{\perp}=\{f(X):\ f(X)h(X)=0\bmod(X^N-e)\}$；

(b) 如果 $h(X)=h_0+h_1X+\cdots+h_{N-1}X^{N-1}$，那么奇偶校验矩阵 $\boldsymbol{H}$ 具有(2.6.9)的形式；

(c) 对偶码$\mathscr{C}^{\perp}$是一个 $\dim\mathscr{C}^{\perp}=r$ 的循环码，并且$\mathscr{C}^{\perp}=(h^{\dagger}(X))$，其中 $h^{\dagger}(X)=h_k^{-1}X^{N-k}h(X^{-1})=h_k^{-1}(h_kX^{N-k}+h_1X^{N-k+1}+\cdots+h_{N-k})$。 □

问题 2.19 考虑 Hamming$[2^\ell-1,\ 2^\ell-\ell-1]$二元码的奇偶校验矩阵 $\boldsymbol{H}$。通过给矩阵 $\boldsymbol{H}$ 增加一个零列和一个全一行生成$[2^\ell,\ 2^\ell-\ell-1]$码的奇偶校验矩阵 $\boldsymbol{H}'$。所产生的码的对偶被称为一阶 Reed-Muller 码。证明一个一阶 Reed-Muller 码可以纠正每个码字的$2^{\ell-2}-1$个错误。

这样的码在 1972 年被用于 Mariner 飞行器拍摄的火星照片。码率是多少？为什么有可能比信道的容量低很多？

解答 问题中的码是$[2^\ell,\ \ell+1,\ 2^{\ell-1}]$码，其中 $\ell=5$。信息速率等于 $6/32\approx5$。我们验证所有除了 0 和 1 之外的码字的权重为 $2^{\ell-1}$。对于 $\ell\ge1$ 的情况，码$\mathscr{R}(\ell)$通过递归形式来定义

$$\mathscr{R}(\ell+1)=\{xx\mid x\in\mathscr{R}(\ell)\}\cup\{x,x+1\mid x\in\mathscr{R}(\ell)\}$$

所以，$\mathscr{R}(\ell+1)$中的码字的长度很明显是 $2^{\ell+1}$。因为$\{xx\mid x\in\mathscr{R}(\ell)\}$和$\{x,\ x+1\mid x\in\mathscr{R}(\ell)\}$是不相交的，码字的数量翻倍，也就是 $\#\mathscr{R}(\ell+1)=2^{\ell+2}$。最后，假设$\mathscr{R}(\ell)$中除 0 和 1 之外所有码字的权重为 $2^{\ell-1}$，考虑一个码字 $y\in\mathscr{R}(\ell+1)$。如果 $y=xx$ 与码字 0 和 1 不同，那么 $x\neq0$ 或 1，并且因此有 $w(y)=2w(x)=2\times2^{\ell-1}=2^\ell$。

如果 $y=x,\ x+1$，我们必须考虑某些情况。如果 $x=0$，那么 $y=01$，权重为 2^ℓ。如果 $x=1$，那么 $y=10$，权重依然为 2^ℓ。最后，如果 $x\neq0$ 或 1，那么 $w(x+1)=2^\ell-2^{\ell-1}=2^{\ell-1}$并且 $w(y)=2\times2^{\ell-1}=2^\ell$。现在 $w(x)=2^{\ell-1}$的码字 xx 和 $x,\ x+1$ 与奇偶校验矩阵 $\boldsymbol{H}'$ 的行相正交就很明显了。

达到 7 比特可能会出现错误，因此传输错误概率 p_e(一个二进制对称无记忆信道的错

误概率为 p)服从

$$p_e \leqslant 1-\sum_{0\leqslant i\leqslant 7}\binom{32}{i}p^i(1-p)^{32-i}$$

当 p 很小时上式很小。(对于一个可以接受的估计 p，我们可以求解 $1+p\log p+(1-p)\log(1-p)=26/32$。)如果块的长度是固定的(或者很小)，对于很小的 p 值，我们不能接近它的容量。

265

确实，对于 $\ell=5$，这个码是[32，6，15]，可以检测 15 个错并且可以纠正 7 个错。也就是说，这个码可以纠正的部分大于全部 32 个数字的 1/5。信息速率为 6/32，如果(无记忆)信道的容量是 $C=1-\eta(p)$(其中 p 表示错误符号概率)，我们要求 $C>6/32$；即对于可靠传输来说，$\eta(p)+6/32<1$。这会导致 $|p-1/2|>|p^*-1/2|$，其中 $p^*\in(0,1)$，可求解 $26/32=\eta(p^*)$。很明显 $0\leqslant p<1/5$ 和 $4/5<p\leqslant 1$ 也以此类推，但事实上错误概率要低得多。 □

问题 2.20 证明任意二进制[5，M，3]码一定有 $M\leqslant 4$。证明恰好存在一个使等号成立的是[5，4，3]码。

解答 根据 Plokin 界，如果 d 是奇数并且 $d>\frac{1}{2}(N-1)$，那么

$$M_2^*(N,d)\leqslant 2\left\lfloor\frac{d+1}{2d+1-N}\right\rfloor$$

事实上

$$M_2^*(5,3)\leqslant 2\left\lfloor\frac{4}{6+1-5}\right\rfloor=2\cdot 2=4$$

所有的[5，4，3]等价于 00000，00111，11001，11110。 □

问题 2.21 令$\mathscr{C}$为一个生成矩阵为 $\boldsymbol{G}$ 的二进制[N，k，d]线性码。证明我们可能假设 $\boldsymbol{G}$ 的第一行是 1…1 0…0 的形式，其中有 d 个 1，可写为

$$\boldsymbol{G}=\begin{pmatrix}1\cdots1 & 0\cdots0\\ \boldsymbol{G}_1 & \boldsymbol{G}_2\end{pmatrix}$$

证明如果 d_2 是生成矩阵为 $\boldsymbol{G}_2$ 的码的距离，那么 $d_2\geqslant d/2$。

解答 令$\mathscr{C}$为[N，k，d]。我们总是可以构造一个 X 的生成矩阵 $\boldsymbol{G}$，其中第一行是一个码字 x，满足 $w(x)=d$；通过排列组合 $\boldsymbol{G}$ 中的列我们可以得到形如 $\underbrace{1\cdots1}_{d}\underbrace{0\cdots0}_{N-d}$ 的第一行。所以，等号成立

$$\boldsymbol{G}=\begin{pmatrix}1\cdots1 & 0\cdots0\\ \boldsymbol{G}_1 & \boldsymbol{G}_2\end{pmatrix}$$

假设 $d(\boldsymbol{G}_1)<d/2$，那么为不失一般性，我们可以假设存在一个($\boldsymbol{G}_1\boldsymbol{G}_2$)的行，其中数字 $d+1$，…，N 中 1 的个数$<d/2$，那么这一行数字 1，…d 中的 1 的个数$>d/2$，因为它的总码重$\geqslant d$。然后将这一行和 1…10…0 加起来得到一个码重$<d$ 的码字。所以，$d(\boldsymbol{G}_2)\geqslant d/2$。 □

265

问题 2.22(Gilbert Varshamov 限) 证明如果 $p^k<2^N/v_{N-1}(d-2)$，那么存在一个 p 进制线性[N，k，d]码。因此，如果 p^k 是满足这个不等式的 p 的最大幂，那么我们有 $M_p^*(N,d)\geqslant p^k$。

解答 我们通过选取长度为 $N-k$ 的 N 列来构造一个奇偶校验矩阵，其中这 N 列中没有 $d-1$ 列是线性相关的。第一列可以是 $\mathbb{Z}_p^{N-k}$ 中的任意非零串，在第 i 步($i\geqslant 2$)我们一定选取一个不是前面已选列中任意 $d-2$(或者更少)列的线性组合的列。这种线性组合(系数不为零)的个数为

$$S_i=\sum_{j=1}^{d-2}\binom{i-1}{j}(p-1)^j$$

所以，当且仅当 $S_{N+1}<p^{N-k}$ 时，可以构造奇偶校验矩阵。最终，我们发现 $S_{N+1}<v_{N-1}(d-2)$。就是说如果 $2^r<32/5$，那么存在 $[5,2^r,3]$ 码，所以 $k=2$ 并且 $M_2^r(5,3)\geq 4$，实际上这是很明显的。 □

问题 2.23 如果一个元素 $b\in\mathbb{F}_q^*$ 的阶（也就是满足 $b^k=1 \bmod q$ 的最小的 k）是 $q-1$，那么这个元素被称为本原的。不妨在乘法群 $\mathbb{F}_q^*$ 中明确地找到一个本原元素。考虑素因子分解

$$q-1=\prod_{j=1}^{s}p_j^{\lambda_j}$$

对于任意 $j=1,\cdots,s$，选择 $a_j\in\mathbb{F}_q$ 使得 $a_j^{(q-1)/p_j}\neq e$。令 $b_j=a_j^{(q-1)/p_j^{\lambda_j}}$，验证 $b=\prod_{j=1}^{s}b_j$ 的阶为 $q-1$。

解答 确实，b_j 的阶为 $p_j^{\lambda_j}$。接下来，如果对某个 n，$b^n=1$，那么 $n\equiv 0 \bmod p_j^{\lambda_j}$，因为 $b_j^{\prod_{i\neq j}p_i^{\lambda_i}}=1$ 表明 $b_j^{\prod_{i\neq j}p_i^{\lambda_i}}=1$，即 $n\prod_{i\neq j}p_i^{\lambda_i}\equiv 0 \bmod p_j^{\lambda_j}$。因为 p_j 是独特的素数，对于任意 j，它都有 $n\equiv 0 \bmod p_j^{\lambda_j}$，因此 $n=\prod_{j=1}^{s}p_j^{\lambda_j}$。 □

问题 2.24 具有一个本原根的最小多项式被称为一个本原多项式。验证在次数为 4 的不可约二进制多项式（见式(2.5.9)）中，$1+X+X^4$ 和 $1+X^3+X^4$ 是本原多项式，$1+X+X^2+X^3+X^4$ 不是。验证所有的次数为 5 的六个不可约二进制多项式（见式(2.5.15)）是本原多项式；在实际中，人们会偏向于用 $1+X^3+X^5$ 来作为计算的模，这个多项式唯能会变组一些。验证式(2.5.16)中所有次数为 6 的不可约多项式中有六个本原多项式；它们位于最上面三行。证明对于每个给定的次数都存在一个原始多项式。 267

解答 对于最后一部分的解，见 3.1 节。 □

问题 2.25 一个长度为 N 的循环码 $\mathscr{C}$，生成多项式 $g(X)$ 的次数 $d=N-k$ 可以按照 $g(X)$ 的根来表示，即元素 $\alpha_1,\cdots,\alpha_N$，使得 $g(\alpha_j)=0$。这些元素被称为码 $\mathscr{C}$ 的零元素，属于 Galois 域 $\mathbb{F}_{2^r}$。因为 $g(X)\mid(1+X^N)$，它们也是 $1-X^N$ 的根。那样就有 $\alpha_j^N=1$，$1\leq j\leq N-k$，即 α_j 是第 N 个单位根。剩余的 k 个单位根 $\alpha_1',\cdots,\alpha_k'$ 被称为 $\mathscr{C}$ 的非零根。当且仅当在 Galois 域 $\mathbb{F}_{2^r}$ 中 $a(\alpha_j')=0$，$1\leq j\leq N-k$ 时，多项式 $a(X)\in\mathscr{C}$。

(a) 证明如果 $\mathscr{C}^\perp$ 是对偶码，那么 $\mathscr{C}^\perp$ 的零元素是 $\alpha_1^{-1},\cdots,\alpha_k^{-1}$，即 $\mathscr{C}$ 的非零元素的倒数。

(b) 如果对于所有的 $x=x_0\cdots x_N\in\mathscr{C}$，码字 $x_{N-1}\cdots x_0\in\mathscr{C}$，那么生成多项式为 $g(X)$ 的循环码 $\mathscr{C}$ 是可逆的。证明当且仅当 $g(\alpha)=0$ 时 $\mathscr{C}$ 是可逆的，这意味着 $g(\alpha^{-1})=0$。

(c) 证明一个长度为 N 的 q 进制循环码 $\mathscr{C}$，其中 $(q,N)=1$，通过排列组合这些数字使得 $\pi_q(i)=qi \bmod N$（也就是 $x\to x^q$）的情况下是不变的。如果 $s=\mathrm{ord}_N(q)$，那么两种排列 $i\to i+1$ 和 $\pi_q(i)$ 在码的自同构群 $\mathrm{Aut}(\mathscr{C})$ 中产生一个阶数为 Ns 的子群。

解答 的确，因为 $a(x^q)=a(x)^q$ 与相同的生成多项式成正比，所以它和 $a(x)$ 属于相同的循环码。 □

问题 2.26 证明有 129 个长度为 128 不相等的循环二元码（包含平凡码 $\{0\cdots 0\}$ 和 $\{0,1\}^{128}$）。找出所有长度为 7 的循环二元码。

解答 长度为 2^r 的循环码的等价类一一对应着 $1+X^{2^r}$ 的因子，它们的数目等于 2^r+1。此外，这里列出了 X^7-1 的除数作为生成多项式的 8 个码

$$X^7-1=(1+X)(1+X+X^3)(1+X^2+X^3)$$ □ 268

编码理论的深层主题

3.1 有限域入门

在这部分，我们将介绍有限域的概要，通过之后的章节和下面的标准文本(见文献[92]、[93]、[131])所需要的材料来限制我们的讨论范围。有限域是一个(有限)集合 $\mathbb{F}$，集合 $\mathbb{F}$ 有两个不同的元素，0(零)元素和 e(单位)元素，并配备与标准分配律相关的加和乘(其中对于所有 $b\in\mathbb{F}$ 有 $0\cdot b=0$)的两个交换群运算。

域 $\mathbb{F}$ 中的向量空间是一个(有限)集合 $\mathbb{V}$，配备一个加的可交换群运算和一个与 $\mathbb{F}$ 中的元素进行标量乘的运算，也服从标准的分配律。$\mathbb{V}$ 的维度 $\dim\mathbb{V}$ 是最小数 d，它使得不同元素 v_1，…，$v_{d+1}\in\mathbb{V}$的任意集合是线性相关的，即可以找到不全等于 0 的元素 k_1，…，$k_{d+1}\in\mathbb{F}$ 使得 $k_1v_1+\cdots+k_{d+1}v_{d+1}=0$。那么存在元素为 b_1，…，$b_d\in\mathbb{V}$ 的集合，若能使得每一个 $v\in\mathbb{V}$ 都可以写成一个线性组合 $a_1b_1+\cdots+a_db_d$，其中 a_1，…，a_d 是由 v 所确定的 $\mathbb{F}$ 中的元素(唯一的)，则称为基础。除非有相反的规定，否则我们认为域是同构的。

域的一个重要的参数是它的特征值，即使得 $pe=e+\cdots+e(p\text{ 次})=0$ 的最小整数 $p\geqslant 1$。这个数字用符号 $\mathrm{char}(\mathbb{F})$表示，根据标准鸽巢原理可证明它的存在性。而且，特征值是一个素数，如果 $p=q_1q_2$，那么 $pe=(q1q2)e=(q1e)(q2e)=0$，这表明 $q_1e=0$ 或者 $q_2e=0$，导致矛盾。

例子 3.1.1 令 p 为一个素数。一个加性循环群 $\mathbb{Z}_p=\{0,1,\cdots,p-1\}$，生成子为 1，通过乘法$(qe)(q'e)=(qq')e$ 成为一个域。这个域的特征等于 p。

令 $\mathbb{K}$ 和 $\mathbb{F}$ 是两个域。如果 $\mathbb{F}\subset\mathbb{K}$，我们可以说 $\mathbb{K}$ 是一个 $\mathbb{F}$ 的扩展，那么 $\mathbb{K}$ 也是 $\mathbb{F}$ 上的一个向量空间，它的维度表示为$[\mathbb{K}:\mathbb{F}]$。

269

引理 3.1.2 令 $\mathbb{K}$ 为 $\mathbb{F}$ 的扩展，并且 $d=[\mathbb{K}:\mathbb{F}]$，那么 $\#\mathbb{K}=(\#\mathbb{F})^d$。

证明 在 $\mathbb{F}$ 中，令 b_1，…，b_d 是 $\mathbb{K}$ 的一组基。对于所有的 $k\in\mathbb{K}$ 有唯一的表达式 $k=\sum_{1\leqslant j\leqslant d}a_jb_j$。那么对于所有的 j，我们对于 a_j 有 $\#\mathbb{F}$ 个可能的选择。所以，恰好一共存在 $(\#\mathbb{F})^d$ 种方法去写所有的组合。 □

引理 3.1.3 对于某个整数 $d\geqslant 1$，如果 $\mathrm{char}(\mathbb{F})=p$，那么 $\#\mathbb{F}=p^d$。

证明 已知元素 0，e，$2e$，…，$(p-1)e$，它们组成$\mathbb{Z}_p$，即$\mathbb{Z}_p\subseteq\mathbb{F}$，那么由引理 3.1.2 可知，$\#\mathbb{F}=p^d$。 □

推论 3.1.4 对于一个有限域 $\mathbb{F}$，其包含元素的数量一定是 $q=p^r$，其中 $p=\mathrm{char}(\mathbb{F})$，$r\geqslant 1$ 是一个自然数。

从现在开始，除非特别地声明，p 代表一个素数，$q=p^r$ 代表素数幂。

引理 3.1.5 (Freshman's Dream)对于所有 a，$b\in\mathbb{F}$，$n\geqslant 1$，如果 $\mathrm{char}(\mathbb{F})=p$，那么

$$(a\pm b)^{p^n}=a^{p^n}+(\pm b)^{p^n} \tag{3.1.1}$$

证明 利用归纳法，当 $n=1$，

$$(a\pm b)^p=\sum_{0\leqslant k\leqslant p}\binom{p}{k}a^k(\pm b)^{p-k}$$

对于 $1\leqslant k\leqslant p-1$，$\binom{p}{k}$ 是 p 的倍数，(利用有限域特征的性质)与 p 倍数相关的项不存在。所以 $(a\pm b)^p=a^p-(\pm b)^p$，归纳的步骤用相同的方法完成，用 $a^{p^{r-1}}$ 和 $(\pm b)^{p^{r-1}}$ 代替 a 和 $\pm b$ 可证。 □

引理 3.1.6 从大小为 q 的域 $\mathbb{F}$ 中取非零元素具有乘法运算的群 $\mathbb{F}^*$，$\mathbb{F}^*$ 与循环群 $\mathbb{Z}_{q-1}$ 是同构的。

证明 对于任何因子 $d\mid(q-1)$，群 $\mathbb{F}^*$ 含有 $\phi(d)$ 个乘阶为 d 的元素，其中 ϕ 是欧拉函数。(回忆 $\phi(d)=\#\{k:\ k<d,\ \gcd(k,\ d)=1\}$)。可以发现阶数为 d 的所有元素都有 $a^{\frac{r}{d}}$ 的形式，其中 a 是一个本原元，$r\leqslant d$，r，d 互素。实际上，$q-1=\sum_{d,d\mid(q-1)}\phi(d)$，$\mathbb{F}^*$ 至少有一个阶数为 $q-1$ 的元素，这暗示着 $\mathbb{F}^*$ 是循环的且阶数为 $q-1$。 □

令 $a\in\mathbb{F}^*$，阶数为 d，其中 $d\mid(q-1)$。取循环子群 $\{e,\ a,\ \cdots,\ a^{d-1}\}$，此子群中的每个元素都有可被 d 整除的乘法阶，即这些元素是多项式 X^d-e 的解(第 d 个单位根)。但是在 $\mathbb{F}$ 中 X^d-e 有 $\leqslant d$ 个不同的根(因为 $\mathbb{F}$ 是一个域)。所以，在 $\mathbb{F}$ 中，$\{e,\ a,\ \cdots,\ a^{d-1}\}$ 是所有多项式 X^d-e 解的集合。特别地，$\mathbb{F}$ 中的每个阶为 d 的元素都属于 $\{e,\ a,\ \cdots,\ a^{d-1}\}$。 [270]
观察到循环群 $\mathbb{Z}_d$ 恰好有 $\phi(d)$ 个阶数为 d 的元素，所以，整个 $\mathbb{F}^*$ 正好有 $\phi(d)$ 个阶数为 d 的元素；换句话说，如果 $\psi(d)$ 是 $\mathbb{F}$ 中阶数为 d 元素的数量，那么 $\psi(d)=0$ 或者 $\psi(d)=\phi(d)$，而且

$$q-1=\sum_{d,d\mid n}\psi(d)\leqslant\sum_{d,d\mid n}\phi(d)=q-1$$

这也说明对于所有 $d\mid n$ 都有

$$\psi(d)=\phi(d)$$

定义 3.1.7 $\mathbb{F}^*$ 的(乘法)发生元(即一个乘法阶为 $q-1$ 的元素)被称为集 $\mathbb{F}$ 中的本原元。虽然这样的元素不是唯一的，但是通常我们会单独选一个这样的元素并用 ω 来表示；当然 ω 的一个 r 次幂 ω^r 也是一个本原元，其中 r 与 $q-1$ 互素。

如果 $a\in\mathbb{F}^*$，并且 $\#\mathbb{F}^*=q-1$，那么 $a^{q-1}=e$(每个元素的阶数都能整除群的阶数)。所以，$a^q=a$，其中 a 是 $\mathbb{F}$ 中多项式 X^q-X 的根，但是多项式 X^q-X 只可能含有 $\leqslant q$ 个根(包括 0)，所以 $\mathbb{F}$ 是所有 X^q-X 根的集合。

定义 3.1.8 已知域 $\mathbb{K}$ 和 $\mathbb{F}$，且 $\mathbb{F}\subseteq\mathbb{K}$，域 $\mathbb{K}$ 称为多项式 $g(X)$ 的分裂域，$g(X)$ 的系数取自 $\mathbb{F}$，如果(a)$\mathbb{K}$ 含有 $g(X)$ 的所有根，(b)不存在同时满足 $\mathbb{F}\subset\widetilde{\mathbb{K}}\subset\mathbb{K}$ 和(a)的域 $\widetilde{\mathbb{K}}$。对于 $g(X)$ 的分裂域，我们记为 $\mathrm{Spl}(g(X))$。

所以，如果 $\#\mathbb{F}=q$，那么 $\mathbb{F}$ 包含多项式 X^q-X 所有的根，同时也是此多项式的分裂域。

引理 3.1.9 $\mathbb{F}$ 上多项式 $g(X)$ 的任意两个分裂域 $\mathbb{K}$，$\mathbb{K}'$ 相同。

证明 实际上，求交集 $\mathbb{K}\cap\mathbb{K}'$，交集包含了 $\mathbb{F}$ 且是 $\mathbb{K}$，$\mathbb{K}'$ 的子集，该子集必然与 $\mathbb{K}$，$\mathbb{K}'$ 中的一个重合。 □

推论 3.1.10 对于任意素数 p 和自然数 $s\geqslant1$，存在至多一个具有 p^s 个元素的域。

证明 满足上述条件的每一个域都是 $\mathbb{Z}_p$ 上的多项式 X^q-X，$q=p^s$ 的分裂域。因此，任意两个这样的分裂域都相同。 □

另一方面，我们会在后面证明下列定理。

定理 3.1.11 对于 $\mathbb{F}$ 上的任何非常数多项式，存在分裂域。 [271]

推论 3.1.12 对于任何素数 p 和自然数 $s\geq 1$，存在一个恰好含有 p^s 个元素的域。

证明 再次利用 $\mathbb{Z}_p$ 上的多项式 X^q-X，$q=p^s$。通过定理 3.1.11，存在一个分裂域 $\mathrm{Spl}(X^q-X)$，其中 $(X^q-X)=X(X^{q-1}-e)$ 可以分解为多个线性多项式。所以，$\mathrm{Spl}(X^q-X)$ 含有 X^q-X 的根，并且特征为 p（因为它含有 $\mathbb{Z}_p$）。

但是，X^q-X 的根组成了一个子域：如果 $a^q=a$，$b^q=b$，可得 $(a\pm b)^q=a^q+(\pm b^q)$（引理 3.1.5），这与计算 $a\pm b$ 的情形一样。同时，$(ab^{-1})^q=a^q(b^q)^{-1}=ab^{-1}$。这个域不能严格地包含于 $\mathrm{Spl}(X^q-X)$，所以它与 $\mathrm{Spl}(X^q-X)$ 重合。

接下来需要检查多项式 X^q-X 的根是否不同：检验域的阶 $\#\mathrm{Spl}(X^q-X)$ 是否等于 q。实际上，如果 X^q-X 有重根，那么多项式会有一个公共因子，且求导为 $\partial_X(X^q-X)=qX^{q-1}-e$。但是，在 $\mathrm{Spl}(X^q-X)$ 中 $qX^{q-1}=0$，所以不可能有这样的因子。 □

总结而言，对于有限域我们有以下两个特征定理。

定理 3.1.13 所有有限的域的大小都可表示为 p^s，其中 p 是一个素数，s 是大于等于 i 的整数。对于所有这样的 p，s，存在唯一一个大小恰好为 p^s 的域。

具有大小 $q=p^s$ 的域用 $\mathbb{F}_q$ 表示（另一运用广泛的表示方法是 $GF(q)$（一个 Galois 域））。对于最基本的域 $\mathbb{F}_p=\{0,1,\cdots,p-1\}$（$p$ 是一个素数），我们用 1 代替 e 来表示单位元。

定理 3.1.14 可将所有有限域有序排列（至"塔"状），对于一个素数 p 和正整数 s_1，s_2，$\cdots$，有

$$\begin{tikzcd}
\cdots \\
\cdots \\
\cdots \\
\mathbb{F}_{p^{s_1+s_2+\cdots}} \arrow[u] \\
\cdots \arrow[u] \\
\mathbb{F}_{p^{s_1+s_2}} \arrow[u] \\
\mathbb{F}_{p^{s_1}} \arrow[u] \\
\mathbb{F}_p \simeq \mathbb{Z}_p \arrow[u]
\end{tikzcd}$$

272 此处每一个箭头代表一个专门定义的单射同态。

例子 3.1.15 在 2.5 节中我们讨论了多项式域 $\mathbb{F}_2[X]/\langle q(X)\rangle$，其中 $q(X)$ 是一个不可约的二元多项式；可参考定理 2.5.32 和 2.5.33。继续讨论例子 2.5.35(c)，考虑由 $\mathbb{F}_2[X]/\langle 1+X^3+X^4\rangle$ 生成的一个域 $\mathbb{F}_{16}$。该域的结构如下：

X 的幂 mod $1+X^3+X^4$	多项式（字符串）	向量（字符串）
	0	0000
X^0	1	1000
X	X	0100
X^2	X^2	0010
X^3	X^3	0001
X^4	$1+X^3$	1001

X 的幂 mod $1+X^2+X^4$	多项式(字符串)	向量(字符串)
X^0	$1+X+X^2$	1101
X^1	$1+X+X^2+X^3$	1111
X^2	$1+X+X^3$	1119
X^3	$X+X^2+X^3$	0111
X^4	$1+X^2$	1010
X^5	$X+X^3$	0101
X^6	$1+X^2+X^3$	1011
X^7	$1+X$	1100
X^8	$X+X^2$	0110
X^9	X^2+X^3	0011

(3.1.2)

最后，如果我们选择确定$\mathbb{F}_2[X]/(1+X+X^2+X^3+X^4)$对应的表格，计算过程会很复杂(内容与结构也会有所不同)。值得指出的是，由于$X^5=1 \bmod (1+X+X^2+X^3+X^4)$，单项式$X$在这种情况下不是一个本原元。事实上，乘法群的生成元可表示为单项式的和，即$1+X$。

举例 3.1.16 (a) $\mathbb{F}_2$ 包含多项式 X^2+X+1 和 X^2+X+1 的所有根，试问$\mathbb{F}_2$ 最小扩域中元素数量是多少？

(b) 计算$\mathbb{F}_{1024}$，$\mathbb{F}_{128}$子域的数量。找出$\mathbb{F}_7$，$\mathbb{F}_9$，$\mathbb{F}_{16}$中所有的本元原。计算$(\omega^{10}+\omega^5)(\omega^4+\omega^2)$，其中$\omega$是$\mathbb{F}_{16}$的本元原。

解答 (a) 容易得到该数量为5^6。

(b) $\mathbb{F}_{1024}=\mathbb{F}_{2^{10}}$有四个子域：$\mathbb{F}_2$，$\mathbb{F}_4$，$\mathbb{F}_{32}$和$\mathbb{F}_{1024}$。$\mathbb{F}_{128}=\mathbb{F}_{2^7}$有四个子域：$\mathbb{F}_2$，$\mathbb{F}_8$，$\mathbb{F}_{32}$，$\mathbb{F}_{128}$。$\mathbb{F}_7$ 有 2 个本元原：ω，ω^2(其中$(\omega^3)^2=\omega$)。 273

F_9 有四个本元原，由ω，ω^2，ω^3，ω^7 组成。F_{16}有 8 个本元原：ω，ω^2，ω^4，ω^7，ω^8，ω^{11}，ω^{13}，ω^{14}。

利用$\mathbb{F}_2[X]/(1+X+X^4)$的表格(参考例子 2.5.35(c))，可得

$$(\omega^{10}+\omega^5)(\omega^4+\omega^2)=\omega^{14}+\omega^9+\omega^{12}+\omega^7$$
$$=1001+0101+1111+1101=1110=\omega^{10}$$

但是，利用$\omega^7=\omega^2$，RHS 为

$$\omega^8+\omega^5+\omega^3+\omega^4=1010+0001+0101+1100=0010=\omega^2=(\omega^7)^{11}$$

从现在开始，我们将关注有限域的多项式表示。一般化 2.5 节中引入的概念，考虑以下问题。 □

定义 3.1.17 $\mathbb{F}_q$ 上的所有多项式的集合构成了一个交换环$\mathbb{F}_q[X]$。对一个固定的多项式$g(X)\in\mathbb{F}_q[X]$取模运算得到一个商环$\mathbb{F}_q[X]/(g(X))$。

定义 3.1.18 一个多项式$g(X)\in\mathbb{F}_q[X]$被称为$\mathbb{F}_q$上的既约多项式，如果它不可再做如下形式的分解

$$g(X)=g_1(X)g_2(X)$$

其中$g_1(X)$，$g_2(X)\in\mathbb{F}_q[X]$。

定理 2.5.32 的一般化的表述为如下定理 3.1.9。

定理 3.1.19 令$g(X)\in\mathbb{F}_q[X]$的阶数为$\deg g(X)=d$，那么当且仅当$g(X)$不可约时，$\mathbb{F}_q[X]/(g(X))$是一个域$\mathbb{F}_{q^d}$。

证明 令$g(X)$是一个$\mathbb{F}_q$上的一个既约多项式，为了证明$\mathbb{F}_q[X]/(g(X))$是一个域，我们应该检验其中每个非零元素$f(X)\in\mathbb{F}_q[X]/(g(X))$存在逆。考虑一个由$f(X)h(X) \bmod$

$g(X)$组成的多项式集合 $\mathbb{F}(f)$，其中 $h(X)\in\mathbb{F}_q[X]/\langle g(X)\rangle$（由 $f(X)$生成的主理想）。如果 $\mathbb{F}(f)$包含单位元 $e\in\mathbb{F}_q[X]$（等于 e 的常数多项式），那么相应的 $h(X)=f(X)^{-1}$。如果 $\mathbb{F}(f)$不包含单位元 $e\in\mathbb{F}_q[X]$，从$\mathbb{F}_q[X]/\langle g(X)\rangle$到它本身的映射 $h(X)\to f(X)h(X)\bmod g(X)$不是满射。也就是说，对于某些不同的 $h_1(X)$，$h_2(X)$，$f(X)h_1(X)=f(X)h_2(X)$，即

[274]

$$f(X)(h_1(X)-h_2(X))=r(X)g(X)$$

那么当 $g(X)$不可约时有 $g(X)\nmid f(X)$或者 $g(X)\mid(h_1(X)-h_2(X))$。因此，要么 $f(X)=0\bmod g(X)$（矛盾），要么 $h_1(X)=h_2(X)\bmod g(X)$。所以，$\mathbb{F}_q[X]/\langle g(X)\rangle$是一个域。

逆命题也可以同样被证明：如果 $g(X)$不可约，那么$\mathbb{F}_q[X]/\langle g(X)\rangle$含有非零的 $g_1(X)$，$g_2(X)$，$g_1(X)g_2(X)=0$。那么$\mathbb{F}_q[X]/\langle g(X)\rangle$不可能是一个域。

$[\mathbb{F}_q[X]/\langle g(X)\rangle:\mathbb{F}_q]$的维数等于 d，即 $g(X)$的阶，所以$\mathbb{F}_q[X]/\langle g(X)\rangle=\mathbb{F}_{q^d}$。 □

举例 3.1.20 证明当且仅当 $\gcd(g(X),X^N-e)=e$ 时，$g(X)$在多项式环$\mathbb{F}_q[X]/\langle X^N-e\rangle$中存在逆。

解答 考虑一个$\mathbb{F}_q[X]/\langle X^N-e\rangle\to\mathbb{F}_q[X]/\langle X^N-e\rangle$的映射 $h(X)\mapsto h(X)g(X)\bmod\langle X^N-e\rangle$。如果它是满射，则存在 $h(X)$满足 $h(X)g(X)=e$ 且 $h(X)=g(X)^{-1}$。假设它不是满射，那么存在 $h^{(1)}(X)\neq h^{(2)}(X)\bmod\langle X^N-e\rangle$使得 $h^{(1)}(X)g(X)=h^{(2)}(X)g(X)\bmod\langle X^N-e\rangle$，即

$$(h^{(1)}(X)-h^{(2)}(X))g(X)=s(X)\langle X^N-e\rangle$$

因为$\langle X^N-e\rangle\nmid(h^{(1)}(X)-h^{(2)}(X))$，这意味着 $\gcd(g(X),X^N-e)=d(X)\neq e$。

相反地，如果 $\gcd(g(X),X^N-e)=d(X)\neq e$，那么基于等式 $h(X)g(X)=e\bmod\langle X^N-e\rangle$可得

$$h(X)g(X)=e+q(X)\langle X^N-e\rangle$$

其中，$d(X)\mid\mathrm{LHS}$，$d(X)\mid q(X)\langle X^N-e\rangle$，所以，$d(X)\mid e$，构成一个矛盾。所以$g(X)^{-1}$不存在。 □

例子 3.1.21 （继续探讨例子 2.5.19）此处有 6 个阶数位为 5 的二元既约多项式：

$$1+X^2+X^5,1+X^3+X^5,1+X+X^2+X^3+X^5$$
$$1+X+X^2+X^3+X^5,1+X+X^3+X^4+X^5$$
$$1+X^2+X^3+X^4+X^5 \tag{3.1.3}$$

那么存在 9 个阶数为 6 的既约多项式，可继续这种推演。计算高阶既约多项式是一项艰难的任务，显然有关这些多项式的大量表格已经可以在网上找到既约多项式。

下面证明定理 3.1.11。

定理 3.1.11 的证明 一个重要的事实是任何一个非常数多项式 $g(X)\in\mathbb{F}_q[X]$在$\mathbb{F}_q$的扩域中存在一个根。不失一般性，假设 $g(X)$是不可约的，阶数是 d。令$\mathbb{F}_q[X]/\langle g(X)\rangle=\mathbb{F}_{q^d}$为一个扩域。在这个域中，$g(\alpha)=0$，其中 α 是多项式 $X\in\mathbb{F}_q[X]/\langle g(X)\rangle$，因此 $g(X)$有一个根。

[275]

在$\mathbb{F}_{q^d}$中，可以用 $g(X)$去除 $X-\alpha$，利用相同的构造方法证明在某个$\mathbb{F}_{q^t}$，$t<d$ 的扩域中，$g_1(X)=g(X)/(X-\alpha)$有一个根。最后，构造一个域包含 $g(X)$的所有 d 个根，即构造一个分裂域 $\mathrm{Spl}(g(X))$。 □

定义 3.1.22 已知一个域 $\mathbb{F}\subset\mathbb{K}$ 和一个元素 $\gamma\in\mathbb{K}$，我们用 $\mathbb{F}(\gamma)$表示包含 $\mathbb{F}$ 和 γ 的最小域（显然 $\mathbb{F}\subset\mathbb{F}(\gamma)\subset\mathbb{K}$）。类似地，$\mathbb{F}(\gamma_1,\cdots,\gamma_t)$是包含 $\mathbb{F}$ 和元素 $\gamma_1,\cdots,\gamma_t\in\mathbb{K}$ 的最小的域。对于 $\mathbb{F}=\mathbb{F}_q$ 和 $\alpha\in\mathbb{K}$，令

$$M_{\alpha,\mathbb{F}}(X)=(X-\alpha)(X-\alpha^q)\cdots(X-\alpha^{q^{d-1}}) \tag{3.1.4}$$

其中 d 是使得 $\alpha^{q^d}-\alpha$ 的最小正整数(在引理 3.1.24 中证明将会证明这样 d 的总会存在)。

首一多项式是以 1 为最高次项系数的多项式。$\mathbb{F}$ 上关于 $\alpha\in\mathbb{K}$ 的极小多项式是一个唯一的首一多项式 $M_\alpha(X)(=M_{\alpha,\mathbb{F}}(X))\in\mathbb{F}[X]$，使得 $M_\alpha(\alpha)=0$，并且对于每一个满足 $g(\alpha)=0$ 的 $g(X)=\mathbb{F}[X]$，都有 $M_\alpha(X)\mid g(X)$。当 ω 是 $\mathbb{K}$(产生 $\mathbb{K}^*$)的一个本原元，$M_\omega(X)$ 称为 $\mathbb{F}$ 中的本原多项式。多项式 $P(X)\in\mathbb{F}[X]$ 的阶是使得 $P(X)\mid(X^n-e)$ 的最小的 n。

例子 3.1.23 (继续讨论例子 3.1.21)这个例子中，我们处理 $\mathbb{F}_2$ 上的多项式。既约多项式 X^3+X+1 是本原多项式且阶是 3。既约多项式 X^3+X+1 和 X^3+X^2+1 是本原多项式且阶是 7。X^4+X^3+1 和 X^4+X+1 也是本原多项式，阶为 15，然而 $X^4+X^3+X^2+X+1$ 不是本原多项式，阶是 5。(值得注意的是当 $d=4$ 时，X^3+X^2+1 和 X^4+X+1 的阶是 2^d-1；另一方面，在域 $\mathbb{F}_2[X]/(1+X+X^2+X^3+X^4)$ 中，元素 X 的阶等于 5，但是在 $\mathbb{F}_2[X]/(1+X+X^5)$ 中，其阶为 15。)所有在式(3.1.3)中罗列的六个多项式都是本原多项式，阶数是 31(即都出现在 $X^{31}+1$ 的因式分解项中)。

引理 3.1.24 *令 $\mathbb{F}_q\subset\mathbb{F}_{q^d}$，$\alpha\subset\mathbb{F}_{q^d}$。令 $M_\alpha(X)\in\mathbb{F}[X]$ 为 α 的极小多项式，次数是 $\deg M_\alpha(X)=d$，那么：*

(a) $M_\alpha(X)$ 是 $\mathbb{F}_q[X]$ 上在 α 处有根的唯一既约多项式。

(b) $M_\alpha(X)$ 是 $\mathbb{F}_q[X]$ 上在 α 处有根且次数为 d 的唯一的首一多项式。

(c) $M_\alpha(X)$ 的形式如式(3.1.4)所示。

276

证明 论断(a)，(b)可由定义得到。为了证明(c)，假设 $\gamma\in\mathbb{K}$ 是 $\mathbb{F}[X]$ 上多项式 $f(X)=a_0+a_1X+\cdots+a_dX^d$ 的一个根，即 $\sum_{0\le i\le d}a_i\gamma^i=0$。因为 $a_i^q=a_i$(对于所有 $a\in\mathbb{F}$ 都成立)，由引理 3.1.5

$$f(\gamma^q)=\sum_{0\le i\le d}a_i\gamma^{iq}=\sum_{0\le i\le d}(a_i\gamma^i)^q=\Big(\sum_{0\le i\le d}(a_i\gamma^i)\Big)^q=0$$

所以 γ^q 是一个根。类似地，$(\gamma^q)^q=\gamma^{q^2}$ 是一个根，以此类推。 □

对于 $M_\alpha(X)$，它产生的 α，α^q，α^{q^2}，…都是根。当第一次满足 $\alpha^{q^s}=\alpha$ 时，这推断会停止(这样就证明了 s 的存在性)。最后当 α，α^q，…，$\alpha^{q^{s-1}}$ 都不同的时候，$s=d$：如果并非不同，那么有 $\alpha^{q^i}=\alpha^{q^j}$，其中 $i<j$。将 $\alpha^{q^i}=\alpha^{q^j}$ 两侧取 q^{d-i} 次幂，我们得到 $\alpha^{q^{d+j-i}}=\alpha^{q^d}=\alpha$。所以，$\alpha$ 是多项式 $P(X)=X^{q^{d+j-i}}-X$ 的根，$\mathrm{Spl}(P(X))=\mathbb{F}_{q^{d+j-i}}$。另一方面，$\alpha$ 是一个次数为 d 的既约多项式的根，$\mathrm{Spl}(M_\alpha(X))=\mathbb{F}_{q^d}$。所以，$d\mid(d+i-j)$ 或者 $d\mid(i-j)$ 都是不可能的。这意味着对于所有根 α^{q^i}，$i<d$ 都是不同的。

定理 3.1.25 *对于任何域 $\mathbb{F}_q$ 和整数 $d\geq1$，存在一个 d 次既约多项式 $f(X)\in\mathbb{F}_q[X]$。*

证明 取本原元 $\omega\in\mathbb{F}_{q^d}$。那么含有 ω 的 $\mathbb{F}_q$ 最小扩域 $\mathbb{F}_q(\omega)$ 与 $\mathbb{F}_{q^d}$ 一致，$\mathbb{F}_q$ 上向量空间 $\mathbb{F}_q(\omega)$ 的维数 $[\mathbb{F}_q(\omega):\mathbb{F}_q]$ 等于 $[\mathbb{F}_{q^d}:\mathbb{F}_q]=d$。$\mathbb{F}_q$ 上关于 ω 的极小多项式 $M_\omega(X)$ 有不同的根 ω，ω^q，…，$\omega^{q^{d-1}}$，因此其次数为 d。 □

虽然没有一般化的方法来证明已知多项式的不可约性，但关于一个已知次数的既约多项式的数量问题却可由简洁的 Möbius 方程求解。

定义 3.1.26 在集合 $\mathbb{Z}$ 上的 Möbius 方程由下式给出

如果 n 可以被一个素数的平方整除，$\mu(1)=1$，$\mu(n)=0$。

如果 n 是 k 个不同素数的乘积，$\mu(n)=(-1)^k$。

定理 3.1.27 *在多项式环 $\mathbb{F}_q[X]$ 中，n 次既约多项式的数量 $N_q(n)$ 由下式给定*

277

$$N_q(n)=\frac{1}{n}\sum_{d|n}\mu(d)q^{n/d} \tag{3.1.5}$$

举例说，$N_q(20)$等于

$$\frac{1}{20}(\mu(1)q^{20}+\mu(2)q^{10}+\mu(4)q^{5}+\mu(5)q^{4}+\mu(10)q^{2}+\mu(20)q)$$

$$=\frac{1}{20}(q^{20}-q^{10}-q^{5}+q^{2})$$

证明　首先，我们构造加性 Mobius 反演公式，令 Ψ 和 Ψ 是两个从$\mathbf{Z}_+$到加性 Abel 群 G 的函数，那么下面的等式是互相等价的

$$\Psi(n)=\sum_{d|n}\psi(d) \tag{3.1.6}$$

$$\psi(n)=\sum_{d|n}\mu(d)\Psi\left(\frac{n}{d}\right) \tag{3.1.7}$$

当我们观察下列两种情况时可导出等价性，即(a)当 $n>1$ 时 $\sum_{d|n}\mu(d)$ 等于 0，当$n=1$时该和为 1。

(b) 对于所有 n

$$\sum_{d,d|n}\mu(d)\Psi(n/d)=\sum_{d,d|n}\mu(d)\sum_{c,c|n/d}\Psi(c)$$

$$=\sum_{c,c|n}\Psi(c)\sum_{d,d|n/c}\mu(d)=\Psi(n)$$

为了检验(a)，令 p_1，…，p_k 是分解 n 的不同素因子，那么

$$\sum_{d|n}\mu(d)=\mu(1)+\sum_{i=1}^{k}\mu(p_i)+\cdots+\mu(p_1\cdots p_k)$$

$$=1+\binom{k}{1}(-1)+\binom{k}{2}(-1)2+\cdots+\binom{k}{k}(-1)^k=0$$

把式(3.1.7)应用到 $G=\mathbf{Z}$($\mathbf{Z}$ 是整数加法群)，利用 $\Psi(n)=nN_q(n)$和 $\Psi(n)=q^n$，可以得到式(3.1.5)。 □

现在把多项式 $X^{q^n}-X$ 分解为既约多项式的乘积。那么式(3.1.6)当 $X^{q^n}-X$ 的阶数 q^n 与所有既约多项式(阶数整除 n)的阶数和相同时，式(3.1.6)成立。确实，我们简单地把 $X^{q^n}-X$ 写为所有既约多项式的乘积，可以观察到当且仅当多项式的次数能够整除 n 时，多项式可以分解(参见推论 3.1.30)。

278

举例 3.1.28　在$\mathbb{F}_3$ 中找到所有次数为 2 和 3 的多项式，计算它们的阶。

解答　在$\mathbb{F}_3=\{0,1,2\}$上，有 3 个 2 次既约多项式：X^2+1，其阶为 4 且满足

$$(X^4-1)/(X^2+1)=X^2-1$$

X^2+X+2 和 X^2+2X+2 的阶为 8 且满足

$$(X^8-1)/(X^2+X+2)(X^2+2X+2)=X^4-1$$

接着，在$\mathbb{F}_3$ 中，存在$(3^3-3)/3=8$ 个 3 次既约多项式，其中有四个的阶为 13(所以，它们不是本原的)：

$$X^3+2X+2,X^3+X^2+2,X^3+X^2+X+2,X^3+2X^2+2X+2$$

剩下的四个的阶为 26(所以它们是本原的)：

$$X^2+2X+1, X^3+X^2+2X+1, X^3+2X^2+1, X^3+2X^2+X+1$$

确实，如果 $p(X)$ 表示前四个多项式的乘积，那么 $(X^{32}-1)/p(X)=X-1$。另一方面，如果 $r(X)$ 代表剩下四个多项式的乘积，那么 $(x^{32}-1)/r(X)$ 等于

$$(X-1)(X+1)(X^5+2X+2)(X^5+X^3+2)$$
$$\times(X^2+X^2-X+2)(X^5+2X^2+2X+2) \quad \square$$

定理 3.1.29 如果一个 d 阶多项式 $g(X)\in F_q[X]$ 是不可约的并且 α 是 $g(X)$ 的一个根，那么分裂域 $\mathrm{Spl}(g(X))$ 和最小扩域 $F_q(\alpha)$ 与 F_{q^d} 是一致的。

证明 已知 $g(X)=M_{\alpha,F_q}(X)=\mathrm{irr}_{\alpha,F_q}(X)$（由引理 3.1.24，$g(X)$ 是不可约的）。我们可以得到 $F_q\subset F_q(\alpha)=F_{q^d}\subseteq\mathrm{Spl}(g(X))$。接下来检验 $g(X)$ 在 $F_q(\alpha)$ 中的任何根 γ：这暗示着 $\mathrm{Spl}(g(X))\subseteq F_q(\alpha)$。由定理 3.1.13，含有 q^d 个元素的唯一 Galois 域 $F_q(\alpha)=F_{q^d}$ 是分裂域 $\mathrm{Spl}(X^{q^d}-X)$，即包含 $X^{q^d}-X$ 所有的根（其中有一个是 α）。那么，因为 $g(X)=M_{\alpha,F_q}(X)$，可知 $g(X)\mid(X^{q^d}-X)$。所以，$g(X)$ 所有的根是 $X^{q^d}-X$ 的根，所以这些根都在 $F_q(\alpha)$ 中。 $\square$

推论 3.1.30 假设 $g(X)\in F_q[X]$ 是一个阶数为 d 的既约多项式。那么，当且仅当 $d\mid n$ 时，$g(X)\mid(X^{q^n}-X)$。

证明 已知分裂域 $\mathrm{Spl}(g(X))=F_{q^d}$ 和 $\mathrm{Spl}(X^{q^n}-X)=F_{q^n}$。由定理 3.1.29，当且仅当 $d\mid n$ 时，$\mathrm{Spl}(g(X))\subseteq\mathrm{Spl}(X^{q^n}-X)$。

如果 $g(X)\mid(X^{q^n}-X)$，$g(X)$ 中的每一个根都是 $(X^{q^n}-X)$ 的根。那么 $\mathrm{Spl}(g(X))\subseteq\mathrm{Spl}(X^{q^n}-X)$，所以 $d\mid n$。 $\square$ [279]

相反地，如果 $d\mid n$，即 $\mathrm{Spl}(g(X))\subseteq\mathrm{Spl}(X^{q^n}-X)$，那么 $g(X)$ 的每个根都在 $\mathrm{Spl}(X^{q^n}-X)$ 中。但是，$\mathrm{Spl}(X^{q^n}-X)$ 是 $(X^{q^n}-X)$ 根的一个完整的集合，所以 $g(X)$ 的每个根也是 $(X^{q^n}-X)$ 的根，那么 $g(X)\mid(X^{q^n}-X)$。

定理 3.1.31 如果 $g(X)\in F_q[X]$ 是一个阶数为 d 的一个不可约多项式，那么 $\alpha\in\mathrm{Spl}(g(X))=F_{q^d}[X]$ 是它的一个根，$g(X)$ 的所有根是 α，α^q，…，$\alpha^{q^{d-1}}$。并且，d 是满足 $\alpha^{q^d}=\alpha$ 的最小正整数。

证明 如引理 3.1.24 的证明，α，α^q，…，$\alpha^{q^{d-1}}$ 是不同的根。所以，所有的根已经列举出来，d 是满足以上性质的最小正整数。 $\square$

推论 3.1.32 $\mathrm{Spl}(g(X))$ 中，阶数为 $\deg g(X)=d$ 的一个不可约多项式 $g(X)\in F_q[X]$ 的所有根含有相同的乘阶，乘阶能整除 q^d-1 并且给定了多项式 $g(X)$ 的阶数（参见定义 3.1.22）。

不可约多项式 $g(X)$ 的乘阶用 $\mathrm{ord}(g(X))$ 表示。

举例 3.1.33 (a) 证明：若自然数 n，q 满足 $\mathrm{lcm}(n, q)=1$，那么存在一个自然数 s 使得 $n\mid(q^s-1)$。(b) 证明：如果一个多项式 $g(X)\in F_2[X]$ 是不可约的，当且仅当 $\mathrm{ord}(g(X))\mid n$ 时，满足 $g(X)\mid(X^n-1)$。

解答 (a) 集合 $q^l-1=nq_l+b_l$，其中 $b_l\leq n$，$l=1, 2, \cdots$。利用鸽洞效应，对于某个 $l_1<l_2$，满足 $b_{l_1}<b_{l_2}$。那么 $n\mid q^{l_1}(q^{l_2-l_1}-1)$，利用条件 $\mathrm{lcm}(n, q)=1$，$n\mid(q^s-1)$，其中 $s=l_2-l_1$。(b) 对于一个不可约多项式 $g(X)$，它的乘阶在定义 3.1.22 中给出。

$$\mathrm{ord}(g(X))=\min[n: g(X)\mid(X^n-1)]$$

首先，我们的目标是检查 $m=\mathrm{ord}(g(X))$ 是否成立，若成立，那么当且仅当 $g(X)\mid(X^n-1)$ 时，满足 $m\mid n$。实际上，假设 $m\mid n$：$n=mr$，那么 $X^n-1=(X^m-1)(1+x^m+\cdots+X^{m(r-1)})$。由

于 $g(X)\mid(X^n-1)$，这暗示着 $g(X)\mid(X^m-1)$。

相反地，如果 $g(X)\mid(X^m-1)$，那么 $g(X)$ 的根 $\alpha_1,\cdots,\alpha_d$ 落在 $\mathrm{Spl}(X^m-X)$ 的 X^m-1 中。所以，在 $\mathrm{Spl}(X^m-1)$ 中，$\alpha_j^m=\alpha_j^n=1$，$1\leqslant j\leqslant d$。若 $n=mb+a$，其中 $0\leqslant a\leqslant m$。那么 $\alpha_j^n=\alpha_j^{bm}\alpha_j^a=\alpha_j^a=1$，即每一个 α_j 都是 X^a-1 的根。所以，如果 $a>0$，那么 $g(X)\mid(X^a-1)$，此处矛盾。所以 $a=0$ 且 $m\mid n$。 □

给定一个根 $\alpha\in\mathbb{F}_{q^n}$，计算一个不可约多项式 $g(X)\in\mathbb{F}_q[X]$（特别地，对于一个最小多项式 $M_{\alpha,\mathbb{F}_q}(X)$）是不容易的。这是因为 q，n，α 和 $d=\deg M_\alpha(X)$ 的关系是复杂的。然而，在 $\omega^{q^n-1}=\varepsilon$，$\omega^{q^d}=\omega$ 的条件下，如果 $\alpha=\omega$ 是 $\mathbb{F}_{q^d}$ 中的一个本元元素，那么 $d=n$，并且 n 是满足这个性质的最小正整数。在此情况下 $M_\alpha(X)=\prod_{b\in\mathbb{F}_{q^d}}(X-b)$。

280

对于一个一般的既约多项式，共轭性质的作用可参见下面的定义 3.1.34。这个概念在 2.5 节中对于集 $\mathbb{F}_{2^n}$ 被非正式地介绍或运用过。

定义 3.1.34 对于两个元素 α，$\alpha^r\in\mathbb{F}_{q^n}$，若 $M_{\alpha,\mathbb{F}_q}(X)=M_{\alpha^r,\mathbb{F}_q}(X)$，那么 α 和 α^r 在 $\mathbb{F}_q$ 中被称作共轭的。

总结上面的内容，我们得到下面的推论。

定理 3.1.35 *在 $\mathbb{F}_q$ 上 $\alpha\in\mathbb{F}_{q^n}$ 的共轭包括 α，α^q，$\cdots$，$\alpha^{q^{d-1}}\in\mathbb{F}_{q^n}$，其中 d 与上文定义一致。特别地，$\prod_{0\leqslant j\leqslant d-1}(X-\alpha^{q^j})$ 中的所有系数取自 $\mathbb{F}_q$，而且它是 $\mathbb{F}_q[X]$ 上唯一的具有根 α 的既约多项式。它也是 $\mathbb{F}_q[X]$ 上唯一的具有根 α 的最低次的首一多项式。*

例子 3.1.36 继续探讨例子 3.1.28，我们利用 $\mathbb{F}_2[\omega]$ 来确定 $\mathbb{F}_{16}$，其中 ω 是一个阶数为 4 的本原多项式的根而 $\mathbb{F}_2(\omega)$ 是包含这个根 ω 的最小域。所以，如果我们选择 $1+X^3+X^4$，ω 满足 $\omega^4=1+\omega^3$，如果我们选择 $1+X^3+X^4$，ω 满足 $\omega^4=1+\omega^2$。在这两种情况下，共轭元素是 ω，ω^2，ω^4，ω^8。

相应地，(3.1.2)的表格有如下形式：

ω 的幂	$1+X+X^4$ 向量(编字)	$1+X^3+X^4$ 向量(编字)
—	0000	0000
0	1000	1000
1	0100	0100
2	0010	0010
3	0001	0001
4	1100	1001
5	0110	1101
6	0011	1111
7	1101	1110
8	1010	0111
9	0101	1010
10	1110	0101
11	0111	1011
12	1111	1100
13	1011	0110
14	1001	0011

(3.1.8)

根据左边表格的加法规则，幂 ω^i 的极小多项式 $M_{\omega^i}(X)$，在 $i=1,2,4,8$ 时是 $1+X+X^4$，在 $i=7,14,13,11$ 时是 $1+X^3+X^4$，在 $i=3,6,12,9$ 时是 $1+X+X^2+$

281

X^3+X^4，最后在 $i=5$，10 时是 $1+X+X^2$。根据右边表格的加法规则，我们需交换多项式 $1+X+X^4$ 和 $1+X^3+X^4$。多项式 $1+X+X^4$，$1+X^3+X^4$ 的阶数是 15。多项式 $1+X+X^4$，$1+X^3+X^4$ 的阶数是 5，多项式 $1+X+X^2$ 的阶数是 3。

一个获得答案的简洁办法是求 $(\omega^7)^i$ 经由 1，ω，$(\omega^7)^2$，$(\omega^7)^3$ 的线性表达式。比如：对于 ω^7，从左边表格可得

$$(\omega^7)^4=\omega^{28}=\omega^3+\omega^2+1$$
$$(\omega^7)^5=\omega^{35}=\omega^3+\omega^2$$

容易看出 $(\omega^7)^4=1+(\omega^7)^5$，由它得到 $1+X^5+X^4$。为了完整性，写下什对 $(\omega^7)^2$ 的未使用表达式

$$(\omega^7)2=\omega^{14}=\omega^{10}\omega^3=(1+\omega)^2\omega^3=(1+\omega+\omega^2+\omega^3)\omega^3$$
$$=\omega^3+\omega^4+\omega^5+\omega^6=\omega^3+\omega^2+1+\omega+(1+\omega)\omega=1+\omega^3$$

对于 $M_{\omega^3}(X)$，用标准的办法可以得到一个捷径

$$M_{\omega^3}(X)=(X-\omega^3)(X-\omega^{10})=X^2+(\omega^3+\omega^{10})X+\omega^{13}=X^2+X+1$$

所以，对于 F_{32} 的所有极小多项式可罗列如下

$$M_{\omega^3}(X)=1+X,M_{\omega}(X)=1+X+X^4$$
$$M_{\omega^5}(X)=1+X+X^2+X^3+X^4$$
$$M_{\omega^7}(X)=1+X+X^2,M_{\omega^9}(X)=1+X^3+X^4$$

例子 3.1.37 对于域 $F_{32}\cong F_2[X]/\langle 1+X^2+X^5\rangle$，加法表格由下面计算得出。极小多项式是

(i) 对于共轭 $\{\omega,\omega^2,\omega^4,\omega^8,\omega^{16}\}$，$1+X^2+X^5$。

(ii) 对于 $\{\omega^3,\omega^6,\omega^{12},\omega^{24},\omega^{17}\}$，$1+X^2+X^3+X^4+X^5$。

(iii) 对于 $\{\omega^5,\omega^{10},\omega^{20},\omega^9,\omega^{18}\}$，$1+X+X^2+X^4+X^5$。

(iv) 对于 $\{\omega^7,\omega^{14},\omega^{28},\omega^{25},\omega^{19}\}$，$1+X+X^2+X^3+X^5$。

(v) 对于 $\{\omega^{11},\omega^{22},\omega^{13},\omega^{26},\omega^{21}\}$，$1+X+X^2+X^4+X^5$。

(vi) 对于 $\{\omega^{15},\omega^{30},\omega^{29},\omega^{27},\omega^{23}\}$，$1+X^3+X^5$。 282

所有的极小多项式的阶数都是 31。

ω 的幂	向量(码字)	ω 的幂	向量(码字)
	00000	15	11111
0	10000	16	11011
1	01000	17	11001
2	00100	18	11000
3	00010	19	01100
4	00001	20	00110
5	10100	21	00011
6	01010	22	10101
7	00101	23	11110
8	10110	24	01111
9	01011	25	10011
10	10001	26	11101
11	11100	27	11010
12	01110	28	01101
13	00111	29	10010
14	10111	30	01001

(3.1.9)

定义 3.1.38 $\mathbb{F}_q$ 上 $\mathbb{F}_{q^s}$ 的自同构(简记为，$(\mathbb{F}_{q^s}, \mathbb{F}_q)$-自同构)是一个双射 $\sigma: \mathbb{F}_{q^s} \to \mathbb{F}_{q^s}$：(a) $\sigma(a+b)=\sigma(a)+\sigma(b)$；(b) $\sigma(ab)=\sigma(a)\sigma(b)$；(c) $\sigma(c)=c$，对于所有 $a, b \in \mathbb{F}_{q^s}$，$c \in \mathbb{F}_q$ 成立。

定理 3.1.39 $(\mathbb{F}_{q^s}, \mathbb{F}_q)$-自同构的集合和循环群 $\mathbb{Z}_s$ 是同构的，其中 Frobenius 映射 $\sigma_q(a)=a^q$，$a \in \mathbb{F}_{q^s}$ 生成。

证明 令 $\omega \in \mathbb{F}_{q^s}$ 是本原元，那么 $\omega^{q^s-1}=e$ 和 $M_\omega(X) \in \mathbb{F}_q[X]$ 的解是 $\omega, \omega^q, \omega^{q^2}, \cdots, \omega^{q^{s-1}}$。$(\mathbb{F}_{q^s}; \mathbb{F}_q)$-自同构 τ 可确定 $M_\omega(X)$ 的参数，所以它对所有根做了一个置换，对于某些 j，$0 \leqslant j \leqslant s-1$，有 $\tau(\omega)=\omega^{q^j}$。但是由于 ω 是本原元，τ 完全由 $\tau(\omega)$ 确定。那么由于 $\sigma_q^j(\omega)=\omega^{q^j}=\tau(\omega)$，可得 $\tau=\sigma_q^j$。 □

本节的余下部分都将研究单位根，即，$\mathbb{F}_q$ 上多项式 X^n-e 的解，其中 $q=p^s$，$p=\mathrm{char}(\mathbb{F}_q)$。为了不失一般性，从现在起假设

$$\gcd(n,q)=1, \quad \text{即 } n \text{ 和 } q \text{ 是互素的} \tag{3.1.10}$$

事实上，如果 n 和 q 不是互素的，可得 $n=mp^k$，那么由引理 3.1.5 有

$$X^n-e=X^{mp^k}-e=(X^m-e)^{p^k}$$

283 这减少了我们对多项式 X^n-e 的分析。

定义 3.1.40 多项式 $X^n-e \in \mathbb{F}_q[X]$ 在分裂域 $\mathrm{Spl}(X^n-e)=\mathbb{F}_{q^s}$ 上的根称为 $\mathbb{F}_q$ 上的 n 次单位根(或者 $(n, \mathbb{F}_q)$-单位根)。所有 $(n, \mathbb{F}_q)$-单位根的集合可以用 $\mathbb{E}^{(n)}$ 表示。可以证明 s 的值是 $s \geqslant 1$ 中能够满足 $q^s \equiv 1 \bmod n$ 的最小整数。那么(参见下面的定理 3.1.44)。这个事实可以通过利用 $\mathrm{ord}_n(q)$ 表示值 s 以及称它的阶数为 $q \bmod n$ 反映。

在假设(3.1.10)下，不存在多重根(因为微分 $\partial_X(X^n-e)=nX^{n-1}$ 没有根在 $\mathrm{Spl}(X^n-e)=\mathbb{F}_{q^s}$ 中)。所以，$\#\mathbb{E}^{(n)}=n$。

定理 3.1.41 $\mathbb{E}^{(n)}$ 是 $\mathbb{F}_{q^s}^*$ 的一个循环子群。

证明 假设 $\alpha, \beta \in \mathbb{E}^{(n)}$，那么 $(\alpha\beta^{-1})^n=\alpha^n(\beta^n)^{-1}=e$，即 $\alpha\beta^{-1} \in \mathbb{E}^{(n)}$。所以，$\mathbb{E}^{(n)}$ 是循环群 $\mathbb{F}_{q^s}^*$ 的一个子群，所以也是循环的。 □

定义 3.1.42 群 $\mathbb{E}^{(n)}$ 的生成元(比如一个 n 次单位根，它的乘阶等于 n)称作本原 $(n, \mathbb{F}_q)$-单位根；它可以用 β 表示。

推论 3.1.43 本原 $(n, \mathbb{F}_q)$-单位根的数目正好是 $\phi(n)$。特别地，本原 $(n, \mathbb{F}_q)$-单位根对于任何与 q 互素的 n 都存在。

上述性质允许我们在分裂域 $\mathbb{F}_{q^s}=\mathrm{Spl}(X^n-e)$ 中计算 s。如果 β 是一个本元 $(n, \mathbb{F}_q)$-单位根，那么它的乘阶等于 n，因为 $\omega \neq 0$，如果 $\beta^q=\beta$，可得 $\omega \in \mathbb{F}_q$，即 $\beta^{q-1}=e$，当且仅当 $n \mid (q-1)$ 时上述关系成立。但是 s 是满足 $\mathbb{F}_{q^s} \ni \omega$ 的最小 r。

定理 3.1.44 $\mathrm{Spl}(X^n-e)=\mathbb{F}_{q^s}$，其中 $s=\mathrm{ord}_n(q)$ 是 $\geqslant 1$ 的最小整数且满足 $n \mid (q^s-1)$，即 $s \geqslant 1$ 中满足 $q^s \equiv 1 \bmod n$ 的最小整数。

在满足 $s=\mathrm{ord}_n(q)$ 的域 $\mathbb{F}_{q^s}$ 中，有必要去强调一下本原与本原 $(n, \mathbb{F}_q)$-单位根的异同。一个本原域元素 ω 生成乘性循环群 $\mathbb{F}_q^*$：$\mathbb{F}_q^*=\{e, \omega, \cdots, \omega^{q^s-2}\}$；其乘阶等于 q^s-1。一个本元单位根生成乘性循环群 $\mathbb{E}^{(n)}$：$\mathbb{E}^{(n)}=\{e, \beta, \cdots, \beta^{n-1}\}$；其乘阶等于 n。(另一方面，β 作为一个域元素生成 $\mathbb{F}_{q^s}$：$\mathbb{F}_{q^s}=\mathbb{F}_q(\beta)=\mathbb{F}_q(\mathbb{E}^{(n)})$。)这说明当且仅当 $n=q^s-1$ 时 $\beta=\omega$。实际上，我们需要考虑在什么条件下幂 ω^k 是一个本元单位根。在举例 3.1.33 中，当 $n \mid (q^s-1)$ 时上述论断成立，即 $q^s-1=nr$。实际上，如果 $k \geqslant 1$ 满足

$$\gcd(k, nr)=\gcd(k, q^s-1)=r$$

那么元素 ω^k 是一个本元单位根，因为它的乘阶等于

$$\frac{q^r-1}{\gcd(k,q^r-1)}=\frac{nr}{r}=n$$

284

当 $k=cu$ 且 u 与 n 互素时上式成立。相反，如果是 ω^k 一个本元单位根，那么 $\gcd(k,q^r-1)=(q^r-1)/n$。所以，可得下面的定理。

定理 3.1.45 令 $(P)^{(n)}$ 为 $(n,\mathbb{F}_q)$-本元单位根的集合，$\mathbb{T}^{(n)}$ 是 $\mathbb{F}_q=\mathrm{Spl}(X^n-e)$ 中本原元的集合。那么以下情况有一个成立(i) $\mathbb{P}^{(n)}\cap\mathbb{T}^{(n)}=\varnothing$ 或者(ii) $\mathbb{P}^{(n)}=\mathbb{T}^{(n)}$；当且仅当 $n=q^r-1$ 时情况(ii)成立。

现在我们可以将 $\mathbb{F}_q$ 上的多项式 (X^n-e) 因式分解为多个有关 $(n,\mathbb{F}_q)$-单位根的不同极小多项式的乘积：

$$X^n-e=\mathrm{lcm}(M_\beta(X):\beta\in\mathbb{E}^{(n)}) \tag{3.1.11}$$

如果从一个本原元 $\omega\in\mathbb{F}_{q^s}$ 开始，其中 $s=\mathrm{ord}_n(q)$，那么 $\beta=\omega^{(q^s-1)/n}$ 是一个本原单位根并且 $\mathbb{E}^{(n)}=\{e,\beta,\cdots,\beta^{n-1}\}$。

这让我们可以计算极小多项式 $M_\beta(X)$。对于所有的 $i=0,\cdots,n-1$，β^i 的共轭是 β^i，β^{iq}，$\cdots$，$\beta^{iq^{d-1}}$，其中对于 $\beta^{iq^d}=\beta^i$，$d(=d(i))$ 是使得 $\beta^{iq^{d+1}}=e$ 成立的最小正整数。这等价于 $n\mid(iq^d-i)$，即 $iq^d=i \bmod n$。所以

$$M_\beta(X)(=M_{\beta^i}(X))=(X-\beta^i)(X-\beta^{iq})\cdots(X-\beta^{iq^{d-1}}) \tag{3.1.12}$$

定义 3.1.46 集合 $i,iq,\cdots,iq^{d-1}$ 被称为(对 i)的分圆陪集并记为 $C_i(=C_i(n,q))$，其中 $d(=d(i))$ 是令 $iq^d=i \bmod n$ 的最小正整数(另一种表述方法，C_i 定义为非零域元素 ω^i，ω^{iq}，$\omega^{iq^{d-1}}$ 的集合)。

举例 3.1.47 验证多项式 X^2+X+2 和 X^3+2X^2+1 是 $\mathbb{F}_3$ 上的本原多项式，计算得到由这些多项式生成的有关域 $\mathbb{F}_9$ 和 $\mathbb{F}_{27}$ 的列表。

解答 域 $\mathbb{F}_9$ 和 $\mathbb{F}_3[X]/(X^2+X+2)$ 是同构的。$\omega\sim X$ 的幂如下

$$\omega^2\sim 2X+1,\omega^3\sim 2X+2,\omega^4\sim 2$$
$$\omega^5\sim 2X,\omega^6\sim X+2,\omega^7\sim X+1,\omega^8\sim 1$$

ω 的分圆陪集是 $\{\omega,\omega^3\}$，(因为 $\omega^3=\omega$)，那么极小多项式

$$\begin{aligned}M_\omega(X)&=(X-\omega)(X-\omega^3)=X^2-(\omega+\omega^3)X+\omega^4\\&=X^2-2X+2=X^2+X+2\end{aligned}$$

所以，X^2+X+2 是本原多项式。

285

接下来，$\mathbb{F}_{27}\sim\mathbb{F}_3[X]/(X^3+2X^2+1)$，$\omega\sim X$，可得

$$\omega^2\sim X^2,\omega^3\sim X^2+2,\omega^4\sim X^2+2X+2,\omega^5\sim 2X+2$$
$$\omega^6\sim 2X^2+2X,\omega^7\sim x^2+1,\omega^8\sim X^2+X+2$$
$$\omega^9\sim 2X^2+2X+2,\omega^{10}\sim X^2+2X+1,\omega^{11}\sim X+2$$
$$\omega^{12}\sim X^2+2X,\omega^{13}\sim 2,\omega^{14}\sim 2X,\omega^{15}\sim 2X^2,\omega^{16}\sim 2X^2+1$$
$$\omega^{17}\sim 2X^2+X+1,\omega^{18}\sim X+1,\omega^{19}\sim X^2+X$$
$$\omega^{20}\sim 2X^2+2,\omega^{21}\sim 2X^2+2X+1,\omega^{22}\sim X^2+X+1$$
$$\omega^{23}\sim 2X^2+X+2,\omega^{24}\sim 2X+1,\omega^{25}\sim 2X^2+X,\omega^{26}\sim 1$$

$\mathbb{F}_{27}$ 上 ω 的分圆陪集是 $\{\omega,\omega^3,\omega^9\}$，最终可得本原多项式是

$$\begin{aligned}M_\omega(X)&=(X-\omega)(X-\omega^3)(X-\omega^9)\\&=X^3-(\omega+\omega^3+\omega^9)X^2+(\omega^4+\omega^{10}+\omega^{12})X-\omega^{13}\end{aligned}$$

$$=X^3+2X^2-1$$ □

举例 3.1.48 (a) 考虑一个$\mathbb{F}_2$上的多项式 $X^{15}-1$($n=15$，$q=2$)，那么 $m=2$，$s=\mathrm{ord}_{15}(2)=4$，$\mathrm{Sp}(X^{15}-1)=\mathbb{F}_2^4=\mathbb{F}_{16}$。

多项式 $g(X)=1+X+X^4$ 是本原的：

在$\mathbb{F}_{16}$中，多项式的所有根β都是本原的。所以，本原(15，$\mathbb{F}_2$)单位根是

$$\beta=\omega^{(2^4-1)/15}=\omega$$

所以，多项式 $X^{15}-1$ 的根是 $1,\beta,\cdots,\beta^{14}$。对于它们的极小多项式已经在例 3.1.36 中计算出。所以，我们可以得到因式分解

$$X^{15}-1=(1+X)(1+X+X^4)(1+X+X^2+X^3+X^4)$$
$$\times(1+X+X^2)(1+X^2+X^4)$$

(b) 已知一个分圆陪集，我们可以证明关于 X^n-a 的一个特殊分解中含有不可约的因子。用$\mathbb{F}_2$上的多项式 $X^{63}-1$($n=63$，$q=2$)为例来具体说明。这里存在三个分圆陪集

$$C_0=\{0\},C_1=\{1,2,4,8,7,5\},C_3=\{3,6\}$$

相应的极小多项式的阶数分别为 1，6 和 2，即

$$1+X,1+X+X^6 \quad 和 \quad 1+X+X^2$$

由此可得

$$X^{63}-1=(1+X)(1+X+X^2)(1+X^2+X^6)$$

(c) 检查下面$\mathbb{F}_2$上多项式的本原性

$$f(X)=1+X+X^6$$

其中 $n=6$，$q=2$。这里，$2^6-1=63=3^2\cdot 7$。因为 $63/3=21$，且 $3^2\mid\mathrm{ord}(f(X))\Leftrightarrow X^{21}-1\neq 0 \bmod(1+X+X^6)$。但是 $X^{21}-1=X+X^2+X^4+X^5+X^6\neq 1 \bmod(1+X+X^6)$，所以 $3^2\mid\mathrm{ord}(f(X))$。

接下来，因为 $63/7=9$，$7\mid\mathrm{ord}(f(X))\Leftrightarrow X^9-1\neq 0 \bmod(1+X+X^6)$。但是 $X^9-1=X^3+X^5\neq 1 \bmod(1+X+X^6)$，所以 $7\mid\mathrm{ord}(f(X))$。因此，$\mathrm{ord}(f(X))=63$，$f(X)$是本原的。下面的定理 3.1.53 说明，因为 $2^6=1 \bmod 63$，任何阶为 63 的既约多项式的次数为 6。

(d) 下面再次考虑$\mathbb{F}_2$上的多项式

$$g(X)=1+X+X^2+X^3+X^6$$

(其中 $n=6$，$q=2$ 和前一个例子一样)。同样，$3^2\mid\mathrm{ord}(g(X))\Leftrightarrow X^9\neq 1 \bmod(1+X+X^2+X^3+X^6)$，然而，在$\mathbb{F}_2$中

$$X^{21}-1=(1+X)(1+X+X^2)(1+X+X^3)(1+X^2+X^3)$$
$$\times(1+X+X^2+X^3+X^6)(1+X^2+X^3+X^5+X^6)$$

于是有 $X^{21}-1=0 \bmod(1+X+X^2+X^3+X^6)=1$，所以 3^2 不能整除 $\mathrm{ord}(g(X))$。

接下来，$3\mid\mathrm{ord}(g(X))\Leftrightarrow X^3\neq 1 \bmod(1+X+X^2+X^3+X^6)$。由于 $X^3=(X+X^2+X^3+X^6)\neq 1 \bmod(1+X+X^2+X^3+X^6)$，所以 3 是 $\mathrm{ord}(g(X))$的一个因子。

最后，$7\mid\mathrm{ord}(g(X))\Leftrightarrow X^7\neq 1 \bmod(1+X+X^2+X^3+X^6)$，且 $X^7=1+X^2+X^3\neq 1 \bmod(1+X+X^2+X^3+X^6)$，7 是 $\mathrm{ord}(g(X))$的一个因子，所以 $\mathrm{ord}(g(X))=21$。

下面总结一下极小多项式和单位根的一些结论。从定理 3.1.25 可知，对于所有整数 $d\geq 1$ 和所有 $q=p^s$，其中 p 是一个素数，s 是大于 1 的整数，存在一个阶为 d 的本原多项式，例如 $M_\omega(X)$，其中 ω 是在域$\mathbb{F}_p$中的一个本原元，另一方面，对于所有阶为 d 的既约多项式 $f(X)\in\mathbb{F}_q[X]$，它的根都在域 $\mathrm{Sp}(f(X))=\mathbb{F}_p$中并且和 $\mathrm{ord}(f(X))$具有相同的乘阶。

285

定理 3.1.49 考虑阶为 d 的既约多项式 $f(X)\in\mathbb{F}_q[X]$ 且 $\mathrm{ord}(f(X))=\ell$。那么：

(a) $\ell\mid(q^d-1)$。

(b) $f(X)\mid(X^\ell-e)$。

(c) 当且仅当 $f(X)\mid(X^n-e)$ 时，$\ell\mid n$。

(d) ℓ 是使得 $f(X)\mid(X^\ell-e)$ 成立的最小正整数。

287

证明 (a) $\mathrm{Spl}(f(X))=\mathbb{F}_{q^d}$，所以 $f(X)$ 的每一个根 α 也是 $X^{\ell-1}-e$ 的根，所以有 $\mathrm{ord}(\alpha)\mid(q^d-1)$。

(b) $\mathrm{Spl}(f(X))$ 上 $f(X)$ 的每一个根 α 有 $\mathrm{ord}(\alpha)=\ell$，$(X^\ell-e)$ 中的根也是一样，所以，$f(X)\mid(X^\ell-e)$。

(c) 如果 $f(X)\mid(X^n-e)$，那么 $f(X)$ 的每一个根也是 X^n-e 的根，也就是说，$\mathrm{ord}(\alpha)\mid n$。所以，$\ell\mid n$。相反地，若 $n=k\ell$，那么由(b)可得 $(X^\ell-e)\mid(X^n-e)$，$f(X)\mid(X^n-e)$。

(d) 与(c)证法一致。 □

定理 3.1.50 若 $f(X)\in\mathbb{F}_q[X]$ 是一个次数为 d 阶数为 ℓ 的既约多项式，那么有 $d=\mathrm{ord}_\ell(q)$。

证明 如果 $\alpha\in\mathbb{F}_{q^d}$ 且 $f(\alpha)=0$，根据定理 3.1.29 可得 $\mathbb{F}_q(\alpha)=\mathbb{F}_{q^d}=\mathrm{Spl}(f(X))$。但是 α 也是一个本原$(\ell,\mathbb{F}_{q^d})$-单位根，所以 $\mathbb{F}_q(\alpha)=\mathbb{F}_q(\omega^{q^s})=\mathrm{Spl}(X^\ell-e)=\mathbb{F}_{q^s}$，其中 $s=\mathrm{ord}_\ell(q)$。所以，$d=\mathrm{ord}_\ell(q)$。 □

举例 3.1.51 利用 Frobenius 映射 σ：$a\mapsto a^q$ 证明每个元素 $a\in\mathbb{F}_q$ 有唯一的 q^j 根，$j=1,\cdots,n-1$。

假设 $q=p^r$ 是奇数，证明 $\mathbb{F}_q$ 中正好有一半的非零元素有平方根。

解答 Frobenius 映射是一个双射 $\mathbb{F}_q\to\mathbb{F}_q$。所以，对于所有 $b\in\mathbb{F}_q$，存在唯一的 a 满足 $a^q=b$(q 次根)。映射的 j 次迭代 σ^j：$a\mapsto a^{q^j}$ 也是一个双射，所以对于所有 $b\in\mathbb{F}_q$，存在一个唯一的 a 满足 $a^{q^j}=b$，观察得到，对于所有 $c\in\mathbb{F}_q$ 都满足 $c^{(q^j)}=c$。

现在考虑一个乘性同态映射 τ：$a\mapsto a^2$，$\mathbb{F}_q^*\to\mathbb{F}_q^*$。如果 q 是奇数，那么 $\mathbb{F}_q^*\cong\mathbb{Z}_{q-1}$ 含有偶数个元素 $q-1$。我们想证明若 $\tau(a)=b$，那么 $\tau^{-1}(b)$ 由两个元素组成，即 a 和 $-a$。实际上，$\tau(-a)=b$。同样，若 $\tau(a')=b$，那么 $\tau(a'a^{-1})=e$。

所以，我们想分析 $\tau^{-1}(e)$。显然，$\pm e\in\tau^{-1}(e)$。另一方面，如果 ω 是一个本原元，那么 $\tau(\omega^{(q-1)/2})=\omega^{q-1}=e$，而 $\tau^{-1}(e)$ 由 $e=\omega^0$ 和 $\omega^{(q-1)/2}$ 组成。所以 $\omega^{(q-1)/2}=-e$。

现在，如果 $\tau(a'a^{-1})=e$，可得 $a'a^{-1}=\pm e$ 和 $a'=\pm a$。所以，τ 精确地将两个元素 a 和 $-a$ 映射到相同的像，所以它的范围 $\tau(\mathbb{F}_q^*)$ 是 $\mathbb{F}_q^*$ 的一半。 □

定理 3.1.52 (参考文献[92]，定理 3.56)考虑阶数为 n 的既约多项式 $p(X)\in\mathbb{F}_q[X]$。令 $m=\gcd(d,n)$。那么 $m\mid n$ 并且 $p(X)$ 可以在 $\mathbb{F}_q$ 中因式分解为 m 个各自阶数为 $n\mid m$ 的既约多项式。所以，当且仅当 $m=1$ 时，$p(X)$ 是 $\mathbb{F}_q$ 中不可约的。

定理 3.1.53 (参考文献[92]，定理 3.5)令 $\gcd(d,q)=1$。如果 $\ell\geq 2$，次数为 d 的阶数为 ℓ 的首一既约多项式的个数等于 $\phi(\ell)/d$ 并且有 $d=\mathrm{ord}_\ell(q)$；当 $\ell=d=1$，这个数为 2；对于其他所有情况，这个数为 0。 288

特别地，一个 ℓ 阶既约多项式的次数总是等于 $\mathrm{ord}_\ell(q)$，即满足 $q^s\equiv 1 \bmod \ell$ 的最小 s，这里 $\phi(\ell)$ 是 Euler 函数。

我们略去定理 3.1.52 和 3.1.53 的证明(参考文献[92])。我们只对定理 3.1.53 做一个简短的点评。如果，$p(0)\neq 0$，d 次既约多项式 $p(X)$ 的阶和乘法群 $\mathbb{F}_q^*$ 中多项式根的阶一致。所以，当且仅当 $d=\mathrm{ord}_\ell(q)$，阶是 ℓ，$p(X)$ 整除所谓的循环多项式

$$Q_\ell(X)=\prod_{\omega\in\mathbb{F}_q,\ \mathrm{ord}_\ell(\omega)=d}(X-\omega)$$

实际上，循环多项式可以分解为次数为 $d=\mathrm{ord}_\ell(q)$ 的既约多项式的乘积，数量是 $\phi(\ell)/d$。（在 $d=\ell-1$ 情况下，多项式 $p(X)-X$ 也应被考虑进去。）

下面对以上有限域基本理论做简要总结。

总结 3.1.54 一个域是一个环，它的非零元素在乘法运算下构成一个交换群。(i)任何有限域 $\mathbb{F}$ 含有 $q=p^r$ 个元素，其中 p 是一个素数，域的特征值 $\mathrm{char}(\mathbb{F})=p$。(ii)任何元素数量相同的两个有限域是同构的。所以，对于给定 $q=p^r$，存在(同构下)唯一的包含 q 个元素的域；这个域可以用 $\mathbb{F}_q$ 表示(通常它也被称作大小为 q 的 Galois 域)。当 q 是素数时，域 $\mathbb{F}_q$ 与元素个数为 p 的加法循环群 $\mathbb{Z}_p$ 是同构的。(iii) $\mathbb{F}_q$ 中的非零元素构成乘法群 $\mathbb{F}_q^*$，$\mathbb{F}_q^*$ 与元素个数为 $q-1$ 的加法循环群 $\mathbb{Z}_{q-1}$ 同构。(iv)当且仅当 $r\mid q$ 时，域 $\mathbb{F}_q$ 包含子域 $\mathbb{F}_r$；在这种情况下，$\mathbb{F}_q$ 与 $\mathbb{F}_r$ 上(即线性组合系数取自 $\mathbb{F}_r$ 中)的线性空间同构，其维数是 $\log_p(q/r)$，所以，每一个素数 p 可引出一系列有限域 $\mathbb{F}_{p^s}$，$s=1,2,\cdots$。一个生成乘法群 $\mathbb{F}_q^*$ 的元素 $\omega\in\mathbb{F}_q$ 称为 $\mathbb{F}_q$ 的本原元。

总结 3.1.55 $\mathbb{F}_q$ 上的多项式环可以用 $\mathbb{F}_q[X]$ 表示；如果考虑对取自 $\mathbb{F}_q[X]$ 中的多项式 $g(X)$ 求余，对应得到的多项式环可以表示为 $\mathbb{F}_q[X]/(g(X))$。(i)当且仅当 $g(X)$ 在 $\mathbb{F}_q$ 上不可约时，环 $\mathbb{F}_q[X]/(g(X))$ 是一个域(即，不存在一个分解 $g(X)=g_1(X)g_2(X)$，其中 $\deg(g_1(X))$，$\deg(g_2(X))<\deg(g(X))$)。(ii)对于任意 q 和正整数 d，存在 $\mathbb{F}_q$ 上次数为 d 的既约多项式 $g(X)$。(iii)如果 $g(X)$ 不可约且 $\deg g(X)=d$，那么域 $\mathbb{F}_q[X]/(g(X))$ 包含 q^d 个元素，即 $\mathbb{F}_q[X]/(g(X))$ 和 $\mathbb{F}_{q^d}$ 是同构的，归属于与 $\mathbb{F}_q$ 相同的域系(即，$\mathrm{char}(\mathbb{F}_{q^d})=\mathrm{char}(\mathbb{F}_q)$)。

289

总结 3.1.56 添加有限族元素 $\alpha_1,\cdots,\alpha_u$ 于 $\mathbb{F}_q$ 的扩域(包含于相同的系列的一个更大的域中)是包含 $\mathbb{F}_q$ 和 α_i，$1\leqslant i\leqslant u$ 的最小域。这样的扩域表示为 $\mathbb{F}_q(\alpha_1,\cdots,\alpha_u)$。(i)对于任意首一多项式 $p(X)\in\mathbb{F}_q[X]$，存在一个更大的与 $\mathbb{F}_q$ 同系的域 $\mathbb{F}_q'$ 使得 $p(X)$ 可以在 $\mathbb{F}_q'$ 中分解：

$$p(X)=\prod_{1\leqslant j\leqslant u}(X-\alpha_j),u=\deg p(X),\alpha_1,\cdots,\alpha_u\in\mathbb{F}_q' \tag{3.1.13}$$

含有此性质的最小域 $\mathbb{F}_{q^d}$(即 $\mathbb{F}_q(\alpha_1,\cdots,\alpha_u)$)被称为 $p(X)$ 的一个分裂域；我们也可以说 $p(X)$ 在 $\mathbb{F}_q(\alpha_1,\cdots,\alpha_u)$ 上分裂。$p(X)$ 的分裂域可用 $\mathrm{Spl}(p(X))$ 表示；当且仅当 $p(\alpha)=0$ 时，元素 $\alpha\in\mathrm{Spl}(p(X))$ 参与分解。分裂域 $\mathrm{Spl}(p(X))$ 可以用一个集合 $\{g(\alpha_j)\}$ 描述，其中 $j=1,\cdots,u$，$g(X)\in\mathbb{F}_q[X]$ 是次数小于 $\deg(p(X))$ 的多项式。(ii)域 $\mathbb{F}_q$ 为多项式 X^q-X 分裂。(iii)如果 d 次多项式在 $\mathbb{F}_q$ 中不可约，其 α 是 $p(X)$ 在域 $\mathrm{Spl}(p(X))$ 中的一个根，那么 $\mathbb{F}_{q^d}\simeq\mathbb{F}_q[X]/(p(X))$ 和 $\mathbb{F}_q(\alpha)$ 是同构的，$p(X)$ 在域 $\mathrm{Spl}(p(X))$ 中的所有根可以用共轭元素 $\alpha,\alpha^q,\alpha^{q^2},\cdots,\alpha^{q^{d-1}}$ 表示。所以，d 是满足 $\alpha^{q^d}=\alpha$ 的最小正整数。(iv)假设对于一个给定的域 $\mathbb{F}_q$，一个首一多项式 $p(X)\in\mathbb{F}_q[X]$ 和一个更大域中的元素 α，我们可得 $p(\alpha)=0$。那么，存在唯一的极小多项式 $M_\alpha(X)$ 且具有性质 $M_\alpha(\alpha)=0$(即其他满足 $p(\alpha)=0$ 的多项式 $p(X)$ 都可以被 $M_\alpha(X)$ 整除)。多项式 $M_\alpha(X)$ 是 $\mathbb{F}_q$ 中唯一在 α 处取零的既约多项式，它同时也是唯一在 α 处取零的最低次多项式。我们称 $M_\alpha(X)$ 为 α 在域 $\mathbb{F}_q$ 上的极小多项式。如果 ω 是 $\mathbb{F}_{q^d}$ 的一个本原元，那么 $M_\omega(X)$ 是 $\mathbb{F}_q$ 中 $\mathbb{F}_{q^d}$ 的本原多项式。如果元素 $\alpha,\beta\in\mathbb{F}_{q^d}$ 拥有相同的极小多项式，那么它们在 $\mathbb{F}_q$ 是共轭的。(v)在 $\mathbb{F}_q$ 中 $\alpha\in\mathbb{F}_{q^d}$ 的共轭为 $\alpha,\alpha^q,\alpha^{q^2},\cdots,\alpha^{q^{d-1}}$，其中 d 是满足 $\alpha^{q^d}=\alpha$ 的最小正整数。当 $\alpha=\omega^j$ 并且 ω 是一个本原元时，共轭集与分圆陪集 $C_\omega=\{\omega^j,\omega^q,\cdots,\omega^{q^{d-1}}\}$ 关联。

总结 3.1.57 现在假设 n 和 $q=p^r$ 互素，并考虑 X^n-e，X^n-e 在分裂域 $\mathrm{Spl}(X^n-e)$ 上的根称作 $\mathbb{F}_q$ 中的 n 次单位根。所有 n 次单位根的集合用 $\mathbb{E}_n$ 表示。(i)集合 $\mathbb{E}_n$ 是域 $\mathrm{Spl}(X^n-e)$ 中一个乘法群的 n 阶循环子群。生成元的 n 次单位根称为本原单位根。(ii)如果 $\mathbb{F}_{q^s}$ 是 $\mathrm{Spl}(X^n-e)$，那么 s 是满足 $n\mid(q^s-1)$ 的最小正整数。(iii)令 $\prod_n$ 为 $\mathbb{F}_q$ 中 n 次本原单位根的集合，Φ_n 是分裂域 $\mathbb{F}_{q^s}=\mathrm{Spl}(X^n-e)$ 的本原元的集合，那么 $\prod_n\cap\Phi_n\neq\varnothing$ 或 $\prod_n=\Phi_n$，当且仅当 $n=q^s-1$ 时后面的式子成立。 290

3.2 Reed-Solomon 编码，再论 BCH 编码

从现在开始，我们考虑有限域 $\mathbb{F}_q$ 及其同构，但在某些情况下也会涉及一个特定的关于域的表格(例如，经典指定 $\mathbb{F}_q$ 为 $\mathbb{F}_p[X]/(g(X))$，其中 $P(X)\in\mathbb{F}_p[X]$ 是 s 次既约多项式)。

在定义 2.5.37 中，我们介绍了狭义二进制 BCH 编码。在本节中，我们会继续讨论长度为 N 的 q 进制 BCH 码 $\mathscr{C}^{\mathrm{BCH}}_{q,N,\delta,\omega,b}$，其距离为 δ 和在元素 $\omega^b,\cdots,\omega^{b+\delta-1}$ 处取零；参见下面的定义 3.2.7。在这之前，我们先讨论 BCH 码中一类有趣的特例，这类码实际上由 Reed-Solomon(RS)码构成；我们将会看到，由于 RS 码是 MDS 的(最大距离可分的)，这个事实有助我们的分析。

定义 3.2.1 给定 $q\geq3$，一个 q 进制 Reed-Solomon 码由一个长度为 $N=q-1$ 的循环编码定义，其生成多项式是

$$g(X)=(X-\omega^b)(X-\omega^{b+1})\cdots(X-\omega^{b+\delta-1}) \tag{3.2.1}$$

其中 δ 和 b 是整数，$1\leq\delta,b<q-1$，ω 是 $\mathbb{F}_q$ 中的本原元(或者等价地说，一个本原 N 次单位根)。这样的编码可以用 $\mathscr{C}^{\mathrm{RS}}$ $(=\mathscr{C}^{\mathrm{BCH}}_{q,N,\delta,\omega,b})$ 表示。

根据定义 3.2.7，RS 码可用 $\mathscr{C}^{\mathrm{RS}}_{q,q-1,\delta,\omega,b}$ 表示，即一个 $(q-1)$ 长的 q 进制 BCH 码，距离是 δ。当码长为 $q-1-1$ 时，不存在合理定义的二进制 RS 编码。注意在域 $\mathbb{F}_q$ 字母表中，非零元素的个数是 $q-1$。并且，对于 $N=q-1$，可得

$$X^N-e=X^{q-1}-e=\prod_{\alpha\in\mathbb{F}_q^*}(X-\alpha)$$

(因为分裂域 $\mathrm{Spl}(X^q-X)$ 正是 $\mathbb{F}_q$。)更进一步可得，ω 是一个本原 $(q-1,\mathbb{F}_q)$-单位根(或 $\mathbb{F}_q$ 中的本原元)，对于所有 $i=0,\cdots,N-1$，极小多项式 $M_i(X)$ 就是 $X-\omega^i$。

一个重要的性质是 RS 码是最大距离可分的。实际上，$\mathscr{C}^{\mathrm{RS}}_{q,N,\delta,\omega,b}$ 的生成多项式 $g(X)$ 中 $\deg g(X)=\delta-1$。所以，秩 k 由下式给出

$$k=\dim\mathscr{C}^{\mathrm{RS}}_{q,N,\delta,\omega,b}=N-\deg g(X)=N-\delta+1 \tag{3.2.2}$$

由广义 BCH 界(参见下面的定理 3.2.9)可得最小距离

$$d(\mathscr{C}^{\mathrm{RS}}_{q,N,\delta,\omega,b})\geq\delta=N-k-1$$

但是 Singleton 界指出 $d(\mathscr{C}^{\mathrm{RS}})\leq N-k-1$。所以，

$$d(\mathscr{C}^{\mathrm{RS}}_{q,N,\delta,\omega,b})=N-k+1=\delta \tag{3.2.3}$$ 291

所以在所有码长度为 $q-1$，维数为 $k=q-\delta$ 的 q 进制码中，RS 码具有最大的最小码距离。总结来说，我们得到

定理 3.2.2 编码 $\mathscr{C}^{\mathrm{RS}}_{q,N,\delta,\omega,b}$ 是 MDS，具有所需 δ 和秩 $q-\delta$。

BCH 码的对偶并不总是 BCH 码。但有如下定理。

定理 3.2.3 RS 码的对偶码也是 RS 码。

证明　证明是很直接的，因为$(\mathscr{C}^{RS}_{q,\delta,\omega,b})^{\perp}=\mathscr{C}^{RS}_{q,q-\delta,\omega,b+(\delta-1)}$。 □

定理 3.2.4　令$\mathscr{C}^{RS}$为$[N,k,\delta]$RS码，那么它的奇偶校验扩展是一个$[N+1,k,\delta+1]$码，距离比$\mathscr{C}^{RS}$大。

证明　令$c(X)=c_0+c_1X+\cdots+c_{N-1}X^{N-1}\in\mathscr{C}^{RS}$，重量为$w(c(X))=\delta$，它的扩展$\tilde{c}(X)=c(X)+c_NX^N$，$c_N=-\sum_{0\leqslant i\leqslant N-1}c_i=-c(e)$。我们想证明$c(e)\neq0$，所以有$w(\tilde{c}(X))=\delta+1$。

为了简化标记，假设$b=1$，令$\mathscr{C}^{RS}$的生成多项式为$g(X)=(X-\omega)(X-\omega^2)\cdots(X-\omega^{\delta-1})$。对于某个$p(X)$，令$c(X)=g(X)p(X)$，进而得到$c(e)=p(e)g(e)$。显然，对于所有$i=1,\cdots,\delta-1$，因为$\omega^i\neq e$，所以有$g(e)\neq0$。如果$p(e)=0$，多项式$g_1(X)=(X-e)g(X)$能够整除$c(X)$，那么$c(X)\in\langle g_1(X)\rangle$，其中$g_1(X)=(X-e)(X-\omega)\cdots(X-\omega^{\delta-1})$。也就是说，$\langle g_1(X)\rangle$是BCH码，距离大于$\delta+1$。但是这和$c(X)$的选择矛盾。 □

RS码有专属的(也是优美的)编解码步骤。考虑一个长为$N=q-1$的$[N,k,\delta]$码$\mathscr{C}^{RS}$，对于消息序列$a_0\cdots a_{k-1}$，令$a(X)=\sum_{0\leqslant i\leqslant k-1}a_iX^i$并将其编码为$c(X)=\sum_{0\leqslant j\leqslant N-1}a(\omega^j)X^j$。为了证明$c(X)\in\mathscr{C}^{RS}$，必须检验$c(\omega)=\cdots=c(\omega^{\delta-1})=0$。将$a(X)$视为满足$a_i=0$的多项式$\sum_{0\leqslant i\leqslant N-1}a_iX^i$，$i\geqslant k$，利用下面的引理

引理 3.2.5　令$a(X)=a_0+a_1X+\cdots+a_{N-1}X^{N-1}\in\mathbb{F}_q[X]$，$\omega$是$\mathbb{F}_q$中的一个$(N,\mathbb{F}_q)$本原单位根，$N=q-1$。那么

$$a_i=\frac{1}{N}\sum_{0\leqslant j\leqslant N-1}a(\omega^j)\omega^{-ij} \tag{3.2.4}$$

我们把对上述引理的证明放在引理3.2.12的后面。

根据引理3.2.5

292

$$a_i=\frac{1}{N}\sum_{0\leqslant j\leqslant N-1}a(\omega^j)\omega^{-ij}=\frac{1}{N}c(\omega^{-i})=\frac{1}{N}c(\omega^{N-i})$$

可得$c(\omega^j)=Na_{N-j}$。对于$0\leqslant j\leqslant\delta-1=N-k$，$c(\omega^j)=Na_{N-j}=0$。因此，$c(X)\in\mathscr{C}^{RS}$。此外，原始的消息可以很容易地从$c(X)$：$a_i=\frac{1}{N}c(\omega^{N-i})$中恢复。

对接收到的码字$u(X)=c(X)+e(X)$进行解码，记

$$u_i=c_i+e_i=e_i+a(\omega^i),0\leqslant i\leqslant N-1$$

然后得到

$$\begin{aligned}
u_0&=e_0+a_0+a_1+\cdots+a_{k-1}\\
u_1&=e_1+a_0+a_1\omega+\cdots+a_{k-1}\omega^{k-1}\\
u_2&=e_2+a_0+a_1\omega^2+\cdots+a_{k-1}\omega^{2(k-1)}\\
&\vdots\\
u_{N-1}&=e_{N-1}+a_0+a_1\omega^{N-1}+\cdots+a_{k-1}\omega^{(N-1)(k-1)}
\end{aligned}$$

如果没有误码，即$e_0=\cdots=e_{N-1}=0$，这些等式中的任何k都可对k个未知量求解，因为对应的矩阵是Vandermond矩阵。实际上，对于任意错误向量，由任意k个等式构成的子系统都可以求解(而这个解是否能给出正确的序列$a_0,\cdots,a_{k-1}$则是另外一个问题)。

现在假设出现t个错误，$t<N-k$。称$e_i=0$为好的等式，$e_i\neq0$为坏的等式。那么我们有t个坏等式，$N-t$个好的等式。如果我们求解所有由k个等式构成的子系统，那么含有k个好等式的$\binom{N-t}{k}$个子系统将给出a_is的正确值。并且，一个给定的错误解不会满足

任意 k 个好等式，它最多可以满足 $k-1$ 个正确的等式。此外，它至多可以满足 t 个不正确的等式。所以，它是 $\leqslant t+k-1$ 个等式的解，即在由 k 个等式构成的子系统中它可被求解 $\leqslant \binom{t+k-1}{k}$ 次。所以，如果

$$\binom{N-t}{k} > \binom{t+k-1}{k}$$

大部分从 $\binom{N}{k}$ 个解中得到的解能给出 $a_i s$ 的正确值。当且仅当 $N-t>t+k-1$ 时，即 $\delta=N-k+1>2t$ 时，最后一个不等式成立。所以我们有：

定理 3.2.6 对于一个 $[N, k, \delta]$ RS 码 $\mathscr{C}^{RS}$，大数逻辑译码算法可以纠正至多 $t<\delta/2$ 个错误，代价是必须求解 $\binom{N}{k}$ 个包含 $k\times k$ 个等式的系统。

Reed-Solomon 码在 1960 年由 Irving S. Reed 和 Gustave Solomon 发现，他们那时都在 MIT 的林肯实验室工作。他们合作的文章发表后，针对这些编码的有效率译码算法尚未出现。在 1969 年由 Elwyn Berlekamp 和 James Massey 共同发现了译码算法，从此命名为 Berlekamp-Massey 译码算法（参见[20]），详见 3.3 节。之后，其他的算法也陆续被提出：连续分式算法和欧几里得算法（参见[112]）。 293

在整个 20 世纪 70 年代和 80 年代，Reed-Solomon 码经常与其他一些码结合，在美国宇宙飞船传输数码照片方面扮演了重要角色。虽然 turbo 码的问世为编译码提供了更多的选择，这些编码在现代空间宇航任务中依然起到显著的作用。

Reed-Solomon 码同时也对 CD 和数码游戏的生产起到了关键作用。其中涉及的编解码方法，可以纠正至多 4000 个连续突发错误（这相当于 CD 表面上 2.5mm 的长度）。

定义 3.2.7 一个参数为 q，N，δ，ω，b 的 BCH 码 $\mathscr{C}^{BCH}_{q,N,\delta,\omega,b}$ 是 N 长 q 进制循环码 $\mathscr{C}_N=(g(X))$，它距离为 δ，生成多项式是

$$g(X)=\operatorname{lcm}(M_{\omega^b}(X), M_{\omega^{b+1}}(X), \cdots, M_{\omega^{b+\delta-2}}(X)) \tag{3.2.5}$$

即

$$\mathscr{C}^{BCH}_{q,N,\delta,\omega,b}=\{f(X)\in \mathbb{F}_q[X]\bmod(X^N-1): f(\omega^{b+i})=0, 0\leqslant i\leqslant \delta-2\}$$

如果 $b=1$，这是一个狭义 BCH 码。如果 ω 是第 N 个本原单位根，即多项式 X^N-1 的一个本原根，那么 BCH 码被称为本原的。（在条件 $\gcd(q, N)=1$ 下，这些根组成一个 N 阶可交换乘性群，其中 ω 是这个群的生成元。）

引理 3.2.8 BCH 码 $\mathscr{C}^{RS}_{q,N,\delta,\omega,b}$ 的最小距离 $\geqslant\delta$。

证明 为了不失一般性，我们考虑一个狭义码，奇偶检验矩阵 $(\delta-1)\times N$ 是

$$\boldsymbol{H}=\begin{pmatrix} 1 & \omega & \omega^2 & \cdots & \omega^{N-1} \\ 1 & \omega^2 & \omega^4 & \cdots & \omega^{2(N-1)} \\ \vdots & \vdots & & \ddots & \vdots \\ 1 & \omega^{\delta-1} & \omega^{2(\delta-1)} & \cdots & \omega^{(\delta-1)(N-1)} \end{pmatrix}$$

$\mathscr{C}$ 中的码字满足矩阵 $\boldsymbol{H}$ 列之间的线性相关关系。引理 2.5.40 表明矩阵 $\boldsymbol{H}$ 的任意 $\delta-1$ 列都是线性独立的。实际上，选择首项为 ω^{k_1}，…，$\omega^{k_{\delta-1}}$ 的列，其中 $0\leqslant k_1<\cdots<k_{\delta-1}\leqslant N-1$。它们构成一个 $(\delta-1)\times(\delta-1)$ 的方阵 294

$$D=\begin{vmatrix} \omega^{b_1}\cdot 1 & \omega^{b_2}\cdot 1 & \cdots & \omega^{b_{\delta-1}}\cdot 1 \\ \omega^{b_1}\cdot\omega^{b_1} & \omega^{b_2}\cdot\omega^{b_2} & \cdots & \omega^{b_{\delta-1}}\cdot\omega^{b_{\delta-1}} \\ \vdots & \vdots & \ddots & \vdots \\ \omega^{b_1}\cdot\omega^{b_1(\delta-2)} & \omega^{b_2}\cdot\omega^{b_2(\delta-2)} & \cdots & \omega^{b_{\delta-1}}\cdot\omega^{b_{\delta-1}(\delta-2)} \end{vmatrix}$$

因为在第 s 列前面存在因子 ω^{b_s}，上述矩阵和 Vandermonde 矩阵不一样，于是 $\boldsymbol{D}$ 的行列式有如下乘积表示

$$\det\boldsymbol{D}=\left(\prod_{s=1}^{\delta-1}\omega^{b_s}\right)\begin{vmatrix} 1 & 1 & \cdots & 1 \\ \omega^{b_1} & \omega^{b_2} & \cdots & \omega^{b_{\delta-1}} \\ \vdots & \vdots & \ddots & \vdots \\ \omega^{b_1(\delta-2)} & \omega^{b_2(\delta-2)} & \cdots & \omega^{b_{\delta-1}(\delta-2)} \end{vmatrix}$$

$$=\left(\prod_{s=1}^{\delta-1}\omega^{b_s}\right)\times\left(\prod_{i>j}(\omega^{b_i}-\omega^{b_j})\right)\neq 0$$

矩阵 $\boldsymbol{H}$ 中任意 $\delta-1$ 的列确实都是线性独立的。反过来说，这表明 $\mathscr{C}$ 中的任意非零码字的重量至少是 δ。因此，$\mathscr{C}$ 的最小距离至少大于等于 δ。 □

定理 3.2.9 （广义的 BCH 界）令 ω 是第 N 个本原单位根，$b\geqslant 1$，$r\geqslant 1$ 和 $\delta>2$ 是整数，$\gcd(r,N)=1$。考虑一个 N 长循环码 $\mathscr{C}=\langle g(X)\rangle$，其中 $g(X)$ 是满足 $g(\omega^b)=g(\omega^{b+r})=\cdots=g(\omega^{b+(\delta-2)r})=0$ 的最低次首一多项式。证明 $d(\mathscr{C})\geqslant\delta$。

证明　因为 $\gcd(r,N)=1$，ω^r 是一个本原单位根。所以，我们可以重复上述证明，用 bru 代替 b，其中 ru 从 $ru+Nv=1$ 得到。另外一种解法：$(\delta-1)\times N$ 矩阵

$$\begin{vmatrix} 1 & 1 & \cdots & 1 \\ \omega^{b} & \omega^{b+r} & \cdots & \omega^{b+(\delta-2)r} \\ \omega^{2b} & \omega^{2(b+r)} & \cdots & \omega^{2(b+(\delta-2)r)} \\ \vdots & \vdots & \ddots & \vdots \\ \omega^{(N-1)b} & \omega^{(N-1)(b+r)} & \cdots & \omega^{(N-1)(b+(\delta-2)r)} \end{vmatrix}$$

检查码 $X-\langle g(X)\rangle$，选取任意一个它的 $(\delta-1)\times(\delta-1)$ 子矩阵，可以考虑其中的列为 $i_1<i_2<\cdots<i_{\delta-1}$。用 $\boldsymbol{D}=(\boldsymbol{D}_{ij})$ 表示该子矩阵，那么

$$\det\boldsymbol{D}=\prod_{1\leqslant i\leqslant\delta-1}\omega^{(i_i-1)b}\det(\omega^{r(i_i-1)(j-1)})$$

$$=\prod_{1\leqslant i\leqslant\delta-1}\omega^{(i_i-1)b}\det(\text{Vandermonde})\neq 0$$

295 因为 $\gcd(r,N)=1$，所以 $d(X)\geqslant\delta$。 □

举例 3.2.10　ω 是 $\mathbb{F}_q$ 扩域中第一个 n 次本原单位根，$a(X)=\sum_{0\leqslant i\leqslant n-1}a_iX^i$ 是次数至多为 $n-1$ 的多项式。Mattson-Solomon 多项式定义如下

$$a_{\mathrm{MS}}(X)=\sum_{j=1}^{n}a(\omega^j)X^{n-j} \tag{3.2.6}$$

令 $q=2$ 而 $a(X)\in\mathbb{F}_2[X]/(X^n-1)$，证明 Mattson-Solomon 多项式 $a_{\mathrm{MS}}(X)$ 是幂等的，即在 $\mathbb{F}_2(X)/(X^n-1)$ 中 $a_{\mathrm{MS}}(X)^2=a_{\mathrm{MS}}(X)$。

解答　令 $a(X)=\sum_{0\leqslant i\leqslant n-1}a_iX^i$，那么由引理 3.2.5 可知 $na_i=a_{\mathrm{MS}}(\omega^i)$，$0\leqslant i\leqslant n-1$。在 $\mathbb{F}_2$ 中，$(na_i)^2=na_i$，所以 $a_{\mathrm{MS}}(\omega^i)^2=a_{\mathrm{MS}}(\omega^i)$。对于多项式，$b^{(2)}(X)$ 代表 $\mathbb{F}_2[X]$ 中的平方，$b(X)^2$ 代表 $\mathbb{F}_2[X]/(X^n-1)$ 中的平方：

$$b^{(2)}(X)=c(X)(X^n-1)-b(X)^2$$

那么

$$a_{MS}(X)\restriction_{X=\omega^i}=(a_{MS}(X)\restriction_{X=\omega^i})^2=a_{MS}^{(2)}(X)\restriction_{X=\omega^i}$$
$$=a_{MS}(X)^2\restriction_{X=\omega^i}$$

即多项式 $a_{MS}(X)^2$ 和 $a_{MS}(X)$ 都满足 $\omega^i=e,\ \omega,\ \cdots,\ \omega^{n-1}$。把上式写成矩阵形式，$a_{MS}(X)=a_{1,MS}X+\cdots+a_{n-1,MS}X^{n-1}$，$a_{MS}(X)^2=a_{1,MS}^2X+\cdots+a_{n-1,MS}^2X^{n-1}$：

$$(a_{MS}-a_{MS}^{(2)})\begin{pmatrix} e & e & \cdots & e \\ e & \omega & \cdots & \omega^{n-1} \\ \vdots & \vdots & \ddots & \vdots \\ e & \omega^{n-1} & \cdots & \omega^{(n-1)^2} \end{pmatrix}=0$$

因为矩阵是 Vandermonde 矩阵，它的行列式是

$$\prod_{0\leqslant i<j\leqslant n-1}(\omega^j-\omega^i)\neq 0$$

且 $a_{MS}=a_{MS}^{(2)}$。所以，$a_{MS}(X)=a_{MS}(X)^2$。 □

定义 3.2.11 令 $v=v_0v_1\cdots v_{N-1}$ 为 $\mathbf{F}_q$ 中的向量，ω 是 $\mathbf{F}_q$ 中本原$(N,\ \mathbf{F}_q)$-单位根。向量 v 的 Fourier 变换是向量 $V=V_0V_1\cdots V_{N-1}$，其中的每一项是

$$V_j=\sum_{i=0}^{N-1}\omega^{ij}v_i,\ j=0,\cdots,N-1 \tag{3.2.7}$$

[295]

引理 3.2.12 （反演公式）利用如下公式，向量 v 可以从它的 Fourier 变换 V 中恢复为

$$v_i=\frac{1}{N}\sum_{j=0}^{N-1}\omega^{-ij}V_j \tag{3.2.8}$$

证明 在任意 $X^N-1=(X-1)(X^{N-1}+\cdots+X+1)$ 域中，ω 的阶是 N，对于任意 r，ω^r 是（上述多项式）LHS 的零点。所以，对于所有 $r\neq 0 \bmod$，ω^r 是（上述多项式）最后一项的一个零，即

$$\sum_{j=0}^{N-1}\omega^{rj}=0 \bmod N$$

另一方面，对于 $r=0$ 有

$$\sum_{j=0}^{N-1}\omega^{rj}=N \bmod p$$

如果 N 不是域特征 p 的倍数，上式就不为零。但是 $q-1=p^r-1$ 是 N 倍数，所以 N 不是 p 倍数。所以，$N\neq 0 \bmod p$。最后，改变加和的顺序可以得到

$$\frac{1}{N}\sum_{j=0}^{N-1}\omega^{-ij}V_j=\frac{1}{N}\sum_{k=0}^{N-1}v_k\sum_{j=0}^{N-1}\omega^{(k-i)j}=v_i$$

证明引理 3.2.5 令 $a(X)=a_0+a_1X+\cdots+a_{N-1}X^{N-1}\in\mathbf{F}_q[X]$，$\omega$ 是本元单位根$(N,\ \mathbf{F}_q)$。那么

$$N^{-1}\sum_{0\leqslant i\leqslant N-1}a(\omega^i)\omega^{-i}=N^{-1}\sum_{0\leqslant i\leqslant N-1}\sum_{0\leqslant k\leqslant N-1}a_k\omega^{ik}\omega^{-i}$$
$$=N^{-1}\sum_{0\leqslant k\leqslant N-1}a_k\sum_{0\leqslant j\leqslant N-1}\omega^{(k-i)j}$$
$$=N^{-1}\sum_{0\leqslant k\leqslant N-1}a_kN\delta_{ki}=a_i$$

这里，对于 $1\leqslant t\leqslant N-1$，$\omega^t\neq 1$ 且

$$\sum_{0\le i\le N-1}\omega^{ij}=\sum_{0\le i\le N-1}(\omega^j)^i=(e-(\omega^j)^N)(e-\omega^j)^{-1}=0$$

得到

297

$$a_j=\frac{1}{N}\sum_{0\le i\le N-1}a(\omega^i)\omega^{-ij} \tag{3.2.9}$$

□

举例 3.2.13 给出 BCH 界的另外一种证明方法：令 ω 是为本原$(N,\mathbb{F}_q)$-单位根，$b\ge1$，$\delta\ge2$ 是整数，令 $\mathscr{C}_N=(g(X))$ 是一个循环编码，其中 $g(X)\in\mathbb{F}_q[X]/(X^N-e)$ 是最低次首一多项式且具有根 $\omega^b,\omega^{b+1},\cdots,\omega^{b+\delta-2}$，那么 $\mathscr{C}$ 的最小距离至少是 δ。

解答 令 $a(X)=\sum_{0\le i\le N-1}a_iX^i\in\mathscr{C}_N$ 满足 $g(X)\mid a(X)$ 和 $a(\omega^i)=0(i=b,\cdots,b+\delta-2)$。考虑 $a(X)$ 的 Mattson-Solomon 多项式 $c_{MS}(X)$：

$$\begin{aligned}c_{MS}(X)&=\sum_{0\le i\le N-1}a(\omega^{-i})X^i=\sum_{0\le i\le N-1}a(\omega^{N-i})X^i\\&=\sum_{1\le i\le N}a(\omega^i)X^{N-i}\\&=\sum_{1\le i\le b-1}a(\omega^i)X^{N-i}+0+\cdots+0(\text{共自 }\omega^b,\cdots,\omega^{b+\delta-2})\\&\quad+a(\omega^{b+\delta-1})X^{N-b-\delta+1}+\cdots+a(\omega^N)\end{aligned} \tag{3.2.10}$$

乘以 X^{b-1} 并合并同类项得到

$$\begin{aligned}X^{b-1}c_{MS}(X)&=a(\omega)X^{N-b+1}+\cdots+a(\omega^{b-1})X^N\\&\quad+a(\omega^{b+\delta-1})X^{N-\delta}+\cdots+a(\omega^N)X^{b-1}\\&=X^N[a(\omega)X^{b-2}+\cdots+a(\omega^{b-1})]\\&\quad+[a(\omega^{b+\delta-1})X^{N-\delta}+\cdots+a(\omega^N)X^{b-1}]\\&=X^Np_2(X)-q(X)\\&=(X^N-e)p_2(X)+p_1(X)+q(X)\end{aligned}$$

可以看出当且仅当 $p_1(\omega^i)+q(\omega^i)=0$ 时，$c_{MS}(\omega^i)=0$。但是 $p_1(X)+q(X)$ 是一个次数 $\le N-\delta$ 的多项式，所以它至多有 $N-\delta$ 个根。所以，$c_{MS}(X)$ 至多有 $N-\delta$ 个形为 ω^i 的根。

所以，反演式(3.2.8)表明重量 $w(a(X))$(即系数序列 $a_0\cdots a_{N-1}$ 的重量)遵循

$$w(a(X))\ge N-c_{MS}(X)\text{ 中具有形如 }\omega^i\text{ 的根的数量} \tag{3.2.11}$$

即

$$w(a(X))\ge N-(N-\delta)=\delta$$

□

在本节结束之前，我们简单讨论一下一种列表译码 Reed-Solomon 码的算法，即 Guruswami-Sudan 算法。首先，我们需要用另外一种描述方法表示 Reed-Solomon 码(在 Reed 和 Solomon 联合发表的文章中已经完成了这项工作)。为了简洁，考虑 $b=1$(但是可以扩展定义到任意 $N>q-1$)。

298

给定 $N\le q$，令 $S=\{x_1,\cdots,x_N\}\subset\mathbb{F}_q$ 是 $\mathbb{F}_q$ 中 N 个取值不同点的集合(一个支撑集)。令 Ev 表示赋值映射

$$Ev:f\in\mathbb{F}_q[X]\to Ev(f)=(f(x_1),\cdots,f(x_N))\in\mathbb{F}_q^N \tag{3.2.12}$$

考虑

$$L=\{f\in\mathbb{F}_q[X]:\deg f<k\} \tag{3.2.13}$$

那么 N 长 q 进制 Reed-Solomon 码的维数 k 可以定义为

$$\mathscr{C}=Ev(L) \tag{3.2.14}$$

它的最小距离是 $d=d(\mathscr{C})=N-k+1$，至多可以纠正 $\left\lfloor\frac{d-1}{2}\right\rfloor$ 个错误。对信源消息 $u=u_0\cdots u_{k-1}\in\mathbb{F}_q^k$ 的编码需要计算多项式 $f(X)=u_0+u_1X+\cdots+u_kX^{k-1}$ 在点 $x_i\in S$ 的值。

定义 3.2.1（其中 $\mathscr{C}$ 是多项式 $c(X)=\sum_{0\leqslant i\leqslant N-1}c_iX^i\in\mathbb{F}_q[X]$ 的集合，满足 $c(\omega)=c(\omega^2)=\cdots=c(\omega^{\delta-1})=0$）当 $N=q-1$，$k=N-\delta+1=q-\delta$ 时使用，支集 $S=\{e,\omega,\cdots,\omega^{N-1}\}$ 和参数 $c_0,c_1,\cdots,c_{N-1}$ 利用下面的式子与多项式 $f(X)$ 建立等式关系

$$c_i=f(\omega^i),0\leqslant i\leqslant N-1$$

这唯一确定了表达式 $f(X)=\sum_{0\leqslant i\leqslant N-1}f_iX^i$ 中的参数 f_i，通过离散的 Fourier 逆变换可得

$$Nf_i=c(\omega^{N-i}),\quad 或者\ Nf_{N-i}=c(\omega^{i}),\quad i=0,\cdots,N-1$$

特别地，确保了 $f_k=\cdots=f_{N-1}=0$。

给定 $f\in\mathbb{F}_q[X]$ 和 $y=y_1\cdots y_N\in\mathbb{F}_q^N[X]$，令

$$\mathrm{dist}(f,y)=\sum_{1\leqslant i\leqslant N}\mathbf{1}(f(x_i)\neq y_i)$$

现在假设 $y=y_1\cdots y_N$ 是一个接收码字并记为 $t=\left\lfloor\frac{d-1}{2}\right\rfloor$。上述“传统的”的译码算法（Berlekamp-Massey 算法，连续分式算法和欧几里得算法）都遵循相同的规则：算法可以找到唯一的 f 使得 $\mathrm{dist}(f,y)\leqslant t$，或者说明 f 不存在。另一方面，给定 $s>t$，列表译码试图找到所有满足 $\mathrm{dist}(f,y)\leqslant s$ 的 f；如果足够幸运，具有这个性质的码字是唯一的，那么我们可以纠正 s 个错误，这超过了“传统”纠正 t 个错误的极限。 299

这个想法可以追溯到 Shannon 有界距离解码：当收到字 y 后，检查 y 附近的 Hamming 球，直到发现最接近 y 的码字（或者最接近的码字集合）为止。当然，我们得确定两种条件，即：(i)当取比 t 稍微大的 s 时，找到两个或更多距离在 s 内的码字可能性很小，(ii)算法的计算复杂度合理。

例子 3.2.14 对于 $\mathbb{F}_{32}$ 上的[32，8]RS 码，$d=25$，$t=12$。如果我们想取 $s=13$，则接收到的字 y 的 Hamming 球可能包含两个码字。不过，假设权重为 13 的所有误差向量 e 近似相等，那么这件事情发生的概率是 2.08437×10^{-11}。

用 Guruswami-Sudan 列表译码算法（见[59]）在一个多项式时间内寻找相隔距离为 s 的多个码字，$t\leqslant s\leqslant t_{GS}$，其中

$$t_{GS}=n-1-\left\lfloor\sqrt{(k-1)n}\right\rfloor$$

并且 t_{GS} 比 t 大得多。

在上面的案例中，$t_{GS}=17$。对于速率为 R 的 RS 码，传统的解码算法将纠正 $(1-R)/2$ 的错误，然而 GS 算法能纠正的比率达到 $1-\sqrt{R}$。在半径为 $s\leqslant t_{GS}$ 的球上，期望码字数量可以被评估出来（假设误差向量均匀分布）。

Guruswami-Sudan 算法不仅仅可以用于 RS 码。在最初的 GS 文章中，这个算法对于不同类型码字也有很好的效果；之后，它也被用于解决 RM 码（参见[7]）。

3.3 再论循环码，BCH 解码

我们重新回顾循环码和 BCH 码。像以前一样，假设 $\gcd(N,q)=1$（所以如果 $q=2$，N 是奇数），用 $x_0\cdots x_{N-1}$ 表示字 $x\in\mathscr{H}_{N,q}$。注意对于所有 $x=x_0\cdots x_{N-1}\in\mathscr{C}$，如果循环移动 $\tilde{x}=x_{N-1}x_1\cdots x_{N-2}\in\mathscr{C}$，那么线性码 $\mathscr{C}\subseteq\mathscr{H}_{N,q}$ 被称为是循环的。对于每一个字 $c=c_0\cdots c_{N-1}$，我们用它们表示一个多项式 $c(X)\in\mathbb{F}_q[X]$：

$$c(X) = c_0 + c_1X + \cdots + c_{N-1}X^{N-1}$$

映射 $c \leftrightarrow c(X)$ 在 $\mathscr{C}$ 和 $\mathbb{F}_q[X]$ 的线性空间之间是同构的。

引理 3.3.1 当且仅当在剩环 $\mathbb{F}_q[X]/\langle X^N - e\rangle$ 中，$\mathscr{C}$ 的像在上述的同构上是理想的，则这时码 $\mathscr{C}$ 是循环的。

[300] 证明 循环位移对应 X 乘以一个多项式 $c(X)$。因此任何多项式相乘都保留 $\mathscr{C}$。 □

在 $\mathbb{F}_q[X]/\langle X^N - e\rangle$ 中，我们认为 $\mathscr{C}$ 是理想的，并且考虑将所有的多项式 $\bmod\langle X^N - e\rangle$，这种方法能得到很好的效果。此外，$\mathbb{F}_q[X]/\langle X^N - e\rangle$ 是一个主理想环；它的每一个理想形式是

$$\langle g(X)\rangle = \{f(X): f(X) = g(X)h(X), h(X) \in \mathbb{F}_q[X]/\langle X^N - e\rangle\} \qquad (3.3.1)$$

在这里 $g(X)$ 是固定多项式。

定理 3.3.2 如果码 $\mathscr{C} \subset \mathscr{C}_{N,q}$ 是周期的，那么，存在特殊首一多项式 $g(X) \in \mathscr{C}$，它符合

(i) $\mathscr{C} = \langle g(X)\rangle$。

(ii) 在所有的多项式 $f(X) \in \mathscr{C}$ 中，$g(X)$ 的度最小。而且：

(a) $g(X) \mid \langle X^N - e\rangle$。

(b) 如果 $\deg g(X) = d$，那么 $\dim \mathscr{C} = N - d$。

(c) $\mathscr{C} = \{f(X): f(X) = g(X)h(X)\}$，$h(X) \in \mathbb{F}_q[X]$，$\deg h(X) < N - d\}$。

(d) 如果 $g(X) = g_0 + g_1X + g_2X^2 + \cdots + g_dX^d$，$g_d = e$，那么 $g_0 \neq 0$，而且

$$\boldsymbol{G} = \begin{bmatrix} g_0 & g_1 & g_2 & \cdots & g_d & 0 & 0 & \cdots & 0 \\ 0 & g_0 & g_1 & \cdots & g_d & g_d & 0 & \cdots & 0 \\ & & & \cdots & & & & \cdots & \\ 0 & 0 & 0 & \cdots & g_1 & g_2 & & \cdots & g_d \end{bmatrix}$$

对 $\mathscr{C}$ 来说是一个生成矩阵，第 i 行是第 $i-1$，$i=2, \cdots, N-d$ 行的循环移动。

相反，对于任意一个多项式 $g(X) \mid \langle X^N - e\rangle$，集合 $\langle g(X)\rangle = \{f(X): f(X) = g(X)h(X), h(X) \in \mathbb{F}_q[X]/\langle X^N - e\rangle\}$ 在 $\mathbb{F}_q[X]/\langle X^N - e\rangle$ 中是理想的，即对于循环码 $\mathscr{C}$，上述性质(b)—(d)成立。

证明 在 $\mathscr{C}$ 中取有最小度的非零多项式 $g(X) \in \mathbb{F}_2[X]$。取 $p(X) \in \mathscr{C}$，

$$p(X) = q(X)g(X) + r(X), \deg r(X) < \deg g(X)$$

那么 $r(X) \bmod \langle X^N - 1\rangle$ 属于 $\mathscr{C}$，除非 $r(X) = 0$，否则这将和 $g(X)$ 相矛盾。因此，$g(X) \mid p(X)$ 可以证明(i)是成立的。选取 $p(X) = X^N - 1$ 可以证明了(ii)。最后，如果 $g(X)$ 和 $\tilde{g}(X)$ 同时满足(i)和(ii)，那么 $g(X) \mid \tilde{g}(X)$ 和 $\tilde{g}(X) \mid g(X)$ 意味着 $g(X) = \tilde{g}(X)$。 □

推论 3.3.3 在一个点对点通信中，长度为 N 的循环码中有 $X^N - e$ 这个元素，换句话说，映射

$$\{\text{长度为 } N \text{ 的循环码}\} \to \{X^N - 1 \text{ 的因子}\}$$
$$\mathscr{C} \mapsto g(X)$$

[301] 是个双向单射。

随着识别鉴定

$$\mathbb{F}_2[X]/\langle\langle X^N - 1\rangle\rangle = \{f \in \mathbb{F}_2[X]: \deg(f) < N\} = \mathbb{F}_2^N$$

的出现，在多项式环 $\mathbb{F}_2[X]/\langle X^N - 1\rangle$ 中，循环码成为最佳状态。这些循环码在包含 $X^N - 1$ 的 $\mathbb{F}_2[X]$ 中有着点对点通信的理想状态。因为 $\mathbb{F}_2[X]$ 是 Euclid 几何学域，在 $\mathbb{F}_2[X]$ 中的所有理念都很重要，形式为 $\{f(X)g(X): f(X) \in \mathbb{F}_2[X]\}$。事实上，在 $\mathbb{F}_2[X]/\langle X^N - 1\rangle$ 上的

所有理念都是重要的理念。

定义 3.3.4 多项式 $g(X)$ 被称为循环码 X 的最小度生成多项式(简单来说就是生成器)，度为 $N-\deg g(X)$ 的比值 $h(X)=(X^N-e)/g(X)$ 被称为循环码 $\mathscr{X}=\langle g(X)\rangle$ 的校验多项式。

例子 3.3.5 $X=e$ 生成奇偶校验码 $\{x:\ \sum_i x_i=0\}$ 以及 $e+X+\cdots+X^{N-1}$ 这个重复码 $\{a\cdots a,\ a\in F_q\}$，$X=e$ 生成 $\mathscr{X}=\mathscr{H}$。

举例 3.3.6 (a) 如果 $c_0\cdots c_{N-1}\in\mathscr{X}$ 也表示 $c_{N-1}\cdots c_0\in\mathscr{X}$，那么长度为 N 的循环码 $\mathscr{X}=\langle g(X)\rangle$ 被称为是可逆的。证明当且仅当 $g(\alpha)=0$ 意味着 $g(\alpha^{-1})=0$ 时，$\mathscr{X}$ 是可逆的。

(b) 对于一些 r/N，如果每个码字 $c\in\mathscr{X}$ 是长度为 r 的字符串 $c'c'\cdots c'$ 中 N/r 个字符的串联。证明当且仅当它的校验多项式是 $h(X)\mid(X^r-1)$ 时，X 是退化的。(提示：证明生成多项式 $g(X)=a(X)(1+X^r+X^{2r}+\cdots+X^{N-r})$。)

解答 (a) 如果码 $\mathscr{X}=\langle g(X)\rangle$ 是可逆的，$g=g_0\cdots g_{N-1}0\cdots0$，那么 $X^{N-1}g(X^{-1})\sim0\cdots0g_{N-1}\cdots g_0\in\mathscr{X}$，即 $X^{N-1}g(X^{-1})=g(X)q(X)$。因此，如果 $g(\alpha)=0$，那么 $\alpha^{N-1}g(\alpha^{-1})=0$，即 $g(\alpha^{-1})=0$。

相反，$g(\alpha)=0$ 意味着 $g(\alpha^{-1})=0$。假设 $c(X)\in\mathscr{X}$，$g(X)\mid c(X)$。而且，$X^{N-1}c(X^{-1})$ 中存在 $g(X)$ 所有的零点根，所以，$X^{N-1}c(X^{-1})$ 属于 $\mathscr{X}$。但是 $X^{N-1}c(X^{-1})\sim c_{N-1}\cdots c_0$，所以 $\mathscr{X}$ 是可逆的。

(b) 条件 $g=a'\cdots a'$ 意味着 $g(X)=a(X)(e+X^r+X^{2r}+\cdots+X^{N-r})$。另一方面，

$$X^N-e=(X^r-e)(X^{N-r}+\cdots+X^r+e)=h(X)g(X)$$

因此，如果 $\mathscr{X}=\langle g(X)\rangle$ 是退化的，那么 $X^r-e=h(X)a(X)$，即 $X^r-e=h(X)\mid(X^r-e)$。 [302]

相反，如果 $h(X)\mid(X^r-e)$，那么 $X^r-e=a(X)h(X)$，

$$\begin{aligned}X^N-e&=(X^r-e)(X^{N-r}+\cdots+X^r+e)\\&=h(X)a(X)(X^{N-r}+\cdots+X^r+e)\end{aligned}$$

并且

$$g(X)=a(X)(X^{N-r}+\cdots+X^r+e)$$

即 $g=a'\cdots a'$。此外，当 $\deg q(X)\leqslant N-\deg g(X)$ 时，形式为 $c(X)=q(X)g(X)$ 的任意一个 $c(X)\in\mathscr{X}$。$c(X)$ 可表示为

$$c(X)=q(X)g(X)=a(X)q(X)(X^{N-r}+\cdots+X^r+e)$$

从中我们可以总结得出，$\deg a(X)q(X)<r$(在乘以 X^{N-r} 后，它的度不会超过 $N-1$)，并且当 $c'\sim a(X)q(X)$ 时，$c=c'\cdots c'$ 是 $c'\cdots c'$ 的串联。 □

举例 3.3.7 [7，4]Hamming 码是校验多项式为 X^3+X^2+X+1 的循环码，它的生成多项式是什么？Hamming 原始码也会等价于它的偶子码吗？

解答 在 F_2 中，我们有

$$X^7-1=(X^3+X+1)(X^4+X^2+X+1)$$

生成多项式为 X^3+X+1 的循环码的校验多项式是 X^4+X^2+X+1。该码的奇偶校验矩阵是

$$\begin{bmatrix}1&0&1&1&1&0&0\\0&1&0&1&1&1&0\\0&0&0&0&1&1&1\end{bmatrix}$$

这个矩阵的所有列都是属于 F_2^3 的非零元素。所以，它和[7，4]Hamming 码等价。

[7，4]Hamming 码的对偶生成多项式是 X^4+X^3+X+1($h(X)$ 的逆)。因为 $X^6+X^3+X+1=(X+1)g(X)$，所以它是[7，4]Hamming 码的子码。 □

举例 3.3.8 定义 ω 是本原第 N 个单位根，$\mathscr{C}=\langle g(X)\rangle$ 是长度为 N 的循环码，证明维度 $\dim(\mathscr{C})$ 等于第指数 ω^i 的值，且 $g(\omega^i)\neq 0$。

解答 定义 $E=\{\omega,\omega^2,\cdots,\omega^N\}$，$\dim(\langle g(X)\rangle)=N-d$，$d=\deg g(X)$。但是 $g(X)=\prod_{\omega_i\in\mathscr{C}}(X-\omega^i)$，其中 $\omega^1,\cdots,\omega^d$ 是 $\langle g(X)\rangle$ 的零点。因此 ω^i 的剩余 $N-d$ 个单位根满足条件 $g(\omega^i)\neq 0$。 □

[333]

我们需要非常注意的是，循环码 $\mathscr{C}=\langle g(X)\rangle$ 的生成多项式是不唯一的。特别是，这里存在一种特殊的多项式 $i(X)\in\mathscr{C}$，使 $i(X)^2=i(X)$ 以及 $\mathscr{C}=\langle i(X)\rangle$（幂等产生器）。

定理 3.3.9 如果 $\mathscr{C}_1=\langle g_1(X)\rangle$ 和 $\mathscr{C}_2=\langle g_2(X)\rangle$ 是生成器为 $g_1(X)$ 和 $g_2(X)$ 的循环码，那么

(a) 当且仅当 $g_2(X)\mid g_1(X)$ 时，$\mathscr{C}_1\subset\mathscr{C}_2$；

(b) $\mathscr{C}_1\cap\mathscr{C}_2=\langle \mathrm{lcm}(g_1(X),g_2(X))\rangle$。

(c) $\mathscr{C}_1+\mathscr{C}_2=\langle \gcd(g_1(X),g_2(X))\rangle$。

定理 3.3.10 令 $h(X)$ 为 $\mathscr{C}$ 的校验多项式，那么

(a) $\mathscr{C}=\{f(X):f(X)h(X)\equiv 0 \bmod (X^N-e)\}$。

(b) 如果 $h(X)=h_0+h_1X+\cdots+h_{N-r}X^{N-r}$，那么 $\mathscr{C}$ 的奇偶检验矩阵 $\boldsymbol{H}$ 是

$$\boldsymbol{H}=\begin{pmatrix} h_{N-r} & h_{N-r-1} & \cdots & h_1 & h_0 & 0 & 0 & \cdots & 0 \\ 0 & h_{N-r} & \cdots & \cdots & h_1 & h_0 & 0 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & h_{N-r} & h_{N-r-1} & \cdots & \cdots & \cdots & h_0 \end{pmatrix}$$

(c) 对偶码 $\mathscr{C}^\perp$ 是 $\dim\mathscr{C}^\perp=r$ 的循环码，$\mathscr{C}^\perp=\langle g^\perp(X)\rangle$，其中 $g^\perp(X)=h_0^{-1}X^{N-r}h(X^{-1})=h_0^{-1}(h_rX^{N-1}+h_1X^{N-(r-1)}+\cdots+h_{N-r})$。

对于 X^N-e 的因式分解而言，循环码的生成器 $g(x)$ 被指定为最小多项式 $M_\omega(X)$ 的一个子积：

$$X^N-e=\mathrm{lcm}(M_\omega(X):\omega\in E^{(N)}) \tag{3.3.2}$$

一个简单的方法就是用 $g(X)$ 的根来描述循环码。如果 ω 是 $M_\omega(X)$ 在扩展域 $\mathbb{F}_q(\omega)$ 的根，那么 $M_\omega(X)$ 就是在 $\mathbb{F}_q$ 上 ω 的最小多项式。对于任何一个多项式 $f(X)\in\mathbb{F}_q(\omega)$，当且仅当 $f(X)=a(X)M_\omega(X)$ 时有 $f(\omega)=0$。另外，如果 $f(X)\in\mathbb{F}_q[X]/(X^N-e)$，那么当且仅当 $f(X)=\langle M_\omega(X)\rangle$ 时 $f(\omega)=0$。因此我们能够得到下述定理。

定理 3.3.11 令 $g(X)=g_1(X)\cdots g_t(X)$ 是 X^N-e 的不可约因子的乘积，$\omega_1,\cdots,\omega_k$ 是在 $\mathbb{F}_q$ 上 $\mathrm{Spl}(x^N-e)$ 的 $g(X)$ 中的根。那么

$$\langle g(X)\rangle=\{f(X)\in\mathbb{F}_q[X]/(X^N-e):f(\omega_1)=\cdots=f(\omega_k)=0\} \tag{3.3.3}$$

而且，从每个不可约元素中选取一个根是完全可以的；如果 ω_i^j 是 $M_\omega(X)$ 的任意一个根，其中 $1\leq j\leq t$，那么

[334]

$$\langle g(X)\rangle=\{f(X)\in\mathbb{F}_q[X]/(X^N-e):f(\omega_1^j)=\cdots=f(\omega_t^j)=0\} \tag{3.3.4}$$

相反，如果 $\omega_1,\cdots,\omega_k$ 是 X^N-e 根的集合，那么码 $\{f(X)\in\mathbb{F}_q/(X^N-e):f(\omega_1)=\cdots=f(\omega_k)=0\}$ 会有一个 $\omega_1,\cdots,\omega_k$ 的最小多项式的 lcm 生成器。

定义 3.3.12 生成器 $g(X)$ 的根被称为循环码 $\langle g(X)\rangle$ 的零点。其他单位根通常被称为码的非零点。

用 $\{\omega_1,\cdots,\omega_k\}$ 表示在扩展域 $\mathbb{F}_{q^l}$ 上 X^N-e 的根集。从之前可知，l 是最小整数，从而 $N\mid q^l-1$。如果 $f(X)=\sum f_iX^i$ 是在 $\mathbb{F}_q[X]/(X^N-e)$ 上的多项式，那么当且仅当 $\sum_{0\leq i\leq N}f_i\omega_j^i=0$

时 $f(\omega_i)=0$。$\mathbb{F}_2^l$ 表示维度为 l 的 $\mathbb{F}_2$ 上的向量空间，我们将 ω_i 和在 $\mathbb{F}_2$ 上长度为 l 的(列)向量 $\vec{\omega}_i$ 结合起来，最后的等式为 $\sum_i f_i\vec{\omega}_i=\overline{\sum_i f_i\omega_i}=0$。所以这个 $(nl)\times N$ 矩阵

$$\vec{H}^{\mathrm{T}}=\begin{bmatrix}\vec{\omega}_1^0 & \vec{\omega}_1^1 & \cdots & \vec{\omega}_1^{N-1}\\ \vec{\omega}_2^0 & \vec{\omega}_2^1 & \cdots & \vec{\omega}_2^{N-1}\\ \vdots & \vdots & \ddots & \vdots\\ \vec{\omega}_n^0 & \vec{\omega}_n^1 & \cdots & \vec{\omega}_n^{N-1}\end{bmatrix} \tag{3.3.5}$$

可以被认为是零点为 $\omega_1,\cdots,\omega_n$ 的码的校验矩阵(附带条件：它的行可能不是线性独立的)。

定理 3.3.13 对于 $q=2$，$[2^l-1,\ 2^l-l-1,\ 3]$ Hamming 码与循环码 $(M_\omega(X))=\prod_{0\leqslant i\leqslant l-1}(X-\omega^{2^i})$ 是等价的，其中，ω 是 $\mathbb{F}_{2^l}$ 上的本原元素。

证明 设 ω 是初始 $(N,\ \mathbb{F}_2)$ 单位根，分裂域 $\mathrm{Spl}(X^N-e)$ 是 $\mathbb{F}_{2^l}$(因为 $\mathrm{ord}_N(2)=l$)，所以 ω 是 $\mathbb{F}_{2^l}$ 的基本元素。取度为 l 的 $M_\omega(X)=(X-\omega)(X-\omega^2)\cdots(X-\omega^{2^{l-1}})$，幂指数 $\omega^0=e$，$\omega,\cdots,\omega^{N-1}$ 构成 $\mathbb{F}_2^l$，非零元素的列表和 $l\times N$ 矩阵的列

$$\boldsymbol{H}=(\vec{\omega}^0,\vec{\omega},\cdots,\vec{\omega}^{N-1}) \tag{3.3.6}$$

包含长度为 l 的所有非零二进制向量。因此 $[2^l-1,\ 2^l-l-1,\ 3]$ Hamming 码等价于循环码 $(M_\omega(X))$，其中它的零点包含一个 ω 的 $(2^l-1;\ \mathbb{F}_2)$ 本原单位根以及最小多项式的所有其他根。 □

定理 3.3.14 如果 $\gcd(l,\ q-1)=1$，那么 q 进制 $\left(\frac{q^l-1}{q-1},\ \frac{q^l-1}{q-1}-l,\ 3\right)$ Hamming 码等价于循环码。

305

证明 记 $\mathrm{Spl}(X^N-e)=\mathbb{F}_{q^l}$，这里 $l=\mathrm{ord}_N(q)$，$N=\frac{q^l-1}{q-1}$。为了证明 l 的取值，我们看到 $\frac{q^l-1}{N}=q-1$ 时，l 是最小正整数并且 $\frac{q^l-1}{q-1}>q^{l-1}-1$。

因此，$\mathrm{Spl}(X^N-e)=\mathbb{F}_{q^l}$。选择初始值 $\beta\in\mathbb{F}_{q^l}$，那么 $\omega=\beta^{(q^l-1)/N}=\beta^{q-1}$ 是本原 $(N,\ \mathbb{F}_q)$ 单位根。就像之前一样，选择最小多项式 $M_\omega(X)=(X-\omega)(X-\omega^q)\cdots(X-\omega^{q^{l-1}})$，使用零点为 ω 的循环码 $(M_\omega(X))$(特别是 $\omega^0,\cdots,\omega^{q^{l-1}}$)。再次考虑 $l\times N$ 矩阵(3.3.6)。我们想验证一下 $\boldsymbol{H}$ 的任意两个不同的列是线性独立的。如果不是，则存在 $i<j$，ω^i 和 ω^j 都是元素 $\omega^{r-1}\in\mathbb{F}_q$ 的标量倍数。但是，在 $\mathbb{F}_q$ 中，$(\omega^{r-1})^{q-1}=\omega^{(r-1)(q-1)}=e$。因为 ω 是本原第 N 个单位根，当且仅当 $(j-i)(q-1)=0 \bmod N$ 时它是成立的。

$$N=\frac{q^l-1}{q-1}=1+\cdots+q^{l-1}$$

因为对于所有 $\gamma\geqslant 1$，$(q-1)\mid(q^\gamma-1)$，我们有 $q^\gamma=(q-1)v_\gamma+1$，v_γ 为自然数。在 $0\leqslant r\leqslant l-1$ 上，假设它服从

$$N=(q-1)\sum_r v_r+l \tag{3.3.7}$$

因为 $\gcd(q-1,\ l)=1$，有 $\gcd(q-1,\ N)=1$。但是，等式 $(j-i)(q-1)=0 \bmod N$ 是不可能成立的。 □

所以，奇偶校验矩阵为 $\boldsymbol{H}$ 的码长度为 N，秩 $k\geqslant N-l$，距离 $d\geqslant 3$。但是 Hamming 界表明

$$q^k \leqslant q^N \left[\sum_{i=0}^{E}\binom{N}{m}(q-1)^m\right]^{-1}, E=\left\lfloor\frac{d-1}{2}\right\rfloor$$

因为球的体积 $v_{N,q}(E)\geqslant q^t$，这意味着事实上 $k=N-1$，$E=1$，$d=3$。所以，这个码等价于 Hamming 码。

接下来，我们可以看到 BCH 码校正多个错误的更多细节。回顾一下，如果 $\omega_1,\cdots,\omega_s\in\mathbb{F}_{q^N}$ 是 $(N,\mathbb{F}_q)$ 的单位根，那么

$$\mathscr{C}_N=\{f(X)\in\mathbb{F}_q[X]/(X^N-\varepsilon):f(\omega_1)=\cdots=f(\omega_s)=0\}$$

是循环码 $(g(X))$，其中

$$g(X)=\mathrm{lcm}(M_{\omega_1,\mathbb{F}_q}(X),\cdots,M_{\omega_s,\mathbb{F}_q}(X)(X)) \tag{3.3.8}$$

是在 $\mathbb{F}_q$ 上，$\omega_1,\cdots,\omega_s$ 对应的不同最小多项式乘积。特别地，如果 $q=2$，$N=2^t-1$，ω 是在 $\mathbb{F}_{2^t}$ 上的本原元素，那么根为 $\omega,\omega^2,\cdots,\omega^{2^{t-1}}$ 的循环码（和只有单个根 ω 是一样的）与 $(M_\omega(X))$ 是一致的，并且它等价于 Hamming 码。我们可以试着用 $\mathscr{C}$ 零点的其他可能性来看它是否能产生一个有趣的案例。这是发现 BCH 码[21-23]的一种方法。

306

回顾最小多项式 $M_i(X)(=M_{\omega^i,\mathbb{F}_q}(X))$ 的因式分解

$$X^N-1=\mathrm{lcm}(M_i(X);i=0,\cdots,t) \tag{3.3.9}$$

其中，ω 是单位 $(N,\mathbb{F}_q)$ 本原元素，$M_i(X)$ 的根是其幂的，即它符合 $\omega^i,\omega^{iq},\cdots,\omega^{iq^{d(i)-1}}$ 这些形式，$d(=d(i))$ 是大于等于 1 的最小整数，因此 $iq^d=i\bmod N$，集合 $C_i=\{i,iq,\cdots,iq^d\}$ 是第 i 个 $q\bmod N$ 分圆陪集。所以，

$$M_i(X)=\prod_{j\in C_i}(X-\omega^j) \tag{3.3.10}$$

在 3.2 节中，我们要求生成器 $g(X)$ 有 $(\delta-1)$ 个连续根（连续指数），得到最小距离大于等于 δ 的循环码，和下面的定理 3.3.16 比较。

例子 3.3.15 二进制 Hamming 码是设计距离为 $\delta=3$ 的二进制本原狭义 BCH 码。

通过引理 3.2.8，得到距离 $d(\mathscr{C}_{q,N,\varepsilon}^{\mathrm{NCH}})=\delta$。因为 $\mathrm{Spl}(X^N-\varepsilon)\mathbb{F}_q$，其中 $s=\mathrm{ord}_N(q)$，有

$$\deg M_{\omega^{(s)}}(X)\leqslant s \tag{3.3.11}$$

因此，秩 $(\mathscr{C}_{q,N,\delta}^{\mathrm{BCH}})=N-\deg(g(X))\geqslant N-(\delta-1)s$。所以有如下定理。

定理 3.3.16 q 进制 BCH 码的距离 $\mathscr{C}_{q,N,\delta}^{\mathrm{BCH}}\geqslant\delta$，秩 $\deg(g(X))\geqslant N-(\delta-1)\mathrm{ord}_N(q)$。

像以前一样，我们将 $\omega^b,\omega^{b+1},\cdots,\omega^{b+\delta-2}$ 组成奇偶校验矩阵 $\mathscr{C}^{\mathrm{BCH}}$，将它们的幂指数组成来自 $\mathbb{F}_q^s$ 的向量，其中 $s=\mathrm{ord}_N(q)$。设置

$$\widetilde{\boldsymbol{H}}^{\mathrm{T}}=\begin{bmatrix}\vec{e} & \vec{e} & \cdots & \vec{e}\\ \vec{\omega}^{b} & \vec{\omega}^{b+} & \cdots & \vec{\omega}^{b+\delta-2}\\ \vec{\omega}^{2b} & \vec{\omega}^{2(b+1)} & \cdots & \vec{\omega}^{2(b+\delta-2)}\\ \vec{\omega}^{(N-1)b} & \vec{\omega}^{(N-1)(b+1)} & \cdots & \vec{\omega}^{(N-1)(b+\delta-2)}\end{bmatrix} \tag{3.3.12}$$

可以通过移除冗余行来获得合适的奇偶校验矩阵 $\boldsymbol{H}$。

二进制 BCH 码是最容易处理的。定义 $C_i=\{i,2i,\cdots,i2^{d-1}\}$ 为第 i 个分圆陪集 $(d(=d(i)))$，它是最小非零整数，因此 $i\cdot 2^d=i\bmod N$，并且当且仅当 $2u\bmod N\in C_i$，$u\in C_i$ 成立。所以，$M_i(X)=M_{2i}(X)$，对于所有的 $i\geqslant 1$，多项式为

307

$$g_{2i-1}(X)=g_{2i}(X)=\mathrm{lcm}(M_1(X),M_3(X),\cdots,M_{2i}(X))$$

我们立刻推断出 $\mathscr{C}_{q,N,\delta-1}^{\mathrm{BCH}}=\mathscr{C}_{q,N,\delta}^{\mathrm{BCH}}$。所以，通过观察距离 $\delta=2E+1$ 为奇数的狭义 BCH

码，我们可以使定理 3.3.16 得到改善。

定理 3.3.17 距离为$\mathscr{C}^{\mathrm{BCH}}_{2,N,2E+1}$的二进制 BCH 码的秩$\geqslant N-E\,\mathrm{ord}_N(2)$。

准确求出 BCH 码的最小距离的问题仅仅只有部分得到解决(尽管在文献中存在许多结论)。我们在没有证明的情况下提出以下定理。

定理 3.3.18 二进制本原狭义 BCH 码的最小距离是奇数。

在一些特殊的例子中，之前得出的结果可能会变得更好。

举例 3.3.19 证明$\log_2(N+1)>1+\log_2(E+1)!$ 意味着

$$(N+1)^E<\sum_{0\leqslant i\leqslant E+1}\binom{N}{i} \tag{3.3.13}$$

解答 对于$i\leqslant E+1$，我们可以知道$i!\leqslant E+1<(N+1)/2$。因此，式(3.3.13)会随着

$$(N+1)^{E+1}\leqslant 2\sum_{1\leqslant i\leqslant E+1}N(N-1)\cdots(N-i+1)=S(E) \tag{3.3.14}$$

发生。对于$E=0$，不等式(3.3.14)成立，并且这个等式通过归纳法可以得到验证。将式(3.3.14)的 RHS 记成$S(E+1)=S(E)+N(N-1)\cdots(N-E)$。接下来通过归纳法假设$S(E)>(N+1)^{E+1}$，并证明

$$N(N+1)^{E+1}\leqslant 2N(N-1)\cdots(N-E)(N-E-1),$$
$$\text{对于 } N+1>2(E+2)! \tag{3.3.15}$$

取多项式$(y+1)^{E+1}-2(y-1)\cdots(y-E)(y-E-1)$，并且将度为$E+1$和$E$的单项式组合起来。很明显，它们否定了$y>2(E+2)!$。继续这个步骤，总结可得式(3.3.13)成立。

定理 3.3.20 设$N=2^s-1$。如果$2^{sE}<\sum_{0\leqslant i\leqslant E+1}\binom{N}{i}$，那么本原二进制狭义 BCH 码$\mathscr{C}^{\mathrm{BCH}}_{2,2^s-1,2E+1}$距离为$2E+1$。

证明 通过定理 3.3.18，我们可以知道距离是奇数，所以，$d(\mathscr{C}^{\mathrm{BCH}}_{2,2^s-1,2E-1})\neq 2E+2$。假设距离$\geqslant 2E+3$，我们可以看到$\mathscr{C}^{\mathrm{BCH}}_{2,2^s-1,2E+1}\geqslant N-sE$，Hamming 边界

$$2^{N-sE}\sum_{0\leqslant i\leqslant E+1}\binom{N}{i}\leqslant 2^N,\quad 即 2^{sE}\geqslant\sum_{0\leqslant i\leqslant E+1}\binom{N}{i}$$

它们产生的矛盾意味着$\mathscr{C}^{\mathrm{BCH}}_{2,2^s-1,2E+1}=2E+1$。 □ 338

推论 3.3.21 如果$N=2^s-1$，$s>1+\log_2(E+1)!$，那么$d(\mathscr{C}^{\mathrm{BCH}}_{2,2^s-1,2E+1})=2E+1$。特别地，若让$N=31$，$s=5$，我们能轻易验证得到，当$E=1$，2 或者 3 时，

$$2^{5E}<\sum_{0\leqslant i\leqslant E+1}\binom{31}{i}$$

该式可以证明$\mathscr{C}^{\mathrm{BCH}}_{2,31,\delta}$的真实距离等于$\delta$，其中$\delta=3$，5，7。

证明 $s>1+\log_2(E-1)!$ 意味着$2^{sE}<\sum_{1\leqslant i\leqslant E+1}\binom{N}{i}$。 □

定理 3.3.22 如果$\delta|N$，设计距离为δ的本原二进制狭义 BCH 码的最小距离等于δ。

证明 设$N=\delta m$，那么

$$X^N-1=X^{\delta m}-1=(X^m-1)(1+X^m+\cdots+X^{(\delta-1)m})$$

因为当$j=1,\cdots,\delta-1$，$\omega^{jm}\neq 1$时，$\omega^m,\omega^{2m},\cdots,\omega^{(\delta-1)m}$中没有一个是$X^m-1$的根。所以，它们必定是$1+X^m+\cdots+X^{(\delta-1)m}$的根。这个多项式给定一个权重为$\delta$的码字。所以，$\delta$是距离。

更多关于 BCH 码的最小距离的结论在定理 3.3.23 和 3.3.25 中被提到。完整的证明超出了这本书的范围，因此不再详述。

定理 3.3.23 令 $N=q^t-1$。设计距离为 q^k-1 的本原 q 进制狭义 BCH 码 $\mathscr{X}^{\mathrm{BCH}}_{q,q^t-1,q^k-1,1}$ 的最小距离是 q^k-1。

定理 3.3.24 设计距离为 δ 的本原 q 进制狭义 BCH 码 $\mathscr{X}^{\mathrm{BCH}}=\mathscr{X}^{\mathrm{BCH}}_{q,q^t-1,\delta,1}$ 的最小距离最大是 $q\delta-1$。

证明 选取 k 为大于等于 1 的正整数，其中 $q^{k-1}\leqslant\delta\leqslant q^k-1$。令 $\delta'=q^k-1$，另外考虑到 $\mathscr{X}'(=\mathscr{X}^{\mathrm{BCH}}_{q,q^t-1,\delta',1})$，它是有着相同的长度 $N=q^t-1$，设计距离为 δ 的本原 q 进制狭义 BCH 码。$\mathscr{X}$ 的生成器的根都在这些 $\mathscr{X}'$ 中，所以 $\mathscr{X}'\subset\mathscr{X}$。但是根据定理 3.3.22 得知，$d(\mathscr{X}')=\delta'$，其中它小于等于 $\delta q-1$。 □

接下来的结果表明，BCH 码并不能"很好地逼近"。然而，对于小 N(几千或者更少)，BCH 是已知最好的编码之一。

309

定理 3.3.25 长度为 N 的 q 进制本原 BCH 码 $\mathscr{X}^{\mathrm{BCH}}_N$ 不存在无穷序列，因此 $d(X_N)/N$ 和秩$(\mathscr{X}_N)/N$ 的下界都大于 0。

BCH 码能够通过使用所谓的 Berlekamp-Massey 算法解码。首先，考虑一个长度为 $N=2^t-1$，设计距离为 5 的二进制本原狭义 BCH 码 $\mathscr{X}^{\mathrm{BCH}}(=\mathscr{X}^{\mathrm{BCH}}_{2,N,5})$。当 $F=2$ 以及 $s\geqslant4$ 时，不等式 $2^{sF}<\sum_{i=0,\cdots,\lfloor sF/2\rfloor}\binom{N}{i}$ 成立，根据定理 3.3.20 可知，距离 $d(\mathscr{X}^{\mathrm{BCH}})$ 等于 5。因此，BCH 码能纠正两种错误。而且，根据定理 3.3.17 可知，$\mathscr{X}^{\mathrm{BCH}}$ 的秩 $\geqslant N-2s$。[对于 $s=4$，秩等于 $N-2s=15-8=7$。]所以，$\mathscr{X}^{\mathrm{BCH}}$ 是$[2^t-1,\geqslant2^t-1-2s,5]$。

定义零点为 ω，ω^2，ω^3，ω^4，其中 ω 是在 $\mathbb{F}_2$ 上的本原第 N 个单位根(这也是 $\mathbb{F}_{2^t}$ 上的一个原始元素 ω)。我们知道 ω 和 ω^3 足以定义为零点：$\mathscr{X}^{\mathrm{BCH}}=\{c(X)\in\mathbb{F}_2[X]/(X^N):c(\omega)=c(\omega^3)=0\}$，所以，在(3.3.12)中的校验矩阵 $\widetilde{\boldsymbol{H}}$ 能从下式中获得

$$\widetilde{\boldsymbol{H}}^{\mathrm{T}}=\begin{bmatrix}\vec{e} & \vec{\omega} & \vec{\omega}^2 & \cdots & \vec{\omega}^{N-1}\\ \vec{e} & \vec{\omega}^3 & \vec{\omega}^6 & \cdots & \vec{\omega}^{3(N-1)}\end{bmatrix}\tag{3.3.16}$$

比较二进制$[2^t-1,2^t-1-l]$Hamming 码 $\mathscr{X}^{(H)}$ 的情况是很有意义的。在 $\mathscr{X}^{\mathrm{BCH}}$ 码的方案中，假设码 $c(X)\in\mathscr{X}$ 被发送出去，接收的字 $r(X)$ 出现的错误会少于 2 个。若 $r(X)=c(X)+e(X)$，误差多项式 $e(X)$ 的权重 $\leqslant2$。这里我们可能会考虑三种可能：$e(X)=0$，$e(X)=X^i$ 或者 $e(X)=X^i+X^j$，$0\leqslant i\neq j\leqslant N-1$。如果 $r(\omega)=r_1$，$r(\omega^3)=r_3$，那么 $e(\omega)=r_1$ 以及 $e(\omega^3)=r_3$。在没有误差($e(X)=0$)的方案里，$r_1=r_3=0$，反之亦然。在只有一个误差的方案中($e(X)=X^i$)，

$$r_3=e(\omega^3)=\omega^{3i}=(\omega^i)^3=(e(\omega))^3=r_1^3\neq0$$

相反，如果 $r_3=r_1^3\neq0$，那么 $e(\omega^3)=e(\omega)^3$，如果 $e(X)=X^i+X^j$，$i\neq j$，那么

310

$$\omega^{3i}+\omega^{3j}=(\omega^i+\omega^j)^3=\omega^{3i}+\omega^{2i}\omega^j+\omega^i\omega^{2j}+\omega^{3j}$$

即，$\omega^{2i}\omega^j+\omega^i\omega^{2j}=0$ 或者 $\omega^i+\omega^j=0$，这意味着 $i=j$，同时这就与之前的假设相矛盾了。所以，当且仅当 $r_1=r_3^3\neq0$ 时，只出现一个误差的情况发生，错误的数字是 i，因此 $r_1=\omega^i$。所以，在只有一个误差的方案上，我们确定 $\widetilde{\boldsymbol{H}}$ 的一列，即一对$(\omega^i,\omega^{3i})=(r_1,r_3)$，改变在 $r(X)$ 中 i 位置的数字。这和 Hamming 码的解码步骤完全相似。

在出现两个误差的方案中(其中 $e(X)=X^i+X^j$，$i\neq j$)，本着 Hamming 码的思想，我们试着寻找一对列(ω^i,ω^{3i})和(ω^j,ω^{3j})，使得和$(\omega^i+\omega^j,\omega^{3i}+\omega^{3j})=(r_1,r_3)$，也就是使等式满足

$$r_1=\omega^i+\omega^j,r_3=\omega^{3i}+\omega^{3j}$$

接着寻找 i 和 j，因此 $y_1=\omega^i$，$y_2=\omega^j$(y_1,y_2 被称为错误定位器)。如果这样的 i 和 j(或者

错误定位器 y_1 与 y_2)被找到，我们就可以知道在位置 i 和 j 上发生误差了。

引入根为 y_1^{-1}，y_2^{-1} 的错误发生器 $\sigma(X)$ 会很简便：

$$\begin{aligned}\sigma(X) &= (1-y_1X)(1-y_2X) = 1-(y_1+y_2)X+y_1y_2X^2 \\ &= 1-r_1X+(r_3r_1^{-1}-r_1^2)X^2\end{aligned} \tag{3.3.17}$$

因为 $y_1+y_2=r_1$，我们需证明 $y_1y_2=r_3r_1^{-1}-r_1^2$。事实上，

$$r_3=y_1^3+y_2^3=(y_1+y_2)(y_1^2+y_1y_2+y_2^2)=r_1(r_1^2+y_1y_2)$$

如果 N 不大，$\sigma(X)$ 的根可能通过试验 $\mathbb{F}_2^l$ 的 2^l-1 个所有非零元素而被找到。(二次多项式根的标准公式不能应用在 $\mathbb{F}_2$ 上)。因此，提出了接下来的定理：

定理 3.3.26 令 $N=2^l-1$，考虑一个长度为 N，设计距离为 5 的两误差校正二进制本原狭义 BCH 码 $\mathscr{C}$(相当于 $\mathscr{C}^{(BCH)}$)，它的校验矩阵由下式产生

$$\overline{H}^T=\begin{bmatrix} e & \omega & \omega^2 & \cdots & \omega^{N-1} \\ e & \omega^3 & \omega^6 & \cdots & \omega^{3(N-1)} \end{bmatrix}$$

其中，ω 是 $\mathbb{F}_{2^l}$ 的基本元素。(码的长 $\geqslant N-2l$，对于 $l\geqslant 4$，距离等于 5，即 $\mathscr{C}$ 是 $[2^l-1,\geqslant 2^l-1-2l,5]$。它校正两个错误。)假设在接收到字 $r(X)$ 时，最多发生两个错误。令 $r(\omega)=r_1$，$r(\omega^3)=r_3$。那么：

(a) 如果 $r_1=0$，那么 $r_3=0$，没有误差发生。

(b) 如果 $r_1=r_1^3\neq 0$，那么在 $r_1=\omega^i$ 的位置 i 处，有单一误差发生。

(c) 如果 $r_1\neq 0$，$r_3\neq r_1^3$，那么两个误差发生：误差定位多项式 $\sigma(X)=1-r_1X+(r_3r_1^{-1}-r_1^2)X^2$ 有两个不同的根 ω^{N-i-1}，ω^{N-j-1}，在位置 i 和 j 上发生误差。

对于有着通用设计距离为 $\delta(\delta=2t+1$ 是奇数)的二进制 BCH 码，我们顺着相同的思路：对于接收到的 $r(X)=c(X)+e(X)$，计算

$$r_1=e(\omega),r_3=e(\omega^3),\cdots r_{\delta-2}=e(\omega^{\delta-2})$$

假设在位置 $i_1,\cdots,i_t$ 发生误差。那么

$$e(X)=\sum_{1\leqslant j\leqslant t}X^{i_j}$$

311

就像以前一样，考虑到系统中

$$\sum_{1\leqslant j\leqslant t}\omega^{i_j}=r_1,\sum_{1\leqslant j\leqslant t}\omega^{3i_j}=r_3,\cdots,\sum_{1\leqslant j\leqslant t}\omega^{(\delta-2)i_j}=r_{\delta-2}$$

使用误差定位器 $y_j=\omega^{i_j}$：

$$\sum_{1\leqslant j\leqslant t}y_j=r_1,\sum_{1\leqslant j\leqslant t}y_j^3=r_3,\cdots,\sum_{1\leqslant j\leqslant t}y_j^{\delta-2}=r_{\delta-2}$$

误差定位多项式

$$\delta(X)=\prod_{1\leqslant j\leqslant t}(1-y_jX)$$

的根是 y_j^{-1}。$\sigma(X)=\sum_{0\leqslant i\leqslant t}\sigma_iX^i$ 中的系数 σ_i 能从下面的等式中得到

$$\begin{bmatrix} 1 & 0 & 0 & 0 & 0 & \cdots & 0 \\ r_1 & r_0 & 1 & 0 & 0 & \cdots & 0 \\ r_3 & r_2 & r_1 & r_0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ r_{\delta-4} & r_{\delta-5} & \vdots & \vdots & \vdots & \ddots & r_{t-2} \\ r_{\delta-2} & r_{\delta-3} & \cdots & \cdots & \cdots & \cdots & r_{t-1} \end{bmatrix}\begin{bmatrix} \sigma_0 \\ \sigma_1 \\ \sigma_2 \\ \vdots \\ \sigma_{\delta-3} \\ \sigma_{\delta-1} \end{bmatrix}=\begin{bmatrix} r_1 \\ r_3 \\ r_5 \\ \vdots \\ r_{\delta-3} \\ r_{\delta-1} \end{bmatrix}$$

这要求在计算 k 时，r_k 为奇数，这是因为

$$r_{2i}=\varepsilon(\omega^{2i})=\varepsilon(\omega^{i})^2=r_i^2$$

一旦 σ_i 被求出，根 y_i^{-1} 能通过实换和误差求出。

例子 3.3.27 考虑$\mathscr{X}^{BCH}_{1,16,9,3}$，$\omega$ 是$\mathbb{F}_{15}^*$的基本元素。我们都知道，初始多项式是 $M_1(X)=X^4+X+1$ 以及 $M_3(X)=X^4+X^3+X^2+X+1$。因此，码的生成式为

[312]

$$g(X)=M_1(X)M_3(X)=X^8+X^7+X^6+X^4+1$$

我们在第4个和第12个位置，通过采取 $a(X)=X^{12}+X^6+X^2+X^0+1$ 的方式，在码 $c=100010111100000000$ 上产生两个错误。那么

$$r_1=a(\omega)=\omega^{12}+\omega^6+\omega^2+\omega^0+1=\omega^7,$$

$$r_3=a(\omega^3)=\omega^{36}+\omega^{18}+\omega^{6}+\omega^{0}+1=\omega^{0}+\omega^{3}+1=\omega^4$$

既然 $r_3\neq r_1^3$，考虑到位置多项式

$$\sigma(X)=1+\omega^7X+(\omega^{12}+\omega^{12})X^2$$

那么我们可以直接得出 $l(X)$的根是 ω^4 和 ω^{12}。因此，我们可以发现在第4个和第12个位置的误差。

3.4 MacWilliams 标识和线性规划界

线性码的 MacWilliams 标识可以处理所谓的加权计算多项式 $W_{\mathscr{X}}(z)$和 $W_{\mathscr{X}^\perp}(z)$，这里$\mathscr{X}$和$\mathscr{X}^\perp$是给定长度为 N 的一对对偶码，多项式 $W_{\mathscr{X}}(z)$和 $W_{\mathscr{X}^\perp}(z)$被定义为

$$W_{\mathscr{X}}(z)=\sum_{0\le k\le N}A_kz^k \quad 和 \quad W_{\mathscr{X}^\perp}(z)=\sum_{0\le k\le N}A_k^\perp z^k \tag{3.4.1}$$

其中，$A_k(=A_k(\mathscr{X}))$等于在$\mathscr{X}$中权重为 k 的码字数量，$A_k^\perp(=A_k^\perp(\mathscr{X}^\perp))$ 是在$\mathscr{X}$中的码字数量。q 进制码的特性为

$$W_{\mathscr{X}^\perp}(z)=\frac{1}{\#\mathscr{X}}[1+(q-1)z]^NW_{\mathscr{X}}\left(\frac{1-z}{1+(q-1)z}\right),z\in\mathbb{C} \tag{3.4.2}$$

在二进制方案($q=2$)中，得到一个特殊简洁的公式：

$$W_{\mathscr{X}^\perp}(z)=\frac{1}{\#\mathscr{X}}(1+z)^NW_{\mathscr{X}}\left(\frac{1-z}{1+z}\right) \tag{3.4.3}$$

抽象 MacWilliams 标识的简短推导是代数形式的。因为只有线性码的详细说明会在之后用到，所以在它第一次出现时，推导可能会被跳过。

定义 3.4.1 令$(G,+)$组合起来，复杂数字 $S=\{z\in\mathscr{L}:|z|=1\}$的乘法组中的同态 χ：G 被称为G 的(一维)特性。由于 χ 是同态

$$\chi(g_1+g_2)=\chi(g_1)\chi(g_2),\quad \chi(0)=1 \tag{3.4.4}$$

[313] 如果 $\chi(\cdot)=1$，我们认为 χ 是平凡的(或者是主要的)。

一般来说，域 F(不一定是有限的)上群 G 的一个线性表示 D 被定义为，从 G 到域 F 上的有限维空间 V 的可逆线性映射群组 $GL(V)$的同态。

$$D:G\to GL(V):g\to D(g) \tag{3.4.5}$$

向量空间 V 被称为表示空间，它的维度 $\dim(V)$被称为表示维度。

用 D 表示群组 G，那么映射

$$\mathscr{X}^D:G\to F:g\to\sum d_{ii}(g)=\mathrm{trace}(D(g)) \tag{3.4.6}$$

这里取$\mathscr{X}^D(g)$对应 $g\in G$，$D(g)=(d_{ij}(g))$的迹，被称为 D 的特征。复杂数值在域 $\mathbb{C}$ 上的表示和特征被称为普通的。基于域 F 上时它是有限的，它们被称为模拟的。

在我们的方案 $G=\mathbb{F}_q$ 中，使用加法群运算，固定本原第 q 个单位根 $\omega=e^{2\pi i/q}\in\mathbb{S}^1$，对于任意一个 $j\in\mathbb{F}_q$，定义群组 $\mathbb{F}_q$ 的一个一维表示如下：

$$\chi^{(j)}:\mathbb{F}_q\to\mathbb{S}^1:u\mapsto\omega^{ju}$$

对于 $j\neq0$，字符 χ^j 是非平凡的。事实上，$\mathbb{F}_q$ 的所有字符都能用这种方式得到，但是我们省略了这种定理的证明。

接下来，我们定义了群 $G'=\mathbb{F}_q^N$ 的一个特征。固定一个非平凡一维普通字符 $\chi:\mathbb{F}_q\to\mathbb{S}^1$ 以及非零元素 $v=\mathbb{F}_q^N$，定义加法群 $G'=\mathbb{F}_q^N$ 的一个特征如下：

$$\chi_{(v)}:\mathbb{F}_q^N\to\mathbb{S}^1:u\mapsto\chi(v\cdot u) \tag{3.4.7}$$

其中 $v\cdot u$，就像之前提到的一样，是点乘的意思。

引理 3.4.2 令 χ 表示有限群 G 的一个非平凡字符(即 $\chi\neq1$)，那么

$$\sum_{g\in G}\chi(g)=0 \tag{3.4.8}$$

如果 χ 是平凡的，那么 $\sum_{g\in G}\chi(g)=\#G$。

证明 因为 χ 是非平凡的，存在一个元素 $h\in G$ 使 $\chi(h)=1$。从

$$\chi(h)\sum_{g\in G}\chi(g)=\sum_{g\in G}\chi(hg)=\sum_{g\in G}\chi(g)$$

中，我们能够得到 $(\chi(h)-1)\sum_{g\in G}\chi(g)=0$。因此，$\sum_{g\in G}\chi(g)=0$。 □ 314

在方案 $G=\mathbb{F}_q^N$ 中，$\sum_{x\in\mathbb{F}_q^N}\chi(x)=q^N$ 是可忽略不计的。

定义 3.4.3 在 $\mathbb{F}_q^N$ 上，函数 f 的离散 Fourier 变换(简化为 DFT)被定义为

$$\hat{f}=\sum_{v\in\mathbb{F}_q^N}f(v)\chi(v) \tag{3.4.9}$$

有时候，码 $\mathscr{C}$ 的权重计算多项式被定义为两个正式变量 x，y 的函数：

$$W_{\mathscr{C}}(x,y)=\sum_{v\in\mathscr{C}}x^{w(v)}y^{N-w(v)} \tag{3.4.10}$$

(如果设定 $x=z$，$y=1$，那么(3.4.10)和(3.4.1)同时发生)。所以，我们将 DFT 应用在函数上(这时认为 x，$y\in\mathbb{S}^1$ 也是对的)

$$g:\mathbb{F}_q^N\to\mathscr{C}[x,y]:v\mapsto x^{w(v)}y^{N-w(v)} \tag{3.4.11}$$

引理 3.4.4 (抽象 MacWilliams 标识)对于 $v\in\mathbb{F}_q^N$，令

$$g:\mathbb{F}_q^N\to\mathscr{C}[x,y]:v\mapsto x^{w(v)}y^{N-w(v)} \tag{3.4.12}$$

那么

$$\hat{g}(u)=(y-x)^{w(u)}(y+(q-1)x)^{N-w(u)} \tag{3.4.13}$$

证明 令 χ 代表加法群 $G=\mathbb{F}_q$ 的一个非平凡字符。假设 $u\in\mathbb{F}_q$，如果 $u=0$，设 $|u|=0$，否则设 $|u|=1$。随后对于所有的 $u\in\mathbb{F}_q^N$，计算

$$\begin{aligned}\hat{g}(u)&=\sum_{v\in\mathbb{F}_q^N}\chi((v,u))g(v)\\&=\sum_{v\in\mathbb{F}_q^N}\chi((v,u))x^{w(v)}y^{N-w(v)}\\&=\sum_{v_0\in\mathbb{F}_q}\cdots\sum_{v_{N-1}\in\mathbb{F}_q}\chi\Big(\sum_{i=0}^{N-1}v_iu_i\Big)x^{|v_0|+\cdots+|v_{N-1}|}y^{(1-|v_0|)+\cdots+(1-|v_{N-1}|)}\\&=\sum_{v_0\in\mathbb{F}_q}\cdots\sum_{v_{N-1}\in\mathbb{F}_q}\prod_{i=0}^{N-1}\chi(v_iu_i)x^{|v_i|}y^{1-|v_i|}\end{aligned}$$

$$-\prod_{i=0}^{N-1}\sum_{g\in G}\chi(gu_i)x^{|g|}y^{1-|g|}$$

如果 $u_i=0$，$\chi(gu_i)=\chi(0)=1$，所以

[315]

$$\sum_{g\in G}x^{|g|}y^{1-|g|}=y+(q-1)x$$

如果 $u_i\neq 0$，那么

$$\sum_{g\in G}\chi(gu_i)x^{|g|}y^{1-|g|}=y+\sum_{g\in G\setminus 0}\chi(gu_i)x=y-\chi(0)x=y-x$$

□

引理 3.4.5 （线性码的 MacWilliams 标识）如果 $\mathscr{X}$ 是 $\mathbb{F}_q$ 上的线性 $[N,k]$ 码，那么

$$\sum_{x\in\mathscr{X}}\hat{f}(x)=q^k\sum_{y\in\mathscr{X}^\perp}f(y) \tag{3.4.14}$$

证明 考虑下面的求和过程

$$\begin{aligned}\sum_{x\in\mathscr{X}}\hat{f}(x)&=\sum_{x\in\mathscr{X}}\sum_{v\in\mathbb{F}_q^N}\chi_{(v)}(x)f(v)\\&=\sum_{v\in\mathbb{F}_q^N}\sum_{x\in\mathscr{X}}\chi(\langle v,x\rangle)f(v)\\&=\sum_{v\in\mathscr{X}^\perp}\sum_{x\in\mathscr{X}}\chi(\langle v,x\rangle)f(v)\\&\quad+\sum_{v\in\mathbb{F}_q^N\setminus\mathscr{X}^\perp}\sum_{x\in\mathscr{X}}\chi(\langle v,x\rangle)f(v)\end{aligned}$$

在第一次求和中，对于所有 $v\in\mathscr{X}^\perp$ 和 $x\in\mathscr{X}^\perp$，我们有 $\chi(\langle v,x\rangle)=\chi(0)=1$。在第二次求和中，我们知道线性形式

$$\mathscr{X}\to\mathbb{F}_q;x\mapsto\langle v,x\rangle$$

因为 $v\in\mathbb{F}_q^N\setminus\mathscr{X}^\perp$，所以这个线性形式是满射的，因此，它的核维度为 $k-1$，即，对于任何一个 $g\in\mathbb{F}_q$，存在 q^{k-1} 个向量 $x\in\mathscr{X}$ 使 $\langle v,x\rangle=g$。这意味着

$$\begin{aligned}\sum_{x\in\mathscr{X}}\hat{f}(x)&=q^k\sum_{y\in\mathscr{X}^\perp}f(y)-q^{k-1}\sum_{v\in\mathbb{F}_q^N\setminus\mathscr{X}^\perp}f(v)\sum_{g\in G}\chi(g)\\&=q^k\sum_{y\in\mathscr{X}^\perp}f(y)\end{aligned}$$

和引理 3.4.2 中的第二个式子被忽略后一样。 □

引理 3.4.6 在 $\mathbb{F}_q$ 上的 $[N,k]$ 码 $\mathscr{X}$ 权重计算与它的对偶权重计算有关，表示如下：

$$W_{\mathscr{X}^\perp}(x,y)=q^{-k}W_{\mathscr{X}}(y-x,y+(q-1)x) \tag{3.4.15}$$

证明 根据引理 3.4.5，令 $g(v)=x^{w(v)}y^{N-w(v)}$

$$\begin{aligned}W_{\mathscr{X}^\perp}(x,y)&=\sum_{v\in\mathscr{X}^\perp}g(v)=q^{-k}\sum_{v\in\mathscr{X}}\hat{g}(v)\\&=q^{-k}W_{\mathscr{X}}(y-x,y+(q-1)x)\end{aligned}$$

[316] 将变量替换成 $x=z$，$y=1$，我们可以得到式(3.4.3)。 □

例子 3.4.7 (i) 对于所有的码 $\mathscr{X}$，$W_{\mathscr{X}}(0)=A_0=1$，$W_{\mathscr{X}}(1)=\#\mathscr{X}$。当 $\mathscr{X}=\mathbb{F}_q^N$，$W_{\mathscr{X}}(z)=[1+z(q-1)]^N$。

(ii) 对于一个二进制重复码 $\mathscr{X}=\{0000,1111\}$，$W_{\mathscr{X}}(x,y)=x^4+y^4$，那么，

$$W_{\mathscr{X}^\perp}(x,y)=\frac{1}{2}((y-x)^4+(y+x)^4)=y^4+6x^2y^2+x^4$$

(iii) 令$\mathscr{X}$为[7, 4]Hamming 码。对偶码$\mathscr{X}^{\perp}$有 8 个码字；除了 **0** 的所有码字的权重是 4。因此 $W_{\mathscr{X}^{\perp}}(x, y)=x^7+7x^4y^3$，根据 MacWilliams 标识，可知

$$W_{\mathscr{X}}=\frac{1}{2^3}W_{\mathscr{X}^{\perp}}(x-y, x+y)=\frac{1}{2^3}((x-y)^7+7(x-y)^4(x+y)^3)$$
$$=x^7+7x^4y^3+7x^3y^4+y^7$$

因此，X 有 7 个权重为 3 和 4 的码字，它和 **0** 还有 **1** 的码字一起构成了[7, 4]Hamming 码的 16 个码字。

获得等式(3.4.1)的另一个方法是使用和群代数以及 Hamming 空间$\mathbb{F}_q^{\times N}$(这是在维度为 N 的领域$\mathbb{F}_q$上的线性空间)字符转换相关的抽象结果。为简便说明，下标 q 和上标(N)将经常被省略。

定义 3.4.8 空间$\mathbb{F}^{\times N}$上的(复数)群代数 $\mathbb{C}\mathbb{F}^{\times N}$被称为复数函数 G：$x\in\mathbb{F}^{\times N}\to G(x)\in\mathbb{C}$ 的线性空间，其中这个复数函数由复数乘方(共轭)和乘法运算组成。因此，函数 $G(x)$有 4 个运算式：加法和标量(复数)乘法都是标准(逐点)方法，其中$(G+G')(x)=G(x)+G'(x)$，$(aG)(x)=aG(x)$，G，$G'\in\mathbb{C}\mathbb{F}^{\times N}$，$a\in\mathbb{C}$，$x\in\mathbb{F}^{\times N}$。乘方仅仅是(逐点式)复数共轭：$G^*(x)=G(x)^*$；它是个幂等运算，满足 $G^{**}=G$。不管怎样，乘法(用 * 表示)表示卷积：

$$(G*G')(x)=\sum_{y\in\mathbb{F}^{\times N}}G(y)G'(x-y), x\in\mathbb{F}^{\times N} \tag{3.4.16}$$

这就将使 $\mathbb{C}\mathbb{F}^{\times N}$变成了一个交换环，同时它也是维度 $\dim\mathbb{C}\mathbb{F}^{\times N}=q^N$ 的(复数)线性空间。(代数，即交换环和线性空间的集合。)$\mathbb{C}\mathbb{F}^{\times N}$中的特性基础由 Dirac(或者说 Kronecker)的 δ 函数δ^x构成，其中 $\delta^x(x)=1(x=y)$，x，$y\in\mathscr{H}$。 317

如果$\mathscr{X}\subseteq\mathbb{F}^{\times N}$是线性码，我们就假设 $G_{\mathscr{X}}(x)=1(x\in\mathscr{X})$。

我们需要证明乘法运算规则(3.4.16)。如果我们重新在对称式 $\sum_{y,y'\in\mathbb{F}^{\times N},(y+y'=x}G(y)G(y')$ 中写下 RHS，(这使得 * 乘法运算很突出)，接着还会出现多项式乘法的类比。事实上，如果 $A(t)=a_0+a_1t+\cdots+a_{l-1}t^{l-1}$，$A'(t)=a_0'+a_1't+\cdots+a_{l'-1}'t^{l'-1}$ 这两个多项式的系数字符串分别是$(a_0, \cdots, a_{l-1})$和$(a_0', \cdots, a_{l'-1}')$，那么乘积 $B(t)=A(t)A'(t)$的字符串系数$(b_0, \cdots, b_{l-1+l'-1})$满足 $b_k=\sum_{m,m'\geq 0,m+m'=k}a_ma_{m'}'$。

从这个可以看出，代(3.4.16)有一些多项式类型的乘法运算。当然，度$\leq n-1$ 的多项式构成纬度为 n(复数)线性空间。但是，它们并不能构成一个群(甚至是半群)。为了构成一个群，我们应该加入倒数 $1/t$，$1/t^2$ 等，或者考虑加入无限序列以及约束 $t^n=1$(即将 t 当作循环群中的一个元素，而不是“自由”变量)。相似的结构可用于由几个变量构成的多项式，但是，在这里，变量之间存在很多约束。

回到我们的群代数 $\mathbb{C}\mathscr{H}$，我们将采取以下步骤：

(i) 生成一个 Hamming 群$\mathscr{H}$的“乘法模型版本”。即，取由 $x\in\mathscr{H}$标志的“正式”变量 $t^{(x)}$集合，对所有的 x，$x'\in\mathbb{C}\mathscr{H}$，假设 $t^{(x)}t^{(x')}=t^{(x+x')}$。

(ii) 接着考虑由所有(复数)线性组合 $G=\sum_{x\in\mathscr{H}}\gamma_xt^{(x)}$ 构成的集合 $\mathbb{T}\mathscr{H}$。引用(ii1)加法运算 $G+G'=\sum_{x\in\mathscr{H}}(\gamma_x+\gamma_x')t^{(x)}$ (ii2) 标量乘法 $aG=\sum_{x\in\mathscr{H}}(a\gamma_x)t^{(x)}$，$G$，$G'\in\mathbb{T}\mathscr{H}$，$a\in\mathbb{C}$。再次获得维度为 q^N 的线性空间，它是由“基本”组合 $t^{(x)}$，$x\in\mathscr{H}$构成的基。很明显，$\mathbb{T}\mathscr{H}$和 $\mathbb{C}\mathscr{H}$像线性空间一样是同构的，其中 $G\leftrightarrow g$。

(iii) 现在，去除 $t^{(x)}$中的括号(但依然遵守 $t^xt^{x'}=t^{x+x'}$)，考虑到 $g(t)$是个由多个遵守

上述规则的变量 t 构成的函数(事实上是个"多项式"),像 $g(t)$ 一样写下 $\sum_{x\in\mathscr{H}}\gamma_x t^x$。最后,参考在 $\mathbb{T}\mathscr{H}$ 中的多项式乘法 $g(t)g'(t)$。因而 $\mathbb{T}\mathscr{H}$ 和 $\mathbb{C}\mathscr{H}$ 不仅是个同构的线性空间也是个同构环,即代数形式。

上述步骤非常好,它不仅可以用于 $\mathscr{H}_N$,而是能够用于任何群。它的能力在 MacWilliams 等式的推分中得到了证明。

所以,我们将 $\mathbb{C}\mathscr{H}$ 看成遵守求幂法则 $t^{x+y}=t^x t^y$ 的自变量 t 的函数集

318

$$g(t)=\sum_{x\in\mathscr{H}_n}\gamma_x t^x \tag{3.4.17}$$

其满足多项式的加法规则以及乘法规则。

和式(3.4.17)一样,对于一个线性码 $\mathscr{X}\subset\mathscr{H}_n$,我们令

$$g_{\mathscr{X}}(t)=\sum_{x\in\mathscr{X}}t^x \tag{3.4.18}$$

其中,$g_{\mathscr{X}}(t)$ 通常被称作 $\mathscr{X}$ 的生成函数。

定义 3.4.3 证实,一个对任何非重要字符 $\mathscr{X}$:$\mathscr{F}\to\mathscr{S}$ 的简单归纳是对的。需要注意 Fourier 变换(以及其他常用转换方法的类型(也就是在群论中的 Hadamard 的变换))的相似性。

定义 3.4.9 定义群代数 $\mathbb{C}\mathscr{H}_n$ 中的字符变换 $g\mapsto\hat{g}$ 为

$$\hat{g}(t)=\sum_{x\in\mathscr{H}_n}X_x(g)t^x \tag{3.4.19a}$$

其中 $g\sim(\gamma_x,\ x\in\mathscr{H}_n)$

$$X_x g=\sum_{y\in\mathscr{H}_n}\gamma_y\chi(x\cdot y) \tag{3.4.19b}$$

这里,$x\cdot y$ 是在 $\mathscr{H}_n$ 中的点乘 $\sum_{1\le j\le n}x_j y_j$ 运算。

现在定义群代数元素 $g\in\mathbb{C}\mathscr{H}$ 的权重枚举为变量 s(它可能被认为是复数变量)的多项式 $W_g(s)$:

$$W_g(s)=\sum_{x\in\mathscr{H}}\gamma_x s^{w(x)}=\sum_{k=0}^{n}\Big[\sum_{x:w(x)=k}\gamma_x\Big]s^k=\sum_{0\le k\le n}A_k s^k,s\in\mathbb{C} \tag{3.4.20}$$

其中

$$A_k=\sum_{x\in\mathscr{H}_n:w(x)=k}\gamma_x \tag{3.4.21}$$

对于线性码 $\mathscr{X}$,生成函数为 $g_{\mathscr{X}}(t)$(参见式(3.4.18)),A_k 为权重为 k 的码字数量:

$$A_k=\#\{x\in\mathscr{X}:w(x)=k\} \tag{3.4.22}$$

$g\sim(\gamma_x,\ x\in\mathscr{H})$ 为字符变换 $\hat{g}$ 的权重 $W_{\hat{g}}(s)$,它为

319

$$W_{\hat{g}}(s)=\sum_{x\in\mathscr{H}}X_x(g)s^{w(x)}=\sum_{0\le k\le n}\Big[\sum_{x:w(x)=k}X_x(g)\Big]s^k=\sum_k\hat{A}_k s^k \tag{3.4.23}$$

其中

$$\hat{A}_k=\sum_{x\in\mathscr{H}_n:w(x)=k}X_x(g) \tag{3.4.24}$$

这个"抽象"MacWilliams 标识会在接下来结论中展示。

定理 3.4.10 我们有

$$W_{\hat{g}}(s)=(1+(q-1)s)^n W_g\Big(\frac{1-s}{1+(q-1)s}\Big) \tag{3.4.25}$$

证明 和引理 3.4.4 基本一样。 □

就系数 A_k 和 $\hat{A}_k$ 而言，重新定义式(3.4.25)

$$\sum_{k=1}^{n} \hat{A}_k s^k = \sum_{k=1}^{n} A_k (1-s)^k (1+(q-1)s)^{n-k} \tag{3.4.26}$$

它可以扩展为

$$(1-s)^k (1+(q-1)s)^{n-k} = \sum_{i=0}^{n} K_i(k) s^i \tag{3.4.27}$$

这里，$K_i(k)(=K_i(k, n, q))$是 Kravchuk 多项式：对于所有 i，$k=0, 1, \cdots, n$

$$K_i(k) = \sum_{j=0 \vee (i+k-n)}^{i \wedge k} \binom{k}{j} \binom{n-k}{i-j} (-1)^j (q-1)^{i-j}$$

$$0 \vee (i+k-n) = \max[0, i+k-n], i \wedge k = \min[i, k] \tag{3.4.28}$$

那么

$$\sum_{0 \le k \le n} \hat{A}_k s^k = \sum_{0 \le k \le n} A_k \sum_{0 \le i \le k} K_i(k) s^i = \sum_{0 \le k \le n} \sum_{0 \le i \le k} A_k K_i(k) s^i$$

$$= \sum_{0 \le k \le n} \sum_{k \le i \le n} A_i K_k(i) s^k$$

即

$$\hat{A}_k = \sum_{k \le i \le n} A_i K_k(i) \tag{3.4.29}$$

□

引理 3.4.11 对于任意(线性)码$\mathscr{X} \subset \mathscr{H}_n$，其生成函数表示为 $g_{\mathscr{X}} \sim 1(x \in \mathscr{X})$，字符转换系数通过

$$X_u(g_{\mathscr{X}}) = \# \mathscr{X} 1(u \in \mathscr{X}^{\perp}) \tag{3.4.30}$$

相关联，并且字符转换表示为

$$\hat{g}_{\mathscr{X}} = \# \mathscr{X} g_{\mathscr{X}^{\perp}} \tag{3.4.31}$$

其中$\mathscr{X}^{\perp}$是对偶码。 320

证明 根据引理 3.4.2 可得

$$X_u(g_{\mathscr{X}}) = X_u\left(\sum_{x \in \mathscr{X}} t^x\right) = \sum_{x \in \mathscr{X}} \chi(y \cdot u) = \# \mathscr{X} 1(u \in \mathscr{X}^{\perp})$$

事实上，当且仅当$u \in \mathscr{X}^{\perp}$时，字符 $y \in \mathscr{X} \to \chi(y \cdot u)$是首要的，从而可得

$$\hat{g}(t) = \sum_{x \in \mathscr{H}} X_u(g_{\mathscr{X}}) t^x = \sum_{x \in \mathscr{H}} \# \mathscr{X} 1(x \in \mathscr{X}^{\perp}) t^x$$

$$= \# \mathscr{X} \sum_{x \in \mathscr{X}^{\perp}} t^x = \# \mathscr{X} g_{\mathscr{X}^{\perp}}(t)$$

因而有

$$W_{g_{\mathscr{X}}}(s) = \# \mathscr{X} W_{g_{\mathscr{X}^{\perp}}}(s) \tag{3.4.32}$$

□

对于线性码我们能得到 MacWillams 等式：

定理 3.4.12 令$\mathscr{X} \subset \mathscr{H}_n$为线性码，$\mathscr{X}^{\perp}$为它的对偶码，则可以得到

$$W_{\mathscr{X}}(s) = \sum_{k=0}^{n} A_k s^k, W_{\mathscr{X}^{\perp}}(s) = \sum_{k=0}^{n} A_k^{\perp} s^k \tag{3.4.33}$$

上面二式分别是$\mathscr{X}$和$\mathscr{X}^{\perp}$的 w 枚举元素，其中 $A_k = \#\{x \in \mathscr{X}: w(x) = k\}$，然后可得下式

$$W_{\mathscr{X}^{\perp}}(s) = \frac{1}{\# \mathscr{X}} (1+(q-1)s)^n W_{\mathscr{X}}\left(\frac{1-s}{1+(q-1)s}\right), s \in \mathbb{C} \tag{3.4.34}$$

上式等价于

$$A_k = \frac{1}{\#\mathscr{X}}\sum_{0\le i\le n} A_i K_k(i) \tag{3.4.35}$$

其中 $K_k(i)$ 是 Kravchuk 多项式(参见式(3.4.28))。

对于二元码，即 $q=2$，式(3.4.34)可以表示为式(3.4.3)的形式，有时权重枚举值被定义为

$$W_{\mathscr{X}^\perp}(s,r) = \sum_k A_k s^k r^{n-k} \tag{3.4.36}$$

则 MacWilliams 标识式(3.4.33)表示为

$$W_{\mathscr{X}^\perp}(s,r) = \frac{1}{\#\mathscr{X}} W_{\mathscr{X}}(r-s, r+(q-1)s) \tag{3.4.37}$$

MacWilliams 等式是非常强有力的结论，特别是当码是自对偶码时，它能够深入解析(线性)码的结构。

321

MacWilliams 等式帮助建立了一个关于线性码的有趣界限，称为线性规划(LP)界限。首先，我们讨论这个等式的一些即时结果。如果$\mathscr{X}\subset\mathscr{H}_{n,q}$是大小为 M 的码，令

$$B_k = \frac{1}{M}\#\{(x,y): x,y\in\mathscr{X}, d(x,y)=k\}, k=0,1,\cdots,N$$

(每对 x，y 值都被计算两次)。数字 B_0，B_1，…，B_N 构成码$\mathscr{X}$的距离分布。表达式

$$B_{\mathscr{X}}(s) = \sum_{0\le k\le N} B_k s^k \tag{3.4.38}$$

被称为$\mathscr{X}$的距离枚举。很明显，线性码的 w-和 d-分布一致。而且，我们有如下引理

引理 3.4.13 *$[N, M]$码$\mathscr{X}$ d-枚举和如下群代数元素的 w-枚举相一致*

$$h_{\mathscr{X}}(s) = \frac{1}{M^2}\zeta_{\mathscr{X}}(s)\zeta_{\mathscr{X}}(s^{-1})) \tag{3.4.39}$$

其中$\mathscr{X}$的生成函数是

$$\zeta_{\mathscr{X}}(s) = \sum_{x\in\mathscr{X}} s^x \tag{3.4.40}$$

证明 利用符号$(s^{-1})^x$，可得

$$h_{\mathscr{X}}(s) = \frac{1}{M}\sum_{x\in\mathscr{X}} s^x \sum_{y\in\mathscr{X}} s^{-y} = \frac{1}{M}\sum_{x,y\in\mathscr{X}} s^{x-y}$$

因此有

$$\begin{aligned} W_{h_{\mathscr{X}}}(s) &= \frac{1}{M}\sum_{0\le k\le N}\sum_{x,y\in\mathscr{X}} \mathbf{1}(w(x-y)=k)s^k = \sum_{0\le k\le N} B_k s^k \\ &= B_{\mathscr{X}}(s) \end{aligned}$$

□

现在通过利用 MacWilliams 标识，对于给定的非平凡字符 χ 以及相应的变换 $\zeta\to\hat{\zeta}$，我们可得如下定理

定理 3.4.14 *对于上文中的 $h_{\mathscr{X}}(s)$ 来说，如果$\hat{h}_{\mathscr{X}}(s)$是字符变换，且 $W_{\hat{\zeta}_{\mathscr{X}}}(s)$是它的 w-枚举，其中*

322

$$W_{\hat{\zeta}_{\mathscr{X}}}(s) = \sum_{0\le k\le N} \hat{B}_k s^k = \sum_{0\le k\le N}\Big[\sum_{w(x)=k}\chi_x(h_{\mathscr{X}})\Big]s^k$$

则有

$$\hat{B}_k = \sum_{0\le i\le N} B_i K_k(i)$$

其中 $K_k(i)$ 是 Kravchuk 多项式。

接下来的说明是直接明了的。

引理 3.4.15 下列等式成立：$\chi_\lambda(\zeta_{\mathscr{X}}(s^{-1}))=\overline{\chi_\lambda(\zeta_{\mathscr{X}}(s))}$，其中横线表示复共轭。

根据定理 3.4.15，我们可得如下等式

$$\chi_\lambda(h_{\mathscr{X}}(t))=\frac{1}{M}\chi_\lambda(\zeta_{\mathscr{X}}(s)\zeta_{\mathscr{X}}(s^{-1}))=\frac{1}{M}\chi_\lambda(\zeta_{\mathscr{X}}(s))\chi_\lambda(\zeta_{\mathscr{X}}(s^{-1}))$$

$$=\frac{1}{M}\chi_\lambda(\zeta_{\mathscr{X}}(s))\overline{\chi_\lambda(\zeta_{\mathscr{X}}(s))}=\frac{1}{M}|\chi_\lambda(\zeta_{\mathscr{X}}(s))|^2$$

所以，

$$\hat{B}_k=\sum_{x\in\mathscr{X},w(x)=k}\chi_\lambda(h_{\mathscr{X}})=\frac{1}{M}\sum_{w(x)=k}|\chi_\lambda(\zeta_{\mathscr{X}})|^2\geqslant 0$$

因此有如下定理：

定理 3.4.16 对于所有的 $[N, M]$ 码 $\mathscr{X}$，且 $k=0, \cdots, N$，可得

$$\sum_{0\leqslant i\leqslant N}B_iK_k(i)\geqslant 0 \tag{3.4.41}$$

现在计算对 $(x, y)\in\mathscr{X}\times\mathscr{X}$ 的数量：

$$\sum_{0\leqslant i\leqslant N}B_i=M^2$$

或者

$$\sum_{0\leqslant i\leqslant N}E_i=M,\quad 其中\ E_i=\frac{1}{M}B_i \tag{3.4.42}$$

（有时 $E_0, E_1, \cdots, E_N$ 被称为 $\mathscr{X}$ 的 d-分布）。根据(3.4.41)—(3.4.42)，可得

$$\sum_{0\leqslant i\leqslant N}E_iK_k(i)\geqslant 0$$

此外根据定义，有 $E_i\geqslant 0$，$0\leqslant i\leqslant N$，$E_0=1$，$E_i=0$，$1\leqslant i<d$。

证明 设 ω 是本原第 q 个单位根，$x\in F_q^N$ 为权重 i 的固定字，则

$$\sum_{y\in F_q^N;w(y)=k}\omega^{(x,y)}=K_k(i) \tag{3.4.43}$$

323

事实上，我们也许会假设 $x=x_1x_2\cdots x_i0\cdots0$，其中坐标 x_i 不等于 0，令 D 为字的集合，它包含给定 k 位置的集合中非零坐标。假设准确的 j 位置 $h_1, \cdots, h_k$ 属于 $[0, i]$，$k-j$ 位置属于 $[i+1, N]$。对此，在 $\binom{i}{j}\binom{N-i}{k-j}$ 中选择一个集合，那么

$$\sum_{y\in D}\omega^{(x,y)}=\sum_{y_{h_1}\in F_q^*}\cdots\sum_{y_{h_k}\in F_q^*}\omega^{x_{h_1}y_{h_1}+\cdots+x_{h_k}y_{h_k}}$$

$$=(q-1)^{k-j}\prod_{i=1}^{j}\sum_{y\in F_q^*}\omega^{x_iy}=(-1)^j(q-1)^{k-j}$$

因此，

$$M\sum_{i=0}^{N}B_iK_k(i)=\sum_{i=0}^{N}\sum_{x,y\in\mathscr{X},d(x,y)=i}\sum_{z\in F_q^N;w(z)=k}\omega^{(x-y,z)}$$

$$=\sum_{z\in F_q^N;w(z)=k}\left|\sum_{x\in\mathscr{X}}\omega^{(x,z)}\right|^2\geqslant 0$$

这就得到在定理 3.4.17 中表述的所谓线性规划(LP)界。 □

定理 3.4.17 (LP 界)下边的不等式成立：

$$M_q^*(N,d) \leqslant \max\Big[\sum_{0\leqslant i\leqslant N} \widetilde{E}_i : \widetilde{E}_i \geqslant 0, \widetilde{E}_0 = 1, \widetilde{E}_i = 0,\quad 1 \leqslant i < d$$

$$\text{和} \sum_{0\leqslant i\leqslant N} \widetilde{E}_i K_k(i) \geqslant 0,\quad 0 \leqslant k \leqslant N\Big] \tag{3.4.44}$$

对于 $q=2$，LP 界将会稍微增加，这将在定理 3.4.19 中说明。首先提出一个辅助结果，它的证明简单明了，将会留作练习。

引理 3.4.18 (a) 如果存在一个二元 $[N, M, d]$ 码，其中 d 是偶数，那么存在一个二元 $[N, M, d]$ 码，其中任意码字的权重 w 是偶数，且所有的码距都是偶数。因而，如果 $q=2$，d 是偶数，我们可以假设对于所有的奇数值 i 有 $E_i=0$。

(b) 对于 $q=2$，有

324

$$K_i(2k) = K_{N-i}(2k)$$

因此，当 d 为偶数时，由于我们假设了 $E_{2i+1}=0$，只有当 $k=0, \cdots, [N/2]$ 时，式(3.4.44)中的约束条件才会被考虑。

(c) 对于所有的 i 来说，$K_0(i)=1$。因此当 $\widetilde{E}_i \geqslant 0$ 时，界 $\sum_{0\leqslant i\leqslant N} \widetilde{E}_i K_0(i) \geqslant 0$ 成立。

根据引理 3.4.18，直接可得以下定理。

定理 3.4.19 (LP 中对 $q=2$ 来说)如果 d 是偶数，则有

$$M_q^*(N,d) \leqslant \max\Big[\sum_{0\leqslant i\leqslant N} \widetilde{E}_i : \widetilde{E}_i \geqslant 0, \widetilde{E}_0 = 1, \widetilde{E}_i = 0 \quad \text{对于 } 1 < i < d,$$

$$\widetilde{E}_i = 0 \quad \text{对于 } i \text{ 是偶数的情况},$$

$$\binom{N}{k} + \sum_{0\leqslant i\leqslant N} \widetilde{E}_i K_k(i) \geqslant 0 \quad \text{对于 } k = 1, \cdots, \Big\lfloor \frac{N}{2} \Big\rfloor\Big] \tag{3.4.45}$$

因为 $M_2^*(N, 2t+1) = M_2^*(N+1, 2t+2)$，所以当 d 为奇数时，定理 3.4.19 同样给出了一个有效的界限。我们将更进一步探讨 MacWilliams 等式。

LP 界限是编码理论中的一个相当通用的工具。比如，Singleton、Hamming、Plotkin 界都可通过 LP 界求得。不过，我们将不详细研究怎样求得的过程。

举例 3.4.20 给定正整数 N 且 $d \leqslant N$，令

$$f(x) = 1 + \sum_{j=1}^{N} f_j K_j(x)$$

为一个多项式，且 $f_j \geqslant 0$，其中 $1 \leqslant j \leqslant N$，$f(i) \leqslant 0$，其中 $d \leqslant i \leqslant N$。证明

$$M_q^*(N,d) \leqslant f(0) \tag{3.4.46}$$

根据式(3.4.46)推导 Singleton 界。

解答 令 $M = M_q^*(N, d)$，且 $\mathscr{X}$ 为 q 进制 $[N, M]$ 码，其中有距离分布 $B_i(\mathscr{X})$，$i=0, \cdots, N$。对于 $d \leqslant i \leqslant N$ 来说，条件 $f(i) \leqslant 0$ 意味着 $\sum_{j=d}^{N} B_j(\mathscr{X}) f(j) \leqslant 0$。当 $k=0$，利用 LP 界限

325

(3.4.45)，可以获得 $K_i(0) \geqslant -\sum_{j=d}^{N} B_j(\mathscr{X}) K_i(j)$。因此

$$f(0) = 1 + \sum_{j=1}^{N} f_j K_j(0) \geqslant 1 - \sum_{k=1}^{N} f_k \sum_{i=d}^{N} B_i(\mathscr{X}) K_k(i)$$

$$= 1 - \sum_{i=d}^{N} B_i(\mathscr{X}) \sum_{k=1}^{N} f_k K_k(i) = 1 - \sum_{i=d}^{N} B_i(\mathscr{X}) (f(i) - 1)$$

$$\geqslant 1+\sum_{i=d}^{N} B_i(\mathscr{X})=M=M_q^*(N,d)$$

为获得 Singleton 界，选择

$$f(x)=q^{N-d+1}\prod_{j=d}^{N}\left(1-\frac{x}{j}\right)$$

然后，根据等式

$$\sum_{i=1}^{j}\binom{N-j}{N-i}K_i(k)=q^j\binom{N-k}{j}$$

其中 $j=d-1$，我们有

$$f_k=\frac{1}{q^N}\sum_{i=1}^{N}f(i)K_i(k)=\frac{1}{q^{d-1}}\sum_{i=1}^{d-1}\binom{N-i}{N-d+1}K_i(k)\Big/\binom{N}{d-1}$$

$$=\binom{N-k}{d-1}\Big/\binom{N}{d-1}\geqslant 0$$

这里，我们使用如下等式

$$\sum_{k=1}^{j}\binom{N-k}{N-j}K_k(x)=q^j\binom{N-x}{j} \tag{3.4.47}$$

显然，当 $d\leqslant i\leqslant N$ 时，$f(i)=0$。因此，$R_q(N,d)\leqslant f(0)=q^{N-d+1}$。利用相似的方法，可以推导出 Hamming 界和 Plotkin 界，参见[97]。 □ 326

举例 3.4.21 利用线性规划界，证明 $M_2^*(13,5)=M_2^*(14,6)\leqslant 64$。将它与 Elias 界进行比较。（提示：$E_6=42$，$E_8=7$，$E_{10}=14$，$E_{12}=E_{14}=0$。你也许需要一个计算机来计算结果。）

解答 线性码的 LP 界是

$$M_2^*(N,d)=\max\sum_{0\leqslant i\leqslant N}E_i$$

$$\text{约束于 } E_i\geqslant 0, E_0=1, E_j=0,\quad 1\leqslant j<d$$

$$E_i=0 \text{ 其中 } j \text{ 是奇数，且}$$

$$\binom{N}{k}+\sum_{\substack{d\leqslant i\leqslant N\\ i\text{ 是偶数}}}E_iK_k(i)\geqslant 0,\quad k=1,\cdots,\left\lfloor\frac{N}{2}\right\rfloor$$

当 $N=14$，$d=6$ 时，约束表示为

$$E_0=1, E_1=E_2=E_3=E_4=E_5=E_7=E_{11}=E_{13}=0,$$

$$E_6,E_8,E_{10},E_{12},E_{14}\geqslant 0,$$

$$14+2E_6-2E_8-6E_{10}-10E_{12}-14E_{14}\geqslant 0,$$

$$91-5E_6-5E_8+11E_{10}+43E_{12}+91E_{14}\geqslant 0,$$

$$364-12E_6+12E_8+4E_{10}-100E_{12}-364E_{14}\geqslant 0,$$

$$1001+9E_6+9E_8-39E_{10}+121E_{12}+1001E_{14}\geqslant 0,$$

$$2002+30E_6-30E_8+38E_{10}-22E_{12}-2002E_{14}\geqslant 0,$$

$$3003-5E_6-5E_8+27E_{10}-165E_{12}+3003E_{14}\geqslant 0,$$

$$3432-40E_6+40E_8-72E_{10}+264E_{12}-3432E_{14}\geqslant 0$$

目标函数 $S=E_6+E_8+E_{10}+E_{12}+E_{14}$ 的最大值为

$$E_6=42, E_8=7, E_{10}=14, E_{12}=E_{14}=0$$

其中 $S=63$，$E_n+S=1+63=64$。所以，LP 界服从

$$M_2^*(13,5)=M_2^*(14,6)\leqslant 64$$

注意到上界是可达的，因为[13，64，5]二进制码的确存在。将 LP 界与 Hamming 界进行对比

$$M_2^*(13,5)\leqslant 2^{13}/(1-13+13\cdot 6)=2^{13}/92=2^{11}/23$$

即

$$M_2^*(13,5)\leqslant 91$$

然后，Singleton 界给出了 $k\leqslant 13-5-1=7$

327

$$M_2^*(13,5)\leqslant 2^7=128$$

同样有趣的是，根据 Elias 界可得

$$M_2^*(13,5)\leqslant \frac{65/2}{s^2-13s+65/2}\frac{2^{13}}{1+13+\cdots+\binom{13}{s}}$$

对全部 $s<13$，$s^2-13s+65/2>0$。

将 s 替换为 $s-2$，则有 $s^2-13s+65/2=4-26+65/2=21/2>0$，且

$$M_2^*(13,5)\leqslant \frac{65}{21}2^{13}/(1+13+13\cdot 6)=2.33277\times 10^4$$

这结果不够好。然后，若 $s=3$ 则有 $s^2-13s+65/2=9-39+65/2=5/2>0$ 且

$$M_2^*(13,5)\leqslant \frac{65}{5}2^{13}/(1+13+13\cdot 6+13\cdot 2\cdot 5)=13\times\frac{2^{13}}{111}\geqslant\frac{13}{66}2^{11}$$

这没有 Hamming 效果好。最终，当看到 $s^2-13\times 4+65/2<0$ 时，步骤结束。 □

3.5 渐近好码

在这一节中，我们简要地讨论一些码族，其中校正误差的数量等于码长乘以一个非零分数。更多详细内容，参见文献[54]、[71]、[131]。

定义 3.5.1 若 k_i/N_i 和 d_i/N_i 的界大于 0，则 $[N_i, k_i, d_i]$ 码的序列是渐近良好的且 $N_i\to\infty$。

定理 3.3.25 表明本原 BCH 码的序列不是渐近良好的(实际上，任何形式的本原 BCH 码的序列都不是渐近良好的)。理论上来说，产生一个渐近的良好的码最好的方法是所谓的 Justensen 码。首先定义一个好的码字，使 $0\neq a\in\mathbb{F}_{2^m}\setminus\mathbb{F}_2$，同时定义集合

$$\mathscr{X}_a=\{(a,am):a\in\mathbb{F}_2^m\} \tag{3.5.1}$$

然后，$\mathscr{X}_a$ 是一个 $[2m, m]$ 线性码，且信息速率为 1/2。因为 $a=ba^{-1}$(在 $\mathbb{F}_{2^m}$ 中分离)，我们可以从任何非零码字 $(a, b)\in\mathscr{X}_a$ 中校正 a。在这里，如果 $a\neq a'$，则 $\mathscr{X}_a\cap\mathscr{X}_{a'}=\{0\}$。

现在，给定 $\lambda=\lambda_n\in(0, 1/2]$，我们需要找到 $a\neq a_n$。因此 $\mathscr{X}_a$ 码的最小权重 $\geqslant 2m\lambda$。

328

由于非零二进制 $(2m)$ 码字可以出现在大多数的 $\mathscr{X}_a$ 中，如果非零 $(2m)$ 二进制码字权重 $<2m\lambda$ 是 $<2^m-1$ 的，那么我们可以找到这样的 a，a 是不同 $\mathscr{X}_a$ 码的数量。也就是说，如果在

$$\sum_{1\leqslant i<2m\lambda-1}\binom{2m}{i}<2^m-1$$

或者更好的条件 $\sum_{1\leqslant i\leqslant 2m}\binom{2m}{i}<2^m-1$ 下，我们可以使用上述定理。现在使用如下引理：

引理 3.5.2 若 $0\leqslant\lambda\leqslant 1/2$，

$$\sum_{0\leqslant k\leqslant[\lambda N]}\binom{N}{k}\leqslant 2^{N\eta(\lambda)} \tag{3.5.2}$$

其中 $\eta(\lambda)$ 是二进制熵。

证明 对于 $\lambda=0$ 以及 $\lambda=1/2$(RHS 等于 2^N)，我们可以看到式(3.5.2)成立(它的左右两边都等于 1)。所以我们假设 $0<\lambda<1/2$。考虑到一个二项分布的随机变量 ξ，满足

$$\mathbb{P}(\xi=k)=\binom{N}{k}(1/2)^N, 0\leqslant k\leqslant N$$

给定的 $t\in\mathbb{R}_+$，使用如下 Chebyshev 不等式

$$\sum_{0\leqslant k\leqslant\lambda N}\binom{N}{k}\left(\frac{1}{2}\right)^N=\mathbb{P}(\xi\leqslant\lambda N)=\mathbb{P}(\exp(-t\xi)\geqslant e^{-t\lambda N})$$

$$\leqslant e^{-t\lambda N}\mathbb{E}e^{-t\xi}=e^{-t\lambda N}\left(\frac{1}{2}+\frac{1}{2}e^{-t}\right)^N \tag{3.5.3}$$

对于 $t>0$，在 $x=e^{-t}$ 中，即 $0<x<1$，最小化式(3.5.3)中的 RHS。它服从最小值 $e^{-t}=\lambda/(1-\lambda)$，最小值为

$$\left(\frac{\lambda}{1-\lambda}\right)^{\lambda N}\left(\frac{1}{2}\right)^N\left(1+\frac{\lambda}{1-\lambda}\right)^N=\lambda^{-\lambda N}\mu^{-\lambda N}\left(\frac{1}{2}\right)^N=2^{N\eta(\lambda)}\left(\frac{1}{2}\right)^N$$

其中 $\mu=(1-\lambda)$。因此，式(3.5.2)表明

$$\sum_{0\leqslant k\leqslant\lambda N}\binom{N}{k}\left(\frac{1}{2}\right)^N\leqslant 2^{N\eta(\lambda)}\left(\frac{1}{2}\right)^N \qquad \square$$

329

由于定理 3.5.2，不等式(3.5.1)成立，当

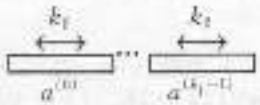
$$2^{2m\eta(\lambda)}<2^m-1 \tag{3.5.4}$$

现在，举例

$$\lambda=\lambda_m=\eta^{-1}\left(1/2-\frac{1}{\log m}\right)$$

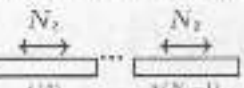
$\left(\text{其中 } 0<\lambda<\frac{1}{2}\right)$，界(3.5.4)变成 $2^{m-2m/\log m}<2^m-1$，当 m 足够大时为真。当 $m\to\infty$ 时，$\lambda_m\to h^{-1}(1/2)>0$。在这里并且接下来，$\eta^{-1}$ 是关于 $\lambda\in(0,1/2]\to\eta(\lambda)$。在固定 α 的码(3.5.1)中，信息速率为 1/2 但无法保证 $d/2m$ 远离 0。此外，没有一种有效的方法找到一个合适的 $\alpha=\alpha_m$。然而，在 1972 年，Justensen 展示了怎样利用 RS 码码字的截断去获得一个好的码字序列。

更确切地，考虑一个二进制 (k_1k_2) 码字 $\underline{a}$，形式为 k_1 间隔的 k_2 大小码字：$\underline{a}=a^{(0)}a^{(1)}\cdots a^{(k_1-1)}$。用图形表示，



$$\underline{a}=\qquad a^{(i)}\in\mathbb{F}_{2^{k_2}}, 0\leqslant i\leqslant k_1-1$$

我们固定一个在 $\mathbb{F}_{2^{k_2}}$ 上的 $[N_1,k_1,d_1]$ 码叫作外码：$\mathscr{C}_1\subset\mathbb{F}_{2^{k_2}}^{N_1}$。然后序列 $\underline{a}$ 编码成一个码字 $\underline{c}=c_0c_1\cdots c_{N_1-1}\in\mathscr{C}_1$。接下来，每个 $c_i\in\mathbb{F}_{2^{k_2}}$ 用码字 b_i 编码，它来自在 $\mathbb{F}_2$ 上的一个码 $\mathscr{C}_2$，叫作内码。结果就是长度为 N_1N_2 的序列 $b=b^{(0)}\cdots b^{(N_1-1)}\in\mathbb{F}_2^{N_1N_2}$：



$$\underline{b}=\qquad b^{(i)}\in\mathbb{F}_{2^{N_2}}, 0\leqslant i\leqslant N_1-1$$

编码过程可以用图表表示：

输入：一个(k, k_2)的数列a，　输出：一个(N, N_2)的码字b

观察到不同符号c_i可以用不同的内码编码。令外码$\mathscr{X}_1$是一个在$\mathbb{F}_{2^m}$上的$[2m-1, k, d]$RS码$\mathscr{X}^{RS}$。记一个二进制$(k2^m)$-码$\underline{a}$为$a^{(0)}\cdots a^{(k-1)}$的级联，其中$a^{(i)}\in\mathbb{F}_{2^m}$。用$\mathscr{X}^{RS}$编码$\underline{a}$产生一个码字$\underline{c}=c_0c_1\cdots c_{N-1}$，其中$N=2^m-1$，$c_i\in\mathbb{F}_{2^m}$。令$\beta$为在$\mathbb{F}_{2^m}$中的原始元素，那么对于所有$j=0, \cdots, N-1=2^m-2$，考虑内码

[330]

$$\mathscr{X}^{(j)}=\{(c, \beta^j c): c\in\mathbb{F}_{2^m}\} \tag{3.5.5}$$

最后的码字(一个"超级码")是

$$\underline{b}=(c_0, c_0)(c_1, \beta c_1)(c_2, \beta^2 c_2)\cdots(c_{N-1}, \beta^{N-1}c_{N-1}) \tag{3.5.6}$$

定义 3.5.3 Justensen 码$\mathscr{X}^{Js}_{m,k}$是用上述方式获得的二进制超级码字$\underline{b}$的集合，以$[2^m-1, k, d]$RS码作为外码$\mathscr{X}^1$，$\mathscr{X}^{(j)}$(见(3.5.5))作为内码，其中$0\leqslant j\leqslant 2^m-2$，码$\mathscr{X}^{Js}_{m,k}$长度为$2m(2^m-1)$，秩为$mk$，所以速率$\frac{k}{2(2^m-1)}<1/2$。

一个描述$\mathscr{X}^{Js}_{m,k}$的参数是$N=2^m-1$，RS外码的长度，我们想当$N\to\infty$时构造序列$\mathscr{X}^{Js}_{m,k}$，但$k/(2m(2^m-1))$和$d/(2m(2^m-1))$有界。所以先固定$R_0\in(0, 1/2)$，选择一个长度为N的RS外码，其中$N=2^m-1$且$k=\lceil 2NR_0\rceil$。$\mathscr{X}^{Js}_{m,k}$的速率为$k/(2N)\geqslant R_0$。

现在考虑最小重量

$$w(\mathscr{X}^{Js}_{m,k})=\min\{w(\underline{x}): \underline{x}\in\mathscr{X}^{Js}_{m,k}, \underline{x}\neq 0\}(=d(\mathscr{X}^{Js}_{m,k})) \tag{3.5.7}$$

对于任何固定的m，如果RS外码$\mathscr{X}^{RS}_N$，$N=2^m-1$，有最小重量d，那么超级码字$\underline{b}=(c_0, c_0)(c_1, \beta c_1)\cdots(c_{N-1}, \beta^{N-1}c_{N-1})\in\mathscr{X}^{Js}_{m,k}$有大于等于$d$个非零起始元素$c_0, \cdots, c_{N-1}$。进一步，任何两个在$\mathscr{X}^{(0)}, \mathscr{X}^{(1)}, \cdots, \mathscr{X}^{(N-1)}$中的内码只有0是相同的。所以，对应的来自不同码的$d$个有序对必须彼此不同。也就是说，超级码字$\underline{b}$有大于等于$d$个不同的非零二进制$(2m)$-序列。所以，我们需要对这样的和建立下界。注意到

$$d=N-k+1=N\left(1-\frac{k-1}{N}\right)\geqslant N(1-2R_0)$$

所以，一个超级码字$\underline{b}\in\mathscr{X}^{Js}_{m,k}$有至少$N(1-2R_0)$个不同的非零二进制$(2m)$-数列。

定理 3.5.4 任意$N(1-2R_0)$个不同的非零二进制$(2m)$-数列的重量之和为

$$\geqslant 2mN(1-2R_0)\left(\eta^{-1}\left(\frac{1}{2}\right)-o(1)\right) \tag{3.5.8}$$

证明 通过引理3.5.2，对于任何$\lambda\in[0, 1/2]$，重量$\leqslant 2m\lambda$的非零二进制$(2m)$-数列的数目为

[331]

$$\leqslant\sum_{1\leqslant i\leqslant 2m\lambda}\binom{2m}{i}\leqslant 2^{2m\eta(\lambda)}$$

去掉这些轻重量的序列，总重量为

$$\geqslant 2m\lambda(N(1-2R_0)-2^{2m\eta(\lambda)})=2mN\lambda(1-2R_0)\left(1-\frac{2^{2m\eta(\lambda)}}{N(1-2R_0)}\right)$$

选择$\lambda_m=\eta^{-1}\left(\frac{1}{2}-\frac{1}{\log(2m)}\right)\in(0, 1/2)$，然后$\lambda_m\to\eta^{-1}(1/2)$，由于$h^{-1}$在$[0, 1/2]$上连续。所以，

$$\lambda_m=\eta^{-1}\left(\frac{1}{2}-\frac{1}{\log(2m)}\right)=\eta^{-1}\left(\frac{1}{2}\right)-o(1)$$

由于$N=2^m-1$，我们有当$m\to\infty$，$N\to\infty$时

$$\frac{2^{m\eta(\delta)}}{N(1-2R_0)}=\frac{1}{1-2R_0}\frac{2^{m-\log(m)\eta(\delta m)}}{2^m-1}$$
$$=\frac{1}{1-2R_0}\frac{2^m}{2^m-1}\frac{1}{2^{\log(m)\eta(\delta m)}}\to 0$$

所以 $N(1-2R_0)$ 个不同的 $(2m)$-序列总重量的界为

$$2mN(1-2R_0)(\eta^{-1}(1/2)-o(1))(1-o(1))=2mN(1-2R_0)(\eta^{-1}(1/2)-o(1))$$

因而得到结论。 □

定理 3.5.4 证明了 $\mathscr{X}_{m,k}^{-\lambda_0}$ 有

$$w(\mathscr{X}_{m,k}^{-\lambda_0})\geqslant 2mN(1-2R_0)(\eta^{-1}\left(\frac{1}{2}\right)-o(1)) \tag{3.5.9}$$

并且

$$\frac{w(\mathscr{X}_{m,k}^{-\lambda_0})}{\mathrm{length}(\mathscr{X}_{m,k}^{-\lambda_0})}\geqslant(1-2R_0)(\eta^{-1}(1/2)-o(1))\to(1-2R_0)\eta^{-1}(1/2)$$
$$\approx c(1-2R_0)>0$$

所以，$k=[2NR_0]$，$N=2^m-1$ 和速率固定 $0<R_0<1/2$ 的数列 $\mathscr{X}_{m,k}^{-\lambda_0}$ 有信息速率 $\geqslant R_0>0$，并有

$$\frac{w(\mathscr{X}_{m,k}^{-\lambda_0})}{\mathrm{length}(\mathscr{X}_{m,k}^{-\lambda_0})}\to c(1-2R_0)>0,c=\eta^{-1}(1/2)>0.3 \tag{3.5.10}$$

在构造中，$R_0\in(0,1/2)$。然而，通过截断可以达到任何给定的速率 $R_0\in(0,1)$。参见文献[110]。

这一部分讨论的下一族码通过交替码形成。交替码是 BCH 码的推广（尽管通常不是循环的）。像 Justesen 码，交替码也构成一族接近完美的码。 332

令 M 为一个在域 $\mathbb{F}_{q^r}$ 上的 $(r\times n)$ 矩阵：

$$\boldsymbol{M}=\begin{bmatrix}c_{11}&\cdots&c_{1n}\\ \vdots&\ddots&\vdots\\ c_{r1}&\cdots&c_{rn}\end{bmatrix}$$

像之前那样，每个 c_{ij} 可以被写作 $\overrightarrow{c_{ij}}\in(\mathbb{F}_q)^m$，在 $\mathbb{F}_q$ 上长度为 m 的列向量。也就是说，我们可以认为 $\boldsymbol{M}$ 是在域 $\mathbb{F}_q$ 上一个 $(mr\times n)$ 的矩阵（再一次被标为 $\boldsymbol{M}$）。

给定元素 $a_1,\cdots,a_n\in\mathbb{F}_{q^r}$，我们有

$$\boldsymbol{M}\begin{bmatrix}a_1\\ \vdots\\ a_n\end{bmatrix}=\begin{bmatrix}c_{11}&\cdots&c_{1n}\\ \vdots&\ddots&\vdots\\ c_{r1}&\cdots&c_{rn}\end{bmatrix}\begin{bmatrix}a_1\\ \vdots\\ a_n\end{bmatrix}=\begin{bmatrix}\sum\limits_{1\leqslant j\leqslant n}a_jc_{1j}\\ \vdots\\ \sum\limits_{1\leqslant j\leqslant n}a_jc_{rj}\end{bmatrix}$$

进一步，如果 $b\in\mathbb{F}_q$，$\vec{c}$，$\vec{d}\in\mathbb{F}_{q^r}$，那么 $b\vec{c}=\overrightarrow{bc}$ 以及 $\vec{c}+\vec{d}=\overrightarrow{(c+d)}$。所以，如果 $a_1,\cdots,a_n\in\mathbb{F}_q$，那么

$$\boldsymbol{M}\begin{bmatrix}a_1\\ \vdots\\ a_n\end{bmatrix}=\begin{bmatrix}\overrightarrow{c_{11}}&\cdots&\overrightarrow{c_{1n}}\\ \vdots&\ddots&\vdots\\ \overrightarrow{c_{r1}}&\cdots&\overrightarrow{c_{rn}}\end{bmatrix}\begin{bmatrix}a_1\\ \vdots\\ a_n\end{bmatrix}=\begin{bmatrix}\sum\limits_{1\leqslant j\leqslant n}\overrightarrow{a_jc_{1j}}\\ \vdots\\ \sum\limits_{1\leqslant j\leqslant n}\overrightarrow{a_jc_{rj}}\end{bmatrix}$$

所以，如果 $\boldsymbol{M}$ 的列在 $\mathbb{F}_{q^r}$ 上为线性独立的 r-向量，它们也是在 $\mathbb{F}_q$ 上线性独立的 (rm)-向

量。也就是 M 的列在 $\mathbb{F}_q$ 上线性独立。

考虑到如果 ω 是单位向量的初始 $(n, \mathbb{F}_{q^r})$ 根，并且 $\delta \geq 2$，那么在 $\mathbb{F}_q$ 上的 $n \times (m\delta)$ Vandermonde 矩阵

$$H^{\mathrm{T}} = \begin{bmatrix} \vec{e} & \vec{e} & \cdots & \vec{e} \\ \vec{\omega} & \vec{\omega}^2 & \cdots & \vec{\omega}^{\delta-1} \\ \vec{\omega}^2 & \vec{\omega}_4 & \cdots & \vec{\omega}^{2(\delta-1)} \\ & & \cdots & \\ \vec{\omega}^{n-1} & \vec{\omega}^{2(n-1)} & \cdots & \vec{\omega}^{(\delta-1)(n-1)} \end{bmatrix}$$

检验一个狭义 BCH 码 $\mathscr{B}^{\mathrm{BCH}}_{q,n,\delta}$（一个正确的奇偶校验矩阵需要经过列消除）。通过在 $\mathbb{F}_{q^r}$ 上取 $n \times r$ 矩阵推广之

$$A = \begin{bmatrix} h_1 & h_1\alpha_1 & \cdots & h_1\alpha_1^{r-2} & h_1\alpha_1^{r-1} \\ h_2 & h_2\alpha_2 & \cdots & h_2\alpha_2^{r-2} & h_2\alpha_2^{r-1} \\ \vdots & \vdots & \ddots & \vdots & \\ h_n & h_n\alpha_n & \cdots & h_n\alpha_n^{r-2} & h_n\alpha_n^{r-1} \end{bmatrix} \tag{3.5.11}$$

333

或者它在 $\mathbb{F}_q$ 上的 $n \times (mr)$ 形式：

$$\vec{A} = \begin{bmatrix} \vec{h_1} & \overrightarrow{h_1\alpha_1} & \cdots & \overrightarrow{h_1\alpha_1^{r-2}} & \overrightarrow{h_1\alpha_1^{r-1}} \\ \vec{h_2} & \overrightarrow{h_2\alpha_2} & \cdots & \overrightarrow{h_2\alpha_2^{r-2}} & \overrightarrow{h_2\alpha_2^{r-1}} \\ \vdots & \vdots & \ddots & \vdots & \\ \vec{h_n} & \overrightarrow{h_n\alpha_n} & \cdots & \overrightarrow{h_n\alpha_n^{r-2}} & \overrightarrow{h_n\alpha_n^{r-1}} \end{bmatrix} \tag{3.5.12}$$

在这里，$r<n$，$h_1, \cdots, h_n$ 是非零元素，$\alpha_1, \cdots, \alpha_n$ 是 $\mathbb{F}_{q^r}$ 中不同的元素。

注意到(3.5.11)中 A 的任意 r 行组成类似于 Vandermonde 矩阵的亚方阵，它的行列式非零，因而 A 的任意 r 行在 $\mathbb{F}_{q^r}$ 和 $\mathbb{F}_q$ 上线性独立。此外式(3.5.12)中 A 的列在 $\mathbb{F}_q$ 上线性独立。然而式(3.5.12)中 $\vec{A}$ 的列可以是线性独立的，需要消除这些列以产生完美的奇偶校验矩阵 H。

定义 3.5.5 令 $\underline{\alpha}=(\alpha_1, \cdots, \alpha_n)$，$\underline{h}=(h_1, \cdots, h_n)$，其中 $\alpha_1, \cdots, \alpha_n$ 不同，$h_1, \cdots, h_n$ 是 $\mathbb{F}_{q^r}$ 上的非零元素。给定 $r<n$，交替码 $\mathscr{B}^{\mathrm{Alt}}_{q,\underline{\alpha},\underline{h}}$ 是(3.5.12)中 $n \times (rm)$ 矩阵 A 的核。

定理 3.5.6 $\mathscr{B}^{\mathrm{Alt}}_{q,\underline{\alpha},\underline{h}}$ 长度为 n，秩为 k，满足 $n-mr \leq k \leq n-r$，最小距离 $d(\mathscr{B}^{\mathrm{Alt}}_{q,\underline{\alpha},\underline{h}}) \geq r+1$。

我们看见交替码的确是 BCH 码的扩展。交替码定理的主要结果是下面的定理 3.5.7（这里不考证明）。

定理 3.5.7 *存在任意长的交替码 $\mathscr{B}^{\mathrm{Alt}}_{q,\underline{\alpha},\underline{h}}$ 满足 Gilbert-Varshamov 界。*

所以，交替码近似完美。更确切地说，一组接近完美的交替码由所谓的 Goppa 码产生。见下面。

Goppa 码是交替码的典型例子。它们由俄罗斯码学家，Valery Goppa 在 1972 年发明，产自来源于解析几何的优雅想法。在这里，我们用本部分采用的方法来演示它的构造。

令 $G(x) \in \mathbb{F}_{q^r}[x]$ 是 $\mathbb{F}_{q^r}$ 上的多项式，考虑 $\mathbb{F}_{q^r}[x]/(G(x))$，多项式环在 $\mathbb{F}_{q^r}$ 上 mod $G(x)$。那么 $\mathbb{F}_{q^r}[x]/(G(x))$ 是一个域当且仅当 $G(x)$ 不可约。但如果，对于一个给定的 $\alpha \in \mathbb{F}_{q^r}$，$G(\alpha) \neq 0$，线性多项式 $X-\alpha$ 在 $\mathbb{F}_{q^r}[x]/(G(x))$ 上可逆。实际上，有

334

$$G(X) = q(X)(X-\alpha) + G(\alpha) \tag{3.5.13}$$

其中 $q(X)\in \mathbb{F}_q[X]$，$\deg q(X)=\deg G(X)-1$。

所以 $q(X)(X-\alpha)=-G(\alpha)\bmod G(X)$ 或者

$$(-G(\alpha)^{-1}q(X))(X-\alpha)=e \bmod G(X)$$

和

$$(X-\alpha)^{-1}=(-G(\alpha)^{-1}q(X))\bmod G(X) \tag{3.5.14a}$$

由于 $q(X)=(G(X)-G(\alpha))(X-\alpha)^{-1}$，我们有

$$(X-\alpha)^{-1}=-(G(X)-G(\alpha))(X-\alpha)^{-1}G(\alpha)^{-1}\bmod G(X) \tag{3.5.14b}$$

我们定义 $(X-\alpha)^{-1}$ 是式(3.5.14a)$\mathbb{F}_{q^m}[x]/(G(x))$ 上的多项式。

定义 3.5.8 固定一个多项式 $G(x)\in \mathbb{F}_q[X]$ 和 $\mathbb{F}_{q^m}$ 上不同元素的集合 $\alpha=(\alpha_1,\cdots,\alpha_n)$，$q^m\geqslant n>\deg G(X)$，其中 $G(\alpha_j)=0$，$1\leqslant j\leqslant n$。给定一个码字 $b=(b_1,\cdots,b_n)$，其中 $b_i\in \mathbb{F}_q$，$1\leqslant i\leqslant n$，设

$$R_b(X)=\sum_{1\leqslant i\leqslant n} b_i(X-\alpha_i)^{-1}\in \mathbb{F}_{q^m}[X]/(G(X)) \tag{3.5.15}$$

q 进制 Goppa 码 $\mathscr{Z}^{\alpha}$（$=\mathscr{Z}^{\alpha}_{q,G}$）定义为如下的集合

$$\{b\in \mathbb{F}_q^n: R_b(X)=0 \bmod G(X)\} \tag{3.5.16}$$

显然，$\mathscr{Z}^{\alpha}_{q,G}$ 是线性码。多项式 $G(x)$ 叫作 Goppa 多项式；如果 $G(X)$ 不可约，我们称 $\mathscr{Z}^{\alpha}$ 不可约。

所以 $b=b_1,\cdots,b_n\in \mathscr{Z}^{\alpha}$ 当且仅当在 $\mathbb{F}_{q^m}[x]$ 上

$$\sum_{1\leqslant i\leqslant n} b_i(G(X)-G(\alpha_i))(X-\alpha_i)^{-1}G(\alpha_i)^{-1}=0 \tag{3.5.17}$$

我们有 $G(X)=\sum_{0\leqslant j\leqslant r} g_jX^j$，其中 $\deg G(X)=r$，$g_r=1$，$r<n$，那么在 $\mathbb{F}_{q^m}[X]$ 中

$$\begin{aligned}(G(X)-G(\alpha_i))(X-\alpha_i)^{-1}&=\sum_{1\leqslant j\leqslant r} g_j(X^j-\alpha_i^j)(X-\alpha_i)^{-1}\\&=\sum_{1\leqslant j\leqslant r}\sum_{0\leqslant k\leqslant j-1} X^k\alpha_i^{j-1-k}\end{aligned}$$

335

所以

$$\begin{aligned}&\sum_{1\leqslant i\leqslant n} b_i(G(X)-G(\alpha_i))(X-\alpha_i)^{-1}G(\alpha_i)^{-1}\\&=\sum_{1\leqslant i\leqslant n} b_i\sum_{1\leqslant j\leqslant r} g_j\sum_{0\leqslant k\leqslant j-1}\alpha_i^{j-1-k}X^kG(\alpha_i)^{-1}\\&=\sum_{0\leqslant k\leqslant r-1} X^k\sum_{1\leqslant i\leqslant n} b_iG(\alpha_i)^{-1}\sum_{k+1\leqslant j\leqslant r} g_j\alpha_i^{j-1-k}\end{aligned}$$

所以，当且仅当在 $\mathbb{F}_{q^m}$ 上时，$b\in \mathscr{Z}^{\alpha}$，

$$\sum_{1\leqslant i\leqslant n} b_iG(\alpha_i)^{-1}\sum_{k+1\leqslant j\leqslant r} g_j\alpha_i^{j-1-k}=0 \tag{3.5.18}$$

对于所有的 $k=0,\cdots,r-1$。

等式(3.5.18)推出 $\mathscr{Z}^{\alpha}$ 的奇偶校验矩阵。首先，我们看到矩阵

$$\begin{pmatrix} G(\alpha_1)^{-1} & G(\alpha_2)^{-1} & \cdots & G(\alpha_n)^{-1}\\ \alpha_1G(\alpha_1)^{-1} & \alpha_2G(\alpha_2)^{-1} & \cdots & \alpha_nG(\alpha_n)^{-1}\\ \alpha_1^2G(\alpha_1)^{-1} & \alpha_2^2G(\alpha_2)^{-1} & \cdots & \alpha_n^2G(\alpha_n)^{-1}\\ \vdots & \vdots & \ddots & \vdots\\ \alpha_1^{r-1}G(\alpha_1)^{-1} & \alpha_2^{r-1}G(\alpha_2)^{-1} & \cdots & \alpha_n^{r-1}G(\alpha_n)^{-1}\end{pmatrix} \tag{3.5.19}$$

它是在 $\mathbb{F}_{q^m}$ 上的 $(n\times r)$ 矩阵，提供了奇偶校验。在之前，矩阵(3.5.19)的任何 r 行在

$\mathbb{F}_{q^m}$上线性独立，列也是这样。同样我们将(3.5.19)写成一个在$\mathbb{F}_q$上的$n\times(mr)$矩阵；经过消除线性相关的列，我们得到奇偶校验矩阵$\boldsymbol{H}$。

我们看到$\mathscr{L}^{\mathrm{GD}}$是一种交替码$\mathscr{L}^{\mathrm{AD}}_{\boldsymbol{a},\boldsymbol{b}}$，其中$\boldsymbol{a}=(\alpha_1,\cdots,\alpha_n)$，$\boldsymbol{b}=(G(\alpha_1)^{-1},\cdots,G(\alpha_n)^{-1})$。所以，定理3.5.6说明

定理3.5.9 $\boldsymbol{\alpha}=(\alpha_1,\cdots,\alpha_n)$，$\deg(G(X))=r<n$的$q$进制Goppa码$\mathscr{L}=\mathscr{L}^{\mathrm{GD}}_{\boldsymbol{\alpha},G}$长度为$n$秩为$k$，满足$n-mr\leqslant k\leqslant n-r$，最小距离$d(\mathscr{L})\geqslant r+1$。

跟之前一样，最小距离的界在二进制的情况下能得到改善。假设一个二进制码字$\boldsymbol{b}=b_1,\cdots,b_n\in\mathscr{L}$，其中$\mathscr{L}$是一个Goppa码$\mathscr{L}^{\mathrm{GD}}_{\boldsymbol{\alpha},G}$，$\boldsymbol{\alpha}\subset\mathbb{F}_{q^m}$以及$G(X)\in\mathbb{F}_q[X]$。假设$w(\boldsymbol{b})=w$，$b_{i_1}=\cdots=b_{i_w}=1$。设$f_{\boldsymbol{b}}(X)=\prod_{1\leqslant j\leqslant w}(X-\alpha_{i_j})$，微分$\partial_X f_{\boldsymbol{b}}(X)$可以写作

$$\partial_X f_{\boldsymbol{b}}(X)=R_{\boldsymbol{b}}(X)f_{\boldsymbol{b}}(X) \tag{3.5.20}$$

其中$R_{\boldsymbol{b}}(X)=\sum_{1\leqslant j\leqslant w}(X-\alpha_{i_j})^{-1}$（参见式(3.5.15)）。由于多项式$f_{\boldsymbol{b}}(X)$和$R_{\boldsymbol{b}}(X)$在任意扩展$\mathbb{F}_{q^m}$上没有共同的根，它们是互质的，所以当且仅当$G(X)$除以$R_{\boldsymbol{b}}(X)$的结果与$G(X)$除以

335 $\partial_X f_{\boldsymbol{b}}(X)$的结果相同，$\boldsymbol{b}\in\mathscr{L}^{\mathrm{GD}}$。对于$q=2$，$\partial_X f_{\boldsymbol{b}}(X)$只有$X$的偶次项（由于它的单项式是$\ell X^{\ell-1}$乘以某个$\alpha_{i_j}$的乘积的形式：当$\ell$是奇数时它会消失）。换句话说，对于某个多项式$h(X)$，$\partial_X f_{\boldsymbol{b}}=h(X^2)=(h(X))^2$。所以如果$g(X)$是被$G(X)$整除的二次型最低度多项式，那么$G(X)$能够整除$\partial_X f_{\boldsymbol{b}}(X)$，当且仅当$g(X)$整除$\partial_X f_{\boldsymbol{b}}(X)$。所以，

$$\boldsymbol{b}\in\mathscr{L}^{\mathrm{GD}}\Leftrightarrow g(X)\mid\partial_X f_{\boldsymbol{b}}(X)\Leftrightarrow R_{\boldsymbol{b}}(X)=0 \bmod g(X) \tag{3.5.21}$$

定理3.5.10 令$\mathscr{L}$是一个二进制Goppa码$\mathscr{L}^{\mathrm{GD}}_{\boldsymbol{\alpha},G}$，如果$g(X)$是被$G(X)$整除的二次型最低度多项式，那么$\mathscr{L}=\mathscr{L}^{\mathrm{GD}}_{\boldsymbol{\alpha},g}$。所以$d(\mathscr{L}^{\mathrm{GD}})\geqslant\deg g(X)+1$。

引理3.5.11 假设Goppa多项式$G(X)\in\mathbb{F}_q[X]$在扩展域上没有多根。那么，$\mathscr{L}^{\mathrm{GD}}_{\boldsymbol{\alpha},G}=\mathscr{L}^{\mathrm{GD}}_{\boldsymbol{\alpha},\tilde{G}}$，最小距离$d(\mathscr{L}^{\mathrm{GD}}_{\boldsymbol{\alpha},G})\geqslant 2\deg G(X)+1$。所以$\mathscr{L}^{\mathrm{GD}}_{\boldsymbol{\alpha},G}$可以纠正$t\geqslant\deg G(X)$个错误。

一个多项式$G(X)$没有多根的二进制Goppa码$\mathscr{L}^{\mathrm{GD}}_{\boldsymbol{\alpha},G}$叫作可分离。

讨论一种适应于交替码和基于欧几里得算法的解码步骤十分有趣，参见2.5节。

解码一个在$\mathbb{F}_q$上的交替码$\mathscr{L}^{\mathrm{AD}}_{\boldsymbol{a},\boldsymbol{b}}$的起始步骤如下。如同在(3.5.12)，通过将在$\mathbb{F}_{q^m}$上的$n\times r$矩阵$\tilde{\boldsymbol{A}}=(b_i\alpha_i^{-j})$的每一项取代为长度为$m$的行，然后从$\tilde{\boldsymbol{A}}$消除线性独立的行。回想$b_1,\cdots,b_n$非零，$\alpha_1,\cdots,\alpha_n$是$\mathbb{F}_{q^m}$上的不同元素。假设接收到一个码字$\boldsymbol{u}=\boldsymbol{c}+\boldsymbol{e}$，其中$\boldsymbol{c}$是正确的码字，$\boldsymbol{e}$是一个错误向量。我们假设$r$是奇数，$t\leqslant r/2$个错误产生在数字$1\leqslant i_1<\cdots<i_t\leqslant n$上。令$\boldsymbol{e}$的第$i_j$项为$e_{i_j}\neq 0$。通过元素$\alpha_{i_j}$很方便检测出错误所在：当对于$i\neq i'$有$\alpha_i\neq\alpha_{i'}$（$\alpha_i$不同），如果我们能确定$\alpha_{i_1},\cdots,\alpha_{i_t}$我们将知道错误的位置。此外，如果我们引进错误位置多项式

$$\ell(X)=\prod_{j=1}^{t}(1-\alpha_{i_j}X)=\sum_{0\leqslant i\leqslant t}\ell_i X^i \tag{3.5.22}$$

其中$\ell_0=1$和根$\alpha_{i_j}^{-1}$，那么已经足够找到$\ell(X)$（也就是系数ℓ_i）。

我们必须计算校验子（我们叫它$\boldsymbol{A}$-校验子）通过矩阵$\boldsymbol{A}$产生：

337 $$\boldsymbol{u}\boldsymbol{A}=\boldsymbol{c}\boldsymbol{A}+0\cdots0e_{i_1}\cdots e_{i_t}0\cdots0\boldsymbol{A}$$

假设$\boldsymbol{A}$-校验子是$s=s_0\cdots s_{r-1}$，其中$s(X)=\sum_{0\leqslant i\leqslant r-1}s_iX^i$。很方便地引进错误评估多项式$\varepsilon(X)$，通过

$$\varepsilon(X)=\sum_{1\leqslant j\leqslant t}b_{i_j}e_{i_j}\prod_{1\leqslant j\leqslant t,j\neq h}(1-\alpha_{i_j}X) \tag{3.5.23}$$

引理 3.5.12 对于所有的 $u=1,\cdots,t$，

$$e_{i_u}=\frac{\varepsilon(\alpha_{i_u}^{-1})}{h_{i_u}\prod\limits_{1\le j\le t,j\ne u}(1-\alpha_{i_j}\alpha_{i_u}^{-1})} \tag{3.5.24}$$

证明 证明过程显而易见。 □

关键的命题是 $\ell(X)$，$\varepsilon(X)$和 $s(X)$通过以下关联

引理 3.5.13 下面的方程为真

$$\varepsilon(X)=\ell(X)s(X)\bmod X^r \tag{3.5.25}$$

证明 写出下面的步骤

$$\begin{aligned}\varepsilon(X)-\ell(X)s(X)&=\sum_{1\le k\le t}h_{i_k}e_{i_k}\prod_{1\le j\le t,j\ne k}(1-\alpha_{i_j}X)-\ell(X)\sum_{1\le i\le r-1}s_iX^i\\&=\sum_k h_{i_k}e_{i_k}\prod_{1\le j\le t,j\ne k}(1-\alpha_{i_j}X)-\ell(X)\sum_i\sum_{1\le j\le t}h_{i_j}\alpha_{i_j}^i e_{i_j}X^i\\&=\sum_k h_{i_k}e_{i_k}\prod_{j\ne k}(1-\alpha_{i_j}X)-\ell(X)\sum_k h_{i_k}e_{i_k}\sum_i\alpha_{i_k}^iX^i\\&=\sum_k h_{i_k}e_{i_k}\Big(\prod_{j\ne k}(1-\alpha_{i_j}X)-\ell(X)\sum_i\alpha_{i_k}^iX^i\Big)\\&=\sum_k h_{i_k}e_{i_k}\prod_{j\ne k}(1-\alpha_{i_j}X)-\Big(1-(1-\alpha_{i_k}X)\sum_i\alpha_{i_k}^iX^i\Big)\\&=\sum_k h_{i_k}e_{i_k}\prod_{j\ne k}(1-\alpha_{i_j}X)\Big(1-(1-\alpha_{i_k}X)\frac{1-\alpha_{i_k}^rX^r}{1-\alpha_{i_k}X}\Big)\\&=\sum_k h_{i_k}e_{i_k}\prod_{j\ne k}(1-\alpha_{i_j}X)\alpha_{i_k}^rX^r=0\bmod X^r\end{aligned}$$ □

引理 3.5.13 展示了解码交替码的方法。我们知道存在一个多项式 $q(X)$使得

$$\varepsilon(X)=q(X)X^r+\ell(X)s(X) \tag{3.5.26}$$

我们也有 $\deg\varepsilon(X)\le t-1<r/2$，$\deg\ell(X)=t\le r/2$，并且 $\ell(X)$和 $\varepsilon(X)$互质的条件是它们在任何扩展中没有共同根。假设我们将欧几里得算法用到已知的多项式 $f(X)=X^r$ 和 $g(X)=s(X)$上来找到 $\ell(X)$和 $\varepsilon(X)$。通过引理 2.5.44，一个典型的步骤产生了余式 338

$$r_k(X)=a_k(X)X^r-b_k(X)s(X) \tag{3.5.27}$$

如果我们想让 $r_k(X)$和 $b_k(X)$给出 $\varepsilon(X)$和 $\ell(X)$，它们的度必须吻合：至少必须有 $\deg r_k(X)<r/2$ 和 $\deg b_k(X)<r/2$。所以，这个算法重复到 $\deg r_k(X)\ge r/2$ 和 $\deg b_k(X)<r/2$。然后，根据引理 2.5.44 的命题(3)，$\deg b_k(X)=\deg X^r-\deg r_{k-1}(X)\le r-r/2=r/2$。这是有可能的，因为这个算法可以循环到 $r_k(X)=\gcd(X^r,s(X))$。但是之后 $r_k(x)\mid\varepsilon(X)$，所以 $\deg r_k(x)\le\deg\varepsilon(X)<r/2$。所以我们假设 $\deg r_k(x)\le r/2$，$\deg b_k(x)\le r/2$。

相关等式为

$$\begin{aligned}&\varepsilon(X)=q(X)X^r+\ell(X)s(X)\\&\deg\varepsilon(X)<r/2,\deg\ell(X)\le r/2\\&\gcd(\varepsilon(X),\ell(X))=1\end{aligned}$$

和

$$r_k(X)=a_k(X)X^r+b_k(X)s(X),\deg r_k(X)<r/2,\deg b_k(X)\le r/2$$

我们想要证明多项式 $r_k(x)$和 $b_k(x)$是 $\varepsilon(X)$和 $\ell(X)$的向量积。排除 $s(X)$得到

$$b_k(X)\varepsilon(X)-r_k(X)\ell(X)=(b_k(X)q(X)-a_k(X)\ell(X))X^r$$

由于

$$\deg b_2(X)\varepsilon(X)=\deg b_2(X)+\deg \varepsilon(X)<r/2+r/2=r$$

和

$$\deg r_2(X)\ell(X)=\deg r_2(X)+\deg \ell(X)<r/2+r/2=r$$

$\deg(b(X)\varepsilon(X)-r_2(x)\ell(X))<r$。所以，$b(X)\varepsilon(X)-r_2(x)\ell(X)$必须是0，即

$$\ell(X)r_2(X)=\varepsilon(X)b_2(X),b_2(X)q(X)=a_2(X)\ell(X)$$

所以，$\ell(X)\mid\varepsilon(X)b_2(X)$和$b_2(X)\mid a_2(X)\ell(X)$。但是$\ell(X)$和$\varepsilon(X)$互质，$a_2(X)$和$b_2(X)$也是这样(通过引理2.5.44的命题(5))。所以$\ell(X)=\lambda b_2(X)$，所以$\varepsilon(X)=\lambda r_2(X)$。由于$\ell(0)=1$，$\lambda=b_2(0)^{-1}$。

总结一下，有如下定理。

定理3.5.14 (交替码的解码算法)假设$\mathscr{Z}_{\alpha,k}^{\mathrm{Alt}}$是一个交替码，$r$为偶数，$t\leqslant r/2$个错误发生在接收的码字为$u$的情况下。然后，在接收到的码字$u$中：

338

(a) 找到A-校验子$uA^{\mathrm{T}}=s_0\cdots s_{r-1}$，对应多项式$s(X)=\sum_i s_i X^i$。

(b) 用欧几里得算法，从$f(X)=X^r$，$g(X)=s(X)$开始，得到$r_i(X)=a_i(X)X^r+b_i(X)s(X)$，其中$\deg r_{i-1}(X)\geqslant r/2$，$\deg r_i(X)<r/2$。

(c) 设$\ell(X)=b_i(0)^{-1}b_i(X)$，$\varepsilon(X)=b_i(0)^{-1}r_i(X)$，

然后$\ell(X)$是错误位置多项式，根为α_{i_1}，…，$\alpha_{i_t}=\alpha_{i_1}$，$i_1$，…，$i_t$为错误数字，$e_{i_j}$的值为

$$e_{i_j}=\frac{\varepsilon(\alpha_{i_j}^{-1})}{h_{i_j}\prod_{l\neq j}(1-\alpha_{i_l}\alpha_{i_j}^{-1})} \tag{3.5.28}$$

这一部分背后的思想建立在解析几何码的深远发展上。解析几何在现代码设计中提供了重要的工具，见文献[98]、[99]、[158]和[160]。

3.6 本章附加问题

问题3.1 定义Reed-Solomon码并证明它们是最大距离分离。证明Reed-Solomon码的对偶是一个Reed-Solomon码。

找到长度为15，秩为11的一个Reed-Solomon码的最小距离和在$\mathbb{F}_{16}$上的生成矩阵$g_1(X)$。用已知的$\mathbb{F}_{16}$域表将$g_1(X)$写成$\omega^3+\omega^5X+\omega^{10}X^2+\cdots$，每个系数为$\mathbb{F}_{16}$中的原始元素的一个幂。

在$\mathbb{F}_{16}$上确定一个两位错误纠正Reed-Solomon码，找到它的长度，秩和生成多项式。

$\mathbb{F}_{11}=\{0,1,2,3,4,5,6,7,8,9,10\}$，满足加法和乘法mod11：

i	0	1	2	3	4	5	6	7	8	9
ω^i	1	2	4	8	5	10	9	7	3	6

$\mathbb{F}_{16}=\mathbb{F}_2^4$的域表为：

340

i	0	1	2	3	4	5	6	7	8
ω^i	0001	0010	0100	1000	0011	0110	1100	1011	0101
i	9	10	11	12	13	14			
ω^i	1010	0111	1110	1111	1101	1001			

解答 设计距离$\delta\leqslant q-1$的一个q进制RS码$\mathscr{Z}^{\mathrm{RS}}$定义为在$\mathbb{F}_q$上长度为$N=q-1$的循环码，其中生成多项式为

$$g(X)=(X-\omega^l)(X-\omega^{l+1})\cdots(X-\omega^{l+\delta-2})$$

$\deg(g(X))=\delta-1$。这里 ω 是单位元素的一个初始($q-1$, $\mathbb{F}_q$)根(即 $\mathbb{F}_q^*$ 的一个初始元素)，$b=0, 1, \cdots, q-2$。幂 $\omega^b, \cdots, \omega^{b+\delta-2}$ 叫作(规定的)零元素和 $\mathscr{X}^{RS}$ 中 ω 非零元素的其余 $N-\delta+1$ 幂。

$\mathscr{X}^{RS}$ 的秩等于 $k=N-\delta+1$。距离 $\geq\delta=N-k+1$，但是由 Singleton 界可知距离应该 $\leq\delta=N-k+1$。所以距离等于 $\delta=N-k+1$，即 $\mathscr{X}^{RS}$ 是最大距离分离。

$\mathscr{X}^{RS}$ 的对偶 $(\mathscr{X}^{RS})^{\perp}$ 是循环码，它的零元素是 $\mathscr{X}^{RS}$ 非零元素的逆：

$$\omega^{j\perp}=(\omega^j)^{-1}, j\neq b, \cdots, b+\delta-2$$

即它们是

$$\omega^{2-b}, \omega^{2-b-1}, \cdots, \omega^{2-b-\delta+1}$$

$(\mathscr{X}^{RS})^{\perp}$ 的生成多项式 $g^{\perp}(X)$ 是

$$g^{\perp}(X)=(X-\omega^{b^{\perp}})(X-\omega^{b^{\perp}-1})\cdots(X-\omega^{b^{\perp}+\delta-2})$$

其中 $b^{\perp}=q-b$。也就是说 $(\mathscr{X}^{RS})^{\perp}$ 是设计距离为 $q-\delta+1$ 的一个 RS 码。

在例子中，长度为 15 意味着 $q=15+1=16$，秩为 11 使得距离 $\delta=15-11+1=5$。对于 $b=1$，在 $\mathbb{F}_{16}=\mathbb{F}_2$ 上的生成多项式 $g_1(X)$ 为

$$\begin{aligned}g_1(X)&=(X-\omega)(X-\omega^2)(X-\omega^3)(X-\omega^4)\\&=X^4-(\omega+\omega^2+\omega^3+\omega^4)X^3\\&\quad+(\omega^3+\omega^4+\omega^5+\omega^5+\omega^6+\omega^7)X^2\\&\quad-(\omega^5+\omega^7+\omega^8+\omega^9)X+\omega^{10}\\&=X^4+\omega^{13}X^3+\omega^4X^2+\omega^9X+\omega^{10}\end{aligned}$$

其中计算是利用 $\mathbb{F}_{16}$ 的域表完成的。 341

类似地，长度为 10 意味着 $q=11$，秩为 6，距离 $\delta=10-6+1=5$。对于 $b=1$ 在 $\mathbb{F}_{11}$ 上的生成多项式 $g_2(X)$ 为

$$\begin{aligned}g_2(X)&=(X-\omega)(X-\omega^2)(X-\omega^3)(X-\omega^4)\\&=X^4-(\omega+\omega^2+\omega^3+\omega^4)X^3\\&\quad+(\omega^3+\omega^4+\omega^5+\omega^5+\omega^6+\omega^7)X^2\\&\quad-(\omega^6+\omega^7+\omega^8+\omega^9)X+\omega^{10}\\&=X^4+3X^3+5X^2+6X-1\end{aligned}$$

其中计算是利用 $\mathbb{F}_{11}$ 的域表完成的。

最后，一个在 $\mathbb{F}_{16}$ 上的两错误纠正 RS 码必须有长度 $N=15$，距离 $\delta=5$，所以秩为 11。因此，它正好是之前的 16 进制[15, 11]RS 码。 □

问题 3.2 令 $\mathscr{X}$ 为一个二进制线性[N, k]码，$\mathscr{X}^{eo}$ 是重量为偶数的 $x\in\mathscr{X}$ 的码字集合。证明下述命题之一：

(i) $\mathscr{X}=\mathscr{X}^{eo}$。

(ii) $\mathscr{X}^{eo}$ 是 $\mathscr{X}$ 的一个[N, k−1]的线性子码。

证明如果 $\mathscr{X}$ 的生成矩阵 G 没有零列向量，那么总重量 $\sum_{x\in\mathscr{X}}w(x)$ 等于 $N2^{k-1}$。(提示：考虑 G 的每一列的贡献。)

记 $\mathscr{X}_{H,\ell}$ 是长度为 $N=2^\ell-1$ 的二进制 Hamming 码，$\mathscr{X}_{H,\ell}^{+}$ 是对偶的单形码，$\ell=3, 4, \cdots$。N 阶向量 1…1(所有数字为 1)是不是 $\mathscr{X}_{H,\ell}$ 中的一个码字？令 A_s 和 A_s^{+} 分别为 $\mathscr{X}_{H,\ell}$ 和 $\mathscr{X}_{H,\ell}^{+}$ 中重量为 s 的码字的数目，其中 $A_0=A_0^{+}=1$，$A_1=A_2=0$。验证

$$A_3=N(N-1)/3!, A_4=N(N-1)(N-3)/4!$$

证明$A_{2^t-1}^{\perp}=2^t-1$(即所有非零码字$x\in$重量为2^{t-1})。利用最后一个事实和二进制码字的MacWilliams恒等式，可以给出A_s关于Kravchuk多项式的值$K_s(2^{t-1})$的表达式：

$$K_s(2^{t-1})=\sum_{j=0\vee s+2^{t-1}-2^t+1}^{s\wedge 2^{t-1}}\binom{2^{t-1}}{j}\binom{2^t-1-2^{t-1}}{s-j}(-1)^j$$

在这里，$0\vee s+2^{t-1}-2^t+1=\max[0,\ s+2^{t-1}-2^t+1]$，$s\wedge 2^{t-1}=\min[s,\ 2^{t-1}]$。验证你的表达式对于$s=N=2^t-1$可给出正解。

342

解答 $\mathscr{X}^{\prime\prime}$总是$\mathscr{X}$的线性子码。事实上，对于二进制码字$x$和$x'$，$w(x+x')=w(x)+w(x')-2w(x\wedge x')$，其中数位$(x\wedge x')_j=x_jx'_j=\min[x_j,\ x'_j]$。如果$x$和$x'$都是偶码字(重量为偶数)或者奇码字(重量为奇数)，那么$x+x'$是偶码字，如果x是偶码字，x'是奇码字，那么$x+x'$是奇码字。所以，如果$\mathscr{X}^{\prime\prime}\neq\mathscr{X}$，那么$\mathscr{X}^{\prime\prime}$是$\mathscr{X}$中序号$[\mathscr{X}^{\prime\prime}:\mathscr{X}^{\prime\prime}]2$的子群。所以，存在两个陪集，并且$\mathscr{X}^{\prime\prime}$是$\mathscr{X}$的一半。所以$\mathscr{X}^{\prime\prime}$是$[N,\ k-1]$码。

令$\boldsymbol{g}^{(j)}=(g_{1j},\ \cdots,\ g_{kj})^{\mathrm{T}}$为$\mathscr{X}$的生成矩阵$\boldsymbol{G}$的第$j$列。集合$\mathscr{W}_j=\{i=1,\ \cdots,\ k:\ g_{ij}=1\}$，其中$\#\mathscr{W}_j=w(\boldsymbol{g}^{(j)})=w_j\geqslant 1$，$1\leqslant j\leqslant N$。和$\sum_{x\in\mathscr{X}}w(x)$中来自$\boldsymbol{g}^{(j)}$的部分等于

$$2^{k-w_j}\times 2^{w_j-1}=2^{k-1}$$

这里2^{k-w_j}代表了补集$\{1,\ \cdots,\ k\}/\mathscr{W}_j$的子集数目，$2^{w_j-1}$则是$\mathscr{W}_j$奇子集的数目。乘以$N$(列的数目)就得到了$N2^{k-1}$。

如果$\boldsymbol{H}=\boldsymbol{H}_{H,\ell}$是Hamming码$\mathscr{X}_{H,\ell}$的奇偶校验矩阵，那么$\boldsymbol{H}$第$j$行的重量等于数字$1,\ \cdots,\ 2^\ell-1$，$1\leqslant j\leqslant\ell$二进制分解中位于$j$的数位1的数目。所以，$w(\boldsymbol{h}^{(j)})=2^{\ell-1}$(数字$1,\ \cdots,\ 2^\ell-1$有一半在$j$处为零，另一半为一)。所以，对于所有$j=1,\ \cdots,\ N$，点乘$\mathbf{1}\cdot\boldsymbol{h}^{(j)}=w(\boldsymbol{h}^{(j)})\bmod 2=0$，即$\mathbf{1}\cdots\mathbf{1}\in\mathscr{X}_{H,\ell}$。

现在$A_3=N(N-1)/3!$，是$\boldsymbol{H}$线性相关的三列的数目(因为可以通过固定两个不同的列进行选择：第三个是它们的和)。接下来，$A_4=N(N-1)(N-3)/4!$，是$\boldsymbol{H}$线性相关的四列的数目(因为可以通过固定：(a)任意两个不同的列，(b)第三列不是它们的和，(c)第四列为前三列的和进行选择)。类似地，$A_5=N(N-1)(N-3)(N-7)/5!$，是$\boldsymbol{H}$线性相关的五列的数目。(这里$N-7$表示当选择第四列时我们应该避免前三列任何一种$2^3-1=7$的线性组合)。

实际上，在对偶码$\mathscr{X}_{H,\ell}^{\perp}$中任何非零码字$x$有$w(x)=2^{\ell-1}$。为了证明这一点，注意到$\mathscr{X}_{H,\ell}^{\perp}$的生成多项式是$\boldsymbol{H}$。所以，将$x$写作$\boldsymbol{H}$行的和，并令$\mathscr{W}$为参与加和的$\boldsymbol{H}$列的集合，其中$\#\mathscr{W}=w\leqslant\ell$。那么$w(x)$等于$1,\ 2,\ \cdots,\ 2^{\ell-1}$中$j$的数目，使得在二进制分解$j=2^0j_0+2^1j_1+\cdots+2^{\ell-1}j_{\ell-1}$中和$\sum_{i\in\mathscr{W}}j_i\bmod 2$等于1。如前所述，它等于$2^{w-1}$($\mathscr{W}$子集为奇数势的数目)。所以，$w(x)=2^{w-1}=2^{\ell-1}$。注意到$\mathscr{X}_{H,\ell}^{\perp}$的秩为$2^\ell-1-(2^\ell-1-\ell)=\ell$，大小为$\#\mathscr{X}_{H,\ell}^{\perp}=2^\ell$。

MacWilliams恒等式(反向形式)为

343

$$A_s=\frac{1}{\#\mathscr{X}_{H,\ell}}\sum_{i=1}^{N}A_iK_s(i)\tag{3.6.1}$$

其中，

$$K_s(i)=\sum_{j=0\vee s+i-N}^{s\wedge i}\binom{i}{j}\binom{N-i}{s-j}(-1)^j\tag{3.6.2}$$

其中$0\vee s+i-N=\max[0,\ s+i-N]$，$s\wedge i=\min[s,\ i]$。在这里，$A_0^{\perp}=1$，$A_{2^\ell-1}^{\perp}=2^\ell-1$

($\mathscr{C}^{\perp}_{n,t}$中非零码字的数目)。所以

$$A_t=\frac{1}{2^t}(1+(2^t-1)K_t(2^{t-1}))$$

$$=\frac{1}{2^t}(1+(2^t-1)\sum_{j=0\vee 2^{t-1}-2^t-1}^{2^{t-1}}\binom{2^{t-1}}{j}\binom{2^t-1-2^{t-1}}{2^t-1-j}(-1)^j)$$

对于$s=N=2^t-1$，A_{2^t-1}可为1(如果2^t-码字11…1在$\mathscr{C}_{n,t}$中)或0(如果不是)。最后给出公式

$$A_{2^t-1}=\frac{1}{2^t}(1+(2^t-1)\sum_{j=2^{t-1}}^{2^{t-1}}\binom{2^{t-1}}{j}\binom{2^t-1-2^{t-1}}{2^t-1-j}$$

$$=\frac{1}{2^t}(1+2^t-1)=1$$

符合$11\cdots1\in\mathscr{C}_{n,t}$的情况。 □

问题 3.3 令ω为$M(X)=X^5+X^2+1$在$\mathbb{F}_{32}$上的一个根；已知$M(X)$是$\mathbb{F}_{32}$的一个初始多项式，ω是一个初始的单位$(31,\mathbb{F}_{32})$-根。利用元素ω，ω^3，ω^4，ω^5构造一个长度为31，设计距离为5的二进制狭义初始BCH码$\mathscr{C}$。确定ω，ω^3，ω^4，ω^5各自的分圆陪集$\{i,2i,\cdots,2^{d-1}i\}$。验证ω和ω^3足够定义$\mathscr{C}$的零点，并且$\mathscr{C}$的实际最小距离等于5。验证$\mathscr{C}$的生成多项式$g(X)$为乘积

$$(X^5+X^2+1)(X^5+X^4+X^3+X^2+1)$$
$$=X^{10}+X^8+X^6+X^5+X^4+X^3+1$$

假设你从使用码$\mathscr{C}$的发送者收到一个码字$u(X)=X^{13}+X^{12}+X^9+X^7+X^6+X^2+1$。验证$u(\omega)=\omega^5$和$u(\omega^3)=\omega^5$，说明$u(X)$应该被解码为

$$c(X)=X^{12}+X^{11}+X^9+X^7+X^6+X^3+X^2+1$$

并证明$c(X)$实际上是$\mathscr{C}$中的一个码字。

我们给出了$\mathbb{F}_{32}=\mathbb{F}_2^5$的域表和$\mathbb{F}_2$上度数等于5的不可约多项式列表来辅助计算。 344

$\mathbb{F}_{32}=\mathbb{F}_2^5$的域表：

i	0	1	2	3	4	5	6	7
ω^i	00001	00010	00100	01000	10000	00101	01010	10100
i	8	9	10	11	12	13	14	15
ω^i	01101	11010	1001	00111	01110	11100	11101	11111

i	16	17	18	19	20	21	22	23
ω^i	11011	10011	00011	00110	01100	11000	10101	01111
i	24	25	26	27	28	29	30	
ω^i	11110	11001	10111	01011	10110	01001	10010	

$\mathbb{F}_2$上度数等于5的不可约多项式列表：

$$X^5+X^2+1,X^5+X^3+1,X^5+X^4+X^2+X+1,$$
$$X^5+X^4+X^3+X+1,X^5+X^4+X^3+X^2+1;$$

它们的阶都为31。多项式X^5+X^2+1是初始的。

解答 由于$M(X)=X^5+X^2+1$是$\mathbb{F}_2[X]$上的一个初始多项式，$M(X)$的任何根ω都是一个初始的单位$(31,\mathbb{F}_2)$-根。此外，$M(X)$是ω的最小多项式。

要构造的BCH码$\mathscr{C}$是一个度数最小的生成多项式，根包含ω，ω^3，ω^4，ω^5的循环码

（即零点构成包含 ω，ω^2，ω^3，ω^4 最小集合的一个循环码）。所以，X 的生成多项式 $g(X)$ 是 ω，ω^2，ω^3，ω^4 最小多项式的最小公倍数。

ω 的分圆陪集为 $C=\{1, 2, 4, 8, 16\}$；所以

$$(X-\omega)(X-\omega^2)(X-\omega^4)(X-\omega^8)(X-\omega^{16})=X^5+X^4+1$$

是 ω，ω^2，ω^4 的最小多项式。ω^3 的分圆陪集为 $C=\{3, 6, 12, 24, 17\}$，ω^3 的最小多项式 $M_3(X)$ 等于

$$\begin{aligned}M_3(X)&=(X-\omega^3)(X-\omega^6)(X-\omega^{12})(X-\omega^{24})(X-\omega^{17})\\&=X^5-(\omega^3+\omega^6+\omega^{12}+\omega^{24}+\omega^{17})X^4+(\omega^9+\omega^{15}+\omega^{27}\\&\quad-\omega^{35}+\omega^{18}+\omega^{30}+\omega^{23}-\omega^{39}+\omega^{34}+\omega^{41})X^3\\&\quad-(\omega^{21}+\omega^{33}+\omega^{45}-\omega^{38}+\omega^{32}+\omega^{44}+\omega^{43}-\omega^{52}\\&\quad+\omega^{47}+\omega^{55})X^2+(\omega^{57}+\omega^{54}+\omega^{52}+\omega^{49}+\omega^{59})X+\omega^{63}\\&=X^5+X^3+X^2+X^1+1\end{aligned}$$

[345] 上式通过直接的域表计算或查找数 2_5 等于 5 的不可约多项式列表得到。

所以，ω 和 ω^3 足够定义 $\mathscr{Z}$ 的零点，生成多项式 $g(X)$ 等于

$$\begin{aligned}&(X^5+X^2+1)(X^5+X^4+X^3+X^2+1)\\&=X^{10}+X^9+X^8+X^6+X^5+X^2+1\end{aligned}$$

如题所要求的一样。换句话说，

$$\begin{aligned}\mathscr{Z}&=\{c(X)\in F_2[X]/(X^{31}-1): c(\omega)=c(\omega^3)=0\}\\&=\{c(X)\in F_2[X]/(X^{31}-1): g(X)\mid c(X)\}\end{aligned}$$

$\mathscr{Z}$ 的秩等于 21，$\mathscr{Z}$ 的最小距离等于 5，即它的最小设计距离。这来自于定理 3.3.20：

令 $N=2^\ell-1$。如果 $2^{\ell E}<\sum_{0\le i\le 2E-1}\binom{N}{i}$，那么设计距离为 $2E-1$ 二进制狄义初始 BCH 码有最小距离 $2E-1$。

在这里，$N=31=2^5-1$，其中 $\ell=5$，$E=2$，即 $2E+1=5$，并且

$$1024=2^{10}<1+31+\frac{31\times 30}{2}+\frac{31\times 30\times 29}{2\times 3}=4992$$

所以，$\mathscr{Z}$ 纠正了两个错误。Berlekamp-Massey 译码步骤要求计算定义零点处的接收多项式的值，从 F_{32} 的域表中我们得到

$$\begin{aligned}u(\omega)&=\omega^{12}+\omega^{11}+\omega^2+\omega^7+\omega^4+\omega^2+1=\omega^3\\u(\omega^3)&=\omega^{26}+\omega^{25}+\omega^{22}+\omega^{19}+\omega^4+1=\omega^9\end{aligned}$$

所以，$u(\omega^3)=u(\omega)^3$。由于 $u(\omega)=\omega^3$，我们知道在数位 3 处有一个错误发生，即 $u(X)$ 译码为

$$c(X)=X^{12}+X^{11}+X^7+X^4+X^3+X^2+1$$

即要求的 $(X^3+1)g(X)$。 □

问题 3.4 定义一个长度为 N，维度为 k，字母表为 F 的线性 $[N, k]$ 码的对偶 $\mathscr{Z}^\perp$。证明或证伪如果 $\mathscr{Z}$ 是一个 N 为奇数的二进制 $[N, (N-1)/2]$ 码，那么 $\mathscr{Z}^\perp$ 由 $\mathscr{Z}$ 的一个基底加上码字 $\{\cdots\}$ 产生。证明或证伪如果一个二进制码 $\mathscr{Z}$ 是自对偶，即 $\mathscr{Z}=\mathscr{Z}^\perp$，那么 N 为偶数，码字 $\{\cdots\}$ 属于 $\mathscr{Z}$。

证明一个二进制自对偶线性 $[N, N/2]$ 码 $\mathscr{Z}$ 对于每个偶数的 N 都存在。相反地，证明如果一个二进制线性 $[N, k]$ 码 $\mathscr{Z}$ 是自对偶的，那么 $k=N/2$。

给出一个非二进制线性自对偶码的例子。

解答 [N，k]线性码$\mathscr{X}$的对偶$\mathscr{X}^{\perp}$为

$$\mathscr{X}^{\perp}=\{\boldsymbol{x}=x_1\cdots x_N\in \mathbb{F}^N;\boldsymbol{x}\cdot\boldsymbol{y}=0,\text{对于所有的 }\boldsymbol{y}\in\mathscr{X}\}$$ [346]

其中$\boldsymbol{x}\cdot\boldsymbol{y}=x_1y_1+\cdots+x_Ny_N$。选取$N=5$，$k=(N-1)/2=2$。

$$\mathscr{X}=\begin{pmatrix}0&0&0&0&0\\0&1&0&0&0\\1&1&0&0&0\\0&0&0&0&0\end{pmatrix}$$

那么$\mathscr{X}^{\perp}$的生成矩阵为

$$\begin{pmatrix}0&0&1&0&0\\0&0&0&1&0\\0&0&0&0&1\end{pmatrix}$$

来自$\mathscr{X}$的向量没有属于$\mathscr{X}^{\perp}$的，所以命题为假。

现在选取一个自对偶码$\mathscr{X}=\mathscr{X}^{\perp}$。如果码字$\mathbf{1}=1\cdots1\notin\mathscr{X}$，那么存在$\boldsymbol{x}\in\mathscr{X}$使得$\boldsymbol{x}\cdot\mathbf{1}\neq0$。但是$\boldsymbol{x}\cdot\mathbf{1}=\sum x_i=w(\boldsymbol{x})\bmod 2$。另一方面，$\sum x_i=\boldsymbol{x}\cdot\boldsymbol{x}$，所以$\boldsymbol{x}\cdot\boldsymbol{x}\neq0$。但是这让$\boldsymbol{x}\notin\mathscr{X}^{\perp}$，所以$\mathbf{1}\in\mathscr{X}$。但是这使得$\mathbf{1}\cdot\mathbf{1}=0$，表明$N$为偶数。

现在令$N=2k$，将数位$1,\cdots,N$分解成k个不相交的对$(\alpha_1,\beta_1),\cdots,(\alpha_k,\beta_k)$，其中$\alpha_i<\beta_i$。然后考虑$k$个长度为$N$，重量为2的二进制码字$\boldsymbol{x}^{(1)},\cdots,\boldsymbol{x}^{(k)}$，其中码字$\boldsymbol{x}^{(i)}$位于$(\alpha_i,\beta_i)$的数位非零。然后构造由$\boldsymbol{x}^{(1)},\cdots,\boldsymbol{x}^{(k)}$所生的[N，k]码。

这个码$\mathscr{X}$是自对偶的。实际上，对于所有的i,i'，$\boldsymbol{x}^{(i)}\cdot\boldsymbol{x}^{(i')}=0$，$\mathscr{X}\subset\mathscr{X}^{\perp}$。相反地，令$\boldsymbol{y}\in\mathscr{X}^{\perp}$。那么对于所有的$i$，$\boldsymbol{y}\cdot\boldsymbol{x}^{(i)}=0$。这表明对于所有的$i$，$\boldsymbol{y}$在$(\alpha_i,\beta_i)$处有0或非零的数位，$\boldsymbol{y}\in\mathscr{X}$，所以$\mathscr{X}=\mathscr{X}^{\perp}$。

现在假设$\mathscr{X}=\mathscr{X}^{\perp}$，$N$为偶数。由秩零化度定理可知维度必须是$k$。

非二进制自对偶码是三进制Golay[12，6]码，其生成矩阵为

$$\boldsymbol{G}=\begin{pmatrix}1&0&0&0&0&0&0&1&1&1&1&1\\0&1&0&0&0&0&1&0&1&2&2&1\\0&0&1&0&0&0&1&1&0&1&2&2\\0&0&0&1&0&0&1&2&1&0&1&2\\0&0&0&0&1&0&1&2&2&1&0&1\\0&0&0&0&0&1&1&1&2&2&1&0\end{pmatrix}$$

这里$\boldsymbol{G}$是行正交的(包括自对偶)，所以$\mathscr{X}\subset\mathscr{X}^{\perp}$，但是$\dim(\mathscr{X})=\dim(\mathscr{X}^{\perp})=6$，所以$\mathscr{X}=\mathscr{X}^{\perp}$。 □

问题3.5 定义一个有q个元素的有限域$\mathbb{F}_q$。证明q的形式必为$q=p^s$，其中p为一个素数，$s\geq1$为正整数。验证p是$\mathbb{F}_q$的特征数。 [347]

证明对于任何上述的p和s存在有p^s元素的有限域$\mathbb{F}_q$，并且这个域唯一同构。

证明$\mathbb{F}_q$非零元素的集合$\mathbb{F}_q^*$是循环群$\mathbb{Z}_{q-1}$。

写出$\mathbb{F}_9$的域表。将一个初始元素$\omega\in\mathbb{F}_9$的幂ω^i确定为$\mathbb{F}_3$上的向量。指出表中所有的向量α使得$\alpha^4=e$。

解答 一个域有q个元素的$\mathbb{F}_q$是一个势为q，并有两个符合分配律的交换群运算+和·的集合，很容易验证$\mathrm{char}(\mathbb{F}_q)=p$是一个素数。所以$\mathbb{F}_p\subset\mathbb{F}_q$，$q=\#\mathbb{F}_q=p^s$，其中$s=[\mathbb{F}_q:\mathbb{F}_p]$是$\mathbb{F}_q$的维度，即一个有$p$个元素的域$\mathbb{F}_p$上的向量空间。

现在，令$\mathbb{F}_q^*$，即来自$\mathbb{F}_q$的非零元素的乘积群，包括一个阶为$q-1=\#\mathbb{F}_q^*$的元素。实际上，每个$b\in\mathbb{F}_q^*$有一个有限的阶$\mathrm{ord}(b)=r(b)$；设$r_0=\max[r(b);\ b\in\mathbb{F}_q^*]$，并固定$a\in\mathbb{F}_q^*$，其中$r(a)=r_0$。那么对于所有$b\in\mathbb{F}_q^*$有$r(b)\mid r_0$。接下来，选取$\gamma$，$r(b)$的质因子，并写出$r(b)=\gamma^s\omega$，$r_0=\gamma^t a$，让我们验证$s\geqslant s'$。实际上，$a^{\gamma^t}$阶为$a$，$b^{\omega}$阶为$\gamma^s$，$a^{\gamma^t}b^{\omega}$阶为$\gamma^t a$。所以，如果$s'>s$，我们得到一个阶大于$r_0$的元素。所以，$s\geqslant s'$对于$r(b)$的任何质因子成立，并且$r(b)\mid r(a)$。

然后对于所有的$b\in\mathbb{F}_q^*$有$b^{r(a)}=e$，即多项式$X^{r_0}-e$能被$(X-b)$整除。所以它必然为乘积$\prod_{b\in\mathbb{F}_q^*}(X-b)$。所以$r_0=\#\mathbb{F}_q^*=q-1$，所以$\mathbb{F}_q^*$最生成矩阵为$a$的循环群。

对于每个素数p和正整数s，对于同构而言最多存在一个域$\mathbb{F}_q$，其中$q=p^s$。实际上，如果$\mathbb{F}_q$和$\mathbb{F}_{q'}$是两个这样的域，那么它们对X^q-X的分裂域（在$\mathbb{F}_q$上，基本域）$\mathrm{Spl}(X^q-X)$同构。

$\mathbb{F}_9=\mathbb{F}_3\times\mathbb{F}_3$的元素$\alpha$，其中$\alpha^4=e$，为$e=01$，$\omega^2=1+2\omega=21$，$\omega^3=02$，$\omega^4=2+\omega=12$，其中$\omega=10$。 □

问题3.6 给出长度为N，字母表为$\mathbb{F}_q$的一个循环码的定义。一个循环码的定义零点是什么，为什么它们总是单位$(N,\mathbb{F}_q)$-根？证明三元Hamming$\left[\frac{3^l-1}{2},\ \frac{3^l-1}{2}-s,\ 3\right]$码等同于一个循环码并且确定这个循环码的定义零点。

一个发送者使用三元[13，10，3]Hamming码，故字母表$\mathbb{F}_3=\{0,1,2\}$，奇偶校验矩阵$\boldsymbol{H}$形式为

$$\begin{bmatrix}1&0&1&2&0&1&2&0&1&2&0&1&2\\0&1&1&1&0&0&0&1&1&1&2&2&2\\0&0&0&0&1&1&1&1&1&1&1&1&1\end{bmatrix}$$

348 接收机接收到码字$x=2120110021120$。怎样解码？

解答 由于$g(X)\mid(X^N-1)$，$g(X)$所有的根是单位N次根。令$\gcd(l,q-1)=1$。我们证明Hamming$\left[\frac{q^l-1}{q-1},\ \frac{q^l-1}{q-1}-l\right]$码等同于一个循环码，其中定义零点$\omega=\beta^{q-1}$，$\beta$是初始单位$(q^l-1)/(q-1)$-根。实际上，设$N=(q^l-1)/(q-1)$。分裂域$\mathrm{Spl}(X^N-1)=F_{q^r}$，其中$r=\mathrm{ord}_N(q)=\min[s;\ N\mid(q^s-1)]$。然后由于$(q^l-1)/(q-1)\mid(q^r-1)$，$r=l$，$l$是这样最小的解。所以，$\mathrm{Spl}(X^N-1)=\mathbb{F}_{q^l}$。

如果β是在$\mathbb{F}_{q^l}$上的初始元素，那么$\omega=\beta^{\frac{q^l-1}{q-1}}=\beta^{q-1}$是在$\mathbb{F}_{q^l}$上的初始单位$N$次根。可得$\mathbb{F}_q\times\cdots\times\mathbb{F}_q$的列向量为$\omega^0=e$，$\omega$，$\omega^2$，…，$\omega^{N-1}$并构造一个$l\times N$的校验矩阵$\boldsymbol{H}$。我们希望验证$\boldsymbol{H}$任何两个不同的列是线性独立的。这个正好在定理3.3.14中完成。

那么校验矩阵为$\boldsymbol{H}$的码中，距离$\geqslant 3$，秩$k\geqslant N-l$，$N=(q^l-1)/(q-1)$，Hamming界（满足$N=(q^l-1)/(q-1)$）

$$q^k\leqslant q^N\left[\sum_{0\leqslant m\leqslant E}\binom{N}{m}(q-1)^m\right]^{-1},\quad 其中\ E=\left\lfloor\frac{d-1}{2}\right\rfloor \tag{3.6.3}$$

表明$d=3$，$k=N-l$。所以，校验矩阵为$\boldsymbol{H}$的循环码等同于Hamming码。

为了解题目中的码，计算校验子$x\boldsymbol{H}^T=202=2\cdot(101)$，表明错误在第6位。所以，$x-2e^{(6)}=y+e^{(6)}$，正确的码字是$c=2120120021120$。 □

问题3.7 计算生成多项式为$g(X)=X^4+X+1$，校验多项式为$h(X)=X^6+X^3+X+1$

的循环码的秩和最小距离。现在令 ω 为域 $\mathbb{F}_8$ 上 $g(X)$ 的一个根。我们接收到码字 $r(X)=X^5+X^4+X(\bmod X^7-1)$。验证 $r(\omega)=\omega^4$，并用最小距离译码来解 $r(X)$。

解答 长度为 N 的一个循环码 $\mathscr{C}$ 有生成多项式 $g(X)\in\mathbb{F}_2[X]$ 和奇偶校验多项式 $h(X)\in\mathbb{F}_2[X]$，其中 $g(X)h(X)=X^N-1$。回想如果 $g(X)$ 度数为 k，即 $g(X)=a_0+a_1X+\cdots+a_kX^k$，$a_k\neq0$，那么 $g(X)$，$Xg(X)$，…，$X^{N-k-1}g(X)$ 构成了 $\mathscr{C}$ 的一个基底。特别地，$\mathscr{C}$ 的秩等于 $N-k$。在这个问题中，$k=3$，$rank(\mathscr{C})=4$。 349

如果 $h(X)=b_0+b_1X+\cdots+b_{N-k}X^{N-k}$，那么码 $\mathscr{C}$ 的奇偶校验矩阵 $\boldsymbol{H}$ 为

$$\underbrace{\begin{bmatrix} b_{N-k} & b_{N-k-1} & \cdots & b_1 & b_0 & 0 & \cdots & 0 & 0 \\ 0 & b_{N-k} & b_{N-k-1} & \cdots & b_1 & b_0 & \cdots & 0 & 0 \\ & & & & & & & & \\ 0 & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & & \\ 0 & 0 & \cdots & 0 & b_{N-k} & b_{N-k-1} & \cdots & b_1 & b_0 \end{bmatrix}}_{N}$$

$\mathscr{C}$ 的码字在矩阵 $\boldsymbol{H}$ 的列之间是线性相关的关系。一个线性码 $\mathscr{C}$ 的最小距离 $d(\mathscr{C})$ 是一个码字的最小非零重量。

在这个问题中，$N=7$，并且

$$\boldsymbol{H}=\begin{bmatrix} 1 & 0 & 1 & 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & 1 & 1 & 0 \\ 0 & 0 & 1 & 0 & 1 & 1 & 1 \end{bmatrix}$$

没有零列 $\Rightarrow$ 没有重量为 1 的码字

没有重复列 $\Rightarrow$ 没有重量为 2 的码字

所以，$d(\mathscr{C})=3$。实际上，$\mathscr{C}$ 等同于 Hamming 原始[7，4]码。

由于 $g(X)\in\mathbb{F}_2[X]$ 是不可约的，码字 $\mathscr{C}\subset\mathfrak{R}=\mathbb{F}_2[X]/(X^7-1)$ 是由 ω 定义的循环码。域 $\mathbb{F}_8$ 非零元素的乘积循环群 $\mathbb{F}_8^*\simeq\mathbb{Z}_7^*$ 是

$$\begin{aligned} \omega^0&=1\\ \omega&\\ \omega^2&\\ \omega^3&=\omega+1\\ \omega^4&=\omega^2+\omega\\ \omega^5&=\omega^3+\omega^2=\omega^2+\omega+1,\\ \omega^6&=\omega^2+\omega^2+\omega=\omega^2+1,\\ \omega^7&=\omega^3+\omega=1 \end{aligned}$$

接下来，$r(\omega)$ 的值为

$$\begin{aligned} r(\omega)&=\omega+\omega^4+\omega^5\\ &=\omega+(\omega+1)+(\omega^2+\omega+1)\\ &=\omega^2+\omega\\ &=\omega^4 \end{aligned}$$ 350

如要求的那样。令 $c(X)=r(X)+X^4\bmod(X^7-1)$。那么 $c(\omega)=0$，即 $c(X)$ 是一个码字。由于 $d(\mathscr{C})=3$，码是 1-错误纠正码。我们刚找到一个码字 $c(X)$ 与 $r(X)$ 距离为 1。那么 $r(X)$ 可写为

$$c(X) = X + X^2 + X^3 + X^5 \bmod (X^7 - 1)$$

并且在最小距离译码准则下用 $c(X)$ 译码。 □

问题 3.8 如果$\mathscr{X}$是一个线性$[N, k]$码，定义它的重量枚举多项式 $W_{\mathscr{X}}(s, t)$。证明：

(a) $W_{\mathscr{X}}(1, 1) = 2^k$。

(b) $W_{\mathscr{X}}(0, 1) = 1$。

(c) $W_{\mathscr{X}}(1, 0)$的值为 0 或 1。

(d) $W_{\mathscr{X}}(s, t) = W_{\mathscr{X}}(t, s)$当且仅当 $W_{\mathscr{X}}(1, 0) = 1$。

解答 如果 $x \in \mathscr{X}$，$\mathscr{X}$的重量 $w(x)$为 $w(x) = \#\{x_i : x_i = 1\}$，定义重量枚举多项式

$$W_{\mathscr{X}}(s, t) = \sum A_j s^j t^{N-j} \tag{3.6.4}$$

其中 $A_j = \#\{x \in \mathscr{X} : w(x) = j\}$。那么：

(a) $W_{\mathscr{X}}(1, 1) = \#\{x : x \in \mathscr{X}\} = 2^{\dim \mathscr{X}} = 2^k$。

(b) $W_{\mathscr{X}}(0, 1) = A_0 = \#\{\mathbf{0}\} = 1$；由于$\mathscr{X}$是子空间，$\mathbf{0} \in \mathscr{X}$。

(c) $W_{\mathscr{X}}(0, 1) = 1 \Leftrightarrow A_N = 1$，即 $11\cdots1 \in \mathscr{X}$，$W_{\mathscr{X}}(1, 0) = 0 \Leftrightarrow A_N = 0$，即 $11\cdots1 \notin \mathscr{X}$

(d) 由(b)可知 $W_{\mathscr{X}}(s, t) = W_{\mathscr{X}}(t, s) \Rightarrow W(0, 1) = W(1, 0) \Rightarrow W_{\mathscr{X}}(1, 0) = 1$ 由(b)，所以，如果 $W_{\mathscr{X}}(1, 0) = 1$，那么

$$\begin{aligned} \#\{x \in \mathscr{X} : w(x) = j\} &= \#\{x + 11\cdots1 : x \in \mathscr{X}, w(x) = j\} \\ &= \#\{y \in \mathscr{X} : w(y) = N - j\} \end{aligned}$$

并且对于所有的 j，$W_{\mathscr{X}}(1, 0) = 1$ 表明 $A_{N-j} = A_j$。所以，$W_{\mathscr{X}}(s, t) = W_{\mathscr{X}}(t, s)$。 □

问题 3.9 描述 MacWilliams 恒等式，它连结了码$\mathscr{X}$和它的对偶$\mathscr{X}^\perp$的重量枚举多项式。证明长度为 $N = 2^r - 1$ 的二进制 Hamming 码$\mathscr{H}_{N,r}$的重量枚举多项式等于

$$W_{\mathscr{H}_r^\perp}(z) = \frac{1}{2^r}\left[(1+z)^{2^r-1} + (2^r - 1)(1 - z^2)^{(2^r-1)/2}(1 - z)\right] \tag{3.6.5}$$

解答 (只解答第二部分)令 A_i 为重量 i 的码字的数目。考虑奇偶校验矩阵 $\boldsymbol{H}$ 的 $i-1$ 列。有三种可能性：

351 (a) 这些列的和为 $\mathbf{0}$。

(b) 这些列的和是其中一个选好的列。

(c) 这些列的和是其中一个剩余的列。

可能(a)发生 A_{i-1} 次；可能(b)发生 iA_i 次，因为选取 $i-1$ 列的组合可以通过把任何重量 i 的码字去掉其中一个非零成员得到。接下来，观察到可能(b)发生 $(N-(i-2))A_{i-2}$ 次。实际上，这种组合可以通过一个重量 $i-2$ 的码字加上剩余 $N-(i-2)$ 列的任何一列得到。然而，我们有 $\binom{N}{i-1}$ 种方法选取 $i-1$ 列。所以，

$$iA_i = \binom{N}{i-1} - A_{i-1} - (N - i - 2)A_{i-2} \tag{3.6.6}$$

如果 $i > N+1$ 是很显然的。如果我们两边乘以 z^{i-1}，在对 i 求和，我们得到一个常微分方程

$$A'(z) = (1+z)^N - A(z) - NzA(z) - z^2 A'(z) \tag{3.6.7}$$

由于 $A(0) = 1$，常微分方程的特殊解为

$$A(z) = \frac{1}{N-1}(1+z)^N + \frac{1}{N+1}(1-z)^{(N-1)/2}(1-z)^{(N+1)/2} \tag{3.6.8}$$

等同于式(3.6.5)。 □

问题 3.10 令$\mathscr{X}$为在$\mathbb{F}_q$上长度为 N、秩为 k 的一个线性码。令 A_i 为重量为 i 的码字的数

目，$i=0,\cdots,N$。定义$\mathscr{X}$的重量枚举多项式为

$$W(\mathscr{X},z)=\sum_{0\le i\le N}A_iz^i$$

令$\mathscr{X}^\perp$为$\mathscr{X}$的对偶码。验证

$$W(\mathscr{X}^\perp,z)=2^{-k}(1+z)^NW\left(\mathscr{X},\frac{1-z}{1+z}\right) \tag{3.6.9}$$

（提示：考虑$g(u)=\sum_{v\in\mathscr{X}}(-1)^{\langle u,v\rangle}z^{w(v)}$，其中$w(v)$为向量$v$在$\mathscr{X}$上的平均重量。）

验证如果$\mathscr{X}$能纠正至少一个错误，那么$\mathscr{X}^\perp$的码字平均重量为$N/2$。

将(3.6.9)应用到 Hamming 码的枚举多项式，

$$W(\mathscr{X}_{\mathrm{Ham}},z)=\frac{1}{N+1}(1+z)^N+\frac{N}{N+1}(1+z)^{(N-1)/2}(1-z)^{(N+1)/2} \tag{3.6.10}$$

以得到极差码的枚举多项式：

$$W(\mathscr{X}_{\mathrm{dsp}},z)=2^{-k}2^N/2^l+2^{-k}(2^l-1)/2^l\times 2^Nz^{2^{l-1}}=1+(2^l-1)z^{2^{l-1}}$$ 352

解答 一个生成矩阵为$\boldsymbol{G}$、奇偶校验矩阵为$\boldsymbol{H}$的线性码$\mathscr{X}$的对偶码$\mathscr{X}^\perp$定义为生成矩阵为$\boldsymbol{H}$的线性码。如果$\mathscr{X}$是一个$[N,k]$码，$\mathscr{X}^\perp$则是一个$[N,N-k]$码，其奇偶校验矩阵为$\boldsymbol{G}$。

同样地，$\mathscr{X}^\perp$是一个在$\mathbb{F}_2^N$上的线性子空间上形成的码，点乘正交于$\mathscr{X}$

$$\langle \boldsymbol{x},\boldsymbol{y}\rangle=\sum_{1\le i\le N}x_iy_i,\quad \boldsymbol{x}=x_1\cdots x_N,\quad \boldsymbol{y}=y_1\cdots y_N$$

由定义得

$$W(\mathscr{X},z)=\sum_{u\in\mathscr{X}}z^{w(u)},W(\mathscr{X}^\perp,z)=\sum_{v\in\mathscr{X}^\perp}z^{w(v)}$$

根据提示，考虑平均

$$\frac{1}{\#\mathscr{X}}\sum_{u\in\mathscr{X}}g(u),\quad 其中\ g(u)=\sum_v(-1)^{\langle u,v\rangle}z^{w(v)} \tag{3.6.11}$$

然后将(3.6.11)写成

$$\frac{1}{\#\mathscr{X}}\sum_v z^{w(v)}\sum_{u\in\mathscr{X}}(-1)^{\langle u,v\rangle} \tag{3.6.12}$$

注意到当$v\in\mathscr{X}^\perp$，和$\sum_{u\in\mathscr{X}}(-1)^{\langle u,v\rangle}=\#\mathscr{X}$。另一方面，当$v\notin\mathscr{X}^\perp$，存在$u_0\in\mathscr{X}$使得$\langle u_0,v\rangle\neq 0$（即$\langle u_0,v\rangle=1$）。所以，如果$v\notin\mathscr{X}^\perp$，那么通过变量替换$u\mapsto u+u_0$，我们得到

$$\begin{aligned}\sum_{u\in\mathscr{X}}(-1)^{\langle u,v\rangle}&=\sum_{u\in\mathscr{X}}(-1)^{\langle u+u_0,v\rangle}\\&=(-1)^{\langle u_0,v\rangle}\sum_{u\in\mathscr{X}}(-1)^{\langle u,v\rangle}=-\sum_{u\in\mathscr{X}}(-1)^{\langle u,v\rangle}\end{aligned}$$

所以在这种情况下$\sum_{u\in\mathscr{X}}(-1)^{\langle u,v\rangle}=0$。我们得到结论，(3.6.11)的和等于

$$\frac{1}{\#\mathscr{X}}\sum_{v\in\mathscr{X}^\perp}z^{w(v)}(\#\mathscr{X})=W(\mathscr{X}^\perp,z) \tag{3.6.13}$$

另一方面，对于$u=u_1\cdots u_N$，

$$\begin{aligned}g(u)&=\sum_{v_1\cdots v_N}\prod_{1\le i\le N}z^{w(v_i)}(-1)^{u_iv_i}=\prod_{1\le i\le N}\sum_{v_i=0,1}z^{w(v_i)}(-1)^{u_iv_i}\\&=\prod_{1\le i\le N}[1+z(-1)^{u_i}]\end{aligned} \tag{3.6.14}$$ 353

在这里，当 $a=0$ 时 $w(a)=0$，当 $a=1$ 时 $w(a)=1$，(3.6.14)的右边等于

$$(1-z)^{w(a)}(1-z)^{N-w(a)}$$

所以，(3.6.11)的另一种表达为

$$\frac{1}{\#\mathscr{X}}(1+z)^N\sum_{x\in\mathscr{X}}\left(\frac{1-z}{1+z}\right)^{w(x)}=\frac{1}{\#\mathscr{X}}(1+z)^N W\left(\mathscr{X},\frac{1-z}{1+z}\right) \tag{3.6.15}$$

将(3.6.13)和(3.6.15)等同得到

$$\frac{1}{\#\mathscr{X}}(1-z)^N W\left(\mathscr{X},\frac{1-z}{1+z}\right)=W(\mathscr{X}^{\perp},z) \tag{3.6.16}$$

于是由于 $\#\mathscr{X}=2^k$ 就有了所要求的等式。

接下来，对(3.6.16)在 $z=1$ 上微分，右边等于

$$\sum_{1\le i\le N}iA_i(\mathscr{X}^{\perp})=(\#\mathscr{X}^{\perp})\times(\text{在}\mathscr{X}^{\perp}\text{上的平均重量})$$

另一方面，左边我们有

$$\frac{\mathrm{d}}{\mathrm{d}z}\left\{\frac{1}{\#\mathscr{X}}\sum_{0\le i\le N}A_i(\mathscr{X})(1-z)^i(1+z)^{N-i}\right\}\Bigg|_{z=1}$$

$$=\frac{1}{\#\mathscr{X}}(N2^{N-1}-A_1(\mathscr{X})2^{N-1})(\text{只有}i=0,1\text{时有贡献})$$

$$=\frac{2^N}{\#\mathscr{X}}\frac{N}{2}(A_1(\mathscr{X})=0(\text{当码至少是1-错误纠正码，其中距离}\geq 3))$$

现在考虑到

$$(\#\mathscr{X})\times(\#\mathscr{X}^{\perp})=2^k\times 2^{N-k}=2^N$$

得到等式

$$\text{在}\mathscr{X}^{\perp}\text{上的平均重量}=\frac{N}{2}$$

极长码的数学多项式由替代可以得到。在这里，平均长度为 $(2^r-1)/2$。 □

问题 3.11 描述长度为15，设计距离为5的二进制狭义BCH码，并找到生成多项式。译码100000111000100。

354

解答 取长度为15，设计距离为5的二进制狭义BCH码，我们有 $\mathrm{Spl}(X^{15}-1)=\mathbb{F}_2^4=\mathbb{F}_{16}$。我们知道 X^4+X+1 是在 $\mathbb{F}_{16}$ 上的一个初始多项式。令 ω 为 X^4+X+1 的一个根，那么

$$M_1(X)=X^4+X+1,M_3(X)=X^4+X^3+X^2+X+1$$

X 的生成多项式 $g(X)$ 为

$$g(X)=M_1(X)M_3(X)=X^8+X^7+X^6+X^4+1$$

取 $g(X)$ 作为一个码字的例子。引入两个错误——在位置4和12——我们有

$$u(X)=X^{12}+X^4+X^7+X^6+1$$

利用 $\mathbb{F}_{16}$ 的域表，我们得到

$$u_1=u(\omega)=\omega^{12}+\omega^6+\omega^7+\omega^6+1=\omega^4$$

以及

$$u_3=u(\omega^3)=\omega^{36}+\omega^{18}+\omega^{15}+1=\omega^7+\omega^3+1=\omega^4$$

由于 $u_1\neq 0$ 和 $u_1^3=\omega^{12}=\omega^3\neq u_3$，推测 ≥ 2 个错误出现。计算定位多项式

$$l(X)=1+\omega^5X+(\omega^{12}+\omega^{12})X^2$$

将 $1,\omega,\cdots,\omega^{14}$ 代入 $l(X)$，验证 ω^5 和 ω^{11} 为根。这证明如果恰好2个错误发生它们的位置在4和12，那么发送码字为100010111000000。 □

问题 3.12 对于一个码字 $x=x_1\cdots x_N\in\mathbb{F}_2^N$，重量 $w(x)$ 是非零数位的数目：$w(x)=\#\{x_i$

$x_i \neq 0\}$。对于一个线性$[N, k]$码$\mathscr{X}$，令A_i为重量为i的码字的数目$(0 \leqslant i \leqslant N)$。定义重量枚举多项式$W(\mathscr{X}, z)=\sum_{i=0}^{N} A_i z^i$。验证如果我们在错误概率为$p$的二进制对称信道上用$\mathscr{X}$，无法检测到错误码字的概率为$(1-p)^N\left(W\left(\mathscr{X}, \frac{p}{1-p}\right)-1\right)$。

解答　假设我们发送了零码字0，那么错误概率为

$$E=\sum_{x \in \mathscr{X} \setminus 0} P(x \mid 发送)=\sum_{i \geq 1} A_i p^i (1-p)^{N-i}=$$

$$(1-p)^N\left(\sum_{i \geq 0} A_i\left(\frac{p}{1-p}\right)^i-1\right)=(1-p)^N\left(W\left(\mathscr{X}, \frac{p}{1-p}\right)-1\right) \qquad \square$$

355

问题 3.13　令$\mathscr{X}$为一个二进制线性$[N, k, d]$码，重量枚举多项式为$W_{\mathscr{X}}(s)$。找到下面关于$W_{\mathscr{X}}(s)$的重量枚举表达式：

(i) 包含所有重量为偶数的码字$x \in \mathscr{X}$的子码$\mathscr{X}^{+} \subset \mathscr{X}$。

(ii) $\mathscr{X}$的奇偶校验扩展$\mathscr{X}^{+}$。

证明如果d是偶数，那么存在一个$[N, k, d]$码使得每个码字重量为偶数。

解答　(i) $\mathscr{X}$的所有重量为偶数的码字属于子码$\mathscr{X}^{+}$，所以，

$$W_{\mathscr{X}^{+}}(s)=\frac{1}{2}\left[W_{\mathscr{X}}(s)+W_{\mathscr{X}}(-s)\right]$$

(ii)很显然，$\mathscr{X}^{+}$上的重量枚举多项式的所有非零系数对应2的偶数幂，且$A_{2i}(\mathscr{X}^{+})=A_{2i}(\mathscr{X})+A_{2i-1}(\mathscr{X})$，$i=1, 2, \cdots$。因此

$$W_{\mathscr{X}^{+}}(s)=\frac{1}{2}\left[(1+s) W_{\mathscr{X}}(s)+(1-s) W_{\mathscr{X}}(-s)\right]$$

如果$\mathscr{X}$是二进制$[N, k, d]$码，那么首先把$\mathscr{X}$截断成$\mathscr{X}^{-}$，然后取奇偶校验扩展$(\mathscr{X}^{-})^{+}$。这使得k和d不变(如果d为偶数)，使得所有码字重量为偶数。 $\square$

问题 3.14　验证多项式$X^4+X^3+X^2+X+1$和X^4-X+1在$\mathbb{F}_2$上不可约。这些多项式在$\mathbb{F}_2$上是初始多项式吗？多项式X^5+X+1，X^4+X^3+1，X^4+X^2+1又是怎样的？

解答　由于在$X=0$或$X=1$上$X^4+X^3+X^2+X+1$和X^4+X+1都不会趋近于零，所以它们不会被X或$X-1$整除。它们也不会被X^2+X+1整除，是度数为2的唯一不可约多项式，或者被X^3+X+1，X^3+X^2+1整除，是度数为3的唯一不可约多项式。所以，它们是不可约的。

多项式$X^4-X^3-X^2-X+1$不可能是初始多项式，因为它能整除X^5-1。让我们验证X^4+X+1是初始的。选取$\mathbb{F}_2[X]/(X^4+X+1)$并利用$\mathbb{F}_2^4$的域表。分圆陪集为$\{\omega, \omega^2, \omega^4, \omega^8\}$(由于$\omega^{16}=\omega$)。所以初始多项式$M_\omega(X)$为

$$\begin{aligned}
&(X-\omega)(X-\omega^2)(X-\omega^4)(X-\omega^8)\\
=&X^4-(\omega+\omega^2+\omega^4+\omega^8)X^3\\
&+(\omega\omega^2+\omega\omega^4+\omega\omega^8+\omega^2\omega^4+\omega^2\omega^8+\omega^4\omega^8)X^2\\
&-(\omega\omega^2\omega^4+\omega\omega^2\omega^8+\omega\omega^4\omega^8+\omega^2\omega^4\omega^8)x+\omega\omega^2\omega^4\omega^8\\
=&X^4-(\omega+\omega^2+\omega^4+\omega^8)X^3+(\omega^3\omega^5+\omega^6+\omega^9+\omega^{10}+\omega^{12})X^2\\
&-(\omega^7+\omega^{11}+\omega^{13}+\omega^{14})X+\omega^{15}=X^4+X+1
\end{aligned}$$

356

X^4+X+1的阶为15；其他阶为15的初始多项式为X^4+X^3+1和X^4+X+1。所以，度数为4的初始多项式只能为X^4+X+1。类似地，度数为3的初始多项式只能为X^3+X+1，X^3+X^2+1，两者的阶皆为7。 $\square$

问题 3.15 假设用到一个二进制缩短 BCH 码，长度为 15，设计距离为 5，接收到的码字为 $X^{11}+X^3+X^2+X+1$。它是怎么解码的？如果接收到的码字是 $X^{11}+X^{10}+X^5+X^3+X^2+X+1$，错误有多少？

解答 假设接收到的码字为

$$r(X)=X^{13}+X^3+X^2+X+1$$

并令 ω 为 $\mathbb{F}_{16}$ 的一个初始元素。那么

$$\begin{aligned}s_1&=r(\omega)=\omega^{13}+\omega^3+\omega^2+\omega+e\\&=0111+0110+0011+0010+0001=0001=e\\s_3&=r(\omega^3)=\omega^{39}+\omega^{13}+\omega^{12}+\omega^3+e\\&=0001+0001+1111+1000+0001=0110=\omega^2\end{aligned}$$

观察到 $s_3\neq s_1^3$，所以有两个错误。错误定位多项式为

$$\sigma(X)=e+s_1X+(s_3s_1^{-1}+s_1^2)X^2=e+X+(\omega^2+e)X^2=e+X+\omega^{12}X^2$$

验证它的根，$\omega^0=e$，ω，ω^2，ω^3，ω^5，ω^7，ω^8 不是，ω^{10} 是。然后相除：

$$(\omega^{12}X^2+X+e)/(X+\omega^{10})=\omega^{10}X+\omega^8=\omega^{12}(X+\omega^{13})$$

并确定第二个根：ω^{13}。所以，错误发生在位置 $15-7=8$ 和 $15-13=2$ 上。解码：

$$r(X)\to X^{13}+X^3+X^2+X+1 \qquad \square$$

问题 3.16 证明长度为 23、生成多项式为 $1+X+X^5+X^6+X^7+X^9+X^{11}$ 的码有最小距离 7，并且是完美的。（提示：如果 $g^{rev}(X)=X^{11}g(1/X)$ 是 $g(X)$ 的逆，那么 $X^{23}+1=(X+1)g(X)g^{rev}(X)\bmod 2$。）

解答 首先，验证这个码是 BCH，设计长度为 5。由引理 3.1.5 fresher's dream 得到，如果 ω 是多项式 $f(X)\in\mathbb{F}_2[X]$ 的一个根，那么 ω^2 也是。所以，如果 ω 是 $g(X)=1+X+X^5+X^6+X^7+X^9+X^{11}$ 的一个根，那么 ω，ω^2，ω^4，ω^8，ω^{16}，ω^{32}，ω^{9}，ω^{18}，ω^{36}，ω^{72}，ω^{45}，ω^{90} 也是，这产生了设计序列 $\{\omega,\omega^2,\omega^4,\omega^{8}\}$。由 BCH 界（定理 2.5.33 和定理 3.2.9）可知，由 $g(X)$ 产生的循环码 $\mathscr{C}$ 的距离 ≥ 5。

357

接下来，奇偶校验扩展 $\mathscr{C}^{+}$ 是自正交的。为了验证这点，我们只需验证 $\mathscr{C}^{+}$ 的生成矩阵的任何两行都是正交的。这些可由中级码字表示：

$$(X^ig(X)\,|\,1)\quad 和\quad (X^jg(X)\,|\,1)$$

它们的点乘等于

$$\begin{aligned}&1+(X^ig(X))(X^jg(X))=1+\sum g_{i+r}g_{j+r}\\&=1+\sum_r g_{i+r}g^{rev}_{11-j+r}\\&=1-X^{(i+j)/}\text{ 的系数在}\underbrace{g(X)\times g^{rev}(X)}_{1+\cdots+X^{22}}\\&=1-1=0\end{aligned}$$

我们得出结论

$\mathscr{C}^{+}$ 上的任何两个码字都是点乘正交的。

接下来，观察到 $\mathscr{C}^{+}$ 上的所有码字重量都能被 4 整除。实际上，经检验，$\mathscr{C}^{+}$ 的生成矩阵的所有行 $(X^ig(X)\,|\,1)$ 重量都为 8。然后，通过归纳包含在和中的列的数目，得到如果 $x\in\mathscr{C}^{+}$ 并且 $g^{(i)}\sim(X^ig(X)\,|\,1)$ 是 $\mathscr{C}^{+}$ 的生成矩阵的一个行，那么

$$w(g^{(i)}+x)=w(g^{(i)})+w(x)-2w(g^{(i)}\wedge x) \tag{3.6.17}$$

其中$(g^{(i)}\wedge x)_l=\min[(g^{(i)})_l,\ x_l]$，$l=1,\cdots,24$。我们知道 8 整除 $w(g^{(i)})$，此外，通过归纳假设，4 整除 $w(x)$。接下来，由(3.6.17)，$w(g^{(i)}\wedge x)$是偶数，所以 $2w(g^{(i)}\wedge x)$能被 4 整除。那么左边 $w(g^{(i)}+x)$能被 4 整除。

所以，$\mathscr{X}^{+}$的距离为 8，因为它$\geqslant 5$并被 4 整除。(很容易看出它不可能$\geqslant 8$，否则它会变成 12。)原始码$\mathscr{X}$的距离等于 7。

码$\mathscr{X}$是完美 3-错误纠正码，由于在F_2^{23}上 3-球的容量等于

$$\binom{23}{0}+\binom{23}{1}+\binom{23}{2}+\binom{23}{3}=1+23+253+1\,771=2\,048=2^{11}$$

并且

$$2^{11}\times 2^{12}=2^{23}$$

这里很显然，12 代表秩，23 代表长度。 □ 358

问题 3.17 利用 MacWilliams 恒等式证明一个距离为 d 的 q-进制 MDS 码的重量分布是

$$\begin{aligned}A_i&=\binom{N}{i}\sum_{0\leqslant j\leqslant i-d}(-1)^j\binom{i}{j}(q^{i-d+1-j}-1)\\&=\binom{N}{i}(q-1)\sum_{0\leqslant j\leqslant i-d}(-1)^j\binom{i-1}{j}q^{i-d-j},d\leqslant i\leqslant N\end{aligned}$$

(提示：为了求解，可以

(a) 写出标准的 MacWilliams 恒等式。

(b) 交换$\mathscr{X}$和$\mathscr{X}^{\perp}$。

(c) 变换 $z\to z^{-1}$。

(d) 乘以 z^{r}。

(e) 取微分 $\mathrm{d}^{r}/\mathrm{d}z^{r}$，$0\leqslant r\leqslant k$(等于 $d(\mathscr{X}^{\perp})-1$)。

利用 Leibniz 法则

$$\frac{\mathrm{d}^{r}}{\mathrm{d}z^{r}}[f(z)g(z)]=\sum_{0\leqslant j\leqslant r}\binom{r}{j}\left(\frac{\mathrm{d}^{j}}{\mathrm{d}z^{j}}f(z)\right)\left(\frac{\mathrm{d}^{r-j}}{\mathrm{d}z^{r-j}}g(z)\right) \tag{3.6.18}$$

利用 $d(\mathscr{X})=N-k+1$ 和 $d(\mathscr{X}^{\perp})k+1$，得到只包含 A_{N-k+1}，…，A_{N-r}的简化等式。接下来，确定 A_{N-k-1}，…，A_{N-r}，变换 r，直到 A_N。)

解答 MacWilliams 恒等式为

$$\sum_{0\leqslant i\leqslant N}A_i z^i=\frac{1}{q^k}\sum_{0\leqslant i\leqslant N}A_i^{\perp}(1-z)^i[1+(q-1)z]^{N-i}$$

交换$\mathscr{X}$和$\mathscr{X}^{\perp}$，变换 $z\mapsto z^{-1}$并乘以 z^{N}。微分 $r\leqslant k$ 次并用 $z=1$ 替换，

$$\frac{1}{q^k}\sum_{0\leqslant i\leqslant N-r}\binom{N-i}{r}A_i=\frac{1}{q^k}\sum_{0\leqslant i\leqslant r}A_i^{\perp}\binom{N-i}{N-r} \tag{3.6.19}$$

(Leibniz 法则(3.6.18)在这里用到。)等式(3.6.19)是起始点。对于一个 MDS 码，$A_i=A_i^{\perp}=1$，并且

$$A_i=0,1\leqslant i\leqslant N-k(=d-1),\quad A_i^{\perp}=0,1\leqslant i\leqslant k(=d^{\perp}-1)$$

然后，

$$\binom{N}{r}\frac{1}{q^k}=\frac{1}{q^k}\sum_{i=N-k+1}^{N-r}\binom{N-i}{r}A_i=\frac{1}{q^r}\binom{N}{N-r}=\frac{1}{q^r}\binom{N}{r} \tag{3.6.19}$$ 359

即

$$\sum_{i=N-d+1}^{N-r}\binom{N-i}{r}A_i=\binom{N}{r}(q^{d-r}-1)$$

当 $r=k$，可得 $0=0$，当 $r=k-1$，有

$$A_{N-d+1}=\binom{N}{k-1}(q-1) \tag{3.6.20}$$

当 $r=k-2$ 则有

$$\binom{k-1}{k-2}A_{N-d+1}+A_{N-d+2}=\binom{N}{k-2}(q^2-1)$$

等等。这是一个关于 A_{N-d+1}，…，A_{N-1} 的三角形方程组。当改变 r 时，我们可以求得 A_{N-d+1}，…，A_{N-1}。得到的结果是

$$\begin{aligned}A_i&=\binom{N}{i}\sum_{0\leqslant j\leqslant i-d}(-1)^j\binom{i}{j}(q^{i-d+1-j}-1)\\&=\binom{N}{i}\left[\sum_{0\leqslant j\leqslant i-d}(-1)^j\binom{i-1}{j}(qq^{i-d-j}-1)\right.\\&\quad\left.-\sum_{1\leqslant j\leqslant i-d+1}(-1)^{j-1}\binom{i-1}{j-1}(q^{i-d+1-j}-1)\right]\\&=\binom{N}{i}(q-1)\sum_{0\leqslant j\leqslant i-d}(-1)^j\binom{i-1}{j}q^{i-d-j},d\leqslant i\leqslant N\end{aligned}$$

事实上，无需计算就可以得到式(3.6.20)：在秩为 k，距离为 d 的 MDS 码中，任意 $k=N-d+1$ 数字唯一地决定码字。而且，对于任意 $N-d$ 位置的选择，在这些位置上存在 q 个包含数位 0 的码字。其中一个码字为零码字，剩余的 $q-1$ 个是重量为 d 的码字。因此

$$A_{N-d+1}=A_d=\binom{N}{d}(q-1)$$ □

问题 3.18 证明 Krawtchuk 多项式 $K_k(i)$ 的下列特性。

(a) 对于所有 q：$(q-1)^i\binom{N}{i}K_k(i)=(q-1)^k\binom{N}{k}K_i(k)$。

(b) 当 $q=2$：$K_k(i)=(-1)^kK_k(N-i)$。

(c) 当 $q=2$：$K_k(2i)=K_{N-k}(2i)$。

368

解答 存在下式

$$K_k(i)=\sum_{0\leqslant j\leqslant\min(k,i)}\binom{i}{j}\binom{N-i}{k-j}(-1)^j(q-1)^{k-j}$$

然后：

(a) 下面的简单方程是正确的：

$$(q-1)^i\binom{N}{i}K_k(i)=(q-1)^k\binom{N}{k}K_i(k)$$

（当交换 $i\leftrightarrow k$ 时，所有的被加数都是不敏感的。）

当 $q=2$，可得 $\binom{N}{i}K_k(i)=\binom{N}{k}K_i(k)$；特别地，

$$\binom{N}{i}K_0(i)=\binom{N}{0}K_i(0)=K_i(0)$$

(b) 同样当 $q=2$ 时：$K_s(i)=(-1)^i K_s(N-i)$（当交换 $i\leftrightarrow i-j$ 后，式子同样是简单的）。

(c) 因此，当 $q=2$：$\binom{N}{2i}K_s(2i)=\binom{N}{k}K_{2i}(k)$，并且等于下式

$$(-1)^k\binom{N}{N-k}K_{2i}(N-k)=\binom{N}{2i}K_{s-k}(2i)，即$$

$$K_s(2i)=K_{s-k}(2i)$$ □

问题 3.19 什么是单位的$(n,\mathbb{F}_q)$根？说明单位的$(n,\mathbb{F}_q)$根的集合$\mathbb{U}^{n,q}$形成一个循环群。如果n和q互质时，证明$\mathbb{U}^{n,q}$的阶等于n。找出最小的s，并满足$\mathbb{U}^{n,q}\subset\mathbb{F}_{q^s}$。

定义单位的原始$(n,\mathbb{F}_q)$根。当n和q互质时，决定单位的原始$(n,\mathbb{F}_q)$根的序号。如果ω是单位原始的$(n,\mathbb{F}_q)$根，试找出满足$\omega\in\mathbb{F}_{q^\ell}$的最小$\ell$。

找出$\mathbb{F}_q$中能够表示为$\mathbb{F}_q$中向量的所有元素代表。找出单位中所有的$(4,\mathbb{F}_3)$根并且是$\mathbb{F}_3$上的向量。

解答 已知任意一个次数为2的不可约多项式的任何根，且满足在域$\mathbb{F}_3=\{0,1,2\}$上，则根均属于$\mathbb{F}_9$。取多项式$f(X)=X^2+1$，令它的根表示为α（两者中的任意一个）。则$\mathbb{F}_9$的所有元素可以表示为$a_0+a_1\alpha$，其中$a_0,a_1\in\mathbb{F}_3$。事实上

$$\mathbb{F}_9=\{0,1,\alpha,1+\alpha,2+\alpha,2\alpha,1+2\alpha,2+2\alpha\}$$ 361

另一个方法表示为：已知$X^8-1=\prod_{1\le i\le 8}(X-\zeta^i)$在域$F_9$中，其中$\zeta$是单位的原始$(8,\mathbb{F}_9)$根，并且可以表示为循环多项式形式，即$X^8-1=Q_1(X)Q_2(X)Q_4(X)Q_8(X)$。在这里$Q_n(x)=\prod_{\omega\text{是}(n,\mathbb{F}_q)\text{根}}(x-\omega^i)$成立，其中$\omega$是单位的原始$(n,\mathbb{F}_q)$根，令$X^8-1=\prod_{d|8}Q_d(x)$，然后计算

$$Q_1(X)=1+X,Q_2(X)=1+X,Q_4(X)=1+X^2$$

$$Q_8(X)=(X^8-1)/(Q_1(X)Q_2(X)Q_4(X))=(X^8-1)/(X^4-1)=X^4+1$$

因为$3^2=1\bmod 8$，根据定理3.1.53，$Q_8(X)$应该在$\mathbb{F}_3$上被分解为次数为2的不可约多项式$\phi(8)/2=2$的乘积。事实上，

$$Q_8(X)=(X^2+X+2)(X^2+2X+2)$$

令ζ表示为X^2+X+2的根，则它是$\mathbb{F}_9$上次数为8的原始根，并且$\mathbb{F}_9=\mathbb{F}_3(\zeta)$，因此，$\mathbb{F}_9=\{0,\zeta,\zeta^2,\zeta^3,\zeta^4,\zeta^5,\zeta^6,\zeta^7,\zeta^8\}$，且$\zeta=1+\alpha$。最后，我们给出索引表：

$$\zeta=1+\alpha,\zeta^2=2\alpha,\zeta^3=1+2\alpha,\zeta^4=2$$

$$\zeta^5=2+2\alpha,\zeta^6=\alpha,\zeta^7=2+\alpha,\zeta^8=1$$

因此，次数为4的根有$\zeta^2,\zeta^3,\zeta^5,\zeta^6$。 □

问题 3.20 在域$\mathbb{F}_q$上定义一个长度为N的循环码，试说明在长度为N的循环码和多项式环$\mathbb{F}_q[X]$中的X^N-e因子之间存在双射。

现在考虑二元循环码。如果N是一个奇整数，我们可以找到一个$\mathbb{F}_2$的有限扩展K，它包含了一个单位ω的第N个原始根。试说明长度为N，且存在定义域$\{\omega,\omega^2,\cdots,\omega^{2^\ell-1}\}$的循环码的最小距离至少为$\delta$。试说明若$N=2^\ell-1$且$\delta=3$，则可得到$\mathrm{Hamming}[2^\ell-1,2^\ell-1-\ell,3]$码。

解答 如果$x_1\cdots x_N\in\mathscr{X}$同时意味着$x_2\cdots x_N x_1\in\mathscr{X}$，则线性码$\mathscr{X}\subset\mathbb{F}_q^{\times N}$同时是循环码。根据推论3.3.3，可以确定循环码和X^N-1的元素是双射。

在二元码中，为了简单令$N=7$，在$\mathbb{F}_2^7$中进行因式分解，得到如下因式分解结果。

$$X^7+1=(X+1)(X^3+X+1)(X^3+X^2+1)=(X+1)f_1(X)f_2(X)$$

假定 ω 是 $f_1(X)$ 的根。因为在 $F_2[X]$ 中存在 $f_1(X)^2=f_1(X^2)$，则

$$f_1(\omega)=f_1(\omega^2)=0$$

若存在定义根为 ω 的循环码 $\mathscr{C}$，则它的生成多项式为 $f_1(X)$，检验多项式为 $(X+1)f_2(X)=X^4+X^3+X+1$。这个性质说明了 Hamming 原始码(相当于等价)的特性。当 ω 是 $f_2(X)$ 的根时，情形是相似的(事实上，仅仅反转每个码字)。对于一般情况下的 $N=2^l-1$，我们取出原始元素 $\omega\in F_{2^l}$，和它的最小多项式 $M_\omega(X)$，则 $M_\omega(X)$ 的根为 ω，ω^2，…，$\omega^{2^{l-1}}$，因此 $\deg M_\omega(X)=l$。故存在定义根为 ω 的码，其秩为 $N-l=2^l-1-l$，就如同 Hamming $[2^l-1,\ 2^l-l-1]_2$ 码。 □

[362]

问题 3.21 写出评论对比 Hamming 码和纠正两差错 BCH 码的解码过程。

解答 为了说明构造 BCH 码的思想，我们首先讨论 Hamming 码。Hamming$[2^l-1,\ 2^l-l-1]$ 码是完美的纠正差错的码，其长度为 $N=2^l-1$。Hamming 码的解码过程表示如下。取字 $y=y_1\cdots y_N$，$N=2^l-1$，构造校验子 $s=yH^{\mathrm{T}}$。如果 $s=0$，利用 y 来解码 y。如果 $s\neq 0$，则 s 在 $\boldsymbol{H}=\boldsymbol{H}_{\mathrm{Ham}}$ 的列中。如果是列 i，利用 $x_i=y+e_i$ 解码 y，其中 $e_i=0\cdots010\cdots0$（1 在第 i 位，其他位是 0）。

我们可以尝试利用下列想法纠正超过一个差错(从二开始)。取出奇偶校验矩阵的 $2l$ 行，并表示为如下形式

$$\widetilde{H}=\begin{pmatrix}\boldsymbol{H}\\ \Pi\boldsymbol{H}\end{pmatrix} \tag{3.6.21}$$

这里通过交换 $\boldsymbol{H}_{\mathrm{Ham}}$ 的列得到 $\Pi\boldsymbol{H}_{\mathrm{Ham}}$（$\Pi$ 是次数为 2^l-1 的排列）。新矩阵 $\widetilde{\boldsymbol{H}}$ 包含 $2l$ 个线性独立行；然后它确定了 $[2^l-1,\ 2^l-1-2l]$ 线性码，校验子是长度为 $2l$ 的字(或者是长度为 l 的一对字)：$y\widetilde{\boldsymbol{H}}^{\mathrm{T}}=(s\ s')$。校验子 $(s,\ s')^{\mathrm{T}}$ 可能是，也可能不是 $\widetilde{\boldsymbol{H}}$ 的列。我们想要能够纠正两个差错的新码，并且解码过程与 Hamming 码解码过程相似。假定发生两个差错，即 y 和码字 x 有两位不相同，表示为 i 和 j。则校验子表示为

$$y\widetilde{\boldsymbol{H}}^{\mathrm{T}}=(s_i+s_j,\ s\Pi_i+s\Pi_j)$$

其中 s_k 是表示 $\boldsymbol{H}$ 中的第 k 列的字。我们构造排列，在知道向量 $(s_i+s_j,\ s_{\Pi i}+s_{\Pi j})$ 情况下，可以得到 i 和 j（或者等价于 s_i，s_j）。即我们能够求解如下等式

$$s_i+s_j=z,\quad s_{\Pi i}+s_{\Pi j}=z' \tag{3.6.22}$$

则对于任意对 $(z,\ z')$，最后它可能表示为存在两个差错情况下的校验子。

自然的猜想是尝试排列 Π，它具有一些代数的意义，例如，$s_{\Pi i}=s_i s_i=(s_i)^2$（坏的选择）或者 $s_{\Pi i}=s_i s_i s_i=(s_i)^3$（好的选择）。或者，一般情况下，$s_{\Pi i}=s_i s_i\cdots s_i$（$k$ 次）。即尝试对乘法取余 $1+X^N$；但是乘法并不能生成域。因为多项式 $1+X^N$ 总是可约的。因此，假定我们构造检验矩阵，表示为

[363]

$$\widetilde{\boldsymbol{H}}=\begin{pmatrix}(1\cdots00) & (1\cdots00)^k\\ & \vdots\\ (1\cdots11) & (1\cdots11)^k\end{pmatrix}$$

然后我们必须处理下列类型的等式

$$s_i+s_j=z,\quad s_i^k+s_j^k=z' \tag{3.6.23}$$

为了求解式(3.6.23)，我们需要了解 Hamming 空间的域结构，即不仅仅需要乘法，也需要相除。长度为 N 的 Hamming 空间的任意域结构与 F_{2^N} 是同构的，且这种结构的具体的实现是

$\mathbb{F}[X]/(c(X))$，即多项式域对一个次数为 N 的不可约多项式 $c(X)$ 取模。这样的多项式总是存在的：它是一个次数为 N 的本原多项式。事实上，公式(3.6.23)的最简单一致系统是

$$s+s'=z, s^3+s'^3=z'$$

它可以简化为单个等式 $zs^2-z^2s+z^3-z'=0$，则问题变成因式分解多项式 $zX^3-z^2X+z^3-z'$。

当 $N=2^l-1$，$l=4$ 时，我们可以得到[15，7，5]码。由于 $\widetilde{\boldsymbol{H}}$ 的列线性独立，则秩为7。需要验证码能够纠正多达两个差错。首先假定我们接收字 $\boldsymbol{y}=y_1\cdots y_{15}$，并在位置 i 和 j 上发生了两个差错，且这两个位置是未知的。为了找到这些位置，需要计算校验子 $\boldsymbol{y}\widetilde{\boldsymbol{H}}^{\mathrm{T}}=(z, z')^{\mathrm{T}}$。已知 z、z' 是长度为4的字；校验子的总长度是8。注意到 $z'\neq z^3$；如果 $z'=z^3$，则只有一个错误发生。写出下列等式

$$z+z'=z, s^3+s'^3=z' \tag{3.6.24}$$

其中 s，s' 是长度为4的字(或者相当于它们的多项式)，且乘法对 $1+X+X^4$ 取模。当发生两个差错时，则式(3.6.24)只有一对解，其中一个向量表示位置 i，另一个向量表示位置 j，且位于矩阵 $\widetilde{\boldsymbol{H}}$ 的上半部(Hamming)列中，并且式(3.6.24)不存在其他的解，这是因为

$$z'=s^3+s'^3=(s+s')(s^2+ss'+s'^2)=z(z^2+ss')$$

即意味着

$$ss'=z^{-1}z'+z^2 \tag{3.6.25}$$

354

根据式(3.6.25)和式(3.6.24)中的第一个等式，得出 s，s' 是下列二次等式的根

$$X^2+zX+(z^{-1}z'+z^2)=0 \tag{3.6.26}$$

(其中 $z^{-1}z'+z^2\neq 0$。)但是在式(3.6.26)LHS多项式不会存在多于两个的不同根(它只会存在两种情况：无根或两个相同根，但是因为假定有两个错误，所以上述情况可以被排除)。当存在一个错误时，可得 $z'=z^3$；在这种情况下 $s=z$ 是唯一的根，且字 z 位于矩阵 $\widetilde{\boldsymbol{H}}$ 上半部的列中。

总结如下，在[15，7]码情况下，解码方案可以表示为：根据接收的字 $\boldsymbol{y}$，构造校验子 $\boldsymbol{y}\widetilde{\boldsymbol{H}}^{\mathrm{T}}=(z, z')^{\mathrm{T}}$，然后有

(i) 如果 z，z' 是零字，则没有错误发生，用 $\boldsymbol{y}$ 自身解码。

(ii) 如果 $z\neq 0$，$z^3=z'$，则发生一个差错，通过在Hamming校验矩阵的列中确定字 z，从而可以找到错误位置。

(iii) 如果 $z\neq 0$，$z^3\neq z'$，构造二次曲面(3.6.24)，如果存在两个相异根 s，s'，则发生两个差错，通过在Hamming校验矩阵的列中确定字 s，s'，则可以找到错误位置。

(iv) 如果 $z\neq 0$，$z^3\neq z'$，且二次曲面(3.6.24)无根，或者 $z=0$，$z'\neq 0$，判定至少存在三个错误。

注意到当 $z\neq 0$，$z^3\neq z'$，且二次曲面(3.6.26)存在一个单根的情况是不会发生的；如果式(3.6.26)有一个根，称为 s，则另一个根 $s'\neq s$，或者 $z=0$ 且只有一个错误发生。

在某些情况下，根据解码过程可知有超过三个差错的情况发生。然而，当存在三个或三个以上的差错时，这个解码过程不可能解出正确的码字。□

365

第 4 章
Information Theory and Coding by Example

信息论的深层主题

在第 4 章中，若同时考虑离散和连续类型的概率分布，这对工作的展开将是很方便的。因此假定所考虑的概率分布均是通过它们的 Radon-Nikodym 推导给出，并且相应的参考测度通常表示为 μ，ν。参考测度的规则可以通过计数测度表现出来，它被离散集或 $\mathbb{R}^N$ 上的 Lebesgue 测度所支持；我们仅仅需要参考测度是局部有限的(即有限的值被分配到紧凑集)。Radon-Nikodym 推导结果将被称为概率质量函数(PMF)；它表示离散情况的概率和连续情况下的概率密度函数(PDF)。

信道容量理论的初步定义是从第 1 章(见 1.4 节)的离散信道发展起来的，通过采用下列符合逻辑的方案，对于具有连续分布的噪声来说信道容量的定义几乎没有改变：

基数 $M=\lceil 2^{NR}\rceil$ 中的信息集合 $\mathscr{U}$

→ 大小为 M 的码本 $\mathscr{C}$，其中码字长度为 N

→ 通过一个有噪信道的可靠速率为 R

→ 信道的容量

然而，为了简化说明，我们从现在起假定编码 $\mathscr{U}\to\mathscr{C}$ 是一一对应的，且通过码本来确定码字。

4.1 Gauss 信道

这里研究的信道具有连续分布的噪声；它们是通信中的基础模型，其中包含有线和无线传输。这种信道最常见的模型是无记忆加性 Gauss 信道(MAGC)，但是其他的连续噪声模型同样很有用。MAGC 的情况具有很大吸引力，这是因为它可以做一些便利和更深一步的计算并且可得到简洁的答案。

366

然而，Gauss(和其他的连续分布)信道提出了一个挑战，但在第 1 章考虑有限字母表的情况时并没有讨论这种情况。也就是说，因为码字(或者，使用一个稍微恰当的术语，码矢)可以从 Euclidean 空间(或噪声向量)中先验取值，当引入功率限制后，则信道容量的定义不得不进行修改。更进一步说，信道容量的取值将会依赖于所谓的局部限制，这会使分析变得困难。在 MAGC 的情况下，Shannon 给出了求解方法，但是他花了几年时间使他的分析更严格。

长度为 N 的输入字(被设计在连续的 N 个时隙中利用信道)可表示为输入 N-向量

$$x(=x^{(N)})=\begin{pmatrix}x_1\\ \vdots\\ x_N\end{pmatrix}$$

我们假定 $x_i\in\mathbb{R}$，因此有 $x^{(N)}\in\mathbb{R}^N$(为了使符号简短，上标(N)经常被省略)。

在加性信道中输入向量 x 被转化为随机向量 $\boldsymbol{Y}^{(N)}=(Y_1\cdots Y_N)^{\mathrm{T}}$，其中 $\boldsymbol{Y}=x+\boldsymbol{Z}$，它的分量形式表示为

$$Y_j=x_j+Z_j,\quad 1\leqslant j\leqslant N \tag{4.1.1}$$

其中

$$\mathbf{Z}=\begin{bmatrix} Z_1 \\ \vdots \\ Z_N \end{bmatrix}$$

是噪声向量，并包含随机变量 Z_1，…，Z_N。因此，噪声可以用联合 PDF $f^{\mathbf{Z}}(\underline{z})\geqslant 0$ 描述，其中

$$\underline{z}=\begin{bmatrix} z_1 \\ \vdots \\ z_N \end{bmatrix}$$

367

且总的积分表示为 $\int f^{\mathbf{Z}}(\underline{z})\mathrm{d}z_1\cdots\mathrm{d}z_N=1$。N 维噪声概率分布利用积分形式定义，这个积分区间定义在关于 $\mathbf{Z}$ 的给定集合上

$$P^{\mathbf{Z}}(\mathbf{Z}\in A)=\int_A f^{\mathbf{Z}}(\underline{z})\,\mathrm{d}z_1\cdots\mathrm{d}z_N,\quad A\subseteq\mathbb{R}^N$$

例子 4.1.1 如果对于每个 N，噪声向量 $(Z_1\cdots Z_N)^T$ 满足多元正态分布，则这个加性信道满足 Gauss 分布(简称为 AGC)；可以与 PSE 1114 页相比较。从现在起假定均值 $\mathbb{E}Z_j=0$。已知零均值的多元正态分布完全由它的协方差矩阵所决定。更确切地说，AGC 的联合 PDF $f_{Z^{(N)}}^{\mathbf{Z}}(\underline{z}^{(N)})$ 会有如下形式

$$\frac{1}{(2\pi)^{N/2}(\det\Sigma)^{1/2}}\exp\left(-\frac{1}{2}\underline{z}^T\Sigma^{-1}\underline{z}\right),\quad \underline{z}=\begin{bmatrix} z_1 \\ \vdots \\ z_N \end{bmatrix}\in\mathbb{R}^N \tag{4.1.2}$$

这里 Σ 是一个 $N\times N$ 的矩阵，假定矩阵是实的、对称的和严格正定性的，其中元素 $\Sigma_{jj'}=\mathbb{E}(Z_jZ_{j'})$ 表示着噪声随机变量 Z_j，$Z_{j'}$ 的协方差，$1\leqslant j$，$j'\leqslant N$。(实的严格正定性定义意味着 Σ 具有 $\boldsymbol{BB}^T$ 的形式，其中 $\boldsymbol{B}$ 是一个 $N\times N$ 的实的可逆矩阵；如果 Σ 是严格正定的，则 Σ 有 N 个相互正交特征向量，且 Σ 的所有 N 个特征向量都大于 0。)特别地，每一个随机变量 Z_j 均是正态的：$Z_j\sim N(0,\sigma_j^2)$，其中 $\sigma_j^2=\mathbb{E}Z_j^2$ 与对角元素 Σ_{jj} 是相同的。(因为严格的正定性，故对于所有的 $j=1$，…，N，有 $\Sigma_{jj}>0$)。

如果随机变量 Z_1，Z_2，…是独立同分布的，则信道被称为无记忆 Gauss(MGC)或者加性 Gauss 白噪声信道。在这种情况下，矩阵 Σ 是对角矩阵：当 $i\neq j$ 时，$\Sigma_{ij}=0$，当 $i=j$ 时，$\Sigma_{ii}>0$。这是重要的模型示例(在理论上和实际中皆如此)，因为它采用了一些良好的终值公式，且是进一步归纳的基础。

因此，MGC 具有独立同分布的噪声随机变量 $Z_j\sim N(0,\sigma^2)$，其中 $\sigma^2=\mathrm{Var}Z_j=\mathbb{E}Z_j^2$。对于正态随机变量，独立相当于去相关，即当所有 j，$j'=1$，…，N 且 $j\neq j'$ 时，等式 $\mathbb{E}(Z_jZ_{j'})=0$ 成立，意味着噪声向量 $\mathbf{Z}^{(N)}$ 的元素 Z_1，…，Z_N 相互独立。从(4.1.2)中推导出：如果 $j\neq j'$，且矩阵 Σ 满足 $\Sigma_{jj'}=0$，则 Σ 是对角的，且 $\det\Sigma=\prod_{1\leqslant j\leqslant N}\Sigma_{jj}$，式(4.1.2)中的联合 PDF 分解为 N 个因子的乘积，其中因子表示 Z_j，$1\leqslant j\leqslant N$ 的独立 PDF。

$$\prod_{1\leqslant j\leqslant N}\frac{1}{(2\pi\Sigma_{jj})^{1/2}}\exp\left(-\frac{z_j^2}{2\Sigma_{jj}}\right) \tag{4.1.3}$$

368

而且，在独立同分布的假设下，存在 $\Sigma_{jj}=\sigma^2>0$，随机变量 $Z_j\sim N(0,\sigma^2)$，且 MGC 的噪声分布用参数 $\sigma>0$ 来表示。更准确地来说，式(4.1.3)中的联合 PDF 可以重写为

$$\left(\frac{1}{\sqrt{2\pi}\sigma}\right)^N\exp\left(-\frac{1}{2\sigma^2}\sum_{1\leqslant j\leqslant N}z_j^2\right)$$

当给出一个无限随机序列$\mathbf{Z}_i=(Z_1, Z_2, \cdots)$，且上述的噪声向量$\mathbf{Z}^{(N)}$是由这个序列的最初$N$个元素组成。在Gauss分布的情况下，$\mathbf{Z}_i$被称为随机Gauss过程；当$\mathbb{E}Z_j=0$，这个过程同样由它的协方差$\Sigma$所决定，其中$\Sigma_{ij}=\mathrm{Cov}(Z_i, Z_j)=\mathbb{E}(Z_iZ_j)$，术语"白Gauss噪声"将这个模型同更一般的有色噪声信道模型区别开来，具体情况如下。

利用在离散情况下采用的模式分析具有连续分布噪声的信道：特别地，如果信道被用来传送$M\sim 2^{NR}$，$R<1$中的一个编码信息，我们需要一个码本，这个码本由M个长度为N的码字组成：$\mathbf{x}^T(i)=(x_1(i), \cdots, x_N(i))$，$1\leqslant i\leqslant M$：

$$\mathscr{X}_{M,N}=\{\mathbf{x}^{(N)}(1), \cdots, \mathbf{x}^{(N)}(M)\}=\left\langle\begin{bmatrix}x_1(1)\\ \vdots\\ x_N(1)\end{bmatrix}, \cdots, \begin{bmatrix}x_1(M)\\ \vdots\\ x_N(M)\end{bmatrix}\right\rangle \tag{4.1.4}$$

当然，假定发送者和接收者都知道码本。发送速率R表示为

$$R=\frac{\log_2 M}{N} \tag{4.1.5}$$

假定发送码矢$\mathbf{x}(i)$。利用选定的译码器解码接收的随机向量$\mathbf{Y}(=\mathbf{Y}(i))=\begin{bmatrix}x_1(i)+Z_1\\ \vdots\\ x_N(i)+Z_N\end{bmatrix}$，即$d: \mathbf{y}\to d(\mathbf{y})\in\mathscr{X}_{M,N}$。从几何学上来说，译码器根据确定的距离（随着译码器不同而改变）寻找最近的码字$\mathbf{x}(k)$；例如，如果我们选择使用欧几里得距离，通过使平方和最小译码向量$\mathbf{Y}$：

$$d(\mathbf{Y})=\arg\min\left[\sum_{1\leqslant j\leqslant N}(Y_j(i)-x_j(l))^2: \mathbf{x}(l)\in\mathscr{X}_{M,N}\right] \tag{4.1.6}$$

当$d(\mathbf{y})\neq\mathbf{x}(i)$时，就发生了一个差错。幸运的是，译码器的选择是建立在最大似然准则的基础上的，推导如下。

369

这里有一个额外的微妙之处：假定对于一个输入的字$\mathbf{x}$，如果有成功译码的机会，那它应该是属于$\mathbb{R}^N$中的确定可传输的域。例如，在MAGC的情况下，利用功率限制

$$\frac{1}{N}\sum_{1\leqslant j\leqslant N}x_j^2\leqslant\alpha \tag{4.1.7}$$

其中$\alpha>0$是一个给定的常量。在无线传输的情形下，N长的输入向量中每个信号振幅平方功率都应该被α约束，否则传输的结果被认为"不可译码"。从几何上说，为了能够进行译码，码本中的输入码字$\mathbf{x}(i)$必须位于中心在$\mathbf{0}\in\mathbb{R}^N$，半径是$r=\sqrt{\alpha N}$的Euclidean球$\mathbb{B}_{\ell_2}^N(\sqrt{\alpha N})$中：

$$\mathbb{B}_{\ell_2}^N(r)=\left\langle\mathbf{x}=\begin{bmatrix}x_1\\ \vdots\\ x_N\end{bmatrix}: \left(\sum_{1\leqslant j\leqslant N}x_j^2\right)^{1/2}\leqslant r\right\rangle$$

下标ℓ_2强调，拥有标准Euclidean距离的$\mathbb{R}^N$被认为是一个Hilbertℓ_2空间。

事实上，并没有要求整个码本$\mathscr{X}_{M,N}$都位于可译码域中：如果一个码字$\mathbf{x}(i)$不在可译码域中，则译码错误概率为1。形象地说，即要求大部分码字，而不是所有的码字都必须都位于$\mathbb{B}_{\ell_2}^N((N\alpha)^{1/2})$中，见图4-1。

存在局部限制(4.1.7)的原因是使空间中的码字相互之间有较大的距离，最终每一个传输速率都会变得稳定。（这意味着信道的容量是无限的；尽管这样的信道不能立即被排除，但在AGC的情形下，无限容量的情形看起来是不切实际的。）

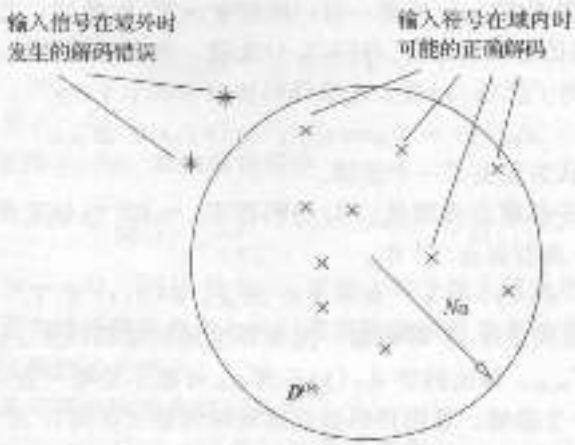


图 4-1

典型的情况是可译码域 $\mathbb{D}^{(N)} \in \mathbb{R}^N$ 被 $\mathbb{R}^N$ 中的一个球所表示，这个球以原点为球心，且与 $\mathbb{R}^N$ 中的特定的距离有关。也就是说，存在指数分布噪声的情况下，很自然地选择

$$\mathbb{D}^{(N)} = \mathbb{B}_1^{(N)}(N\alpha) = \left\{ \boldsymbol{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_N \end{bmatrix} : \sum_{1 \le i \le N} x_i \le N\alpha \right\}$$

即以 ℓ_1 度量的球。当输出信号向量距离码字的距离在 r 以内，这个向量就可以被这个码字译码，如果有(i)如果输出信号精确地落在码字的一定范围内，(ii)这个码字可能位于 $\mathbb{D}^{(N)}$ 中，(iii)这个特定的码字被发送了，即可说解码是正确的。当超过一个码字落在这个范围内，可能会解码错误。 370

在离散的情况下，长度为 N 的接收向量的(条件)概率分布可以表示更通用的信道，假定发送输入字 $\boldsymbol{x}^{(N)} \in \mathbb{R}^{(N)}$：

$$\boldsymbol{P}_{\boldsymbol{x}}^{(N)}(\cdot \mid \boldsymbol{x}^{(N)}) = \boldsymbol{P}_{\boldsymbol{x}}^{(N)}(\cdot \mid \text{码}\boldsymbol{x}^{(N)}\text{ 发送}),\ \boldsymbol{x} \in \mathbb{R}^N \tag{4.1.8}$$

同样 $N=1, 2, \cdots$ 意味着被用来传送的信道时隙数目，且将会考虑极限情况 $N \to \infty$。假定 $\boldsymbol{P}_{\boldsymbol{x}}^{(N)}(\cdot \mid \boldsymbol{x}^{(N)})$ 由 PMF $f_{\boldsymbol{x}}^{(N)}(\boldsymbol{y}^{(N)} \mid \boldsymbol{x}^{(N)})$ 决定，这个 PMF 与 $\mathbb{R}^N$ 中固定测量 $v^{(N)}$ 有关：

$$\boldsymbol{P}_{\boldsymbol{x}}^{(N)}(\boldsymbol{Y}^{(N)} \in A \mid \boldsymbol{x}^{(N)}) = \int_A f_{\boldsymbol{x}}^{(N)}(\cdot \mid \boldsymbol{x}^{(N)})\, \mathrm{d}v^{(N)}(\boldsymbol{y}^{(N)}) \tag{4.1.9a}$$

存在一个典型的假设，即 $v^{(N)}$ 是下式的乘积测度

$$v^{(N)} = v \times \cdots \times v\,(N\text{ 次}) \tag{4.1.9b}$$

例如，$v^{(N)}$ 可以是 $\mathbb{R}^N$ 上的 Lebesgue 测度，它是 $\mathbb{R}$ 上 Lebesgue 测度的乘积：$\mathrm{d}\,\boldsymbol{x}^{(N)} = \mathrm{d}x_1 \times \cdots \times \mathrm{d}x_N$。在离散情况下，数字 x_i 表示输入信道字母表(即在二元情况下，$\mathscr{A} = \{0, 1\}$)中的字母，v 是 $\mathscr{A}$ 上的计数测度，将重量 1 分配到字母表中每个符号。$v^{(N)}$ 是在 $\mathscr{A}^N$ 上的计数测度，其中 $\mathscr{A}^N$ 表示所有长度为 N 的输入字，且每个字被分配重量 1。 371

假定乘积形式的参考测度为 $v^{(N)}$(4.1.9b)，我们利用乘积形式 PMF $f_{\boldsymbol{x}}^{(N)}(\boldsymbol{y}^{(N)} \mid \boldsymbol{x}^{(N)})$ 表示无记忆信道：

$$f_{\boldsymbol{x}}^{(N)}(\boldsymbol{y}^{(N)} \mid \boldsymbol{x}^{(N)}) = \prod_{1 \le i \le N} f_{x_i}(y_i \mid x_i) \tag{4.1.10}$$

在这里 $f_x(y \mid x)$ 表示符号对符号信道的 PMF，它描述了信道单个使用时的影响。对于 MGC，$f_x(y \mid x)$ 满足正态分布 $N(x, \sigma^2)$，即 $f_x(y \mid x)$ 表示随机变量 $Y = x + Z$ 的 PDF，其中 $Z \sim N(0, \sigma^2)$ 表示影响每个独立输入值 x 的白噪声。

接下来我们考虑码本$\mathscr{X}_{M,N}$，它是一对一映射$\mathscr{M}\to\mathbb{R}^N$的描述，其中$\mathscr{M}$是有限的消息集合(最初以消息符号集的形式表现)；与(4.1.4)比较，在离散情况时，ML译码器d_{ML}通过使$x=x^{(N)}\in\mathscr{X}_{M,N}$中的$f_{Y}^{(N)}(y|x)$最大化来译码接收到的字$Y=y^{(N)}$：

$$d_{ML}(y)=\arg\max[f_{Y}^{(N)}(y|x):x\in\mathscr{X}_{M,N}] \tag{4.1.11}$$

当最大值不唯一时就认为发生了一个差错。

另一个有用的例子是联合典型性(JT)译码器$d_{JT}=d_{JT}^{(N)\varepsilon}$(如下所示)：它寻找码字$x$，使得$x$，$y$同时位于$\varepsilon$-典型集合$T_{\varepsilon}^{N}$中：

$$d_{JT}(y)=x \quad 如果\ x\in\mathscr{X}_{M,N}\ 且(x,y)\in T_{\varepsilon}^{N} \tag{4.1.12}$$

通过集合T_{ε}^{N}的特定形式设计JT译码器，用来译生成的随机码$\mathscr{X}_{M,N}$样例。因此对于给定的输出向量y^N和码$\mathscr{X}_{M,N}$，译出的字$d_{JT}(y)\in\mathscr{X}_{M,N}$可能不是唯一定义的(或者根本没有定义)，这会再次引起一个差错。通用译码器应该解释为定义在集合$\mathbb{K}^{(N)}\subseteq\mathbb{R}^N$上的一对一映射，即取点$y^N\in\mathbb{K}^N$则有对应的点$x\in\mathscr{X}_{M,N}$；在集合$\mathbb{K}^{(N)}$外可能没有被正确定义。可译码域$\mathbb{K}^{(N)}$是译码器$d^{(N)}$参数的一部分。在任何情况下，希望能够得到

$$\begin{aligned}\boldsymbol{P}_{\mathscr{X}}^{(N)}(d^{(N)}(\boldsymbol{Y})\neq x\,|\,x\ 发送)&=\boldsymbol{P}_{\mathscr{X}}^{(N)}(\boldsymbol{Y}\notin\mathbb{K}^{(N)}\,|\,x\ 发送)\\&\quad+\boldsymbol{P}_{\mathscr{X}}^{(N)}(\boldsymbol{Y}\in\mathbb{K}^{(N)},d(\boldsymbol{Y})\neq x\,|\,x\ 发送)\to 0\end{aligned}$$

其中$N\to\infty$。在MGC的情况下，对于任意码$\mathscr{X}_{M,N}$，从式(4.1.6)中得到的ML译码器几乎在$\mathbb{R}^N$中任何位置都被唯一定义(但是不需要知道具体是如何定义的)。

输入向量须满足$x^{(N)}\in\mathbb{D}^{(N)}\subseteq\mathbb{R}^N$，当$x^{(N)}\notin\mathbb{D}^{(N)}$时，传输的结果被认为是不可译的(无需考虑使用的译码器质量)。当使用码本$\mathscr{X}_{M,N}$和译码器$d^{(N)}$时，则平均错误概率被定义为

372

$$e^{av}(\mathscr{X}_{M,N},d^{(N)},\mathbb{D}^{(N)})=\frac{1}{M}\sum_{x\in\mathscr{X}_{M,N}}e(x,d^{(N)},\mathbb{D}^{(N)}) \tag{4.1.13a}$$

其中最大错误概率定义为

$$e^{max}(\mathscr{X}_{M,N},d^{(N)},\mathbb{D}^{(N)})=\max[e(x,d^{(N)},\mathbb{D}^{(N)}):x\in\mathscr{X}_{M,N}] \tag{4.1.13b}$$

当码字x被传输时，$e(x,d^{(N)},\mathbb{D}^{(N)})$表示错误概率：

$$e(x,d^{(N)},\mathbb{D}^{(N)})=\begin{cases}1, & x\notin\mathbb{D}^{(N)}\\ \boldsymbol{P}_{\mathscr{X}}^{(N)}(d^{(N)}(\boldsymbol{Y})\neq x\,|\,x), & x\in\mathbb{D}^{(N)}\end{cases} \tag{4.1.14}$$

在(4.1.14)中，码本$\mathscr{X}_{M,N}$中的码字顺序无关紧要；因此$\mathscr{X}_{M,N}$被简单地认为是Euclidean空间$\mathbb{R}^N$中M个点的集合。从几何上说，希望$\mathscr{X}_{M,N}$中的点能够被放置在合适的位置，即位于域$\mathbb{D}^{(N)}$中，从而使得到正确ML-译码的概率最大(这又会引起球-布局问题)。

最后假设数字$R>0$是固定的，码本$\mathscr{X}_{M,N}$的大小是：$M=\lceil 2^{NR}\rceil$。当$N\to\infty$时，我们想定义可靠传输速率，这和在1.4节中工作有点相似。

定义4.1.2 如果存在码本$\mathscr{X}_{M,N}\in R^N$的序列$\{\mathscr{X}_{M,N}\}$和译码器$d^{(N)}$的序列$\{d^{(N)}\}$：$\mathbb{R}^N\to\mathbb{R}^N$且$M=\lceil 2^{NR}\rceil$，则称$R>0$为可靠传输速率，其中存在区域约束$\mathbb{D}^{(N)}$。有

$$\lim_{N\to\infty}e^{av}(\mathscr{X}_{M,N},d^{(N)},\mathbb{D}^{(N)})=0 \tag{4.1.15}$$

备注4.1.3 从平均错误概率$e^{av}(\mathscr{X}_{M,N},d^{(N)},\mathbb{D}^{(N)})$这个角度看，对于最大错误概率$e^{max}(\mathscr{X}_{M,N},d^{(N)},\mathbb{D}^{(N)})$来说很容易验证传输速率$R$是可靠的。事实上，从定义4.1.2角度看，即从平均错误概率意义上看，假定R是可靠的。取相应码本的序列$\{\mathscr{X}_{M,N}\}$，其中$M=\lceil 2^{NR}\rceil$，和相应译码规则序列$\{d_N\}$。将每个码$\mathscr{X}_N$分成两部分，$\mathscr{X}_N^{(1)}$和$\mathscr{X}_N^{(2)}$，并且根据译码错误的概率将这些码字以非递减顺序排列，将最初$M^{(1)}=\lceil M/2\rceil$个码字放在$\mathscr{X}_N^{(1)}$中，剩下的$M^{(2)}=M-M^{(1)}$个码字放在$\mathscr{X}_N^{(2)}$中。则对于码$\{\mathscr{X}_{M,N}^{(i)}\}$序列有：

(i) 当 $N\to\infty$ 且

$$\frac{1}{N}\log M^{(0)} \geqslant R+O(N^{-1})$$

信息速率的值超近于 R。

(ii) 当利用译码规则 d_N 时，最大错误概率

$$P_e^{\max}(\mathscr{X}_N^{(0)},d_N) \leqslant \frac{1}{M(1)}\sum_{x^{(0)}\in\mathscr{X}_N^{(1)}} P_e(x^{(0)},d_N) \leqslant \frac{M}{M(1)}P_e^{\mathrm{av}}(\mathscr{X}_N,d_N)$$ 375

因为 $M/M^{(1)}\leqslant 2$，当 $N\to\infty$ 时，RHS 趋于 0。推测出对于最大错误概率来说 R 是可靠传输速率。反过来说，在最大错误概率意义上的可靠传输速率 R 在平均错误概率意义上同样也是可靠的，可以很明显得到这个结论。

可知信道的容量是可靠传输速率的上确界：

$$C=\sup[R>0;R\text{ 是可靠的}] \tag{4.1.16}$$

且不同信道的容量是不同，并随着约束域的形状而变化。

当(与下面定理 4.1.9 相比)存在平均功率限制门限 α 时，对于 MGC 来说，信道容量 $C(\alpha,\sigma^2)$ 可表示为如下简洁形式：

$$C(\alpha,\sigma^2)=\frac{1}{2}\log_2\left(1+\frac{\alpha}{\sigma^2}\right) \tag{4.1.17}$$

而且，同 1.4 节中类似，若存在随机编码序列，其中码字 $x(i)=(X_1(i),\cdots,X_N(i))$ 中的元素满足独立同分布，即 $X_j(i)\sim N(0,\alpha-\varepsilon_N)$，$j=1,\cdots,N$，$i=1,\cdots,M$，且当 $N\to\infty$ 时 $\varepsilon_N\to 0$，此时可以达到容量 $C(\alpha,\sigma^2)$。但是对于有限 N 来说，这样的随机编码形式上不满足约束条件(4.1.7)，当 $N\to\infty$ 时它违反了消失概率(因为存在适当的 ε_N 时，$\limsup_{N\to\infty}P\left(\max\left[\frac{1}{N}\sum_{1\leqslant j\leqslant N}X_j(i)^2;1\leqslant i\leqslant M\right]\leqslant\alpha\right)=1$)。因此，可知平均错误概率(4.1.13a)趋于 0(当然对于随机编码来说，错误概率本身也是随机的)。

例子 4.1.4 接下来我们讨论存在有色 Gauss 噪声的 AGC 例子。令码向量 $x=(x_1,\cdots,x_N)$ 含有多维元素

$$x_j=\begin{pmatrix}x_{j1}\\ \vdots\\ x_{jk}\end{pmatrix}\in\mathbb{R}^k,\quad 1\leqslant j\leqslant N$$

且噪声向量

$$\boldsymbol{Z}=\begin{pmatrix}Z_1\\ \vdots\\ Z_N\end{pmatrix}$$ 374

的分量 Z_j 同样是维度为 k 的随机向量：

$$Z_j=\begin{pmatrix}Z_{j1}\\ \vdots\\ Z_{jk}\end{pmatrix}$$

例如，$Z_1,\cdots,Z_N$ 可能是独立同分布 $N(0,\Sigma)$(k-变量正态分布)，其中 Σ 是已知的 $k\times k$ 协方差矩阵。

当并行使用由 k 个标量 Gauss 信道组成的系统时，则存在有色的模型。标量信号 x_{j1} 通过信道 1 被发送出去，x_{j2} 通过信道 2 被发送出去，以此类推，假设在每个应用中标量信

道联合生成 Gauss 噪声；不同的信道可能是独立的(矩阵 Σ 是 $k\times k$ 对角矩阵)，或者是非独立的(矩阵 Σ 是一般的 $k\times k$ 正定矩阵)。

仍然存在一个码本，它是一个(有序或无序)的集合 $\mathscr{X}_{M,N}=\{x(1),\cdots,x(M)\}$，其中每个码字 $x(i)$ 是一个多重向量 $(x_1(i),\cdots,x_N(i))^T\in\mathbb{R}^{k\times N}=\mathbb{R}^k\times\cdots\times\mathbb{R}^k$。令 Q 为 $k\times k$ 正定矩阵，且 $\Sigma Q=Q\Sigma$。则功率限制表示为

$$\frac{1}{N}\sum_{1\leqslant i\leqslant N}\langle x_i(i),\ Qx_i(i)\rangle\leqslant\alpha \tag{4.1.18}$$

毫无疑问，存在有色噪声的 AGC 的容量计算公式更复杂。因为存在 $\Sigma Q=Q\Sigma$，则矩阵 Σ 和 Q 可能同时被对角化。令 $\lambda_i,\gamma_i,i=1,\cdots,k$ 分别为 Σ 和 Q 的特征值(对应于相同的特征向量)。则有

$$C(\alpha,Q,\Sigma)=\frac{1}{2}\sum_{1\leqslant i\leqslant k}\log_2\left(1+\frac{(v\gamma_i^{-1}-\lambda_i)_+}{\lambda_i}\right) \tag{4.1.19}$$

其中 $(v\gamma_i^{-1}-\lambda_i)_+=\max(v\gamma_i^{-1}-\lambda_i,0)$。即 $(v\gamma_i^{-1}-\lambda_i)_+$ 是矩阵 $(vQ^{-1}-\Sigma)_+$ 的特征值，且此矩阵表示 Hermitian 矩阵 $vQ^{-1}-\Sigma$ 的正定部分。当存在

$$\mathrm{tr}[(vI-Q\Sigma)_+]=\alpha \tag{4.1.20}$$

$v=v(\alpha)>0$ 成立。

根据下式定义正定部分 $(vI-Q\Sigma)_+$

$$(vI-Q\Sigma)_+=\Pi_+(vI-Q\Sigma))\Pi_+$$

其中 Π_+ 是子空间中的正射投影(在 $\mathbb{R}^k$ 中)，此子空间是由特征值为 $\gamma_i\lambda_i<v$ 的矩阵 $Q\Sigma$ 的特征向量所张成。在式(4.1.20)中 $\mathrm{tr}[(vI-Q\Sigma)_+]\geqslant 0$(因为对于所有的正定矩阵有 $\mathrm{tr}AB\geqslant 0$)，当 $v=0$ 时上式等于 0(因为 $(-Q\Sigma)_+=\mathbf{0}$)且随着 v 单调递增会增至 $+\infty$。因此对于任何已知 $\alpha>0$，式(4.1.20)唯一确定 $v=v(\alpha)$ 的值。

375

尽管式(4.1.19)看起来比式(4.1.17)更复杂，但这两个式子都是下列两个事实的推论：(i)容量可以被定义为(随机)输入和输出信号之间的最大互信息熵，如同在离散情况下的定义(比较 1.3 节和 1.4 节)，(ii)若存在 Gauss 噪声(白色或有色)，当输入信号本身是 Gauss 分布且它的协方差能够解决辅助优化问题，这时可以获得互信息。当在式(4.1.17)情况下这个优化问题相当简单，但是在式(4.1.19)情况下则变得更复杂(但是仍然易懂)。

当随机编码达到容量 $C(\alpha;Q;\Sigma)$ 时，需要满足下列条件：信号 $X_j(i)$，$1\leqslant j\leqslant N$，$i=1,\cdots,M$ 是独立同分布且满足 $X_j(i)\sim N(0,A-\varepsilon_N I)$，其中 A 是 $k\times k$ 正定矩阵，并使行列式 $\det(A+\Sigma)$ 最大且满足约束条件 $\mathrm{tr}QA=\alpha$；这样的矩阵满足形式 $(vQ^{-1}-\Sigma)_+$。为了计算各种模型的容量，随机编码提供了一个方便的工具。我们将在举例中继续讨论多种类似的模型。

与具有连续分布噪声的信道不同，在恰当的时候信息熵应该被微分熵替代。已知 1.5 节中介绍的微分熵的内容。已知随机变量 X，Y，存在与参考测量 $\mu\times\nu$、边缘 PMF $f_X(x)=\int f_{X,Y}(x,y)\nu(\mathrm{d}y)$ 和 $f_Y(y)=\int f_{X,Y}(x,y)\mu(\mathrm{d}x)$ 相关的联合 PMF $f_{X,Y}(x,y)$，其中互信息表示为

$$\begin{aligned}I(X;Y)&=\mathbb{E}\log\frac{f_{X,Y}(X,Y)}{f_X(X)f_Y(Y)}\\&=\int f_{X,Y}(x,y)\log\frac{f_{X,Y}(x,y)}{f_X(x)f_Y(y)}\mu(\mathrm{d}x)\nu(\mathrm{d}y)\end{aligned}$$

当 X，Y 被随机向量 $\boldsymbol{X}=(X_1,\cdots,X_N)$ 和 $\boldsymbol{Y}=(Y_1,\cdots,Y_N)$（甚至可以是例子 4.1.4 中的多重向量，即 X_i 和 Y_i 本身也是向量）替换时，存在相似的互信息定义：

$$I(\boldsymbol{X}^{(N)};\boldsymbol{Y}^{(N)})=\mathbb{E}\log\frac{f_{\boldsymbol{X}^{(N)},\boldsymbol{Y}^{(N)}}(\boldsymbol{X}^{(N)},\boldsymbol{Y}^{(N)})}{f_{\boldsymbol{X}^{(N)}}(\boldsymbol{X}^{(N)})f_{\boldsymbol{Y}^{(N)}}(\boldsymbol{Y}^{(N)})} \tag{4.1.21a}$$

其中 $f_{\boldsymbol{X}^{(N)}}(\boldsymbol{x}^{(N)})$ 和 $f_{\boldsymbol{Y}^{(N)}}(\boldsymbol{y}^{(N)})$ 分别是 $\boldsymbol{X}^{(N)}$，$\boldsymbol{Y}^{(N)}$ 的边缘 PMF（即向量元素的联合 PMF）。 376

特别地，如果 $N=N'$，$\boldsymbol{X}^{(N)}$ 表示随机输入，则 $\boldsymbol{Y}^{(N)}=\boldsymbol{X}^{(N)}-\boldsymbol{Z}^{(N)}$ 表示相应的信道随机输出，与式(4.1.14)相比，（随机）错误概率表示为：

$$E(\boldsymbol{x}^{(N)},\mathbb{D}^{(N)})=\begin{cases}1, & \boldsymbol{x}^{(N)}\notin\mathbb{D}^{(N)}\\ \boldsymbol{P}_{\mathrm{ch}}^{(N)}(d_{\mathrm{ML}}(\boldsymbol{Y}^{(N)})\neq\boldsymbol{x}^{(N)}\mid\boldsymbol{x}^{(N)}), & \boldsymbol{x}^{(N)}\in\mathbb{D}^{(N)}\end{cases}$$

此时期望值表示为

$$\mathscr{E}(\boldsymbol{P}_{\boldsymbol{X}^{(N)}};\mathbb{D}^{(N)})=\mathbb{E}[E(\boldsymbol{X}^{(N)},\mathbb{D}^{(N)})] \tag{4.1.21b}$$

给定 $\varepsilon>0$，根据输入概率分布 $P_{\boldsymbol{X}^{(N)}}$，其中 $\mathscr{E}(\boldsymbol{P}_{\boldsymbol{X}^{(N)}},\mathbb{D}^{(N)})\leqslant\varepsilon$，可以定义每个信号互信息的上确界（即每个信道都只应用一次）：

$$\overline{C}_{\varepsilon,N}=\frac{1}{N}\sup[I(\boldsymbol{X}^{(N)};\boldsymbol{Y}^{(N)}):\mathscr{E}(P_{\boldsymbol{X}^{(N)}},\mathbb{D}^{(N)})\leqslant\varepsilon] \tag{4.1.22}$$

$$C_\varepsilon=\limsup_{N\to\infty}\overline{C}_{\varepsilon,N}\quad C=\lim_{\varepsilon\to 0}\inf C_\varepsilon \tag{4.1.23}$$

需要强调的是式(4.1.22)中的上确界是根据输入字 $\boldsymbol{X}^{(N)}$ 所有概率分布 $P_{\boldsymbol{X}^{(N)}}$ 求得，无论这些分布是离散、连续或混合的（包含离散与连续），输入字 $\boldsymbol{X}^{(N)}$ 的平均错误概率应 $\leqslant\varepsilon$。这使得对 $\overline{C}_{\varepsilon,N}$ 正确求值变得相当困难。然而，极限值 $\overline{C}$ 在一些重要的例子中更易于求解。

我们现在需要证明 Shannon 第二编码定理的相反部分：

定理 4.1.5 （比较定理 1.4.14 和 2.2.10）考虑这样一个信道，已知随机输出字 $\boldsymbol{Y}^{(N)}$ 的概率分布序列 $\boldsymbol{P}_{\mathrm{ch}}(\cdot\mid\boldsymbol{x}^{(N)}$ 发送）序列和可解域 $\mathbb{D}^{(N)}$。则根据式(4.1.22)和式(4.1.23)得到的值 $\overline{C}$ 是容量的上界，即：

$$C\leqslant\overline{C} \tag{4.1.24}$$

证明 令 R 表示可靠传输速率，$\{\mathscr{X}_{M,N}\}$ 是码本序列，其中 $M=\#\mathscr{X}_{M,N}\sim 2^{NR}$，且满足 $\lim_{N\to\infty}e^{(N)}(\mathscr{X}_{M,N},\mathbb{D}^{(N)})=0$。考虑 $(\boldsymbol{x},d_{\mathrm{ML}}(\boldsymbol{y}))$，其中 (i) $\boldsymbol{x}=\boldsymbol{x}_m^{(N)}$ 是随机输入字并且在 $\mathscr{X}_{M,N}$ 中满足均匀分布，(ii) $\boldsymbol{Y}=\boldsymbol{Y}^{(N)}$ 是接收的字，(iii) $d_{\mathrm{ML}}(\boldsymbol{y})$ 是经过传输后，利用 ML 检测准则 377
d_{ML} 评出的码字。字 $\boldsymbol{x}$ 和 $d_{\mathrm{ML}}(\boldsymbol{Y})$ 在 $\mathscr{X}_{M,N}$ 中联合运行，即离散类型的联合分布。然后利用广义 Fano 不等式(1.2.23)，可得

$$\begin{aligned}h_{\mathrm{discr}}(\boldsymbol{X}\mid d(\boldsymbol{Y}))&\leqslant 1+\log(M-1)\sum_{\boldsymbol{x}\in\mathscr{X}_{M,N}}\mathscr{P}(\boldsymbol{x}=\boldsymbol{x},d_{\mathrm{ML}}(\boldsymbol{Y})\neq\boldsymbol{x})\\&\leqslant 1+\frac{NR}{M}\sum_{\boldsymbol{x}\in\mathscr{X}_{M,N}}\boldsymbol{P}_{\mathrm{ch}}(d_{\mathrm{ML}}(\boldsymbol{Y})\neq\boldsymbol{x}\mid\boldsymbol{x}\text{ 发送})\\&=1+NRe^{(N)}(\mathscr{X}_{M,N},\mathbb{D}^{(N)}):=N\theta_N\end{aligned}$$

其中随着 $N\to\infty$，可得 $\theta_N\to 0$。接下来已知 $h(\boldsymbol{X}_m^{(N)})=\log M$，可得 $NR-1\leqslant h(\boldsymbol{X}_m^{(N)})$，因此可得

$$\begin{aligned}R&\leqslant\frac{1+h(\boldsymbol{X}_m^{(N)})}{N}\\&=\frac{1}{N}I(\boldsymbol{X}_m^{(N)};d(\boldsymbol{Y}^{(N)}))+\frac{1}{N}h(\boldsymbol{X}_m^{(N)}\mid d(\boldsymbol{Y}^{(N)}))\\&\quad+\frac{1}{N}\leqslant\frac{1}{N}I(\boldsymbol{x}_m^{(N)};\boldsymbol{Y}^{(N)})+\theta_N\end{aligned}$$

对于任意给定的 $\epsilon>0$，当 N 充分大时，平均错误概率满足 $e^{n}(\mathscr{X}_{M,N},\mathbb{B}^{(N)})<\epsilon$，因此，当 N 充分大时，$R\leqslant\overline{C}_{\epsilon,N}$。（因为存在码本 $\mathscr{X}_{M,N}$ 且 $e^{n}(\mathscr{X}_{M,N},\mathbb{B}^{(N)})$，根据此码本均匀分布能够得到输入分布为 $P_{X^{(N)}}$ 且 $\mathscr{E}(P_{X^{(N)}},\mathbb{B}^{(N)})\leqslant\epsilon$ 的例子。）因此，对于所有的 $\epsilon>0$，当 $R\leqslant\overline{C}_{\epsilon}$ 时，意味着传输速率 $R\leqslant C$。因此，可证得 $C\leqslant C$。 □

在很多有趣的情形下，式(4.1.24)中的界 $C\leqslant C$ 会变得精确(即 $C=\overline{C}$)。而且，$\overline{C}$ 的表达式在一些情况下会简化。例如，对于 MAGC 来说，无需通过改变 N 使互信息 $I(\boldsymbol{X}^{(N)};\boldsymbol{Y}^{(N)})$ 最大，但是当 $I(X;Y)$ 是满足适当约束的单输入输出信号之间的互信息时，使 $I(X;Y)$ 最大化变得可能，即对于 MAGC 存在

$$C=\overline{C}=\sup\{I(X;Y):\mathbb{E}X^2<\alpha\} \tag{4.1.25a}$$

在平方功率限制 α 下，值 $\sup\{I(X;Y):\mathbb{E}X^2\leqslant\alpha\}$ 通常被称为 MAGC 的信息容量。而且对于一般的 AGC，存在

$$C=C=\lim_{N\to\infty}\frac{1}{N}\sup\left[I(\boldsymbol{X}^{(N)};\boldsymbol{Y}^{(N)}):\frac{1}{N}\sum_{1\leqslant j\leqslant N}\mathbb{E}X_j^2\leqslant\alpha\right] \tag{4.1.25b}$$

例子 4.1.6 当存在平均功率约束时(即 $\mathbb{B}^{(N)}=\mathbb{B}^{(N)}((N\alpha)^{1/2})$)(与例子 4.1.1 相比)，即式(4.1.25b)RHS 有界，我们需要估算 MAGC 的容量 $\overline{C}(\alpha,\sigma^2)$，其中 MAGC 包含方差为 σ^2 的加性白 Gauss 噪声。

378

已知输入分布 $P_{X^{(N)}}$，可以得到

$$\begin{aligned}I(\boldsymbol{X}^{(N)};\boldsymbol{Y}^{(N)})&=h(\boldsymbol{Y}^{(N)})-h(\boldsymbol{Y}^{(N)}|\boldsymbol{X}^{(N)})=h(\boldsymbol{Y}^{(N)})-h(\boldsymbol{Z}^{(N)})\\&\leqslant\sum_{1\leqslant j\leqslant N}h(Y_j)-h(\boldsymbol{Z}^{(N)})=\sum_{1\leqslant j\leqslant N}(h(Y_j)-h(Z_j))\end{aligned} \tag{4.1.26}$$

单输入随机变量 X_j 的二阶矩表示为 $\alpha_j^2=EX_j^2$，其中 X_j 是随机输入向量 $\boldsymbol{X}^{(N)}$ 的第 j 项。相应的随机输出变量 Y_j 满足

$$\mathbb{E}Y_j^2=\mathbb{E}(X_j+Z_j)^2=\mathbb{E}X_j^2+2\mathbb{E}X_jZ_j+\mathbb{E}Z_j^2=\alpha_j^2+\sigma^2$$

其中 X_j 和 Z_j 相互独立且 $\mathbb{E}Z_j=0$。

在 Gauss 信道情况下，Y_j 具有连续分布(利用卷积 $\int\phi_j(x-y)\mathrm{d}F_{X_j}(x)$ 可得 PDF $f_{Y_j}(y)$，其中 ϕ_j 是 $Z_j\sim N(0,\sigma^2)$ 的 PDF)。因此在式(4.1.26)中得到的熵(同样表现在式(4.1.25a,b)中)是微分熵，已知 PDF 为 f_{Y_j} 的随机变量 Y_j，在 $\mathbb{E}Y_j^2\leqslant\alpha_j^2+\sigma^2$ 条件下，最大微分熵表示为 $h(Y_j)\leqslant\frac{1}{2}\log_2[2\pi e(\alpha_j^2+\sigma^2)]$。事实上利用 Gibbs 可得

$$\begin{aligned}h(Y_j)&=-\int f_{Y_j}(y)\log_2 f_{Y_j}(y)\mathrm{d}y\\&\leqslant-\int f_{Y_j}(y)\log_2\phi_{\alpha_j^2+\sigma^2}(y)\mathrm{d}y\\&=\frac{1}{2}\log_2[2\pi(\alpha_j^2+\sigma^2)]+\frac{\log_2 e}{2(\alpha_j^2+\sigma^2)}\mathbb{E}Y_j^2\\&\leqslant\frac{1}{2}\log_2[2\pi e(\alpha_j^2+\sigma^2)]\end{aligned}$$

因此有

$$\begin{aligned}I(X_j;Y_j)=h(Y_j)-h(Z_j)&\leqslant\log_2[2\pi e(\alpha_j^2+\sigma^2)]-\log_2(2\pi e\sigma^2)\\&=\log_2\left(1+\frac{\alpha_j^2}{\sigma^2}\right)\end{aligned}$$

379

当且仅当 $Y_j\sim N(0,\alpha_j^2+\sigma^2)$ 时等式成立。

利用大数定理，根据式(4.1.25b)中的界 $\sum_{1\leqslant j\leqslant N}\mathbb{E}X_j^2=\sum_{1\leqslant j\leqslant N}\alpha_j^2<N\alpha$ 可得 $\lim_{N\to\infty}P_{X^{(N)}}(\mathbb{E}^{(N)}(\sqrt{N\alpha}))=1$。而且，对于任意的输入概率分布 $P_{X^{(N)}}$ 且有 $\mathbb{E}X_j^2\leqslant\alpha_j^2$，$1\leqslant j\leqslant N$，可以得到

$$\frac{1}{N}I(\boldsymbol{X}^{(N)};\boldsymbol{Y}^{(N)})\leqslant\frac{1}{2N}\sum_{1\leqslant j\leqslant N}\log_2\left(1+\frac{\alpha_j^2}{\sigma^2}\right)$$

将 Jensen 不等式应用到凹函数 $x\mapsto\log_2(1+x)$，则有

$$\begin{aligned}\frac{1}{2N}\sum_{1\leqslant j\leqslant N}\log_2\left(1+\frac{\alpha_j^2}{\sigma^2}\right)&\leqslant\frac{1}{2}\log_2\left(1+\frac{1}{N}\sum_{1\leqslant j\leqslant N}\frac{\alpha_j^2}{\sigma^2}\right)\\&\leqslant\frac{1}{2}\log_2\left(1+\frac{\alpha}{\sigma^2}\right)\end{aligned}$$

因此在这个例子中，式(4.1.25b)的 RHS，即信息容量 $\overline{C}$ 满足

$$\overline{C}\leqslant\frac{1}{2}\log_2\left(1+\frac{\alpha}{\sigma^2}\right)\tag{4.1.27}$$

在证明定理 4.1.8 之后，可以在式(4.1.17)中证明容量 $C(\alpha,\sigma^2)$ 等于 RHS。

例子 4.1.7 对于有色 Gauss 噪声，式(4.1.26)的界可以重写为：

$$I(\boldsymbol{X}^{(N)};\boldsymbol{Y}^{(N)})\leqslant\sum_{1\leqslant j\leqslant N}[h(Y_j)-h(Z_j)]$$

其中利用了输入和输出信号 X_j，$Y_j=X_j+Z_j$ 随机向量的混合二阶矩：

$$\alpha_j^2=\mathbb{E}\langle X_j,QX_j\rangle,\mathbb{E}\langle Y_j,QY_j\rangle=\alpha_j^2+\mathrm{tr}(Q\Sigma),\frac{1}{N}\sum_{1\leqslant j\leqslant N}\alpha_j^2\leqslant\alpha$$

在计算中同样利用了下列事实：即 X_j，Z_j 相互独立，且期望值满足 $\mathbb{E}Z_j=0$。

在标量的情况下，$\frac{1}{N}I(\boldsymbol{X}^{(N)};\boldsymbol{Y}^{(N)})$ 不会超过差值 $h(Y)-h(Z)$，其中 $Z\sim N(0,\Sigma)$ 是有色噪声向量，当 $Y=X+Z$ 是多元正态分布时，在跟踪限制条件下能使微分熵最大，得出：

$$\frac{1}{N}I(\boldsymbol{X}^{(N)};\boldsymbol{Y}^{(N)})\leqslant\overline{h}(\alpha,Q,\Sigma)-h(Z)$$ 383

其中 $\boldsymbol{K}$ 是信号的协方差矩阵，且

$$\overline{h}(\alpha,Q,\Sigma)=\frac{1}{2}\max\{\log[(2\pi)^k\mathrm{e}\det(\boldsymbol{K}+\Sigma)]:$$

$$\boldsymbol{K}\ \ k\times k\text{ 的正定矩阵满足 }\mathrm{tr}(Q\boldsymbol{K})\leqslant\alpha\}$$

将 Σ 写成对角形式 $\Sigma=\boldsymbol{C}\Delta\boldsymbol{C}^{\mathrm{T}}$，其中 $\boldsymbol{C}$ 是正交矩阵，$\boldsymbol{\Lambda}$ 是由 Σ 特征值构成的 $k\times k$ 对角矩阵：

$$\boldsymbol{\Lambda}=\begin{pmatrix}\lambda_1&0&\cdots&0\\0&\lambda_2&\cdots&0\\0&0&\ddots&0\\0&0&\cdots&\lambda_k\end{pmatrix}$$

写出 $\boldsymbol{C}^{\mathrm{T}}\boldsymbol{K}\boldsymbol{C}=\boldsymbol{B}$，在满足下列约束条件的情况下使 $\det(\boldsymbol{B}+\boldsymbol{\Lambda})$ 最大

$$\boldsymbol{B}\text{ 正定且 }\mathrm{tr}(\boldsymbol{\Gamma B})\leqslant\alpha,\quad\text{其中 }\boldsymbol{\Gamma}=\boldsymbol{C}^{\mathrm{T}}Q\boldsymbol{C}$$

利用举例 1.5.10 中的 Hadamard 不等式，可得 $\det(\boldsymbol{B}+\boldsymbol{\Lambda})\leqslant\prod_{1\leqslant j\leqslant k}(B_{jj}+\lambda_j)$，当且仅当 $\boldsymbol{B}$ 是对角时(即矩阵 Σ 和 $\boldsymbol{K}$ 有相同的特征向量)，且 $B_{11}=\beta_1$，…，$B_{kk}=\beta_k$ 是 $\boldsymbol{K}$ 的特征值时，

等式成立。同样假设 $Q\Sigma=\Sigma Q$，则 $\mathrm{tr}(\Gamma B)=\sum_{1\leqslant i\leqslant k}\gamma_i\beta_i$。因此若想使乘积 $\sum_{1\leqslant i\leqslant k}(\beta_i+\lambda_i)$ 最大，即相当于使下列的和最大

$$\sum_{1\leqslant i\leqslant k}\log(\beta_i+\lambda_i),\quad 约束于\ \beta_1,\cdots,\beta_k\geqslant 0\ 且\ \sum_{1\leqslant i\leqslant k}\gamma_i\beta_i\leqslant \alpha$$

如果忽略区域约束 $\beta_1,\cdots,\beta_k\geqslant 0$，Lagrangian 量

$$\mathscr{L}(\beta_1,\cdots,\beta_k;\kappa)=\sum_{1\leqslant i\leqslant k}\log(\beta_i+\lambda_i)-\kappa\Big(\alpha-\sum_{1\leqslant i\leqslant k}\gamma_i\beta_i\Big)$$

在满足下列条件时

$$\frac{1}{\beta_i+\lambda_i}=\kappa\gamma_i,\quad 即\quad \beta_i=\frac{1}{\kappa\gamma_i}-\lambda_i,\quad i=1,\cdots,k$$

取得最大值。

为了满足区域限制，取

$$\beta_i=\Big(\frac{1}{\kappa\gamma_i}-\lambda_i\Big)_+,\quad i=1,\cdots,k$$

并且选整值 $\kappa>0$。因此有

381

$$\sum_{1\leqslant i\leqslant k}\Big(\frac{1}{\kappa}-\gamma\lambda_i\Big)_+=\alpha \tag{4.1.28}$$

可知信息容量 $\overline{C}(\alpha,Q,\Sigma)$ 满足

$$\overline{C}(\alpha,Q,\Sigma)\leqslant\frac{1}{2}\sum_{1\leqslant i\leqslant k}\log_2\Big(1+\frac{(v\gamma_i^{-1}-\lambda_i)_+}{\lambda_i}\Big) \tag{4.1.29}$$

其中 RHS 从式(4.1.28)中得到且 $v=1/\kappa$。同样能够说明容量 $C(\alpha,Q,\Sigma)$ 等于上一个表达式，最终的答案会在式(4.1.19)确定。

我们现在讨论 Shannon 第二编码定理的直接部分，它适用于有区域限制的一般信道。尽管这个定理的表述与定理 1.4.15 和 2.2.1 仅在码字约束的假设上不同(且下面的证明仅仅是定理 1.4.15 证明的重复)，但把它放在正式内容中是有益的。

定理 4.1.8 已知接收字 $\boldsymbol{Y}^{(N)}$ 的条件概率 $P_{\mathbf{A}}^{(N)}(\cdot\mid x^{(N)}$ 发送)序列和输入向量的解码约束 $x^{(N)}\in\mathcal{D}^{(N)}$ 序列，根据这些条件能够确定信道。假定概率 $P_{\mathbf{A}}^{(N)}(\cdot\mid x^{(N)}$ 发送)是由与参考测量 $\nu^{(N)}$ 有关的 PMF $f_{\mathbf{A}}(y^{(N)}\mid x^{(N)}$ 发送)给定。已知 $c>0$，假定存在输入概率分布 $P_{X^{(N)}}$ 序列，则有

(i) $\lim_{N\to\infty}P_{X^{(N)}}(\mathcal{D}^{(N)})=1$，

(ii) 分布 $P_{X^{(N)}}$ 由与参考测量 $\mu^{(N)}$ 有关的 PMF $f_{X^{(N)}}(x^{(N)})$ 给出，

(iii) 下式依概率收敛成立：对于所有 $\varepsilon>0$，

$$\lim_{N\to\infty}P_{Y^{(N)},X^{(N)}}(T_\varepsilon^N)=1$$

$$T_\varepsilon^N=\Big\{\Big|\frac{1}{N}\log\frac{f_{Y^{(N)},X^{(N)}}(x^{(N)},\boldsymbol{Y}^{(N)})}{f_{X^{(N)}}(x^{(N)})f_{Y^{(N)}}(\boldsymbol{Y}^{(N)})}-c\Big|\leqslant\varepsilon\Big\} \tag{4.1.30a}$$

其中

$$f_{X^{(N)},Y^{(N)}}(x^{(N)},y^{(N)})=f_{X^{(N)}}(x^{(N)})f_{\mathbf{A}}(y^{(N)}\mid x^{(N)}\ 发送)$$

$$f_{Y^{(N)}}(y^{(N)})=\int f_{X^{(N)}}(\tilde{x}^{(N)})f_{\mathbf{A}}(y^{(N)}\mid\tilde{x}^{(N)}\ 发送)\mu^{(N)}(\mathrm{d}\,\tilde{x}^{(N)}) \tag{4.1.30b}$$

则信道容量满足 $C\geqslant c$。

证明 令 $R<c$，考虑随机码本 $\{x^{(N)}(1),\cdots,x^{(N)}(M)\}$，其中 $M\sim 2^{NR}$，它是由独立同分

382 布的码字组成，且每个码字 $x^{(N)}(j)$ 由 $P^N=P_{X^N}$ 表示出。假定(随机)码字 $x^{(N)}(j)$ 被发送，

接收到(随机)字$\boldsymbol{Y}^N=\boldsymbol{Y}^N(j)$，且存在如(4.1.30b)中所示的联合 PMF$f_{\boldsymbol{X}^{(N)},\boldsymbol{Y}^{(N)}}$。令$\epsilon>0$，利用联合典型性解码$\boldsymbol{Y}^N$：

$$d_{\mathrm{JT}}(\boldsymbol{Y}^N)=\boldsymbol{x}^N(i)\ \text{当}\ \boldsymbol{x}^N(i)\ \text{是}$$
$$\boldsymbol{x}^N(1),\cdots,\boldsymbol{x}^N(M)\ \text{中唯一的向量且}(\boldsymbol{x}^N(i),\boldsymbol{Y}^N)\in T_\epsilon^N$$

其中集合T_ϵ^N已在(4.1.30a)中详细说明。

假定随机向量$\boldsymbol{x}^N(j)$已被发送，当存在下列情况时：

(i) $\boldsymbol{x}^N(j)\notin\mathbb{D}^{(N)}$。

(ii) $(\boldsymbol{x}^N(j),\boldsymbol{Y}^N)\notin T_\epsilon^N$。

(iii) 对于一些$i\neq j$，有$(\boldsymbol{x}^N(i),\boldsymbol{Y}^N)\in T_\epsilon^N$。

假定每次都会发生一个差错。

这些概率互相不排斥，但是如果所有的情况都没出现，则有

(a) $\boldsymbol{x}^N(j)\in\mathbb{D}^{(N)}$。

(b) $\boldsymbol{x}(j)$是在$\boldsymbol{x}^N(1)$，…，$\boldsymbol{x}^N(M)$中唯一的字且$(\boldsymbol{x}^N(j),\boldsymbol{Y}^N)\in T_\epsilon^N$。

因此，JT 解码器将会返回正确的结果。考虑到平均错误概率

$$\mathscr{E}_M(P^N)=\frac{1}{M}\sum_{1\le j\le M}E(j,P^N)$$

其中$E(j,P^N)$是上述任意可能事件(i)-(iii)发生的概率：

$$\begin{aligned}E(j,P^N)&=\mathbb{P}(\{\boldsymbol{x}^N(j)\notin\mathbb{D}^{(N)}\}\cup\{(\boldsymbol{x}^N(j),\boldsymbol{Y}^N)\notin T_\epsilon^N\}\\&\quad\cup\{(\boldsymbol{x}^N(i),\boldsymbol{Y}^N)\in T_\epsilon^N\quad\text{对某些}\ i\neq j\})\\&=\mathbb{E}\mathbb{1}(\boldsymbol{x}^N(j)\notin\mathbb{D}^{(N)})\\&\quad+\mathbb{E}\mathbb{1}(\boldsymbol{x}^N(j)\in\mathbb{D}^{(N)},d_{\mathrm{JT}}(\boldsymbol{Y}^N)\neq\boldsymbol{x}^N(j))\end{aligned}\tag{4.1.31}$$

式(4.1.31)中的符号$\mathbb{P}$，$\mathbb{E}$分别表示：(1)独立同分布输入向量$\boldsymbol{x}^N(1)$，…，$\boldsymbol{x}^N(M)$集合(2)和$\boldsymbol{x}^N(j)$有关系的输出向量$\boldsymbol{Y}^N$。因此当$i\neq j$时$\boldsymbol{Y}^N$和$\boldsymbol{x}^N(i)$是相互独立的。为了便于进一步计算，将相应的概率分布$\mathbb{P}$表示为 Cartesian 乘积，例如，当$j=1$时，在式(4.1.31)中将表示为

$$\mathbb{P}=P_{\boldsymbol{x}^N(1),\boldsymbol{x}^N(1)}\times P_{\boldsymbol{x}^N(2)}\times\cdots\times P_{\boldsymbol{x}^N(M)}$$

其中$P_{\boldsymbol{x}^N(1),\boldsymbol{y}^N(1)}$表示输入向量$\boldsymbol{x}^N(1)$和输出向量$\boldsymbol{Y}^N(1)$的联合分布，它是由联合 PMF 决定的

$$f_{\boldsymbol{x}^N(1),\boldsymbol{x}^N(1)}(\boldsymbol{x}^N,\boldsymbol{y}^N)=f_{\boldsymbol{x}^N(1)}(\boldsymbol{x}^N)f_m(\boldsymbol{y}^N\mid\boldsymbol{x}^N\ \text{发送})$$ [383]

根据对称性知$E(j,P^N)$不随j的变化而变化，因此在接下的论证中我们取$j=1$。已知概率$E(1,P^N)$不会超过下列概率的总和

$$\begin{aligned}&\mathbb{P}(\boldsymbol{x}^N(1)\notin\mathbb{D}^{(N)})+\mathbb{P}((\boldsymbol{x}^N(1),\boldsymbol{Y}^N)\notin T_\epsilon^N)\\&\quad+\sum_{i=2}^{M}\mathbb{P}((\boldsymbol{x}^N(i),\boldsymbol{Y}^N)\in T_\epsilon^N)\end{aligned}$$

根据条件知$\lim\limits_{N\to\infty}P_{\boldsymbol{x}^N}(\mathbb{D}^{(N)})=1$，当$N\to\infty$第一个被加数趋于0。根据(4.1.30a)知，当取极限$N\to\infty$时，第二个被加数同样趋于0，只需要估计和$\sum\limits_{i=2}^{M}\mathbb{P}((\boldsymbol{x}^N(i),\boldsymbol{Y}^N)\in T_\epsilon^N)$。

首先，根据对称性知所有的被加数都相等，因此有

$$\sum_{i=2}^{M}\mathbb{P}((\boldsymbol{x}^N(i),\boldsymbol{Y}^N)\in T_\epsilon^N)=(2^{[NR]}-1)\mathbb{P}((\boldsymbol{x}^N(2),\boldsymbol{Y}^N)\in T_\epsilon^N)$$

接下来，根据举例 4.2.3 可得(见下式(4.2.9))

$$\mathbb{P}((\boldsymbol{x}^N(2),\boldsymbol{Y}^N)\in T_\epsilon^N)\le 2^{-N(I-3\epsilon)}$$

因此有

$$\sum_{i=2}^{M}\mathbb{P}((\boldsymbol{x}^N(i),\boldsymbol{Y}^N)\in T_\varepsilon^N)\leqslant 2^{N(R-c+3\varepsilon)}$$

当 $\varepsilon<(c-R)/3$ 时，当 $N\to\infty$ 上式趋于 0。

因此当 $R<c$ 时，有 $\lim\limits_{N\to\infty}\mathscr{E}_M(P^N)=0$。但是 $\mathscr{E}_M(P^N)$ 会有如下表达形式

$$\mathscr{E}_M(P^N)=\mathbb{E}_{P_{X^N(1)}}\times\cdots\times_{P_{X^N(M)}}\left(\frac{1}{M}\sum_{1\leqslant j\leqslant M}E(j)\right)$$

其中值 $E(j)$ 表示在(4.1.14)中定义的错误概率，

$$E(j)=\begin{cases}1, & \boldsymbol{x}^N\notin\mathbb{D}^{(N)}\\ \boldsymbol{P}_{\mathrm{ch}}(d_{\mathrm{IT}}^{(N)\cdot\varepsilon}(\boldsymbol{Y}^N)\neq\boldsymbol{x}^N(j)\mid\boldsymbol{x}^N(j)\text{ 发送}), & \boldsymbol{x}^N\in\mathbb{D}^{(N)}\end{cases}$$

我们总结出若存在样本码本 $\mathscr{X}_{M,N}$ 序列且平均错误概率为

384

$$\frac{1}{M}\sum_{x\in\mathscr{X}_{M,N}}e(x)\to 0$$

其中 $e(x)=e(x,\mathscr{X}_{M,N},\mathbb{D}^{(N)},d_{\mathrm{IT}}^{(N)\cdot\varepsilon})$ 表示码 $\mathscr{X}_{M,N}$ 中输入字 x 的错误概率，存在 JT 译码器，并且根据 $\mathbb{D}^{(N)}$ 规定的区域约束，可得：

$$e(x)=\begin{cases}1, & \boldsymbol{x}^N\notin\mathbb{D}^{(N)}\\ \boldsymbol{P}_{\mathrm{ch}}(d_{\mathrm{IT}}^{(N)\cdot\varepsilon}(\boldsymbol{Y}^N)\neq x\mid x\text{ 发送}), & \boldsymbol{x}^N\in\mathbb{D}^{(N)}\end{cases}$$

因此，根据定义 4.1.2 可知 R 是可靠传输速率，则完成了定理 4.1.8 的证明。 □

我们还顺便证明了如下结论。

定理 4.1.9 *假定定理 4.1.5 的条件在这里同样成立。对于所有 $R<C$，存在长度为 N，大小为 $M\sim 2^{NR}$ 的码 $\mathscr{X}_{M,N}$ 序列，当 $N\to\infty$ 时，最大错误概率趋于 0。*

例子 4.1.10 根据定理 4.1.8 能够说明式(4.1.17)和式(4.1.19)中的表达式是相应容量的真实值(基于 ML 准则)：对于方差为 σ^2 的标量白噪声，在存在平均输入功率约束 $\sum\limits_{1\leqslant i\leqslant N}x_i^2\leqslant N\alpha$ 情况下，则有

$$C(\alpha,\sigma^2)=\frac{1}{2}\log\left(1+\frac{\alpha}{\sigma^2}\right)$$

对于方差为 $\boldsymbol{\sigma}^2=(\sigma_1^2,\cdots,\sigma_k^2)$ 的向量白噪声，在约束 $\sum\limits_{1\leqslant i\leqslant N}x_i^2\leqslant N\alpha$ 条件下，有

$$C(\alpha,\boldsymbol{\sigma}^2)=\frac{1}{2}\sum_{1\leqslant i\leqslant k}\log\left(1+\frac{(v-\sigma_i^2)_+}{\sigma_i^2}\right)$$

其中 $\sum\limits_{1\leqslant i\leqslant k}(v-\sigma_i^2)_+=\alpha^2$。

对于协方差矩阵为 Σ 的有色向量噪声，在存在约束 $\sum\limits_{1\leqslant i\leqslant N}x_i^2\boldsymbol{Q}x_i\leqslant N\alpha$ 条件下，有

$$C(\alpha,\boldsymbol{Q},\Sigma)=\frac{1}{2}\sum_{1\leqslant i\leqslant k}\log\left(1+\frac{(v\gamma_i^{-1}-\lambda_i)_+}{\lambda_i}\right)$$

其中 $\sum\limits_{1\leqslant i\leqslant k}(v-\gamma\lambda_i)_+=\alpha$。

明确地说，对于标量白噪声我们取随机编码，其中信号 $X_i(j)$，$1\leqslant j\leqslant N$，$1\leqslant i\leqslant M-\lceil 2^{NR}\rceil$ 是独立同分布 $N(0,\alpha-\varepsilon)$。在这种情况下需要验证定理 4.1.5 的条件：当 $N\to\infty$ 时

385

(i) $\lim\limits_{N\to\infty}\mathbb{P}(\boldsymbol{x}^{(N)}(i)\in\mathbb{B}^{(N)}(\sqrt{N\alpha})$，对于所有 $i=1,\cdots,M)=1$。

(ii) 依概率 $\lim\limits_{\varepsilon\to 0}\lim\limits_{N\to\infty}\theta_N=C(\alpha,\sigma^2)$，其中

$$\theta_N = \frac{1}{N}\sum_{1\leqslant i\leqslant M}\log\frac{P(X,Y)}{P_X(X)P_Y(Y)}$$

首先探究性质(i)

$$\begin{aligned}
&\mathbb{P}(x^{(N)}(i)\notin\mathbb{B}^{(N)}(\sqrt{N\alpha}),\quad 对某些\ i=1,\cdots,M)\\
&\leqslant\mathbb{P}\Big(\frac{1}{NM}\sum_{1\leqslant i\leqslant M}\sum_{1\leqslant j\leqslant N}X_j(i)^2\geqslant\alpha\Big)\\
&=\mathbb{P}\Big(\frac{1}{NM}\sum_{1\leqslant i\leqslant M}\sum_{1\leqslant j\leqslant N}(X_j(i)^2-\sigma^2)\geqslant\varepsilon\Big)\\
&\leqslant\mathbb{E}(X^2-\sigma^2)^2\Big(\frac{1}{NM\varepsilon^2}\Big)\to 0
\end{aligned}$$

然后探究性质(ii)：因为(X_i, Y_i)满足独立同分布，利用大数定理可以得到

$$\theta_N\to\mathbb{E}\log\frac{P(X,Y)}{P_X(X)P_Y(Y)}=I(X_1;Y_1)$$

但是存在

$$\begin{aligned}
I(X_1;Y_1)&=h(Y_1)-h(Y_1\mid X_1)\\
&=\frac{1}{2}\log[2\pi e(\alpha-\varepsilon+\sigma^2)]-\frac{1}{2}\log(2\pi e\sigma^2)\\
&=\frac{1}{2}\log\Big(1+\frac{\alpha-\varepsilon}{\sigma^2}\Big)\to C(\alpha,\sigma^2),\quad 当\ \varepsilon\to 0\ 时
\end{aligned}$$

因此如同说明那样，容量等于$C(\alpha, \sigma^2)$。有色噪声的情况可以利用同样的方法进行研究。

备注 4.1.11 引入由域$\mathcal{D}$描述的区域约束并不意味着整个码$\mathscr{X}$都应该位于$\mathcal{D}$中。为了保证错误概率$P_e^n(\mathscr{X})\to 0$，我们只需要确保当码字长$N\to\infty$时，码字$x(i)\in\mathscr{X}$的大部分都位于$\mathcal{D}$中。

例子 4.1.12 我们考虑一个非 Gauss 加性信道，其中噪声向量表示为

$$\boldsymbol{Z}=\begin{bmatrix}Z_1\\ \vdots\\ Z_\mu\end{bmatrix}$$

386

双面指数独立同分布元素$Z_j\sim(2)\mathrm{Exp}(\lambda)$，且 PDF 表示为

$$f_{Z_j}(z)=\frac{1}{2}\lambda e^{-\lambda|z|},\quad -\infty<z<\infty$$

其中 Exp 表示指数分布，$\lambda>0$且$\mathbb{E}|Z_j|=1/\lambda$(见附录 PES 1)。接下来计算容量，同样基于 ML 准则，存在区域约束$x^{(N)}\in\mathcal{L}(N\alpha)$，且满足

$$\mathcal{L}(N\alpha)=\{x^{(N)}\in\mathbb{R}^N:\sum_{1\leqslant j\leqslant N}|x_j|\leqslant N\alpha\}$$

首先，如果随机变量X满足$\mathbb{E}|X|\leqslant\alpha$，随机变量$Z$满足$\mathbb{E}|Z|\leqslant\zeta$，则可得$\mathbb{E}|X+Z|\leqslant\alpha+\zeta$。然后利用下列事实：PDF 为$f_Y$且$\mathbb{E}|Y|\leqslant\eta$的随机变量$Y$存在微分熵

$$h(Y)\leqslant 2+\log_2\eta$$

当且仅当$Y\sim(2)\mathrm{Exp}(1/\eta)$时等号成立。

事实上，利用 Gibbs 性质可得

$$\begin{aligned}
h(Y)&=-\int f_Y(y)\log f_Y(y)\mathrm{d}y\\
&\leqslant-\int f_Y(y)\log g^{(2)\mathrm{Exp}(1/\eta)}(y)\mathrm{d}y
\end{aligned}$$

$$
\begin{aligned}
&=1+\frac{1}{\eta}\int f_Y(y)\,|y|\,\mathrm{d}y+\log\eta \\
&=1+\log\eta \leqslant 2+\log\eta+\frac{1}{\eta}\mathbb{E}|Y| \\
&=-\int \phi^{(2\mathbb{E}\log(\lambda)/\eta)}(y)\log\phi^{(2\mathbb{E}\log(\lambda)/\eta)}(y)\,\mathrm{d}y
\end{aligned}
$$

只有当 $f_Y=\phi^{(2\mathbb{E}\log(\lambda)/\eta)}$ 时，等式成立。

然后根据 SSCT 的相反部分，有

$$
\begin{aligned}
\frac{1}{N}I(x^{(N)};Y^{(N)}) &= \frac{1}{N}\sum h(Y_i)-h(Z_i) \\
&\leqslant \frac{1}{N}\sum\left[2+\log_2(\alpha_i+\lambda^{-1})-2+\log_2(\lambda)\right] \\
&= \frac{1}{N}\sum \log_2(1+\alpha_i\lambda) \\
&\leqslant \log_2(1+\omega\lambda)
\end{aligned}
$$

387 根据以前相同的参数可得 RHS 是信道的容量。

举例 4.1.13 接下来我们考虑存在加性均匀噪声的信道，噪声随机变量满足 $Z\sim U(-b,b)$，其中 $b>0$ 表示噪声幅度的范围。令输入信号的区域约束表示为有限集 $\mathscr{A}\subset\mathbb{R}$（输入符号集），即 $\mathscr{A}=\{a,a+b,\cdots,a+(M-1)b\}$，计算信道的信息容量：

$$C^{\mathrm{sd}}=\sup\{I(X;Y):p_X(\mathscr{A})=1,Y=X+Z\}$$

解答 由于平移不变性，我们可以假定有 $a=-A$ 和 $a+Mb=A$，其中 $2A=Mb$ 是输入信号集合的大小。有公式 $I(X;Y)=h(Y)-h(Y|X)$，其中 $h(Y|X)=h(Z)=\ln(2b)$ 表明必须使输出信号的熵 $h(Y)$ 最大。Y 的范围是 $-A-b\leqslant Y\leqslant A+b$，因此 P_Y 必须尽可能接近均匀分布 $U(-A-b,A+b)$。

首先假定 M 是奇数：$\#\mathscr{A}=2m+1$，且

$$\mathscr{A}=\{0,\pm A/m,\pm 2A/m,\cdots,\pm A\}\quad 及\quad b=A/m$$

也就是说 $\mathscr{A}$ 中的点将区间 $[-A,A]$ 等分为 $2m$ 个间隔，每个间隔长度为 A/m；扩展的区间 $[-A-b,A+b]$ 包含了 $2(m+1)$ 个这样的间隔。无需计算就可知道如何使概率分布 P_X 最大化：即 $m+1$ 个点都是等概率出现

$$-A,-A+2b,\cdots,A-2b,A$$

也就是说，我们"剔除"$\mathscr{A}$ 中每间隔两个的字母，使用剩下的字母且是等概率的。

事实上，当 $P_X(-A)=P_X(-A+2b)=\cdots=P_X(A)$ 时，输出信号 PDF f_Y 将值 $[2b(m+1)]^{-1}$ 分配到每个点 $y\in[-A-b,A+b]$，正如要求的那样，存在 $Y\sim U(-A-b,A-b)$，在这种情况下信息容量 C^{sd} 等于

388 $$\ln(2A+2b)-\ln 2b=\ln(1+m) \tag{4.1.32}$$

当 $M=3$ 时（三个输入信号，在 $-A$，0，A 处，且 $b=A$），$C^{\mathrm{sd}}=\ln 2$。当 $M=5$ 时（五个输入信号，在 $-A$，$-A/2$，0，$A/2$，A 处，且 $b=A/2$），$C^{\mathrm{sd}}=\ln 3$。当 $M=13$ 时见图 4-2。

图 4-2

备注 4.1.14 可以证明若(i)噪声随机变量 $Z\sim U(-b,b)$ 与 X 相互独立时(ii)X 在区间 $[-A,A]$ 上有广义分布，且 $b=A/m$，则从式(4.1.32)中可以得到输入和输出信号 X 和 $Y=X+Z$ 之间的最大互信息 $I(X;Y)$。因此根据 Kolmogorov，互信息 $I(X;Y)$ 可以被定义为：

$$I(X;Y)=\sup_{\xi,\eta} I(X_\xi;Y_\eta) \tag{4.1.33}$$

其中上确界取自区间 $[-A,A]$ 和 $[-A-b,A+b]$ 中所有有限分区 ξ 和 η，X_ξ 和 Y_η 分别表示随机变量量化之后的形式。

如果

$$P_X(-A)=P_X(-A+2b)=\cdots=P_X(A-2b)=P_X(A)=\frac{1}{m+1} \tag{4.1.34}$$

在满足假设(i)和(ii)条件下，输入信号分布 P_X 使 $I(X;Y)$ 最大。我们将这种分布表示为 $P_X^{[A,A/m]}$，或者 $P_X^{[m,b]}$。

假如 $M=2m$，即允许的信号数目 $\#\mathscr{A}$ 是偶数，则计算变得更复杂。很明显输出信号 Y 不可能达到均匀分布 $U(-A-b,A+b)$。为了使 $h(Y)=h(X+Z)$，必须限制在区间 $[-A-b,A+b]$ 上的分段常量 PDFs f_Y 类；证明如下。

在区间 $[-A,A]$ 上等分可以由点 $\pm A/(2m-1)$，$\pm 3A/(2m-1)$，…，$\pm A$ 生成；它们可以由公式 $\pm(2k-1)A/(2m-1)$ 描述，$k=1,\cdots,m$。这些点将区间 $[-A,A]$ 分割为 $(2m-1)$ 间隔，长度为 $2A/(2m-1)$。当 $Z\sim U(-b,b)$ 且 $A=b(m-1/2)$，在区间 $[-A-b,A+b]$ 上我们可以得到输出信号 PDF $f_Y(y)$：

$$f_Y(y)=\begin{cases} p_m/(2b), & b(m-1/2)\leqslant y\leqslant b(m+1/2) \\ (p_k+p_{k+1})/(2b), & b(k-1/2)\leqslant y\leqslant b(k+1/2),\ k=1,\cdots,m-1 \\ (p_{-1}+p_1)/(2b), & -b/2\leqslant y\leqslant b/2 \\ (p_k+p_{k+1})/(2b), & b(k-1/2)\leqslant y\leqslant b(k+1/2),\ k=-1,\cdots,-m+1 \\ p-m/(2b), & -b(m+1/2)\leqslant y\leqslant -b(m-1/2) \end{cases}$$

[389]

其中

$$p_{\pm k}=p_X\left(\pm b\left(k-\frac{1}{2}\right)\right)=P\left(X=\pm\frac{(2k-1)A}{2m-1}\right),\quad k=1,\cdots,m$$

代表了输入符号概率。熵 $h(Y)=h(X+Z)$ 写为

$$-\frac{p_m}{2}\ln\frac{p_m}{2b}-\sum_{1\leqslant k\leqslant m-1}\frac{p_k+p_{k+1}}{2}\ln\frac{p_k+p_{k+1}}{2b}-\frac{p_{-1}+p_1}{2}\ln\frac{p_{-1}+p_1}{2b}$$

$$-\sum_{-m\leqslant k\leqslant -1}\frac{p_k+p_{k+1}}{2}\ln\frac{p_k+p_{k+1}}{2b}-\frac{p-m}{2}\ln\frac{p-m}{2b}$$

可以看出最大分布 P_X 为 $p_{-k}=p_k$，$k=1,\cdots,m$。因此，我们面对一个优化问题

$$\text{最大化}\ G(\underline{p})=-p_m\ln\frac{p_m}{2b}-\sum_{1\leqslant k\leqslant m}(p_k+p_{k+1})\ln\frac{p_k+p_{k+1}}{2b}-p_1\ln\frac{p_1}{b} \tag{4.1.35}$$

服从概率约束 $p_k\geqslant 0$ 和 $2\sum_{1\leqslant k\leqslant m}p_k=1$。拉格朗日函数 $\mathscr{L}(P_X;\lambda)$ 如下

$$\mathscr{L}(P_X;\lambda)=G(\underline{p})-\lambda(2p_1+\cdots+2p_m-1)$$

且当下列条件满足时取得最大值。

$$\frac{\partial}{\partial p_k}\mathscr{L}(P_X;\lambda)=0,k=1,\cdots,m$$

因此，我们有 m 个等式，且左端相同：

$$-\ln\frac{p_m(p_{m-1}+p_m)}{4b^2}-2+2\lambda=0,(\text{逐步})p_m(p_{m-1}+p_m)=4b^2e^{2\lambda-2}$$

$$-\ln\frac{(p_{k-1}+p_k)(p_k+p_{k+1})}{4b^2}-2+2\lambda=0(\text{推出})(p_{k-1}+p_k)(p_k+p_{k+1})$$
$$=4b^2e^{\lambda-2},1<k<m$$
$$-\ln\frac{2p_1(p_1+p2)}{4b^2}-2+2\lambda=0(\text{推出})2p_1(p_1+p_2)=4b^2e^{\lambda-2}$$

由此导出

[390]
$$\left.\begin{aligned}p_m&=p_{m-1}+p_{m-2}=\cdots=p_2+p_2=2p_1\\p_m+p_{m-1}&=p_{m-2}+p_{m-3}=\cdots=p_2+p_1\end{aligned}\right\},m\text{是偶数}$$

且

$$\left.\begin{aligned}p_m&=p_{m-1}+p_{m-3}=\cdots=p_2+p_1\\p_m+p_{m-1}&=p_{m-2}+p_{m-3}=\cdots=p_1+p_2=2p_1\end{aligned}\right\},m\text{是奇数}$$

对于 $M=2m$ 的一些值，解是十分直观的，即，对 $M=2$(两个输入符号 $\pm A$ 且 $b=2A$)：$p_1=1/2$且最大输出符号 PDF 为

$$f_Y(y)=\begin{cases}1/(4b), & A\leqslant y\leqslant 3A\\1/(2b), & -A\leqslant y\leqslant A,\ \text{得}\ C^{\text{ef}}=(\ln 2)/2\\1/(4b), & -3A\leqslant y\leqslant -A\end{cases}$$

对于 $M=4$(四个输入符号 $-A$，$-A/3$，$A/3$，A 且 $b=2A/3$)：$p_1=1/6$，$p_2=1/3$ 且最大输入符号 PDF 为

$$f_Y(y)=\begin{cases}1/(6b), & A\leqslant y\leqslant 5A/3\ \text{且}\ -5A/3\leqslant y\leqslant -A\\1/(4b), & 2A/3\leqslant y\leqslant A\ \text{且}\ -A\leqslant y\leqslant -2A/3\\1/(6b), & -2A/3\leqslant y\leqslant 2A/3\end{cases}$$

从而导出 $C^{\text{ef}}=\ln(6^{1/3}4^{1/3}/2)$。

对于 $M=6$(六个输入符号 $-A$，$-3A/5$，$-A/5$，$A/5$，$3A/5$，A 且 $b=2A/5$)：$p_1=1/6$，$p_2=1/12$，$p_3=1/4$。类似地，对于 $M=8$(八个输入符号 $-A$，$-5A/7$，$-3A/7-A/7$，$A/7$，$3A/7$，$5A/7$，A 且 $b=2A/7$)：$p_1=1/10$，$p_2=3/20$，$p_3=1/20$，$p_4=1/5$。

实际上，我们可以将所有的概率用 p_1 表示。即对偶数 m：

$$p_m=2p_1$$
$$p_{m-1}=p_2-p_1$$
$$p_{m-2}=3p_2-p_2$$
$$p_{m-3}=2(p_3-p_1)$$
$$p_{m-4}=4p_1-2p_2$$
$$\vdots$$
$$p_3=\left(\frac{m}{2}-1\right)(p_2-p_1)$$

[391]
$$p_2=\frac{m+2}{m}p_1$$

由此

$$p_3=\frac{m+2}{m}p_1,$$
$$p_2=\frac{m-2}{m}p_1$$
$$p_4=\frac{m+4}{m}p_1$$

$$p_3=\frac{m-4}{m}p_1$$
$$\vdots$$
$$p_{m-2}=\frac{2m-2}{m}$$
$$p_{m-1}=\frac{2}{m}$$
$$p_m=2p_1$$
$$p_1=\frac{1}{2(m+1)} \tag{4.1.36}$$

对应的 PDF f_Y 给出值

$$h(Y)=-\frac{1}{2}\ln\frac{1}{4m(m+1)b^2} \quad 和 \quad C_{\mathscr{A}}^{\mathrm{sq}}=-\frac{1}{2}\ln\frac{1}{4m(m+1)}-\ln 2 \tag{4.1.37}$$

另一方面，对于一般的奇数 m，最大输入符号分布 P_X 为

$$p_1=\frac{m+1}{2m(m+1)}$$
$$p_2=\frac{m-1}{2m(m+1)}$$
$$p_3=\frac{m+3}{2m(m+1)}$$
$$p_4=\frac{m-3}{2m(m+1)}$$
$$\vdots$$
$$p_{m-1}=\frac{1}{2m(m+1)}$$
$$p_m=\frac{m}{m(m+1)} \tag{4.1.38}$$

392

这导出了对于最大熵和约束容量相同的答案

$$h(Y)=-\frac{1}{2}\ln\frac{1}{4m(m+1)b^2} \quad 和 \quad C_{\mathscr{A}}^{\mathrm{sq}}=-\frac{1}{2}\ln\frac{1}{4m(m+1)}-\ln 2 \tag{4.1.39}$$

接下来，我们将(4.1.36)和(4.1.38)中的特定输入符号分布用 $\bar{P}_X^{(A,2A/(2m-1))}$ 表示。

备注 4.1.15 当(i)噪声随机变量 $Z\sim U(-b,b)$ 独立于 X 且(ii)输入符号分布 P_X 约束于 $[-A,A]$ 且 $b=2A/(2m-1)$ 或为任意值(此时 $I(X;Y)$ 定义为(4.1.33))，很自然地，上述公式给出了最大互信息 $I(X;Y)$。进一步推测，最大值的取得在上述假设下(i)和(ii)且对任意 $A>b>0$，A/b 不一定为整数或半整数。这里 $M=2A/b+1$ 也不是整数，但值得作为一个参考变量。

所以当 b 从 A/m 变为 $A/(m+1)$(或者等价地，A 从 bm 增加到 $b(m+1)$，且相应地，M 从 $2m+1$ 增加到 $2m+3$)，最大值 $P_X^{(A,b)}$ 从 $P_X^{(A,A/m)}$ 增加到 $P_X^{(A,A/(m+1))}$；当 $A=b(m+1/2)$(即 $M=2(m+1)$)，分布 $P_X^{(A,b)}$ 可能或可能不与式(4.1.36)和式(4.1.38)中的分布 $\bar{P}_X^{(A,b)}$ 一致。

为了(部分)阐释清楚该问题，考虑 $A/2\leqslant b\leqslant A$(即 $3\leqslant M\leqslant 5$)并假设输入符号分布 P_X 如下

$$P_X(-A)=P_X(A)=p \quad 且 \quad P_X(0)=1-2p$$
$$其中\ 0\leqslant p\leqslant\frac{1}{2} \tag{4.1.40}$$

于是

$$h_Y(Y)=-\frac{1}{b}\left(Ap\ln\frac{p}{2b}+(2b-A)(1-p)\ln\frac{1-p}{2b}+(A-b)(1-2p)\ln\frac{1-2p}{2b}\right) \tag{4.1.41}$$

且等式 $dh(Y)/dp=0$ 等价于

$$p^A=(1-p)^{2b-A}(1-2p)^{2A-b} \tag{4.1.42}$$

对于 $b=A/2$，这导出了 $p^A=(1-2p)^A$，即 $p=1-2p$，由此可得 $p=1/3$；类似地，对于 $b=A$，$p=1/2$，这些和之前得到的结论一致。对于 $b=2A/3$，我们有

$$p^A=(1-p)^{A/3}(1-2p)^{2A/3}$$

即

[393]

$$p^3=(1-p)(1-2p)^2 \tag{4.1.43a}$$

我们对在(0，1/2)之间的解感兴趣(实际上，在(1/3，1/2)之间)。当 $b=3A/4$ 时，等式变为

$$p^A=(1-p)^{A/2}(1-2p)^{A/2}$$

即

$$p^2=(1-p)(1-2p) \tag{4.1.43b}$$

由此 $p=(3-\sqrt{5})/2$。

例子 4.1.16 考虑以下例子，噪声随机变量 Z 有两部分：离散和连续。假设

$$f_Z(z)=q\delta_0+(1-q)\phi(z;\sigma^2)$$

即 $Z=0$ 的概率为 q 且 $Z\sim N(0,\sigma^2)$ 的概率为 $1-q\in(0,1)$。(所以 $1-q$ 给出了总错误概率。)这里我们考虑以下情形

$$f_Z=q\delta_0+(1-q)\frac{1}{2b}\mathbf{1}(|z|\leqslant b)$$

并且输入信号 PMF 的形式为

$$P_X(-A)=p_{-1},P_X(0)=p_0,P_X(A)=p_1 \tag{4.1.44a}$$

[394] 其中

$$p_{-1},p_0,p_1\geqslant 0,p_{-1}+p_0+p_1=1 \tag{4.1.44b}$$

且 $b=A$，$M=3$($-A$，A)中有三个信号等级)。输入信号熵为

$$h(X)=h(p_{-1},p_0,p_1)=-p_{-1}\ln p_{-1}-p_0\ln p_0-p_1\ln p_1$$

输出信号 PMF 的形式为

$$f_Y(y)=q(p_{-1}\delta_{-A}+p_0\delta_0+p_1\delta_A)+(1-q)\frac{1}{2b}$$
$$\times[p_{-1}\mathbf{1}(-2A\leqslant y\leqslant 0)+p_0\mathbf{1}(-A\leqslant y\leqslant A)+p_1\mathbf{1}(0\leqslant y\leqslant 2A)]$$

且其熵 $h(Y)$(相对于 $\mathcal{R}$ 中的参考计量 μ 来计算，其中 μ 的绝对连续部分符合 Lebesgue 且离散部分在 $-A$，0，A 处的取值为 1)由下式给出

$$h(Y)=-q\ln q-(1-q)\ln(1-q)-qh(p_{-1},p_0,p_1)$$
$$-(1-q)A\left[p_{-1}\ln\frac{p_{-1}}{2A}+(p_{-1}+p_0)\ln\frac{p_{-1}+p_0}{2A}\right.$$
$$\left.(p_0+p_1)\ln\frac{p_0+p_1}{2A}+p_1\ln\frac{p_1}{2A}\right]$$

[394] 根据对称性，当 $p_{-1}=p_1=p$，$p_0=1-2p$ 时，$h(Y)$ 取得最大值，且 $q\in(0,1)$ 时，我们必须最大化下式

$$h(Y)=h(q,1-q)-qh(p,p,1-2p)-(1-q)A\left[2p\ln\frac{p}{2A}+(1-2p)\ln\frac{1-2p}{2A}\right]$$

求导得到

$$\frac{\mathrm{d}}{\mathrm{d}p}h(Y)=0\Leftrightarrow\frac{p}{1-2p}=\left(\frac{p}{1-p}\right)^{-(1-q)A/q}$$

如果$(1-q)A/q>1$该等式导出一个唯一解，其确定了最优的输入信号分布P_X，形式为式(4.1.44a)～(4.1.44b)。

如果我们想要用q导出$h(Y)$的最大值(即信息容量若最大值)，则对q求导：

$$\frac{\mathrm{d}}{\mathrm{d}q}h(Y)=0\Leftrightarrow\log\frac{q}{1-q}=(A-1)h(p,p,1-2p)-2A\ln 2A$$

如果考虑一个$[-A,A]$上连续分布的输入信号，且PDF为$f_X(x)$，则输出随机变量$Y=X+Z$的PDF由卷积给出：

$$f_Y(y)=\frac{1}{2b}\int_{(y-b)\vee(-A)}^{(y+b)\wedge A}f_X(x)\mathrm{d}x$$

微分熵$h(Y)=-\int f_Y(y)\ln f_Y(y)\mathrm{d}y$用$f_X$的形式表示为

$$h(X+Z)=-\frac{1}{2b}\int_{-A}^{A}f_X(x)\int_{-b}^{b}\ln\left[\frac{1}{2b}\int_{(z+x-b)\vee(-A)}^{(z+x+b)\wedge A}f_X(x')\mathrm{d}x'\right]\mathrm{d}z\mathrm{d}x$$

PDFf_X最小化微分熵$h(X+Z)$导出解如下

$$0=\int_{-b}^{b}\left(\ln\left[\frac{1}{2b}\int_{(x+z-b)\vee(-A)}^{(x+z+b)\wedge A}f_X(x')\mathrm{d}x'\right]+f_X(x)\right.$$
$$\left.\times\left[\int_{(x+z-b)\vee(-A)}^{(x+z+b)\wedge A}f_X(x')\mathrm{d}x'\right]^{-1}[f_X(x+z+b)-f_X(x+z-b)]\right)\mathrm{d}z$$

当考虑一个单位噪声的二时每信号的信道时，有一个有趣的问题。假设一个输入信号用一个$\mathbb{R}^2$上的点$x=(x_1,x_2)$表示，跟之前一样假设$Z\sim U(-b,b)$且独立于输入信号。则平方$S_b(x)=(x_1-b,x_1+b)\times(x_2-b,x_2+b)$且PDF为$1/(4b^2)$，给定输入信号$X=(x_1,x_2)$下，确定了输出信号$Y$的可能位置。假设我们要处理一个有限输入符号集$\mathscr{A}\subset\mathbb{R}^2$，则输出信号域为有限集$\mathscr{B}=\bigcup_{x\in\mathscr{A}}S(x)$。以上的论述说明了，如果我们能找到一个子集$\mathscr{A}'\subseteq\mathscr{A}$使得$S_b(x)$且$x\in\mathscr{A}'$划分区域$\mathscr{B}$(即覆盖$\mathscr{B}$但不相交)，则对于输入PMF$P_X$且$P_X(x)=1/(\#\mathscr{A}')$(在$\mathscr{A}'$上的均匀分布)，输出向量信号PDF$f_Y$在$\mathscr{B}$上为均匀分布(即$f_Y(y)=1/(\mathscr{B}$的面积$)$)。因此，输出信号熵$h(Y)=\ln(\mathscr{B}$的面积$)$在所有输入信号PMFs$P_X$且$P_X(\mathscr{A})=1$中为最大值(即使在$P_X(\mathscr{A}')=1$且$\mathscr{A}'\subset\mathscr{B}$为一个仿射$\bigcup_{x\in\mathscr{B}}S(x')$位于$\mathscr{B}$中的任意集合，也为最大值)。最后，在信道的信息容量为

395

$$C^{\mathrm{sd}}=\frac{1}{2}\ln\frac{\mathscr{B}\text{的面积}}{4b^2}\text{nats/(标量输入符号)}$$

参考图4.3。

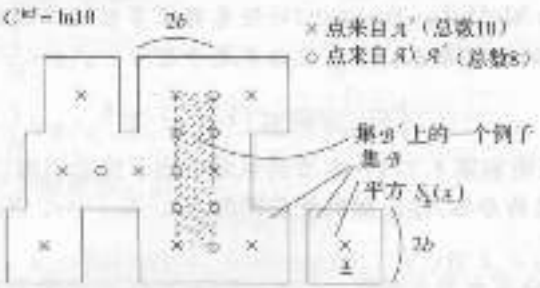


图 4.3

为了有所区别，任意可以被划分为不同的长度为 $2b$ 的区域的有界集 $\mathbb{D}_2 \subset \mathbb{R}^2$ 可导出信息容量

$$C_2^{\text{H}} = \frac{1}{2}\ln\frac{\mathbb{D}^2\text{的面积}}{4b^2}\ \text{nats/(标量输入符号)}$$

加性噪声在 $(-b, b)$ 上均匀分布，且每个任意的输入信号在信道上经过两次，随机向量输入 $x=(X_1, X_2)$ 服从区域限制 $x \in \mathbb{D}_2$。最大输入向量 PMF 对中心形成块分配相等的概率。

一个类似的结论在 $\mathbb{R}^3$ 上仍成立，当每个输入信号在信道上经过三次时，即输入信号是一个三维向量 $x=(x_1, x_2, x_3)$。客观地，当我们用一个 K 维的输入信号 $x=(x_1, \cdots, x_K) \in \mathbb{R}^K$ 且区域限制为 $x \in \mathbb{D}_K \subset \mathbb{R}^K$，其中 $\mathbb{D}_K$ 为一个能被划分为长度为 $2b$ 的不同的立方体的有界区域，对应的信息容量

[396]

$$C_K^{\text{H}} = \frac{1}{K}\ln\frac{\mathbb{D}_K\text{的容积}}{(2b)^K}\ \text{nats/(标量输入符号)}$$

当输入向量信号 PMFP 对中心形成立方体分配相同的权重时达到该容量。

当 $K \to \infty$ 时，C_K 可能收敛于一个极限值 C_∞^{H}，从而导出在区域限制 $\mathbb{D}_K$ 下的平均每个任意输入信号容量。这种情形下的一个例子是，当 $\mathbb{D}_K$ 为一个 K 维的立方体

$$S_n^K = (-2bn, 2bn)^{\times K}$$

于是 $C_K^{\text{H}} = \ln(1+m)$ 不随 K 的变化而变化（且信道是无记忆的）。

4.2 连续时间集的渐近均分性

宁可用智者的错误来制定你的原则，而不是愚者的完美。

William Blake(1757—1821)，英国诗人

本节提供了定理 4.1.6 证明中的一个缺失步骤和额外的例子。我们先从一些渐近均分性的各种形式的说明开始，中心事实基于 Shannon-McMillan-Breiman(SMB)定理，该定理被认为是信息论的基础。该定理给出了平稳遍历过程 $\boldsymbol{X}=(X_n)$ 的信息速率。回顾概率空间 T 的一种转换形式为遍历的，当每个集合 A 满足 $TA=A$ 且 $\mathbb{P}(A)=0$ 或 1。对具有有限期望值的一个平稳遍历信源，Birkhoff 的遍历定理说明了大数定理(概率为 1)：

$$\frac{1}{n}\sum_{i=1}^{n} X_i \to \mathbb{E}X \tag{4.2.1}$$

典型地，对于一个遍历过程的可测函数 $f(X_i)$，

$$\frac{1}{n}\sum_{i=1}^{n} f(X_i) \to \mathbb{E}f(X) \tag{4.2.2}$$

定理 4.2.1 (Shannon-McMillan-Breiman)对任意有限多值的平稳遍历过程 X，信息速率 $R=h$，即，(4.2.3)中的极限存在，当其收敛并趋于 h

[397]

$$-\lim_{n\to\infty}\frac{1}{n}\log p_{X_0^{n-1}}(X_0^{n-1}) = h \tag{4.2.3}$$

需要一些辅助引理来证明定理 4.2.1，本节的末尾给出了这些引理。

举例 4.2.2 (一般渐近均分性)给定随机变量序列 $X_1, X_2, \cdots, X_N$，其中 $N=1, 2, \cdots$，随机向量 $x_1^N = \begin{bmatrix} X_1 \\ \vdots \\ X_N \end{bmatrix}$ 的分布由关于测度 $\mu^{(N)} = \mu \times \mu \cdots \times \mu$ (N 个因子)的 PMF $f_{x_1^N}(x_1^N)$ 决定。

假设 Shannon-McMillan-Breiman 定理是正确的：

$$-\frac{1}{N}\log f_{X^N}(x^N) \text{ 依概率收敛于 } h$$

其中 h 为一个大于 0 的常数(典型地，$h=\lim_{i\to\infty} h(X_i)$)。给定 $\varepsilon>0$，考虑典型集

$$S_\varepsilon^N=\left\{x_1^N=\begin{pmatrix}x_1\\ \vdots\\ x_N\end{pmatrix}: \varepsilon \leqslant \frac{1}{N}\log f_{X^N}(x_1^N)+h \leqslant \varepsilon\right\}$$

集合 S_ε^N 的体积 $\mu^{(N)}(S_\varepsilon^N)=\int_{S_\varepsilon^N}\mu(\mathrm{d}x_1)\cdots\mu(\mathrm{d}x_N)$ 具有如下性质：

$$\mu^{(N)}(S_\varepsilon^N)\leqslant 2^{N(h+\varepsilon)}, \quad \text{对于所有的 } \varepsilon \text{ 和 } N \tag{4.2.4}$$

且，对于 $0<\varepsilon<h$ 和所有的 $\delta>0$，

$$\mu^{(N)}(S_\varepsilon^N)\geqslant(1-\delta)2^{N(h-\varepsilon)}, \quad \text{对于足够大的 } N\text{，依赖于 } \delta \tag{4.2.5}$$

解答　由于 $\mathbb{P}(\mathbb{R}^N)=\int_{\mathbb{R}^N} f_{X_1^N}(x_1^N)\prod_{1\leqslant j\leqslant N}\mu(\mathrm{d}x_j)=1$，我们有

$$\begin{aligned}1&=\int_{\mathbb{R}^N} f_{X_1^N}(x^N)\prod_{1\leqslant j\leqslant N}\mu(\mathrm{d}x_j)\geqslant\int_{S_\varepsilon^N} f_{X_1^N}(x_1^N)\prod_{1\leqslant j\leqslant N}\mu(\mathrm{d}x_j)\\&\geqslant 2^{-N(h+\varepsilon)}\int_{S_\varepsilon^N}\prod_{1\leqslant j\leqslant N}\mu(\mathrm{d}x_j)=2^{-N(h+\varepsilon)}\mu^{(N)}(S_\varepsilon^N)\end{aligned}$$

398

上式给定了上界(4.2.4)。另一方面，给定 $\delta>0$，当 N 充分大时，$\mathbb{P}(S_\varepsilon^N)\geqslant 1-\delta$。此时，对于 $0<\varepsilon<h$，

$$\begin{aligned}1-\delta\leqslant\mathbb{P}(S_\varepsilon^N)&=\int_{S_\varepsilon^N} f_{X_1^N}(x_1^N)\prod_{1\leqslant j\leqslant N}\mu(\mathrm{d}x_j)\\&\leqslant 2^{-N(h-\varepsilon)}\int_{S_\varepsilon^N}\prod_{1\leqslant j\leqslant N}\mu(\mathrm{d}x_j)=2^{-N(h-\varepsilon)}\mu^{(N)}(S_\varepsilon^N)\end{aligned}$$

这就导出了下界(4.2.5)。 □

接下来将渐近均分型扩展到 $\boldsymbol{X}_1^N$ 和 $\boldsymbol{Y}_1^N$ 的联合分布(在实际应用中，$\boldsymbol{X}_1^N$ 可以是某个信道的输入而 $\boldsymbol{Y}_1^N$ 为对应的输出)。给定两个随机变量序列 $X_1, X_2, \cdots$ 和 $Y_1, Y_2, \cdots$，且 $N=1, 2, \cdots$，随机向量 $\boldsymbol{X}_1^N=\begin{pmatrix}X_1\\ \vdots\\ X_N\end{pmatrix}$ 和 $\boldsymbol{Y}_1^N=\begin{pmatrix}Y_1\\ \vdots\\ Y_N\end{pmatrix}$ 的联合分布由测度 $\mu^{(N)}\times\nu^{(N)}$ 的联合 PMF $f_{X_1^N,Y_1^N}$ 决定，其中 $\mu^{(N)}=\mu\times\mu\cdots\times\mu$，$\nu^{(N)}=\nu\times\nu\cdots\times\nu$(各有 N 个因子)。令 $f_{X_1^N}$ 和 $f_{Y_1^N}$ 分别为向量 $\boldsymbol{X}_1^N$ 和 $\boldsymbol{Y}_1^N$ 的(联合)PMF。

正如例 4.2.2 所示，我们假设 Shannon-McMillan-Breiman 定理是正确的，对于 $(\boldsymbol{X}_1^N, \boldsymbol{Y}_1^N)$ 和每一个 $\boldsymbol{X}_1^N$ 和 $\boldsymbol{Y}_1^N$，随着 $N\to\infty$，

$$-\frac{1}{N}\log f_{X^N}(X_1^N)\to h_1, \quad -\frac{1}{N}\log f_{Y_1^N}(\boldsymbol{Y}_1^N)\to h_2$$

$$-\frac{1}{N}\log f_{X_1^N,Y_1^N}(X_1^N,\boldsymbol{Y}_1^N) \text{ 依概率收敛于 } h$$

其中 h_1，h_2 和 h 为大于 0 的常数，且

$$h_1+h_2\geqslant h \tag{4.2.6}$$

典型地，$h_1=\lim_{i\to\infty} h(X_i)$，$h_2=\lim_{i\to\infty} h(Y_i)$，$h=\lim_{i\to\infty} h(X_i, Y_i)$ 且 $h_1+h_2-h=\lim_{i\to\infty} I(X_i; Y_i)$。给定 $\varepsilon>0$，考虑由采样 (x_1^N, y_1^N) 组成的典型集，其中

399

$$x_1^N=\begin{bmatrix}x_1\\ \vdots\\ x_N\end{bmatrix},\quad y_1^N=\begin{bmatrix}y_1\\ \vdots\\ y_N\end{bmatrix}$$

且

$$\begin{aligned}T_\varepsilon^N=\{(x_1^N,y_1^N):&-\varepsilon\leqslant\frac{1}{N}\log f_{X_1^N}(x_1^N)+h_1\leqslant\varepsilon\\ &-\varepsilon\leqslant\frac{1}{N}\log f_{Y_1^N}(y_1^N)+h_2\leqslant\varepsilon\\ &-\varepsilon\leqslant\frac{1}{N}\log f_{X_1^N,Y_1^N}(x_1^N,y_1^N)+h\leqslant\varepsilon\}\end{aligned}\tag{4.2.7}$$

根据上述假设，对于任意的$\varepsilon>0$，有$\lim\limits_{N\to\infty}\mathrm{P}(T_\varepsilon^N)=1$。接下来，定义$T_\varepsilon^N$的体积：

$$\mu^{(N)}\times\nu^{(N)}(T_\varepsilon^N)=\int_{T_\varepsilon^N}\mu^{(N)}(\mathrm{d}\,x_1^N)\nu^{(N)}(\mathrm{d}\,y_1^N)$$

最后，考虑独立组合$(\tilde{X}_1^N,\tilde{Y}_1^N)$，其中$\tilde{X}_1^N$和$X_1^N$的PMF相同，$\tilde{Y}_1^N$和$Y_1^N$的PMF相同。即$\tilde{X}_1^N$和$\tilde{Y}_1^N$的联合PMF具有如下形式

$$f_{\tilde{X}_1^N,\tilde{Y}_1^N}(x_1^N,y_1^N)=f_{X_1^N}(x_1^N)f_{Y_1^N}(y_1^N)\tag{4.2.8}$$

接下来，我们估计T_ε^N的体积及对应的$(\tilde{X}_1^N,\tilde{Y}_1^N)\in T_\varepsilon^N$的概率。

举例4.2.3 （一般联合渐近均分性）(Ⅰ)典型集的体积具有如下性质：

$$\mu^{(N)}\times\nu^{(N)}(T_\varepsilon^N)\leqslant 2^{N(h+\varepsilon)},\quad 对于所有的\varepsilon和N\tag{4.2.9}$$

且，对任意的$\delta>0$，$0<\varepsilon<h$，当N充分大时，

$$\mu^{(N)}\times\nu^{(N)}(T_\varepsilon^N)\geqslant(1-\delta)2^{N(h-\varepsilon)}\tag{4.2.10}$$

(Ⅱ)对独立组合$(\tilde{X}_1^N,\tilde{Y}_1^N)$，

403

$$\mathrm{P}((\tilde{X}_1^N,\tilde{Y}_1^N)\in T_\varepsilon^N)\leqslant 2^{-N(h_1+h_2-h-3\varepsilon)},\quad 对于所有的\varepsilon和N\tag{4.2.11}$$

$$\mathrm{P}((\tilde{X}_1^N,\tilde{Y}_1^N)\in T_\varepsilon^N)\geqslant(1-\delta)2^{-N(h_1+h_2-h+3\varepsilon)},\quad 对于所有的\varepsilon\tag{4.2.12}$$

解答 (Ⅰ)和式(4.2.4)及式(4.2.5)的证明相同，其中用$f_{X_1^N,Y_1^N}$替换对应的部分。

(Ⅱ)对于概率$\mathrm{P}((\tilde{X}_1^N,\tilde{Y}_1^N)\in T_\varepsilon^N)$，根据式(4.2.11)可得：

$$\begin{aligned}\mathrm{P}((\tilde{X}_1^N,\tilde{Y}_1^N)\in T_\varepsilon^N)&=\int_{T_\varepsilon^N}f_{\tilde{X}_1^N,\tilde{Y}_1^N}\mu(\mathrm{d}\,x_1^N)\nu(\mathrm{d}\,y_1^N)\ 由定义\\ &=\int_{T_\varepsilon^N}f_{X_1^N}(x_1^N)f_{Y_1^N}(y_1^N)\mu(\mathrm{d}\,x_1^N)\nu(\mathrm{d}\,y_1^N)\ 代入式(4.2.8)\\ &\leqslant 2^{-N(h_1-\varepsilon)}2^{-N(h_2-\varepsilon)}\int_{T_\varepsilon^N}\mu(\mathrm{d}\,x_1^N)\nu(\mathrm{d}\,y_1^N)\ 由式(4.2.7)\\ &\leqslant 2^{-N(h_1-\varepsilon)}2^{-N(h_2-\varepsilon)}2^{N(h+\varepsilon)}=2^{-N(h_1+h_2-h-3\varepsilon)}\ 由下界(4.2.9)\end{aligned}$$

最后，反转最后两行的不等式，可得

$$\begin{aligned}&\geqslant 2^{-N(h_1+\varepsilon)}2^{-N(h_2+\varepsilon)}\int_{T_\varepsilon^N}\mu(\mathrm{d}\,x_1^N)\nu(\mathrm{d}\,y_1^N)\ 由式(4.2.7)\\ &\geqslant(1-\delta)2^{-N(h_1+\varepsilon)}2^{-N(h_2+\varepsilon)}2^{N(h-\varepsilon)}=(1-\delta)2^{-N(h_1+h_2-h-3\varepsilon)}\ 由下界(4.2.10)\end{aligned}$$

当然，我们假设$0<\varepsilon<h$(基于式(4.2.10)的假设)，但增加ε只会使因子$2^{-N(h_1+h_2-h-3\varepsilon)}$变小。这证明了边界(4.2.12)。 □

渐近均分性的一个更方便(也更广义)的扩展为，假设对比值$f_{X_1^N,Y_1^N}(\boldsymbol{X}_1^N,\boldsymbol{Y}_1^N)/[f_{X_1^N}(\boldsymbol{X}_1^N)f_{Y_1^N}(\boldsymbol{Y}_1^N)]$，Shannon-McMillan-Breiman定理仍成立，即

$$\frac{1}{N}\log\frac{f_{X_1^N,Y_1^N}(\boldsymbol{X}_1^N,\boldsymbol{Y}_1^N)}{f_{X_1^N}(\boldsymbol{X}_1^N)f_{Y_1^N}(\boldsymbol{Y}_1^N)} \text{ 依概率收敛于 } c \tag{4.2.13}$$

431

其中 c 为常数。注意到 $f_{X_1^N,Y_1^N}$ 为联合 PMF，$f_{X_1^N}$ 和 $f_{Y_1^N}$ 分别为随机输入和输出向量 $\boldsymbol{X}^N$ 和 $\boldsymbol{Y}^N$ 的独立 PMF，且参考测度为 $\mu^{(N)}$ 和 $\nu^{(N)}$：

$$f_{X_1^N,Y_1^N}(\boldsymbol{x}_1^N,\boldsymbol{y}_1^N)=f_{X_1^N}(\boldsymbol{x}_1^N)f_{Y_1^N}(\boldsymbol{y}_1^N|\boldsymbol{x}_1^N \text{ 发送})$$

$$f_{Y_1^N}(\boldsymbol{Y}_1^N)=\int f_{X_1^N,Y_1^N}(\boldsymbol{x}_1^N,\boldsymbol{y}_1^N)\mu^{(N)}(\mathrm{d}\boldsymbol{x}_1^N)$$

这里，对 $\varepsilon>0$，考虑典型集

$$T_\varepsilon^N=\left\{(\boldsymbol{X}_1^N,\boldsymbol{y}_1^N):\ \varepsilon\leqslant\frac{1}{N}\log\frac{f_{X_1^N,Y_1^N}(\boldsymbol{x}_1^N,\boldsymbol{y}_1^N)}{f_{X_1^N}(\boldsymbol{x}_1^N)f_{Y_1^N}(\boldsymbol{y}_1^N)}-c\leqslant\varepsilon\right\} \tag{4.2.14}$$

根据假设式(4.2.13)可得 $\lim\limits_{N\to\infty}\mathbf{P}((\boldsymbol{X}_1^N,\boldsymbol{Y}_1^N)\in T_\varepsilon^N)=1$ 对任意 $\varepsilon>0$ 恒成立。

我们将再次考虑独立组合 $(\overline{\boldsymbol{X}}_1^N,\overline{\boldsymbol{Y}}_1^N)$，其中 $\overline{\boldsymbol{X}}_1^N$ 的 PMF 和 $\boldsymbol{X}_1^N$ 相同，$\overline{\boldsymbol{Y}}_1^N$ 的 PMF 和 $\boldsymbol{Y}_1^N$ 相同。

定理 4.2.4 （和联合渐近均分性相反）假设性质(4.2.13)成立。对于独立组合 $(\overline{\boldsymbol{X}}_1^N,\overline{\boldsymbol{Y}}_1^N)$，$(\boldsymbol{X}_1^N,\boldsymbol{Y}_1^N)\in T_\varepsilon^N$ 的概率服从

$$\mathbf{P}((\overline{\boldsymbol{X}}_1^N,\overline{\boldsymbol{Y}}_1^N)\in T_\varepsilon^N)\leqslant 2^{-N(c-\varepsilon)},\quad \text{对于所有的 } \varepsilon \text{ 和 } N \tag{4.2.15}$$

且，对所有的 $\delta>0$，当 N 充分大时，

$$\mathbf{P}((\overline{\boldsymbol{X}}_1^N,\overline{\boldsymbol{Y}}_1^N)\in T_\varepsilon^N)\geqslant(1-\delta)2^{-N(c+\varepsilon)},\quad \text{对于所有的 } \varepsilon \tag{4.2.16}$$

证明 推导出式(4.2.15)的过程如下：

$$\begin{aligned}
\mathbf{P}((\overline{\boldsymbol{X}}_1^N,\overline{\boldsymbol{Y}}_1^N)\in T_\varepsilon^N)&=\int_{T_\varepsilon^N}f_{\overline{X}_1^N,\overline{Y}_1^N}\mu^{(N)}(\mathrm{d}\boldsymbol{X}_1^N)\nu^{(N)}(\mathrm{d}\boldsymbol{y}_1^N)\\
&=\int_{T_\varepsilon^N}f_{X_1^N}(\boldsymbol{x}_1^N)f_{Y_1^N}(\boldsymbol{y}_1^N)\mu(\mathrm{d}\boldsymbol{X}_1^N)\nu(\mathrm{d}\boldsymbol{y}_1^N)\\
&=\int_{T_\varepsilon^N}\exp\left(-\frac{f_{X_1^N,Y_1^N}(\boldsymbol{x}_1^N,\boldsymbol{y}_1^N)}{f_{X_1^N}(\boldsymbol{x}_1^N)f_{Y_1^N}(\boldsymbol{y}_1^N)}\right)\\
&\quad\times f_{X_1^N,Y_1^N}(\boldsymbol{x}_1^N,\boldsymbol{y}_1^N)\mu^{(N)}(\mathrm{d}\boldsymbol{x}_1^N)\nu^{(N)}(\mathrm{d}\boldsymbol{y}_1^N)\\
&\leqslant 2^{-N(c-\varepsilon)}\int_{T_\varepsilon^N}f_{X_1^N,Y_1^N}(\boldsymbol{x}_1^N,\boldsymbol{y}_1^N)\mu(\mathrm{d}\boldsymbol{x}_1^N)\nu(\mathrm{d}\boldsymbol{y}_1^N)\\
&=2^{-N(c-\varepsilon)}\mathbf{P}((\boldsymbol{X}_1^N,\boldsymbol{Y}_1^N)\in T_\varepsilon^N)\\
&\leqslant 2^{-N(c-\varepsilon)}
\end{aligned}$$

402

第一个等式根据定义得到，通过代入式(4.2.8)得到第二个等式，第三个等式由直接计算得到，根据边界(4.2.14)得出第四个等式。 □

最后，通过反转最后两行的不等式，我们得到边界(4.2.16)：

$$\begin{aligned}
&\geqslant 2^{-N(c+\varepsilon)}\int_{T_\varepsilon^N}f_{X_1^N,Y_1^N}(\boldsymbol{x}_1^N,\boldsymbol{y}_1^N)\mu(\mathrm{d}\boldsymbol{x}_1^N)\nu(\mathrm{d}\boldsymbol{y}_1^N)\\
&=2^{-N(c+\varepsilon)}\mathbf{P}((\boldsymbol{X}_1^N,\boldsymbol{Y}_1^N)\in T_\varepsilon^N)\geqslant 2^{-N(c+\varepsilon)}(1-\delta)
\end{aligned}$$

第一个不等式可由式(4.2.14)得到。

举例 4.2.5 令 $\boldsymbol{x}=\{X(1),\cdots,X(n)\}^T$ 为随机变量的向量/集合，$\boldsymbol{x}(C)$ 为子集 $\{X(i):i\in C\}$，其中 C 为一个下标集 $\{1,\cdots,n\}$ 的非空子集。假设满足 $|C|=k$，$1\leqslant k\leqslant n$ 的任意子集 $\boldsymbol{x}(C)$ 的联合分布由关于测度 $\mu\times\cdots\times\mu$（k 个因子）每个对应一个随机变量 $X(i)$ 定

$i\in C$)的联合 PMF $f_{x(C)}$ 给出。类似地，给定 x 的值为向量 $x=\begin{bmatrix}x(1)\\ \vdots \\ x(n)\end{bmatrix}$，将其用参数 $\{x(i): i\in C\}$ 表示为 $x(C)$(取 x 的第 i 列且 $i\in C$)。根据 Gibbs 不等式，集合 $\{1,\cdots,n\}$ 可划分为非空不相交的子集 $C_1,\cdots,C_s$，对于所有的划分 $\{C_1,\cdots,C_s\}$ $(1\leqslant s\leqslant n)$，积分如下

$$\int f_x(x)\log\frac{f_x(x)}{f_{x(C_1)}(x(C_1))\cdots f_{x(C_s)}(x(C_s))}\prod_{1\leqslant j\leqslant n}\mu(\mathrm{d}x(j))\geqslant 0 \tag{4.2.17}$$

当式(4.2.17)中的积分取得最大值时，应当如何进行对应的划分？

解答 问题中的划分有 $s=n$ 个子集，每一个由单个点组成。事实上，考虑将集合 $\{1,\cdots,n\}$ 划分为单点，其对应的积分等于

$$\int f_x(x)\log\frac{f_x(x)}{\prod\limits_{1\leqslant j\leqslant n}f_{x_j}(x_j)}\prod_{1\leqslant j\leqslant n}\mu(\mathrm{d}x(j)) \tag{4.2.18}$$

令 $\{C_1,\cdots,C_s\}$ 为 $\{1,\cdots,n\}$ 的任意划分，用联合 PMF 的乘积 $\prod\limits_{1\leqslant j\leqslant s}f_{x(C_j)}(x(C_j))$ 去除 log 中的因子，则积分(4.2.18)化为如下和的形式

$$\int f_x(x)\log\frac{f_x(x)}{\prod\limits_{1\leqslant j\leqslant s}f_{x(C_j)}(x(C_j))}\prod_{1\leqslant j\leqslant n}\mu(\mathrm{d}x(j))+\text{terms}\geqslant 0$$

403 从而得到答案。 □

举例 4.2.6 和举例 4.2.5 相同，令 $x=\{X(1),\cdots,X(n)\}^{\mathrm{T}}$ 为随机变量的集合，令 Y 为另一个随机变量，假设存在关于测度 $\mu^{(n)}\times\nu$ 的联合 PMF $f_{x,Y}$，其中 $\mu^{(n)}=\mu\times\cdots\times\mu$(乘了 n 次)，给定子集 $C\subseteq\{1,\cdots,n\}$，考虑求和

$$I(x(C);Y)-\mathbb{E}[I(x(\overline{C});Y)|x(C)]$$

其中 $x(C)=\{X(i): i\in C\}$，$x(\overline{C})=\{X(i): i\notin C\}$，$\mathbb{E}[I(x(\overline{C});Y)|x(C)]$ 表示在 $x(C)$ 确定下的 $I(\overline{C};Y)$ 的期望。证明该求和不依赖于集合 C。 □

解答 注意到该问题中的表达式等于 $I(x;Y)$。

在 4.3 节我们需要关于并行(或乘积)信道的如下结论。

举例 4.2.7 (文献[173]中的引理 A，亦可参考文献[174]) 证明以 $(\alpha_j,p^{(j)},\sigma_j^2)$ 为参数的 r 个离散时间 Gauss 信道的级联的容量等于

$$C=\sum_{1\leqslant j\leqslant r}\frac{\alpha_j}{2}\ln\left(1+\frac{p^{(j)}}{\alpha_j\sigma_j^2}\right) \tag{4.2.19}$$

此外，当一些 α_j 等于 $-\infty$ 时，(4.2.19)仍成立。在这种情形下，对应的求和形式为 $p^{(j)}/\sigma_j^2$。

解答 假设多维向量 $x=\{x_1,\cdots,x_r\}$ 通过 r 个容量分别为 $C_1,\cdots,C_r$ 的并行信道传输，每个向量 $x_j=\begin{bmatrix}x_{j_1}\\ \vdots \\ x_{j_{n_j}}\end{bmatrix}\in\mathbb{R}^{n_j}$。假设当 $r\to\infty$ 时，$n_j=\lceil\alpha_j r\rceil$，则该乘积信道的容量等于 $\sum\limits_{1\leqslant j\leqslant r}C_j$。根据定义，考虑 $r=2$ 的情形。对于直接部分，假设 $R<C_1+C_2$ 且给定 $\varepsilon>0$，当 r 充分大时，我们必须找到一种编码，使得对于该乘积信道的码字数量为 $M=e^{nR}$ 且 $P_e<\varepsilon$，设 $\eta=(C_1+C_2-R)/2$。令 $\mathscr{X}^1$ 和 $\mathscr{X}^2$ 分别为信道 1 和信道 2 的编码，且 $M_1\sim e^{(C_1-\eta)r}$，

$M_i \sim e^{(C_i-\varepsilon)n}$，对应的错误概率满足 $P_e^{\mathscr{X}^1}$，$P_e^{\mathscr{X}^2} < \varepsilon/2$。构造一种串联码$\mathscr{X}$，其码字为 $x = x_1^1 x_1^2$ 且 $x^i \in \mathscr{X}^i$，$i=1$，2，则，对于所考虑的乘积信道，对应的编码为$\mathscr{X}^1$ 和$\mathscr{X}^2$，错误概率 $P_e^{\mathscr{X}^1 \times \mathscr{X}^2}$ 分解如下：

$$P_e^{\mathscr{X}^1 \times \mathscr{X}^2} = \frac{1}{M_1 M_2} \sum_{1 \le i \le M_1, 1 \le j \le M_2} \mathrm{P}(\text{在信道 1 或 2 中的错误} \mid x_i^1 x_j^2 \text{ 发送})$$

根据信道的独立性，$P_e^{\mathscr{X}^1 \times \mathscr{X}^2} \le P_e^{\mathscr{X}^1} + P_e^{\mathscr{X}^2} \le \varepsilon$ 从而导出了直接部分。 404

相反部分的证明更加复杂，我们只提供一个大概的证明，有兴趣的读者可参考文献[174]。主体思想是应用所谓的列表解码。假设我们有大小为 M 的编码$\mathscr{Y}$和对应的解码规则 $d = d^{\mathscr{Y}}$。接下来，给定信道的输出向量$\mathscr{Y}$，根据解码规则$\bar{d} = \bar{d}_L^{\mathscr{Y}}$ 生成了一个有 L 个可能的编码向量列表，且如果正确码字在列表中，则解码(根据规则$\bar{d}$)成功。于是，对编码$\mathscr{Y}$的平均错误概率 $P_e = P_e^{\mathscr{Y}}(d)$，有

$$P_e \ge P_e(\bar{d}) P_e^{\mathrm{AV}}(L, d) \tag{4.2.20}$$

其中 $P_e(\bar{d}) = P_e^{\mathscr{Y}}(\bar{d})$为列表解码的错误概率，$P_e^{\mathrm{AV}}(L, d) = P_e^{\mathrm{AV}}(\mathscr{Y}, L, d)$为对编码$\mathscr{Y}$的所有长度为 L 的子编码用解码规则 d 时的错误概率。

现在，考虑边际容量为 C_1 和 C_2 的乘积信道，令 $R > C_1 + C_2$，$\eta = (R - C_1 - C_2)/2$，设列表大小为 $L = e^{R_1 \tau}$，且 $R_1 = C_2 + \eta$。假设我们使用大小为 e^{R} 的编码$\mathscr{Y}$，解码规则为 d，对大小为 L 的列表，其解码器为$\bar{d}$。根据(4.2.20)，有

$$P_e \ge P_e(\bar{d}) P_e^{\mathrm{AV}}(e^{R_1 \tau}, d) \tag{4.2.21}$$

已知 $R_1 > C_2$ 且 $P_e^{\mathrm{AV}}(e^{R_1 \tau}, d)$不为 0。相反部分的证明和举例 4.2.8 中的讨论类似。设 $R_2 < R - R_1$ 并考虑子编码$\mathscr{L} \subset \mathscr{Y}$，且 $\#\mathscr{L} = e^{R_2 \tau}$。假设我们等概率随机选择子编码$\mathscr{L}$。令 $M_2 = e^{R_2 \tau}$ 且 $P_e^{\mathscr{Y}, M_2}(d)$为对所有大小为 $\#\mathscr{L} = e^{R_2 \tau}$的子编码$\mathscr{L} \subset \mathscr{Y}$求平均后的平均错误概率，则

$$P_e(\bar{d}) \ge P_e^{\mathscr{Y}, M_2}(d) - \varepsilon(\tau) \tag{4.2.22}$$

其中随着 $\tau \to \infty$，有 $\varepsilon(\tau) \to 0$。 □

举例 4.2.8 令 $L = e^{R_1 \tau}$，$M = e^{R}$。我们旨在证明，当 $R_2 < R - R_1$ 且 $M_2 = e^{R_2 \tau}$ 时，如下结论成立。给定大小为 M 的编码$\mathscr{X}$，解码规则 d，大小为 L 的列表的解码器为$\bar{d}$，考虑大小为 $\#\mathscr{S} = M_2$ 的等分布子编码$\mathscr{S} \subset \mathscr{X}$的平均错误概率 $P_e^{\mathscr{X}, M_2}(d)$。则 $P_e^{\mathscr{X}, M_2}(d)$和列表错误概率 $P_e^{\mathscr{X}}(\bar{d})$满足

$$P_e^{\mathscr{X}}(\bar{d}) \ge P_e^{\mathscr{X}, M_2}(d) + \varepsilon(\tau) \tag{4.2.23}$$

其中随着 $\tau \to \infty$，$\varepsilon(\tau) \to 0$。

解答 令$\mathscr{X}$，$\mathscr{S}$，d 如上所述，假设我们使用长度为 L 的列表解码器$\bar{d}$。 405

给定包含 M_2 种码字的子编码$\mathscr{S} \subset \mathscr{X}$，我们将使用如下解码方式。令$\mathscr{L}$为解码器$\bar{d}$的输出。如果有且仅有一个元素 $x_i \in \mathscr{S}$包含于$\mathscr{L}$，则$\mathscr{S}$对应的解码为 x_i。否则，解码错误。设$\mathscr{S}$的解码器为 $d^{\mathscr{S}}$。给定发送符号 $x_k \in \mathscr{S}$，在上述解码规则下，对应的错误概率为

$$P_{ek} = \sum_{\mathscr{L}} p(\mathscr{L} \mid x_k) E_{\mathscr{S}}(\mathscr{L} \mid x_k)$$

其中 $p(\mathscr{L} \mid x_k)$为在解码规则 $d^{\mathscr{S}}$下，在发送 x_k 后接收到$\mathscr{L}$的概率，$E_{\mathscr{S}}(\mathscr{L} \mid x_k)$为 $d^{\mathscr{S}}$的错误概率。接下来，令 $E_{\mathscr{S}}(\mathscr{L} \mid x_k) = E_{\mathscr{S}}^1(\mathscr{L} \mid x_k) + E_{\mathscr{S}}^2(\mathscr{L} \mid x_k)$，其中 $E_{\mathscr{S}}^1(\mathscr{L} \mid x_k)$表示 $x_k \notin \mathscr{L}$的概率而 $E_{\mathscr{S}}^2(\mathscr{L} \mid x_k)$表示 $x_k \in \mathscr{L}$但是被$\mathscr{S}$中错误的码向量解码的概率(二者都是以发送

x_k 为条件)。更进一步，将 $E_{\mathscr{S}}^{k}(\mathscr{L}|x_k)$ 化为(条件)概率 $E_{\mathscr{S}}(\mathscr{L},x_j|x_k)$ 的和的形式，当 $j\neq k$ 时，解码器返回 $x_j\in\mathscr{L}$。

令 $P_e^{\mathscr{S}}(d)=P_e^{\mathscr{S},\Lambda\gamma}(d)$ 表示子编码 $\mathscr{S}$ 的平均错误概率。上述的推论可导出

$$P_e^{\mathscr{S}}(d)\leqslant\frac{1}{M_2}\sum_{k,x_k\in\mathscr{S}}\sum_{\mathscr{L}}p(\mathscr{L}|x_k)\left[E_{\mathscr{S}}^{k}(\mathscr{L}|x_k)+\sum_{j\neq k}E_{\mathscr{S}}(\mathscr{L},x_j|x_k)\right]\tag{4.2.24}$$

不等式(4.2.24)对任意子编码 $\mathscr{S}$ 恒成立。从 $\mathscr{X}$ 中等概率地随机选取大小为 M_2 的子编码 $\mathscr{S}$，通过在这种子编码中取平均，我们可以得到平均错误概率 $P_e^{\mathscr{X},M_2}=P_e^{\mathscr{X},M_2}(d)$ 的边界：

$$P_e^{\mathscr{X},M_2}\leqslant P_e^{\mathscr{X}}(\bar{d})+\frac{1}{M_2}\sum_{k=1}^{M_2}\sum_{\mathscr{L}}\sum_{j\neq k}(p(\mathscr{L}|x_k)E_{\mathscr{S}}(\mathscr{L},x_j|x_k))^{\mathscr{X},M_2}\tag{4.2.25}$$

其中 $(\ )^{\mathscr{X},M_2}$ 表示在所选取的所有子编码中取平均。由于 x_j 和 x_k 是随机选取的，

$$(p(\mathscr{L}|x_k)E_{\mathscr{S}}(\mathscr{L},x_j))^{\mathscr{X},M_2}=(p(\mathscr{L}|x_k))^{\mathscr{X},M_2}(E_{\mathscr{S}}(\mathscr{L},x_j))^{\mathscr{X},M_2}$$

于是

$$(p(\mathscr{L}|x_k))^{\mathscr{X},M_2}=\sum_{x\in\mathscr{X}}\frac{1}{M}p(\mathscr{L}|x),(E_{\mathscr{S}}(\mathscr{L},x_j|x_k))^{\mathscr{X},M_2}=\frac{L}{M}$$

从而得到

406

$$P_e^{\mathscr{X},M_2}\leqslant P_e^{\mathscr{X}}(\bar{d})+\frac{1}{M_2}\sum_{k=1}^{M_2}\sum_{\mathscr{L}}\left(\sum_{x\in\mathscr{X}}\frac{1}{M}p(\mathscr{L}|x)\right)\left(\sum_{j\neq k}\frac{L}{M}\right)$$

这说明了

$$P_e^{\mathscr{X},M_2}\leqslant P_e^{\mathscr{X}}(\bar{d})+\frac{M_2L}{M}\tag{4.2.26}$$

由于 $R_0<R-R_1$，随着 $\tau\to\infty$，$M_2L/M=e^{R_2\tau}e^{-(R-R_1)\tau}\to0$，从而证明了不等式(4.2.23)。 □

接下来我们给出定理4.2.1的证明，考虑过程 $\boldsymbol{X}$ 的 k 阶 Markov 近似序列，设

$$p^{(k)}(X_0^{\tau-1})=p_{X_0^{k-1}}(X_0^{k-1})\prod_{i=k}^{\tau-1}p(X_i|X_{i-k}^{i-1})\tag{4.2.27}$$

$$H^{(k)}=\mathbb{E}[-\log p(X_0|X_{-k}^{-1})]=h(X_1|X_{-k}^{-1})\tag{4.2.28}$$

$$\overline{H}=\mathbb{E}[-\log p(X_0|X_{-\infty}^{-1})]=h(X_1|X_{-\infty}^{-1})\tag{4.2.29}$$

该证明基于如下结论：引理4.2.9(三明治引理)，引理4.2.10(Markov 近似引理)和引理4.2.11(无间隙引理)。

引理4.2.9 对任意平稳过程 $\boldsymbol{X}$，

$$\limsup_{n\to\infty}\frac{1}{n}\log\frac{p^{(k)}(X_0^{\tau-1})}{p(X_0^{\tau-1})}\leqslant0\quad\text{a.s.}\tag{4.2.30}$$

$$\limsup_{n\to\infty}\frac{1}{n}\log\frac{p(X_0^{\tau-1})}{p(X_0^{\tau-1}|X_{-\infty}^{-1})}\leqslant0\quad\text{a.s.}\tag{4.2.31}$$

证明 如果 A_n 为 $p_{X_0^{\tau-1}}$ 的支持事件(即，$\mathbb{P}(X_0^{\tau-1}\in A_n)=1$)，有

$$\begin{aligned}\mathbb{E}\frac{p^{(k)}(X_0^{\tau-1})}{p(X_0^{\tau-1})}&=\sum_{x_0^{\tau-1}\in A_n}p(x_0^{\tau-1})\frac{p^{(k)}(x_0^{\tau-1})}{p(x_0^{\tau-1})}\\&=\sum_{x_0^{\tau-1}\in A_n}p^{(k)}(x_0^{\tau-1})\\&=p^{(k)}(A)\leqslant1\end{aligned}$$

407

类似地，如果 $B_n=B_n(X_{-\infty}^{-1})$ 为 $p_{X_0^{\tau-1}|X_{-\infty}^{-1}}$ 的支持事件(即，$\mathbb{P}(X_0^{\tau-1}\in B_n|X_{-\infty}^{-1})=1$)，有

$$\mathbb{E}\frac{p(X_0^{n-1})}{p(X_0^{n-1}\mid X_{-\infty}^{-1})}=\mathbb{E}_{X_{-\infty}^{-1}}\sum_{x_0^{n-1}\in B_n}p(x_0^{n-1}\mid X_{-\infty}^{-1})\frac{p(x_0^{n-1})}{p(x_0^{n-1}\mid X_{-\infty}^{-1})}$$

$$=\mathbb{E}_{X_{-\infty}^{-1}}\sum_{x_0^{n-1}\in B_n}p(x_0^{n-1})=\mathbb{E}_{X_{-\infty}^{-1}}\mathbb{P}(B_n)\leqslant 1$$

根据 Markov 不等式，

$$\mathbb{P}\left(\frac{p^{(k)}(X_0^{n-1})}{p(X_0^{n-1})}\geqslant t_n\right)=\mathbb{P}\left(\frac{1}{n}\log\frac{p^{(k)}(X_0^{n-1})}{p(X_0^{n-1})}\geqslant\frac{1}{t_n}\log t_n\right)\leqslant\frac{1}{t_n}$$

对概率 $\mathbb{P}\left(\frac{p(X_0^{n-1})}{p(X_0^{n-1}\mid X_{-\infty}^{-1})}\geqslant t_n\right)$ 有类似结果。令 $t_n=n^2$，于是 $\sum_n 1/t_n<\infty$，由 Borel-Cantelli 引理即可证明结论。 □

引理 4.2.10 对平稳遍历过程 X，

$$-\frac{1}{n}\log p^{(k)}(X_0^{n-1})\xrightarrow{a.s.}H^{(k)} \tag{4.2.32}$$

$$-\frac{1}{n}\log p(X_0^{n-1}\mid X_{-\infty}^{-1})\xrightarrow{a.s.}\overline{H} \tag{4.2.33}$$

证明 将 $f=-\log p(X_k\mid X_{-k}^{-1})$ 和 $f=-\log p(X_k\mid X_{-\infty}^{-1})$ 分别代入 Birkhoff 遍历定理(参考文献[36]的定理 9.1)，从而导出

$$-\frac{1}{n}\log p^{(k)}(X_0^{n-1})=-\frac{1}{n}\log p(X_k^{n-1})$$

$$-\frac{1}{n}\sum_{i=k}^{n-1}\log p^{(k)}(X_i\mid X_{i-k}^{i-1})\xrightarrow{a.s.}0+H^{(k)} \tag{4.2.34}$$

和

$$-\frac{1}{n}\log p(X_0^{n-1}\mid X_{-\infty}^{-1})=-\frac{1}{n}\sum_{i=0}^{n-1}\log p(X_i\mid X_{-\infty}^{i-1})\xrightarrow{a.s.}\overline{H} \tag{4.2.35}$$

于是，根据引理 4.2.9 和 4.2.10，有

$$\limsup_{n\to\infty}\frac{1}{n}\log\frac{1}{p(X_0^{n-1})}\leqslant\lim_{n\to\infty}\frac{1}{n}\log\frac{1}{p^{(k)}(X_0^{n-1})}=H^{(k)} \tag{4.2.36}$$ [438]

且

$$\liminf_{n\to\infty}\frac{1}{n}\log\frac{1}{p(X_0^{n-1})}\geqslant\lim_{n\to\infty}\frac{1}{n}\log\frac{1}{p(X_0^{n-1}\mid X_{-\infty}^{-1})}=\overline{H}$$

将其重写如下

$$\overline{H}\leqslant\liminf_{n\to\infty}-\frac{1}{n}\log p(X_0^{n-1})\leqslant\limsup_{n\to\infty}-\frac{1}{n}\log p(X_0^{n-1})\leqslant H^{(k)} \tag{4.2.37}$$ □

引理 4.2.11 对任意平稳过程 $\boldsymbol{X}$，$H^{(k)}\searrow\overline{H}=H$

证明 根据平稳性和熵不增的条件，$H^{(k)}\searrow H$ 收敛性成立。于是，证明转化为 $H^{(k)}\searrow\overline{H}$，即 $\overline{H}=H$。利用条件概率的 Doob-Levy 鞅收敛定理，可导出

$$p(X_0=x_0\mid X_{-k}^{-1})\xrightarrow{a.s.}p(X_0=x_0\mid X_{-\infty}^{-1}),k\to\infty \tag{4.2.38}$$

由于 l 的取值集合是有限的，且映射 $p\in[0,1]\to-p\log p$ 有界，当 $k\to\infty$ 时，由有界收敛定理可得

$$H^{(k)}=\mathbb{E}-\sum_{x_0\in I}p(X_0=x_0\mid X_{-k}^{-1})\log p(X_0=x_0\mid X_{-k}^{-1})$$

$$\to\mathbb{E}\left[-\sum_{x_0\in I}p(X_0=x_0\mid X_{-\infty}^{-1})\log p(X_0=x_0\mid X_{-\infty}^{-1})\right]=\overline{H}$$ □

4.3 Nyquist-Shannon 公式

本节将给出功率约束且带宽受限的连续时间信道容量的 Nyquist-Shannon 公式[⊖]的严谨推导，其结果被公认为是信息论的基本事实。我们的推导和文献[173]一致(有些许偏差)。由于该推导过长，我们将本节分解为许多子节，每个子节分别解决整个过程的一个特定步骤。

Harry Nyquist(1889—1976)是一位信息论的先驱者，他和 Ralph Hartley(1888—1970)等人一同创建了信道容量的概念。

409

基本假设如下。设 τ，α，$p>0$ 为已知数，假设编码器每 τ 秒产生一个实编码向量

$$\boldsymbol{x}=\begin{bmatrix}x_1\\ \vdots\\ x_n\end{bmatrix}$$

其中 $n=\lceil \alpha\tau \rceil$。所有由编码器产生的向量 $\boldsymbol{x}$ 位于大小为 $M\sim 2^{n\rho}=e^{\tau\alpha\rho}$ 的码集 $\mathscr{X}=\mathscr{X}_n\subset\mathbb{R}^n$ 中。有时为了强调 M 和 n，我们会将该码集写为 $\mathscr{X}_{M,n}$。将 $\mathscr{X}$ 中的编码向量列为 $\boldsymbol{x}(1)$，…，$\boldsymbol{x}(M)$(顺序随机)，其中

$$\boldsymbol{x}(i)=\begin{bmatrix}x_1(i)\\ \vdots\\ x_n(i)\end{bmatrix},1\leqslant i\leqslant M$$

将编码向量 $\boldsymbol{x}$ 转化为连续时间信号

$$x(t)=\sum_{i=1}^{n}x_i\phi_i(t),\quad 0\leqslant t\leqslant\tau \tag{4.3.1}$$

使用函数 $\phi_i(t)$，$i=1,2,\cdots\left(\int_0^\tau\phi_i(t)\overline{\phi_j(t)}\mathrm{d}t=\delta_{ij}\right)$构成的 $L_2[0,\tau]$ 中的正交偏置，于是元素 x_i 可写为积分形式

$$x_i=\int_0^\tau x(t)\overline{\phi_i(t)}\mathrm{d}t \tag{4.3.2}$$

t 时刻的瞬时信号功率和 $|x(t)|^2$ 相关，于是 $\|\boldsymbol{x}\|^2=\int_0^\tau|x(t)|^2\mathrm{d}t=\sum_{1\leqslant i\leqslant n}|x_i|^2$ 表示在区间[0，τ]内信号的全部能量。发送的全部能量的上界具有如下形式

$$\|\boldsymbol{x}\|^2\leqslant p\tau\quad 或\quad \boldsymbol{x}\in\mathbb{B}_n(\sqrt{p\tau}) \tag{4.3.3}$$

(在波导理论中，将维度 n 称为 Nyquist 数，$W=n/(2\tau)\sim\alpha/2$ 为信道带宽)。

将编码向量 $\boldsymbol{x}(i)$ 通过一个加性信道发送，接收端得到(随机)向量

410

$$\boldsymbol{Y}=\begin{bmatrix}Y_1\\ \vdots\\ Y_n\end{bmatrix},\quad Y_k=x_k(i)+Z_k,1\leqslant k\leqslant n \tag{4.3.4}$$

在此我们假设

$$\boldsymbol{Z}=\begin{bmatrix}Z_1\\ \vdots\\ Z_n\end{bmatrix}$$

⊖ 一些作者在 Shannon-Hartley 定理中提及。

为满足 $Z_i \sim N(0, \sigma^2)$ 且独立同分布的向量。（在实际应用中，为了表明"白噪声"过程 $Z(t)$，通常将其表示为 $Z_i = \int_{i\tau}^{\tau} Z(t)\overline{\phi_i(t)}dt$。）

我们定义：若 $x(i) \in \mathscr{X} \setminus \mathbb{B}_0(\sqrt{p\tau})$，即 $\|x(i)\|^2 > p\tau$，则输出符号向量 $\boldsymbol{Y}$ 是不可解码的。换言之，当 $\|x(i)\|^2 > p\tau$ 时，解码输出向量 $\boldsymbol{Y} = x(i) + \boldsymbol{Z}$ 的概率趋于 0（不考虑有当噪声 $\boldsymbol{Z}$ 很小，输出向量 $\boldsymbol{Y}$ 以一个确定概率接近 $x(i)$ 的情况）。

另一方面，当 $\|x(i)\|^2 \leqslant p\tau$ 时，接收端对输出向量 $\boldsymbol{Y}$ 使用解码规则 $d(=d_{\mathbb{K},\mathscr{X}})$，即映射 $y \in \mathbb{K} \mapsto d(y) \in \mathscr{X}$ 其中 $\mathbb{K} \subset \mathbb{R}^n$ 为"可解码域"（当映射 d 预先定义时）。换言之，如果 $\boldsymbol{Y} \subset \mathbb{K}$，则可用 $d(\boldsymbol{Y}) \subset \mathscr{X}$ 解码 $\boldsymbol{Y}$。当已发送 $x(i)$，而 $\boldsymbol{Y} \notin \mathbb{K}$ 或 $d(\boldsymbol{Y}) \neq x(i)$ 时，解码错误。这导出了计算对输入编码向量 $x(i)$ 错误解码概率的公式：

$$P_e(i,d) = \begin{cases} 1, & \|x(i)\|^2 > p\tau \\ \mathbf{P}_{\mathbb{R}}(\boldsymbol{Y} \notin \mathbb{K} \text{ 或 } d(\boldsymbol{Y}) \neq x(i) \mid x(i) \text{ 发送}), & \|x(i)\|^2 \leqslant p\tau \end{cases} \tag{4.3.5}$$

对编码 $\mathscr{X}$ 的平均错误概率 $P_e = P_e^{\mathscr{X},\infty}(d)$ 定义如下

$$P_e = \frac{1}{M}\sum_{1 \leqslant i \leqslant M} P_e(i,d) \tag{4.3.6}$$

此外，如果对任意的 $\varepsilon > 0$，可以确定 $\tau_0(\varepsilon) > 0$，使得对所有的 $\tau > \tau_0(\varepsilon)$，存在大小为 $\#\mathscr{X} \sim e^{R\tau}$ 的码集 $\mathscr{X}$ 和对应的解码规则 d 满足 $P_e = P_e^{\mathscr{X},\infty}(d) < \varepsilon$，则称 R_{∞}（或 R_{∞}）为可达传输速率。信道容量 C 定义为所有可达传输速率的上界，参考式(4.1.17)，由 4.1 节的参数可导出

$$C = \frac{\alpha}{2}\ln\left(1 + \frac{p}{\alpha\sigma^2}\right) \text{(in nats)} \tag{4.3.7}$$

注意到当 $\alpha \to \infty$ 时，式(4.3.7)的右式趋于 $p/(2\sigma^2)$。 [411]

在连续时间信道中，Shannon（以及在他之前的 Nyquist）类出了式(4.3.7)在带限信号中的应用。更准确地说，设 $W = \alpha/2$，于是下式给出了对频率范围 $[-W, W]$ 且单位时间功率不大于 p 的带限信号 $x(t)$，其通过方差为 $\sigma^2 = \sigma_0^2 W$ 的加性白噪声连续时间信道时，对应的信道容量。

$$C = W\ln\left(1 + \frac{p}{2\sigma_0^2 W}\right) \tag{4.3.8}$$

该命题对一名合格的工程师而言十分直观，对数学家来说，这却是一个难点。该命题需要一些技术方面的讨论来证明它的有效性。对工程师的语言来说，一个满足式(4.3.1)在 $[0, \tau]$ 中的"理想"正交系统是一个 $n \sim 2W\tau$ 个等分割 δ 函数的集合。换言之，将编码向量 $x(i) = (x_1(i), \cdots, x_n(i))$ 表示为以时间 $t \in [0, \tau]$ 为参数的函数 $f_i(t)$

$$f_i(t) = \sum_{1 \leqslant k \leqslant n} x_k(i)\delta\left(t - \frac{k}{2W}\right) \tag{4.3.9}$$

其中 $n = \lceil 2W\tau \rceil$（且 $\alpha = 2W$）。这里 δ 表示在时间 0 附近的"单脉冲"，其图像就是在 $t = 0$ 附近的一个"单峰"。于是 $\delta(t - k/2W)$ 表示移位函数，其为 $t = k/2W$ 附近的脉冲，$f_i(t)$ 的函数图像如图 4-4 所示。

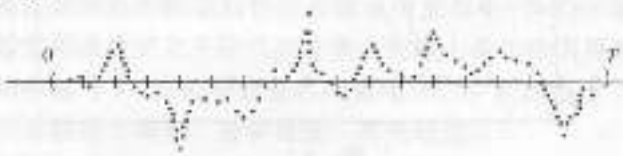


图 4-4

可以认为，编码器每隔 τ 秒处理 $x_i(t)$ 函数，每一个这样的函数都是编码消息 i 后得到的。此外，在每个长度为 τ 的时间内，峰值 $x_i(i)\delta(t-k/2W)$ 每隔时间 1/2W 取得，这里 $\delta(t-k/2W)$ 为时间移位 Dirac 函数。

问题在于，$\delta(t)$ 是一种"概念上的函数"，且 $\delta \notin L_2$。该问题的一种解决方法是，将信号通过一个低通滤波器。于是 $\tilde{f}_i(t)=\tilde{f}_{W,i}(t)$ 取代 $f_i(t)$ 如下

412

$$\tilde{f}_i(t)=\sum_{1\le k\le 2W\tau} x_i(i)\operatorname{sinc}(2Wt-k) \tag{4.3.10}$$

这里

$$\operatorname{sinc}(2Wt-k)=\frac{\sin(\pi(2Wt-k))}{\pi(2Wt-k)} \tag{4.3.11}$$

为对归一化 sinc 函数进行移位和放缩后的函数，归一化 sinc 函数如下

$$\operatorname{sinc}(s)=\begin{cases}\dfrac{\sin(\pi s)}{\pi s}, & s\neq 0\\ 1, & s=0\end{cases} \quad s\in\mathbb{R} \tag{4.3.12}$$

函数图如图 4-5 所示。

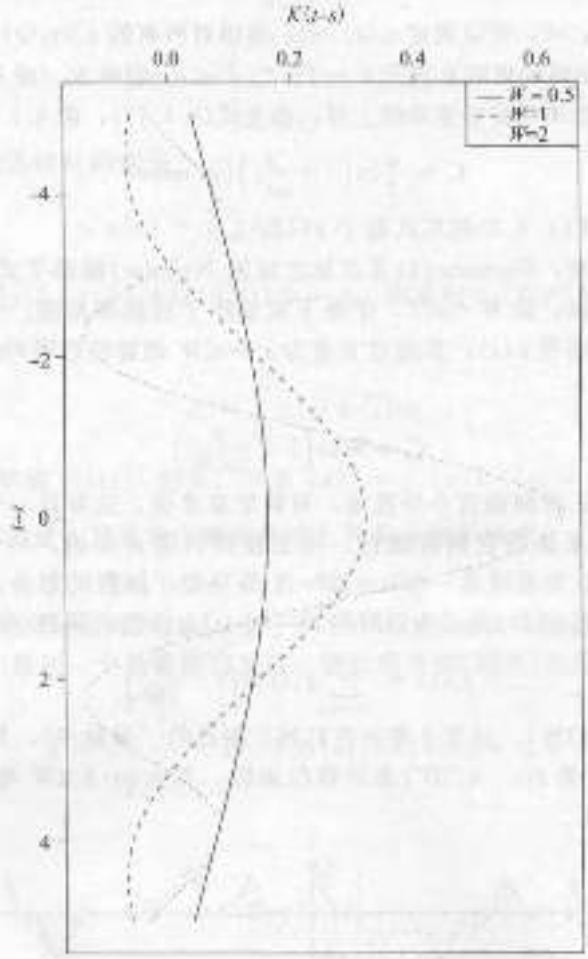


图 4-5

在现代计算机图形学和数字信号处理的其他领域中，移除高频谐波（或者说，高频部

分)并将信号 $f_1(t)$ 替换为它的(近似)低频版本 $\tilde{f}_1(t)$ 的方法得到了广泛应用。

例子 4.3.1 (L_2 中的 Fourier 变换)一个可积函数 ϕ(即 $\int|\phi(x)|\mathrm{d}x<+\infty$)的 Fourier 变换 $\phi\mapsto F\phi$ 定义如下

$$[F\phi](\omega)=\int\phi(x)\mathrm{e}^{\mathrm{i}\omega x}\mathrm{d}x,\omega\in\mathbb{R} \tag{4.3.13}$$

其逆 Fourier 变换可写为逆映射如下

$$[F^{-1}\phi](x)=\frac{1}{2\pi}\int\phi(\omega)\mathrm{e}^{-\mathrm{i}\omega x}\mathrm{d}\omega \tag{4.3.14}$$

值得注意的是，可以将式(4.3.13)和式(4.3.14)扩展用于均方可积函数 $\phi\in L_2(\mathbb{R})$(即 $\|\phi\|^2=\int|\phi(x)|^2\mathrm{d}x<+\infty$)。在此我们不作深入探讨，感兴趣的读者可参考文献[127]。此外，在许多应用中，Fourier 变换极具作用。例如，设 $F\phi=\hat{\phi}$，$F^{-1}\hat{\phi}=\phi$，从式(4.3.13)和式(4.3.14)可得

$$\phi(x)=\frac{1}{2\pi}\int\hat{\phi}(\omega)\mathrm{e}^{-\mathrm{i}\omega x}\mathrm{d}\omega \tag{4.3.15}$$

413 ~ 414

此外，对于任意两个均方可积函数 ϕ_1，$\phi_2\in L_2(\mathbb{R})$，有

$$2\pi\int\phi_1(x)\overline{\phi_2(x)}\mathrm{d}x=\int\hat{\phi}_1(\omega)\overline{\hat{\phi}_2(\omega)}\mathrm{d}\omega \tag{4.3.16}$$

另外，也可以对一般函数定义 Fourier 变换，参见文献[127]。特别地，对于 delta 函数，类似于式(4.3.13)和式(4.3.14)，有

$$\hat{\delta}(\omega)=\frac{1}{2\pi}\int\mathrm{e}^{-\mathrm{i}\omega t}\mathrm{d}\omega=1=\int\delta(t)\mathrm{e}^{\mathrm{i}\omega t}\mathrm{d}t \tag{4.3.17}$$

这说明了 Dirac delta 的 Fourier 变换为 $\hat{\delta}(\omega)=1$。对于移位 delta 函数，可得

$$\hat{\delta}\left(t-\frac{k}{2W}\right)=\frac{1}{2\pi}\int\mathrm{e}^{\mathrm{i}\omega k/(2W)}\mathrm{e}^{-\mathrm{i}\omega t}\mathrm{d}\omega \tag{4.3.18}$$

当信道加上一个“去除”所有谐波 $\mathrm{e}^{\mathrm{i}\omega t}$ 的“滤波器”时，Shannon-Nyquist 公式成立，其中 ω 为在区间 $[-2\pi W,2\pi W]$ 之外的频率。换言之，(4.3.18)中的(移位)单位脉冲 $\delta(t-k/(2W))$ 被它的截取之后的形式所代替，当滤波器去除了 $|\omega|>2\pi W$ 的所有谐波 $\mathrm{e}^{-\mathrm{i}\omega t}$ 该脉冲才产生。

当我们把式(4.3.17)中的 ω 的区间缩减到 $[-\pi,\pi]$，则出现了经典的 sinc 函数(应用数学中的著名函数)：

$$\mathrm{sinc}(t)=\frac{1}{2\pi}\int_{-\pi}^{\pi}\mathrm{e}^{-\mathrm{i}\omega t}\mathrm{d}\omega,\mathbf{1}_{[-\pi,\pi]}(\omega)=\int\mathrm{sinc}(t)\mathrm{e}^{\mathrm{i}\omega t}\mathrm{d}t,\quad t,\omega\in\mathbb{R} \tag{4.3.19}$$

(象征性地，$\mathrm{sinc}=F^{-1}\mathbf{1}_{[-\pi,\pi]}$)在本文中，当 $A>0$ 充分大时，可将函数 $t\mapsto A\,\mathrm{sinc}(At)$ 看作是 $\delta(t)$ 的简单近似，通常，$\mathrm{sinc}(t)$ 在整个 $\mathbb{R}$ 上是不可积的(由于因子 $1/t$)，但它仍是均方可积的：$\int(\mathrm{sinc}(t))^2\mathrm{d}t<\infty$，因此式(4.3.18)的右式应理解为 L_2 下的情形。

然而，这并没有让该理论在数学和物理方面显得更简单(和工程视角相比)。实际上，上面所考虑的将不需要的谐波完全去除的滤波器在物理上是无法实现的。即使有这种完美的器件，我们所得到的信号 $\tilde{f}_1(t)$ 也不再是时间受限于 $[0,\tau]$ 的信号，而是分布在整个时间轴上的函数。为了克服这个难点，需要做进一步地近似。

举例 4.3.2 证明函数

$$t\mapsto(2\sqrt{\pi W})\mathrm{sinc}(2Wt-k),k=1,\cdots,n \tag{4.3.20}$$

在 $L_2(\mathbb{R}^1)$ 空间是正交的，即

415

$$(4\pi W)\int[\mathrm{sinc}(2Wt-k)][\mathrm{sinc}(2Wt-k')]\mathrm{d}t=\delta_{kk'}$$

解答 最简单的方法是由(4.3.19)写出其(在 $L_2(\mathbb{R})$ 中)Fourier 分解：

$$2\sqrt{\pi W}\,\mathrm{sinc}(2Wt-k)=\frac{1}{2\sqrt{\pi W}}\int_{-2\pi W}^{2\pi W}\mathrm{e}^{\mathrm{j}\omega k/(2W)}\mathrm{e}^{-\mathrm{j}\omega t}\mathrm{d}\omega \tag{4.3.21}$$

然后检查该 Fourier 变换对应的函数是正交的。

$$\frac{1}{2\sqrt{\pi W}}1(|\omega|\leqslant 2\pi W)\mathrm{e}^{\mathrm{j}\omega k/(2W)},\quad k=1,\cdots,n$$

即

$$\frac{1}{4\pi W}\int_{-2\pi W}^{2\pi W}\mathrm{e}^{\mathrm{j}(k-k')\omega/(2W)}\mathrm{d}\omega=\delta_{kk'} \tag{4.3.22}$$

其中

$$\delta_{kk'}=\begin{cases}1, & k=k'\\ 0, & k\neq k'\end{cases}$$

为 Kronecker 符号。根据标准的积分过程可验证式(4.3.22)。

因为式(4.3.20)中的函数是正交的，所以我们得到

$$\|x(t)\|^2=(4\pi W)\|\tilde{f}_1\|^2,\quad \|\tilde{f}_1\|^2=\int|\tilde{f}_1(t)|^2\mathrm{d}t \tag{4.3.23}$$

其中 $\tilde{f}_1$ 在式(4.3.10)中已说明过。于是，功率约束可写为

$$\|\tilde{f}_1\|^2\leqslant p\tau/4\pi W=p_0 \tag{4.3.24}$$

实际上，参数 $x_k(i)$ 和函数 $\tilde{f}$ 在时刻 $k/(2W)$，$k=1,\cdots,n$ 计算得到的值 $\tilde{f}_1(k/(2W))$ 相一致，这些点可以看作是采样点。

因此，虽然输入信号 $\tilde{f}_1(t)$ 完全由值 $\tilde{f}_1(k/(2W))=x_k(i)$ 确定，但仍可将其视为连续时间上的信号。于是，考虑在不同时间区间 $(0,\tau)$，$(\tau,2\tau)$，…产生的信号，在忽略 $\mathrm{sinc}(t)$ 函数的末端产生的干扰后，通过他们在采样点的取值可清楚地将其识别出来。

Nyquist-Shannon 假设如下，$\tilde{f}_1(t)$ 在信道中转换为如下形式

416

$$g(t)=\tilde{f}_1(t)+\tilde{Z}(t) \tag{4.3.25}$$

这里 $\tilde{Z}(t)$ 为一个零均值($\mathbb{E}\,\tilde{Z}(t)=0$)的连续时间平稳 Gauss 过程，其自相关函数为

$$\mathbb{E}[\tilde{Z}(s)\,\tilde{Z}(t+s)]=2\sigma_0^2W\,\mathrm{sinc}(2Wt),t,s\in\mathbb{R} \tag{4.3.26}$$

特别地，当 t 为 π/W 的整数倍时(即 t 和采样时刻重合)，随机变量 $\tilde{Z}(s)$ 和 $\tilde{Z}(t+s)$ 是相互独立的。该条件的一种等价形式为，频谱密度满足

$$\Phi(\omega)_t=\int\mathrm{e}^{\mathrm{j}\omega t}\mathbb{E}[\tilde{Z}(0)\,\tilde{Z}(t)]\mathrm{d}t=\sigma_0^2 1(|\omega|\leqslant 2\pi W) \tag{4.3.27}$$

可以看出，接收到的连续时间信号 $y(t)$ 可以通过下式得到对应的值 $y_k=y\left(\frac{k}{2W}\right)$ 从而识别出来

$$y_k=x_k(i)+Z_k\text{ 其中 }Z_k=\tilde{Z}\left(\frac{k}{2W}\right)\text{ 是独立同分布 }\mathrm{N}(0,2\sigma_0^2W)$$

这和 4.1 节考虑的 $p=2Wp_i$，$\sigma^2=2\sigma_0^2W$ 的系统相一致。在工程界，系统容量 C 由式(4.3.8)决定，即低于该容量的传输速率是可靠的，反之是不可靠的。

然后，为了深入理解式(4.3.8)，还有一些待解决的问题。其中一个问题是，只有理想的滤波器才能将信号限制于一个特定频段内。另一个问题是，由于具有形式

$$t \in \mathbb{R} \mapsto \sum_{1 \le k \le n} (x_k(t) + z_k)\operatorname{sinc}(2Wt - k) \tag{4.3.28}$$

的任意采样函数在 t 上是解析的，所以可以通过在一小段时间区间内记录下式(4.3.25)中的输出信号 $g(t)$ 后再重构出来。因此，速率这一术语需要恰当的定义。

最简单的解决方式(在文献[173]中提到)是引入满足下列条件的函数族 $\mathscr{A}(\tau, W, p_0)$

(i) 在单位时间(一秒)内带宽近似受限于 W。

(ii) 时域上的长度为 τ(为了方便，可将其设为 $[-\tau/2, \tau/2]$)。

(iii) 总能量(由 $L_2(\mathbb{R})$ 范数定义)不超过 $p_0\tau$。

这些条件构成了系统的约束。 417

然后，考虑大小为 $M \sim e^{R\tau}$ 的编码 $\mathscr{X}$，即以时间 t 为变量的函数 $\bar{f}_1(t), \cdots, \bar{f}_M(t)$ 的集合。如果给定编码函数 $\bar{f}_i \notin \mathscr{A}(\tau, W, p_0)$，则称其为不可解码的，对应的错误概率为 1。否则，信号 $\bar{f}_i \in \mathscr{A}(\tau, W, p_0)$ 和均值为 $\mathbb{E}\bar{Z}(t)=0$ 且满足(4.3.27)的加性 Gauss 噪声 $\bar{Z}(t)$ 在信道的输出端构成了输出信号 $g(t)=\bar{f}_i(t)+\bar{Z}(t)$(参考(4.3.25))。接收端使用解码规则 d，即映射 $d: \mathbb{K} \to \mathscr{X}$，其中 $\mathbb{K}$ 为之前 d 的定义域，即定义映射 d 的一些函数族。(和之前一样，解码规则 d 可能随着编码的不同而不同，即 $d=d^{\mathscr{X}}$。)同样地，若 $g \notin \mathbb{K}$，则认为该传输出错。最后，如果 $g \in \mathbb{K}$，则由编码函数 $d^{\mathscr{X}}(g)(t) \in \mathscr{X}$ 对接收信号 g 进行解码。当编码器产生的编码信号为 $\bar{f}_i \in \mathscr{X}$，编码 $\mathscr{X}$ 的错误概率为

$$P_e(i) = \begin{cases} 1, & \bar{f}_i \notin \mathscr{A}(\tau, W, p_0) \\ \boldsymbol{P}_{\mathrm{ch}}(\mathbb{K}^c \sqcup \{g : d^{\mathscr{X}}(g) \neq \bar{f}_i\}), & \bar{f}_i \in \mathscr{A}(\tau, W, p_0) \end{cases} \tag{4.3.29}$$

编码 $\mathscr{X}$(和编码器 d)的平均错误概率 $P_e = P_e^{\mathscr{X}, m}(d)$ 为

$$P_e = \frac{1}{M} \sum_{1 \le i \le M} P_e(i, d) \tag{4.3.30}$$

如果对于任意的 $\varepsilon>0$，存在 τ 和大小为 $M \sim e^{R\tau}$ 的编码 $\mathscr{X}$，使得 $P_e<\varepsilon$，则称 $R=R_{\mathrm{ch}}$ 为可靠传输速率。

现在，设 $\eta \in (0, 1)$，函数 $f^{\tau}(t)$ 的集合 $\mathscr{A}(\tau, W, p_0) = \mathscr{A}(\tau, W, p_0, \eta)$ 定义如下

(i) $f^{\tau}(t) = D_{\tau} f$，其中

$$D_{\tau} f(t) = f(t) 1(|t| < \tau/2), t \in \mathbb{R}$$

其中 $f(t)$ 为 $\int e^{i\omega t} f(t) dt$ 的 Fourier 变换，且随着 $|\omega| > 2\pi W$ 趋于 0。

(ii) 比值

$$\frac{\|f^{\tau}\|^2}{\|f\|^2} \geq 1 - \eta$$

(iii) 范数 $\|f^{\tau}\|^2 \leq p_0 \tau$。

换言之，发送信号 $f^{\tau} \in \mathscr{A}(\tau, W, p_0, \eta)$ 在时域上严格受限且在频域上近似为带宽受限。 418

在受限情况下，根据一些推论可以得到 Nyquist-Shannon 公式，最简单的形式为下述的定理 4.3.3。另外一种形式将在稍后的定理 4.3.7 中说明。

定理 4.3.3 受条件(i)-(iii)所描述的 $\mathscr{A}(\tau, W, p_0, \eta)$ 约束的信道对应的信道容量 $C = C(\eta)$ 如下

$$C = W\ln\left(1+\frac{p_0}{2\sigma_0^2 W}\right)+\frac{\eta}{1-\eta}\frac{p_0}{\sigma_0^2} \tag{4.3.31}$$

随着 $\eta\to 0$，有

$$C(\eta)\to W\ln\left(1+\frac{p_0}{2\sigma_0^2 W}\right) \tag{4.3.32}$$

从而导出了式(4.3.8)的 Nyquist-Shannon 公式。

在深入技术细节之前，我们先讨论关于 r 个时间离散 Gauss 信道的乘积，或者说是其并行组合的一些事实。(本质上，该模型已在 4.2 节的末尾讨论过。)这里，输入信号每隔 τ 个时间单位产生，其向量的排序集合如下

$$\{x^{(1)},\cdots,x^{(r)}\}\ 其中 x^{(j)}=\begin{pmatrix} x_1^{(j)} \\ \vdots \\ x_{n_j}^{(j)}\end{pmatrix}\in\mathbb{R}^{n_j},1\leqslant j\leqslant r \tag{4.3.33}$$

且 $n_j=[a_j\tau]$ 中的 a_j 为给定常数(编码 j 的数字乘积的速度)。对每个向量 $x^{(j)}$，考虑特定的功率约束：

$$\|x^{(j)}\|^2\leqslant p^{(j)}\tau,1\leqslant j\leqslant r \tag{4.3.34}$$

输出信号为随机向量的集合

$$\{Y^{(1)},\cdots,Y^{(r)}\}，其中 Y^{(j)}=\begin{pmatrix} Y_1^{(j)} \\ \vdots \\ Y_{n_j}^{(j)}\end{pmatrix} \quad 和 \quad Y_k^{(j)}=x_k^{(j)}+Z_k^{(j)} \tag{4.3.35}$$

[419] 其中 $Z_k^{(j)}$ 为 IID 随机变量，$Z_k^{(j)}\sim N(0,\sigma^{(j)^2})$，$1\leqslant k\leqslant n_j$，$1\leqslant j\leqslant r$。

对于乘积信道来说，信息率为 R 的码本 $\mathscr{X}$ 是一个 M 进制的输入信号：

$$\begin{gathered}\{(x^{(1)}(1),\cdots,x^{(r)}(1))\\ (x^{(1)}(2),\cdots,x^{(r)}(2))\\ \cdots\quad\cdots\quad\cdots\\ (x^{(1)}(M),\cdots,x^{(r)}(M))\}\end{gathered} \tag{4.3.36}$$

每个都和式(4.3.33)有一样的结构。如之前所述，解码器 d 是一个作用在样本输出信号 $\{y^{(1)},\cdots,y^{(r)}\}$ 给定的集合 $\mathbb{K}$ 上的映射，并且该解码器将这些信号映射到 $\mathscr{X}$ 中。

如上所述，对于 $i=1,\cdots,M$，当发送一个输入信号 $\{x^{(1)}(i),\cdots,x^{(r)}(i)\}$ 时，我们为码 $\mathscr{X}$ 定义一个错误概率 $P_e(i,d)$。

如果对于一些 $j=1,\cdots,r$，$|x^{(j)}(i)|^2\geqslant p^{(j)}\tau$ 成立，那么 $P_e(i,d)=1$，而如果对于 $j=1,\cdots,r$，$|x^{(j)}(i)|^2<p^{(j)}\tau$ 成立，那么

$$\begin{aligned}P_e(i,d)=\boldsymbol{P}_{\pm}(&\{Y^{(1)},\cdots,Y^{(r)}\}\notin\mathbb{K}\quad 或\\ &d(\{Y^{(1)},\cdots,Y^{(r)}\})\neq\{x^{(1)}(i),\cdots,x^{(r)}(i)\}\\ &\mid\{x^{(1)}(i),\cdots,x^{(r)}(i)\}\ 发送)\end{aligned}$$

码 $\mathscr{X}$(当使用解码器 d 时)的平均错误概率 $P_e=P_e^{\mathscr{X},\mathbb{K}}(d)$ 由下式给出

$$P_e=\frac{1}{M}\sum_{1\leqslant i\leqslant M}P_e(i,d)$$

和通常情况一样，如果对于所有的 $\varepsilon>0$ 都存在一个 $\varepsilon_0>0$，使得对于所有 $\varepsilon_0>0$ 都存在一个势为 $M\sim e^{R\tau}$ 的码 $\mathscr{X}$ 和一个解码规则 d 使 $P_e<\varepsilon$ 成立，那么称 R 是一个可靠的传输速率。定义联合信道的容量为所有可靠传输速率的上确界。下面的定理在例子 4.2.7 中给出(参考文献[173]引理 A；同见文献[174])。

引理 4.3.4 乘积信道的容量等于

$$C=\sum_{1\le j\le r}\frac{\alpha_j}{2}\ln\left(1+\frac{p^{(j)}}{\alpha_j\sigma_j^2}\right) \tag{4.3.37}$$

此外，当一些 α_j 等于 $+\infty$ 时，式(4.3.37)成立；在这种情况下，相应的被加数形式是 $p^{(j)}/2\sigma_j^2$。 420

我们下一步考虑时间离散 Gauss 信道的联合约束乘积。下面我们讨论联合约束的类型。

情况 1 令 $r=2$，假定 $\sigma_1^2=\sigma_2^2=\sigma_3^2$，用下式代替条件式(4.3.34)

$$|x^{(1)}|^2+|x^{(2)}|^2\le p_0\tau \tag{4.3.38a}$$

此外，如果 $\alpha_1\le\alpha_2$，我们引入 $\beta\in(0,1)$，使下式成立

$$|x^{(2)}|^2\le\beta(|x^{(1)}|^2+|x^{(2)}|^2) \tag{4.3.38b}$$

另一种情况，也就是 $\alpha_2\le\alpha_1$，用下式代替式(4.3.38b)

$$|x^{(1)}|^2\le\beta(|x^{(1)}|^2+|x^{(2)}|^2) \tag{4.3.38c}$$

情况 2 这里我们让 $r=3$，并假定 $\sigma_1^2=\sigma_2^2\ge\sigma_3^2$，$\alpha_3=+\infty$。需要使下式成立

$$\sum_{1\le j\le 2}|x^{(j)}|^2<p_0\tau \tag{4.3.39a}$$

以及

$$|x^{(3)}|^2\le\beta\sum_{1\le j\le 2}|x^{(j)}|^2 \tag{4.3.39b}$$

情况 3 和情况 1 中一样，令 $r=2$，假定 $\sigma_1^2=\sigma_2^2=\sigma_3^2$。并让 $\alpha_3=+\infty$。限制条件现在是

$$|x^{(3)}|^2<p_0\tau \tag{4.3.40a}$$

以及

$$|x^{(3)}|^2<\beta(|x^{(1)}|^2+|x^{(2)}|^2) \tag{4.3.40b}$$

举例 4.3.5 （参见文献[173]定理 1） 证明下面给出的容量就是上述情况 1～3 中联合平行信道的容量。

情况 1 $\alpha_1\le\alpha_2$：

$$C=\frac{\alpha_1}{2}\ln\left(1+\frac{(1-\zeta)p_0}{\alpha_1\sigma_1^2}\right)+\frac{\alpha_2}{2}\ln\left(1+\frac{\zeta p_0}{\alpha_2\sigma_2^2}\right) \tag{4.3.41a}$$

其中

$$\zeta=\min\left[\beta,\frac{\alpha_2}{\alpha_1+\alpha_2}\right] \tag{4.3.41b}$$ 421

如果 $\alpha_2\le\alpha_1$，那么下标 1 和 2 需要在这些等式中互换位置。此外，当 $\alpha_1=+\infty$ 时，使用限定表达式 $\lim_{\alpha\to\infty}(\alpha/2)\ln(1+v/\alpha)=v/2$。特别地，如果 $\alpha_2\le\alpha_1=+\infty$，那么 $\beta=\zeta$，并且容量变为

$$C=\frac{\alpha_1}{2}\ln\left(1+\frac{(1-\beta)p_0}{\alpha_1\sigma_1^2}\right)+\beta\frac{p_0}{2\sigma_2^2} \tag{4.3.41c}$$

这意味着，当在信道 2 上注入的能量等于(4.3.38b)中所允许的最大能量时，我们可以得到一个最好的传输速率。

情况 2

$$C=\frac{\alpha_1}{2}\ln\left(1+\frac{(1-\beta)p_0}{(\alpha_1+\alpha_2)\sigma_1^2}\right)$$

$$+\frac{\alpha_2}{2}\ln\left(1+\frac{(1-\beta)p_2}{(\alpha_1+\alpha_2)\sigma_0^2}\right)+\frac{\beta_*}{2\sigma_0^2} \tag{4.3.42}$$

情况 3

$$C=\frac{\alpha_1}{2}\ln\left(1+\frac{p_1}{\alpha_1\sigma_0^2}\right)+\frac{\beta p_1}{2(1-\beta)\sigma_0^2} \tag{4.3.43}$$

解答 我们现在只给出情况 1 的证明。为了明确起见，我们假定 $\alpha_1<\alpha_2\leqslant+\infty$。首先，给出主要的部分。考虑两信道的并联组合，其中 $p_1=(1-\zeta)p_0$，$p_2=\zeta p_0$，信道各自的输入信号分别为$x^{(1)}$和$x^{(2)}$：

$$\|x^{(1)}\|^2\leqslant p_1\tau,\quad \|x^{(2)}\|^2\leqslant p_2\tau \tag{4.3.44a}$$

当然，式(4.3.44a)蕴含着式(4.3.38a)。接着，如果 $\zeta\leqslant\beta$，那么条件(4.3.38b)也是成立的。根据引理 4.3.4 的主要部分，任何满足 $R<C_1(p_1)+C_2(p_2)$的速率 R 都是可靠的。如下所述，

$$C_i(q)=\frac{\alpha_i}{2}\ln\left(1+\frac{q}{\alpha_i\sigma_0^2}\right),\quad i=1,2 \tag{4.3.44b}$$

这就意味着命题正向可证。

证明逆命题是一个更长久的争论。集合 $C^*=C_1(p_1)+C_2(p_2)$。目标是证明任何 $R>C^*$ 都是不可靠的。我们假定一个相反的例子：存在一个可靠速率 $R=C^*+\varepsilon$。让我们回忆一下它正式的意义。也即存在一系列值 $\tau^{(l)}\to\infty$和(a)一系列由“联合”码向量 $x(i)=\{x^{(1)}(i),x^{(2)}(i)\}$组成、大小为 $M^{(l)}\sim e^{R\tau^{(l)}}$ 的码

422

$$\mathscr{X}^{(l)}=\{x(i)=\{x^{(1)}(i),x^{(2)}(i)\},1\leqslant i\leqslant M^{(l)}\}$$

并且满足均方范数$\|x(i)\|^2=\|x^{(1)}(i)\|^2+\|x^{(2)}(i)\|^2$。(b)一系列解码映射 $d^{(l)}$：$y\in\mathscr{Y}^{(l)}\mapsto d^{(l)}(y)\in\mathscr{X}^{(l)}$ 使得 $P_e\to0$。这里，如前面所述，$P_e=P_e^{\mathscr{X}^{(l)}}(d^{(l)})$代表平均错误概率：

$$P_e=\frac{1}{M^{(l)}}\sum_{i\leqslant i\leqslant M^{(l)}}P_e(i,d^{(l)})$$

它是由各自的错误概率 $P_e(i,d^{(l)})$计算得出：

$$P_e(i,d^{(l)})=\begin{cases}1,\\ \quad 如果\|x(i)\|^2>p_2\tau^{(l)}\ 或者\ \|x^{(2)}(i)\|^2>\beta\|x(i)\|^2\\ \boldsymbol{P}_{\mathscr{A}}(\boldsymbol{Y}\notin\mathscr{A}^{(l)}或者\ d^{(l)}(\boldsymbol{Y})\neq x(i)\mid x(i)\ 发送),\\ \quad 如果\|x(i)\|^2\leqslant p_1\tau^{(l)}\ 且\ \|x^{(2)}(i)\|^2\leqslant\beta\|x(i)\|^2\end{cases}$$

分向量

$$x^{(1)}(i)=\begin{bmatrix}x_1^{(1)}(i)\\ \vdots\\ x_{\lfloor\alpha_1\tau^{(l)}\rfloor}^{(1)}(i)\end{bmatrix}\in\mathbb{R}^{\lfloor\alpha_1\tau^{(l)}\rfloor}\quad 和\quad x^{(2)}(i)=\begin{bmatrix}x_1^{(2)}(i)\\ \vdots\\ x_{\lfloor\alpha_2\tau^{(l)}\rfloor}^{(2)}(i)\end{bmatrix}\in\mathbb{R}^{\lfloor\alpha_2\tau^{(l)}\rfloor}$$

在并行信道联合的各自部分上发送，得到的输出向量

$$\boldsymbol{Y}^{(1)}=\begin{bmatrix}Y_1^{(1)}(i)\\ \vdots\\ Y_{\lfloor\alpha_1\tau^{(l)}\rfloor}^{(1)}(i)\end{bmatrix}\in\mathbb{R}^{\lfloor\alpha_1\tau^{(l)}\rfloor}\quad 和\quad \boldsymbol{Y}^{(2)}=\begin{bmatrix}Y_1^{(2)}(i)\\ \vdots\\ Y_{\lfloor\alpha_2\tau^{(l)}\rfloor}^{(2)}(i)\end{bmatrix}\in\mathbb{R}^{\lfloor\alpha_2\tau^{(l)}\rfloor}$$

生成一个联合输出信号$\boldsymbol{Y}=\{\boldsymbol{Y}^{(1)},\boldsymbol{Y}^{(2)}\}$。$\boldsymbol{Y}^{(1)}$，$\boldsymbol{Y}^{(2)}$向量的元素是下面的和

$$Y_j^{(1)}=x_j^{(1)}(i)+Z_j^{(1)},Y_k^{(2)}=x_k^{(2)}(i)+Z_k^{(2)}$$

其中 $Z_j^{(1)}$ 和 $Z_k^{(2)}$ 是独立同分布的，$N(0,\sigma_0^2)$随机变量。相应地，$\boldsymbol{P}_{\mathscr{A}}$指的是随机变量 $Y_j^{(1)}$

和 $Y_k^{(i)}$ 的联合分布，其中 $1\leqslant j\leqslant\lceil \alpha_1\tau^{(i)}\rceil$，$1\leqslant k\leqslant\lceil \alpha_2\tau^{(i)}\rceil$。

可以看到 $q\mapsto C_1(q)$ 在 $[0,\ p_0]$ 上是关于 q 均匀连续的。因此，我们可以找到一个足够大的整数 J_0 使得对于所有的 $q\in(0,\ \zeta p_0)$，都有下式成立

$$\left|C_1(q)-C_1\left(q-\frac{\zeta p_0}{J_0}\right)\right|<\frac{\varepsilon}{2}$$

然后我们将码字 $\mathscr{X}^{(i)}$ 分割为 J_0 个集（子码）$\mathscr{X}_j^{(i)}$，$j=1,\ \cdots,\ J_0$；如果 423

$$(j-1)\frac{\zeta p_0\tau}{J_0}<1\leqslant\sum_{k\leqslant\lceil\alpha_2\tau^{(i)}\rceil}(x_k^{(i)})^2\leqslant j\frac{\zeta p_0\tau}{J_0}\tag{4.3.45a}$$

则码向量 $(x^{(1)}(i),\ x^{(2)}(i))$ 落在集 $\mathscr{X}_j^{(i)}$ 中。因为可传输的码向量 x 有一个分量 $x^{(2)}$，其中 $\|x^{(2)}\|^2\leqslant\zeta\|x\|^2$，所以这样的 x 落在且仅落在一个集中。（我们约定零码向量属于 $\mathscr{X}_1^{(i)}$）用 $\mathscr{X}_{j_0}^{(i)}$ 来表示包含最多码向量的集 $\mathscr{X}_j^{(i)}$。然后，很明显，$\#\mathscr{X}_{j_0}^{(i)}\geqslant M^{(i)}/J_0$，并且码 $\mathscr{X}_{j_0}^{(i)}$ 的传输速率满足

$$R_c\geqslant R-\frac{1}{\tau^{(i)}}\ln J_0\tag{4.3.45b}$$

在另一方面，子码 $\mathscr{X}_{j_0}^{(i)}$ 最大误差概率小于等于整个码的最大误差概率 $\mathscr{X}^{(i)}$（当使用同一译码器 $d^{(i)}$）；因此，误差概率 $P_e^{\mathscr{X}_{j_0}^{(i)},d^{(i)}}(d^{(i)})\leqslant P_e^{(i)}\to 0$。

已知 $\mathscr{X}^{(i)}$ 划分为确定的 J_0 个集，我们可以至少找到一个 $j_0\in\{1,\ \cdots,\ J_0\}$ 使得对于无穷多的 i，有最多的集合 $\mathscr{X}_{j_0}^{(i)}$ 满足和 $\mathscr{X}_{j_0}^{(i)}$ 一致。将我们的自变量减到 i，我们可能假定对于所有的 i 都有 $\mathscr{X}_{j_0}^{(i)}=\mathscr{X}_{j_0}^{(1)}$。然后，对于所有的 $(x^{(1)},\ x^{(2)})\in\mathscr{X}_{j_0}^{(i)}$，其中

$$x^{(i)}=\begin{bmatrix}x_1^{(i)}\\ \vdots\\ x_{\lceil\alpha_i\tau^{(i)}\rceil}^{(i)}\end{bmatrix},\quad i=1,2$$

通过式(4.4.38a)和式(4.3.45a)

$$\|x^{(1)}\|^2\leqslant\left(1-\frac{(j_0-1)\zeta}{J_0}\right)p_0\tau^{(i)},\ \|x^{(2)}\|^2\leqslant\frac{j_0\zeta}{J_0}p_0\tau^{(i)}$$

就是说，$\{(\mathscr{X}_{j_0}^{(i)},\ d^{(i)})\}$ 对于"标准的"并行信道组合是一个编码/译码序列（参考式(4.3.34)），其中

$$p_1=\left\lceil 1-\frac{(j_0-1)\zeta}{J_0}\right\rceil p_0\text{ 且 }p_2=\frac{j_0\zeta}{J_0}p_0$$

当错误概率 $p_e^{\mathscr{X}_{j_0}^{(i)},d^{(i)}}(d^{(i)})\to 0$ 时，速率 R 对联合信道来说是可靠的。因此速率的值不超过容量。

$$R^*\leqslant C_1\left(\left(1-\frac{(j_0-1)\zeta}{J_0}\right)p_0\right)+C_2\left(\frac{j_0\zeta}{J_0}p_0\right)$$ 424

接下来，我们参考在式(4.3.44b)中给出的 $C_i(\alpha)$ 的定义，也就是说

$$R^*\leqslant C_1((1-\delta)p_0)+C_2(\delta p_0)+\frac{\varepsilon}{2}\tag{4.3.46}$$

其中 $\delta=j_0\zeta/J_0$。

现在，我们可以注意到对于 $\alpha_2\geqslant\alpha_1$，函数

$$\delta\mapsto C_1((1-\delta)p_0)+C_2(\delta p_0)$$

当 $\delta<\alpha_2/(\alpha_1+\alpha_2)$ 时，是关于 δ 的增函数，当 $\delta>\alpha_2/(\alpha_1+\alpha_2)$ 时，该函数为减函数。

因此，根据 $\delta=j_0\zeta/J_0\leqslant\zeta$，我们可以得到

$$C_1((1-\delta)p_0)+C_2(\delta p_0)\leqslant C_1(p_1)+C_2(p_2)=C^*\tag{4.3.47}$$

其中 $\zeta=\min[\beta,\ \alpha_2/(\alpha_1+\alpha_2)]$。相反，这意味着因为有式(4.3.45b)、式(4.3.46)和式(4.3.47)，可知

$$R\leqslant C^{-}+\frac{\varepsilon}{2}+\frac{1}{\tau^{\beta\zeta}}\ln f_{\tau},\quad 或\ R\leqslant C^{-}+\frac{\varepsilon}{2}\quad 当\quad \tau^{\beta\zeta}\to\infty$$

与 $R=C^{-}+\varepsilon$ 矛盾，即逆命题得证。 □

例子 4.3.6 （扁长椭球波函数(PSWF)，见文献[146]、[90]、[91]）对于任意给定的 τ，$W>0$，都存在一系列属于 Hilbert 空间 $L_2(\mathbb{R})$（也就是说满足 $\int\varphi(t)^2\mathrm{d}t<\infty$）关于 $t\in\mathbb{R}$ 的函数 $\Psi_1(t)$，$\Psi_2(t)$，…，这些函数被称为扁长椭球波函数(PSWF)，有以下特性

(a) Fourier 变换 $\hat{\Psi}_n(\omega)=\int\Psi(t)\mathrm{e}^{-\mathrm{i}\omega t}\mathrm{d}t$ 在 $|\omega|>2\pi W$ 时为零；并且，函数 $\Psi_n(t)$ 生成一个 Hilbert 子空间中的正交基，该子空间由满足这个特性的 $L_2(\mathbb{R})$ 中的函数生成。

(b) 函数 $\Psi_n^{\tau}(t)_{\tau}=\Psi_n(t)\mathbb{1}(|t|<\tau/2)$（$\Psi_n(t)$ 定义在 $(-\tau/2,\ \tau/2)$ 区间上）是两两正交的：

$$\int\Psi_n^{\tau}(t)\Psi_{n'}^{\tau}(t)\mathrm{d}t=\int_{-\tau/2}^{\tau/2}\Psi_n(t)\Psi_{n'}(t)\mathrm{d}t=0,\quad n\neq n' \tag{4.3.48a}$$

并且，$\Psi_n(t)$ 在 $L_2(-\tau/2,\ \tau/2)$ 上生成了一个完整的系统；如果对于所有的 $n\geqslant1$，函数 $\varphi\in L_2(-\tau/2,\ \tau/2)$ 满足 $\int_{-\tau/2}^{\tau/2}\varphi(t)\Psi_n(t)\mathrm{d}t=0$，那么在 $L_2(-\tau/2,\ \tau/2)$ 上 $\varphi(t)=0$。

(c) 对于所有的 $n\geqslant1$ 和 $t\in\mathbb{R}$，函数 $\Psi_n(t)$ 满足下式

[425]

$$\lambda_n\Psi_n(t)=2W\int_{-\tau/2}^{\tau/2}\Psi_n(s)\operatorname{sinc}(2W\pi(t-s))\mathrm{d}s \tag{4.3.48b}$$

就是说，函数 $\Psi_n(t)$ 是积分算子 $\varphi\to\int\varphi(s)K(\cdot,s)\mathrm{d}s$ 的特征函数（特征根为 λ_n），积分核

$$\begin{aligned}K(t,s)&=\mathbb{1}(|s|<\tau/2)(2W)\operatorname{sinc}(2W(t-s))\\&=\mathbb{1}(|s|<\tau/2)\frac{\sin(2\pi W(t-s))}{\pi(t-s)},\ -\tau/2\leqslant s,t\leqslant\tau/2\end{aligned}$$

(d) 特征根 λ_n 满足下面的条件

$$对于\ 1>\lambda_1>\lambda_2>\cdots>0,\quad 有\ \lambda_n=\int_{-\tau/2}^{\tau/2}\Psi_n(t)^2\mathrm{d}t$$

可以明确地给出一个等价的方程，其中包含 Fourier 变换 $[F\Psi_n^{\tau}](\omega)=\int\Psi_n^{\tau}(t)\mathrm{e}^{-\mathrm{i}\omega t}\mathrm{d}t$：

$$\frac{1}{2\pi}\int_{-2\pi W}^{2\pi W}|[F\Psi_n^{\tau}](\omega)|^2\mathrm{d}\omega\Big/\int_{-\tau/2}^{\tau/2}|\Psi_n(t)|^2\mathrm{d}t=\lambda_n$$

这意味着 λ_n 给出了对于截断函数 Ψ_n^{τ} 的“频域测重”点。

(e) 可以验证函数 $\Psi_n(t)$（以及特征根 λ_n）取决于 W 和 τ，且仅与二者的乘积 $W\tau$ 有关。并且，对于所有的 $\theta\in(0,\ 1)$，因为 $W\tau\to\infty$，有

$$\lambda_{2W\tau(1-\theta)}\to1\quad 和\quad \lambda_{2W\tau(1-\theta)}\to0 \tag{4.3.48c}$$

也就是说，对于足够大的 τ，大概有 $2W\tau$ 个 λ_n 接近 1，剩下的接近 0。

这个正在探索的推论中很重要的一部分是 Karhunen-loève 分解。假定 $Z(t)$ 是一个 Gauss 随机过程，其谱密度由式(4.3.27)给出。Karhunen-loève 分解定义了对于所有的 $t\in(-\tau/2,\ \tau/2)$，随机变量 $Z(t)$ 都可以被写成收敛级数（在均方意义上收敛）

$$Z(t)=\sum_{n\geqslant1}A_n\Psi_n(t) \tag{4.3.49}$$

其中 $\Psi_1(t)$，$\Psi_2(t)$，…是在例子 4.3.9 中讨论的 PSWF，而 A_1，A_2，…是 IID 随机变

量，并且满足 $A_n \sim N(0, \lambda_n)$，其中 λ_n 是对应的特征根。等价地可以写出 $Z(t) = \sum_{n\geq 1} \sqrt{\lambda_n}\xi_n\Psi_n(t)$，其中 $\xi_n \sim N(0, 1)$是 IID 随机变量。

该证明超出了本书的讨论范围，有兴趣的读者可以参考文献[38]或[103]，144 页。 426

关于定理 4.3.3 中证明的思想将在下面说明。给定 W 和 τ，$\mathscr{A}(\tau, W, p_0, \eta)$中的输入信号 $s^*(t)$可以在 PSWF Ψ_n 被表示为一个 Fourier 级数的形式。在这个级数里，前 $2W\tau$ 个被加数代表限制在频域带宽 $\pm 2\pi W$ 和时域 $\pm\tau/2$ 上的部分信号。相似地，噪声实现 $Z(t)$ 在函数 Ψ_n 级数上被分解。离散时间信道的作用体现在两个联合约束离散时间 Gauss 信道的并行联合。在信道 1 中处理信号分解中前 $2W\tau$ 个的 PSWF，并且 $\alpha_1 = 2W$。信道 2 接到展开式中剩余的部分，并且 $\alpha_2 = +\infty$。功率约束 $\|s\|^2 \leq p_0\tau$ 引出了一个联合约束，即式(4.3.38a)。另外，这也就有了这样一个需求，即分配在频带限制 $\pm 2\pi W$ 或时间限制 $\pm\tau/2$ 外的能量要小一些，这就引出了另一个功率约束，即式(4.3.38b)。在情况 1 中应用例子 4.3.5，就可得出定理 4.3.3。

为了让这些想法更精确，我们首先推导出定理 4.3.7，它给了一个其他的方法来得到 Nyquist-Shannon 公式(公式更复杂，但证明稍微简单(但仍然很冗长))。

定理 4.3.7 下面考虑定理 4.3.3 中模型的修正。存在一组由函数 $t \in \mathbb{R} \mapsto s(t)$组成的符合规则的信号$\mathscr{A}_\tau(\tau, W, p_0, \eta)$，可以使得

(1) $\|s\|^2 = \int |s(t)|^2 dt \leq p_0\tau$。

(2) 当 $|\omega| > 2\pi W$ 时，Fourier 变换$[Fs](\omega) = \int s(t)e^{i\omega t}dt$ 为零。

(3) 比值$\int_{-\tau/2}^{\tau/2} |s(t)|^2 dt / \|s\|^2 > 1 - \eta$。也就是，函数 $s \in \mathscr{A}(\tau, W, p_0, \eta)$ 在频域上是"严格带限"的，在时间上是"近似局限"的。

噪声过程是 Gauss 的，其中当 $|\omega| > 2\pi W$ 时，频率密度为零，而当 $|\omega| \leq 2\pi W$ 是频率密度等于 σ_0^2。

那么这个信道的容量为

$$C = C_\tau = W\ln\left(1 + (1-\eta)\frac{p_0}{2\sigma_0^2 W}\right) + \frac{\eta p_1}{2\sigma_1^2} \tag{4.3.50}$$

因为 $\eta \to 0$，

$$C_\tau \to W\ln\left(1 + \frac{p_0}{2\sigma_0^2 W}\right)$$

可得出 Nyquist-Shannon 公式(4.3.8)。 427

证明 首先，我们证明正向证明部分。令

$$R < W\ln\left(1 + \frac{(1-\eta)p_0}{2\sigma_0^2 W}\right) + \frac{\eta p_1}{2\sigma_1^2} \tag{4.3.51}$$

$\delta \in (0, 1)$，$\xi \in (0, \min[\eta, 1-\eta])$使得 R 仍然小于

$$C^* = W(1-\delta)\ln\left(1 + \frac{(1-\eta+\xi)p_0}{2\sigma_0^2 W(1-\delta)}\right) + \frac{(\eta-\xi)p_1}{2\sigma_1^2} \tag{4.3.52}$$

根据举例 4.3.5，C^* 是并行信道的联合约束离散时间对的容量，如情况 1，满足

$$\alpha_1 = 2W(1-\delta), \alpha_2 = +\infty, \beta = \eta - \xi, p = p_1, \sigma^2 = \sigma_1^2 \tag{4.3.53}$$

参考式(4.3.41a)。我们想为时间连续信道建立一个编码和解码规则，当 $\tau \to \infty$时候得到渐近错误消失概率。离散时间信道参数由式(4.3.53)中给出，假定$(x^{(1)}, x^{(2)})$是对于该离散

信道并行对的一个符合规则的输入信号，时间连续信道的是下列一系列(W, τ)的PSWF：

$$s(t)=\sum_{1\le k\le\lfloor\alpha_1\tau\rfloor} x_k^{(1)}\Psi_k(t)+\sum_{\lfloor\alpha_1\tau\rfloor<k<\infty} x_k^{(2)}\Psi_{k-\lfloor\alpha_1\tau\rfloor}(t) \tag{4.3.54}$$

需要验证的第一件事是式(4.3.54)中的信号属于$\mathscr{A}_s(\tau, W, p_c, \eta)$，也就是说，满足定理4.3.7中的条件(1)~(3)。

为了验证特性(1)，记

$$\|s\|^2=\sum_{1\le k\le\lfloor\alpha_1\tau\rfloor}(x_k^{(1)})^2+\sum_{1\le k<\infty}(x_k^{(2)})^2=\|x^{(1)}\|^2+\|x^{(2)}\|^2\le p_1\tau$$

然后，信号$s(t)$是带宽受限的，并且继承了PSWF$\Psi_k(t)$的特性。所以，(2)成立。

确定特性(3)需要更多相关的参数。因为PDWF $\Psi_k(t)$是在$\mathcal{L}_2(-\tau/2, \tau/2)$上正交的，并且通过值$\lambda_k$(参考式(4.3.48b))的单调性，我们可知

$$1-\int_{-\tau/2}^{\tau/2}|s(t)|^2\mathrm{d}t/\|s\|^2=\frac{\|(1-D_\tau)s\|^2}{\|s\|^2}$$

$$=\sum_{1\le k\le\lfloor\alpha_1\tau\rfloor}\frac{(1-\lambda_k)(x_k^{(1)})^2}{\|x^{(1)}\|^2+\|x^{(2)}\|^2}+\sum_{1\le k<\infty}\frac{(1-\lambda_{k-\lfloor\alpha_1\tau\rfloor})(x_k^{(2)})^2}{\|x^{(1)}\|^2+\|x^{(2)}\|^2}$$

428

$$\le(1-\lambda_{\lfloor\alpha_1\tau\rfloor})\frac{\|x^{(1)}\|^2}{\|x^{(1)}\|^2+\|x^{(2)}\|^2}+\frac{\|x^{(2)}\|^2}{\|x^{(1)}\|^2+\|x^{(2)}\|^2}$$

现在，随着$\tau\to\infty$，值$\lambda_{\lfloor\alpha_1\tau\rfloor}\to 1$(见(4.3.48c))，满足$\|x^{(1)}\|^2/(\|x^{(1)}\|^2+\|x^{(2)}\|^2)\le 1$时，对于足够大的$\tau$，下式成立

$$(1-\lambda_{\lfloor\alpha_1\tau\rfloor})\frac{\|x^{(1)}\|^2}{\|x^{(1)}\|^2+\|x^{(2)}\|^2}\le\xi$$

然后，比值$\|x^{(2)}\|^2/(\|x^{(1)}\|^2+\|x^{(2)}\|^2)\le\eta-\xi$(参考式(4.3.38b))，最后得到

$$1-\int_{-\tau/2}^{\tau/2}|s(t)|^2\mathrm{d}t/\|s\|^2=\frac{\|(1-D_\tau)s\|^2}{\|s\|^2}\le\xi+\eta-\xi=\eta$$

也就是特性(3)。

并且，噪声可以根据Karhunen-Loeve扩展：

$$Z(t)=\sum_{1\le k\le\lfloor\alpha_1\tau\rfloor}Z_k^{(1)}\Psi_k(t)+\sum_{\lfloor\alpha_1\tau\rfloor<k<\infty}Z_k^{(2)}\Psi_{k-\lfloor\alpha_1\tau\rfloor}(t) \tag{4.3.55}$$

这里，$\Psi_k(t)$仍然是PSWF和IID随机变量$Z_k^{(j)}\sim N(0, \lambda_k)$。相应地，输出信号可以记为

$$Y(t)=\sum_{1\le k\le\lfloor\alpha_1\tau\rfloor}Y_k^{(1)}\Psi_k(t)+\sum_{\lfloor\alpha_1\tau\rfloor<k<\infty}Y_k^{(2)}\Psi_{k-\lfloor\alpha_1\tau\rfloor}(t) \tag{4.3.56}$$

其中

$$Y_k^{(j)}=x_k^{(j)}+Z_k^{(j)}, j=1,2, k\ge 1 \tag{4.3.57}$$

所以，连续时间信道等价于联合约束并行组合。正如我们验证的，容量等于在式(4.3.32)中确定的C^*。所以，对于$R<C^*$，我们可以建立速率为R的码和使得错误概率趋近于0的解码规则。

相反地，假设存在一个序列$\tau^{(l)}\to\infty$，一系列在(1)~(3)中描述的可达域$\mathscr{A}_s^{(l)}(\tau^{(l)}, W, p_c, \eta^{(l)})$和一系列大小为$M=\lceil e^{R\tau^{(l)}}\rceil$码$\mathscr{C}^{(l)}$，其中

$$R>W\ln\left(1+\frac{(1-\eta)p_2}{2W\sigma_0^2}\right)-\frac{Wp_2}{\sigma_0^2}$$

和通常情况一样，我们希望证明错误概率$P_e=P_e^{\mathscr{C}^{(l)}, \mathscr{A}_s^{(l)}}(d^{(l)})$不趋近于0。

和之前一样，我们令$\delta\in(0, 1)$，$\xi\in(0, 1-\eta)$以满足$R>C^*$，其中

$$C^{*}=W(1+\delta)\ln\left(1+\frac{(1-\eta-\xi)}{(1-\xi)}\frac{p_0}{2W\sigma_0^2(1+\delta)}\right)+\frac{\eta\phi_0}{(1-\xi)\sigma_0^2}$$ 429

那么，正如在正向证明部分中的论证一样，C^{*} 是类型 1 信道联合约束并行组合的容量，其中

$$\beta=\frac{\eta}{1-\xi},\sigma^2=\sigma_0^2,p=p_0,\alpha_1=2W(1+\delta),\alpha_2=+\infty \tag{4.3.58}$$

令 $s(t)\in\mathscr{B}^{(2)}\cap\mathscr{A}_2^{(2)}(\tau^{(2)},W,p_0,\eta^{(2)})$ 为一个时间连续的码函数。因为 PSWF $\Psi_k(t)$ 生成一个 $L_2(\mathbb{R})$ 中的正交基，我们可以分解

$$s(t)=\sum_{1\leqslant k\leqslant\lceil\alpha_1\tau^{(1)}\rceil}x_k^{(1)}\Psi_k(t)+\sum_{1\leqslant k<\infty}x_k^{(2)}\Psi_{k+\lceil\alpha_1\tau^{(1)}\rceil}(t),t\in\mathbb{R} \tag{4.3.59}$$

我们想证明在式(4.3.38a～c)中所说明的类型 1 联合约束并行组合中，离散时间信号$(x^{(1)},x^{(2)})$代表了一个符合规则的输入信号。根据 $L_2(\mathbb{R})$中 PSWF $\Psi_k(t)$的正交性，我们可以得出

$$\|x\|^2=\|s\|^2\leqslant p_0\tau^{(2)}$$

来保证满足条件(4.3.38a)。并且，根据在 $L_2(-\tau/2,\tau/2)$中 PSW 函数 $\Psi_k(t)$的正交性以及特征根 λ_k 的单调减性，我们可以得到

$$\begin{aligned}1-\int_{-\tau^{(2)}/2}^{\tau^{(2)}/2}|s(t)|^2\mathrm{d}t/\|s\|^2&=\frac{\|(1-D_{\tau^{(2)}})s\|^2}{\|s\|^2}\\&=\sum_{1\leqslant k\leqslant\lceil\alpha_1\tau^{(1)}\rceil}\frac{(1-\lambda_k)(x_k^{(1)})^2}{\|x\|^2}+\sum_{1\leqslant k<\infty}\frac{(1-\lambda_{k+\lceil\alpha_1\tau^{(1)}\rceil})(x_k^{(2)})^2}{\|x\|^2}\\&\geqslant(1-\lambda_{\lceil\alpha_1\tau^{(1)}\rceil})\frac{\|x^{(2)}\|^2}{\|x\|^2}\end{aligned}$$

由于式(4.3.48c)，当 l 足够大的时候，$\lambda_{\lceil\alpha_1\tau^{(1)}\rceil}\leqslant\xi$ 成立。并且，因为 $1-\int_{-\tau^{(2)}/2}^{\tau^{(2)}/2}|s(t)|^2\mathrm{d}t/\|s\|^2\leqslant\eta$，我们可以得到

$$\frac{\|x^{(2)}\|^2}{\|x\|^2}\leqslant\frac{\eta}{1-\xi}$$

也可以推导出性质式(4.3.38b)。

然后，像在正向证明部分中那样，我们可以用噪声 $Z(t)$的 Karhunen-Loève 分解来推断对于每一个时间连续信道的码都对应着一个离散时间信道的联合限制并行组合的码，并且有着相同的速率和错误概率。因为 R 是大于 C^{*} 的，所以离散时间信道的容量与错误概率 $P_e=P_e^{\mathscr{B}^{(l)},\infty}(d^{(l)})$在 $l\to\infty$时是有非零界的。这就得到了相反的结论。 □ 430

定理 4.3.3 的证明 （概要）形式参数和定理 4.3.7 中一样，我们必须通过正向证明与反证法来证明该定理。前面提到过正向证明部分指出了容量是$\geqslant C$ 的，其值在式(4.3.31)中给出，而反证时即假设容量$\leqslant C$。

对于直接部分，例如情况 3，信道可以分解为两个并行信道的乘积，其中

$$\alpha_1=2W(1-\theta),\alpha_2=+\infty,p=p_0,\sigma^2=\sigma_0^2,\beta=\eta-\xi \tag{4.3.60}$$

其中 $\theta\in(0,1)$(参考例子 4.3.6 中 PSWF 的性质(e))，并且 $\xi\in(0,\eta)$是辅助值。

对于相反部分，我们可以再次像情况 3 中那样，将其分解为两个并行信道，其中

$$\alpha_1=2W(1+\theta),\alpha_2=+\infty,p=p_0,\sigma^2=\sigma_0^2,\beta=\frac{\eta}{1-\xi} \tag{4.3.61}$$

这里和之前一样，值 $\theta\in(0,1)$取自 PSWF 的性质(e)，而值 $\xi\in(0,1)$。 □

总结前面的观察，我们可以得到著名的

引理 4.3.8 （Nyquist-Shannon-Kotelnikov-Whittaker 采样引理）令 f 为一个函数 $t\in\mathbb{R}\mapsto$

$f(t)\in\mathbb{R}$，其中$\int|f(t)|\,dt<+\infty$，假定 Fourier 变换

$$[Ff](\omega)=\int e^{i\omega t}f(t)dt$$

当$|\omega|>2\pi W$时为零。然后，对于所有的$x\in\mathbb{R}$，函数f可以通过在点$x+n/(2W)$上的值$f(x+n/(2W))$被唯一地重建，其中$n=0,\pm1,\pm2$。更准确地说，对于所有的$t\in\mathbb{R}$都有

$$f(t)=\sum_{n\in\mathbb{Z}}f\left(\frac{n}{2W}\right)\frac{\sin[2\pi(Wt-n)]}{2\pi(Wt-n)}\tag{4.3.62}$$

举例 4.3.9 通过量子物理学著名的不确定定理，一个函数和其 Fourier 变换不可以同时局限在有限的间隔$[-\tau,\tau]$和$[-2\pi W,2\pi W]$上。什么情况下会有一个函数和它的 Fourier 变换是接近局限的？我们怎么量化在这种情况的不确定性？

解答 假定函数$f\in L_2(\mathbb{R})$并令$\hat{f}=Ff\in L_2(\mathbb{R})$为$f$的 Fourier 变换。($L_2(\mathbb{R})$是由在$\mathbb{R}$上的函数$f$构成，其中$\|f\|^2=\int|f(t)|^2dt<\infty$，并且对于所有的$f,g\in L_2(\mathbb{R})$，内积$\int f(t)\overline{g}(t)dt$是有限的)。我们可以知道，如果

431

$$\int_{-\tau}^{\tau}|f(t)|^2dt\Big/\int_{-\infty}^{\infty}|f(t)|^2dt=\alpha^2\tag{4.3.63}$$

和

$$\int_{-2\pi W}^{2\pi W}|Ff(\omega)|^2d\omega\Big/\int_{-\infty}^{+\infty}|Ff(\omega)|^2d\omega=\beta^2\tag{4.3.64}$$

那么$W\tau\geq\eta$。显然，可以发现$\eta=\eta(\alpha,\beta)$。(不等式将会很清楚，并且得到等式的函数也会很明确。)

线性算子$f\in L_2(\mathbb{R})\mapsto Df\in L_2(\mathbb{R})$和$f\in L_2(\mathbb{R})\mapsto Bf\in L_2(\mathbb{R})$可以通过

$$Df(t)=f(t)\mathbb{1}(|t|\leq\tau/2)\tag{4.3.65}$$

以及

$$Bf(t)=\frac{1}{2\pi}\int_{-2\pi W}^{2\pi W}Ff(\omega)e^{-i\omega t}d\omega=\frac{1}{\pi}\int_{-\infty}^{\infty}f(s)\frac{\sin 2\pi W(t-s)}{t-s}ds\tag{4.3.66}$$

得到。我们对这些算子的乘积很感兴趣，$A=BD$：

$$Af(t)=\frac{1}{\pi}\int_{-\tau/2}^{\tau/2}f(s)\frac{\sin 2\pi W(t-s)}{ts}ds\tag{4.3.67}$$

见例子 4.3.5。A的特征根λ_n遵循$1>\lambda_0>\lambda_1>\cdots$并且随着$n\to\infty$趋近于0；见文献[91]。我们也需要关注$\lambda_0$的特征根：可以证明$\lambda_0$是一个关于乘积$W\tau$的函数。事实上，(4.3.57)的特征函数$(\Psi_j)$产生了一个在$L_2(\mathbb{R})$中的正交基；同时这些函数形成了在$L_2[-\tau/2,\tau/2]$上的一个正交基：

$$\int_{-\tau/2}^{\tau/2}\Psi_j(t)\Psi_l(t)dt=\lambda_j\delta_{jl}$$

通常来讲，在 Hilbert 空间$L_2(\mathbb{R})$上f与g的角度由下式决定

$$\theta(f,g)=\cos^{-1}\left(\frac{1}{\|f\|\|g\|}\mathrm{Re}\int f(t)\overline{g}(t)dt\right)\tag{4.3.68}$$

这两个子空间的夹角是二者中向量之间的夹角的最小值。我们将证明了空间$\mathscr{B}$和$\mathscr{D}$(也即B和D算子的象空间)的正夹角$\theta(\mathscr{B},\mathscr{D})$是存在的。就是说，$\mathscr{B}$是所有有限带宽函数的线性子空间，而$\mathscr{D}$是时域有限函数的线性子空间。并且，

$$\theta(\mathscr{B},\mathscr{D}) = \cos^{-1}\sqrt{\lambda_0} \tag{4.3.69}$$ 432

以及在 $f=\Psi_0$，$g=D\Psi_0$ 时可以得到 $\inf_{f\in\mathscr{B},g\in\mathscr{D}}\theta(f,g)$，其中 Ψ_0 是(唯一)特征根为 λ_0 的特征函数。

为了达到这个目的，我们要验证对于任意的 $f\in\mathscr{B}$ 都有

$$\min_{g\in\mathscr{D}}\theta(f,g) = \cos^{-1}\frac{\|Df\|}{\|f\|} \tag{4.3.70}$$

的确，扩展 $f=f-Df+Df$ 并观察到积分 $\int[f(t)-Df(t)]g(t)\mathrm{d}t=0$(因为 g 和 $f-Df$ 是不相交的)，这意味着

$$\left|Re\int f(t)\overline{g}(t)\mathrm{d}t\right| \leqslant \left|\int f(t)\overline{g}(t)\mathrm{d}t\right| = \left|\int Df(t)\overline{g}(t)\mathrm{d}t\right|$$

所以

$$\frac{1}{\|f\|\|g\|}\mathrm{Re}\int f(t)\overline{g}(t)\mathrm{d}t \leqslant \frac{\|Df\|}{\|f\|}$$

意味着通过令 $g=Df$，式(4.3.70)成立。

然后，相对于 A 的特征函数，我们扩展 $f=\sum_{n=0}^{\infty}a_n\Psi_n$。这可以得到公式

$$\cos^{-1}\frac{\|Df\|}{\|f\|} = \cos^{-1}\left(\frac{\sum_n|a_n|^2\lambda_n}{\sum_n|a_n|^2}\right)^{1/2} \tag{4.3.71}$$

当 $a_n=0$ 对于 $n\geqslant 1$，$f=\Psi_0$ 成立时，式子右边可以得到关于 f 的上确界。我们得到结论，子空间 $\mathscr{B}$ 和 $\mathscr{D}$ 的最小夹角存在，并且正如所需的，这个角度在对 $f=\Psi_0$，$g=D\Psi_0$ 上可以得到。 □

接下来，我们引出如下定理。

引理 4.3.10 *当且仅当 α 和 β 落在下面情况(a)-(d)中的任意一个时，存在函数 $f\in L_2$ 使得 $\|f\|=1$，$\|Df\|=\alpha$ 并且 $\|Bf\|=\beta$。*

(a) $\alpha=0$ 以及 $0\leqslant\beta<1$。

(b) $0<\alpha<\sqrt{\lambda_0}<1$ 以及 $0\leqslant\beta\leqslant 1$。

(c) $\sqrt{\lambda_0}\leqslant\alpha\leqslant 1$ 以及 $\cos^{-1}\alpha+\cos^{-1}\beta\geqslant\cos^{-1}\sqrt{\lambda_0}$。

(d) $\alpha=1$ 以及 $0<\beta\leqslant\sqrt{\lambda_0}$。

证明 给定 $\alpha\in[0,1]$，令 $\mathscr{G}(\alpha)$ 为一组函数 $f\in L^2$，满足范数 $\|f\|=1$，$\|Df\|=\alpha$。接着，定义 $\beta^*(\alpha)_{\pm}=\sup_{f\in\mathscr{G}(\alpha)}\|Bf\|$。

(a) 如果 $\alpha=0$，函数组 $\mathscr{G}(0)$ 可以不包含任何满足 $\beta=\|Bf\|=1$ 的函数。并且，如果对于 $f\in\mathscr{B}$，有 $\|Df\|=0$ 以及 $\|Bf\|=1$，那么对于 $|t|<\pi/2$，f 是解析函数并且 $f(t)=0$，这意味着 $f=0$。为了证明 $\mathscr{G}(0)$ 包含满足所有值 $\beta\in[0,1)$ 的函数，我们令 $\widetilde{f}_n=\frac{\Psi_n-D\Psi_n}{\sqrt{1-\lambda_n}}$。那么范数 $\|B\widetilde{f}_n\|=\sqrt{1-\lambda_n}$。因为存在特征根 λ_n 任意趋近于 0，$\|B\widetilde{f}_n\|$ 会任意 433
趋近于 1。通过函数 $e^{ip}\widetilde{f}(t)$，我们可以得到在点 $\sqrt{1-\lambda_n}$ 之间的所有 β 的值，因为

$$\|Be^{ip}\widetilde{f}\| = \left(\int_{-\infty}^{+\infty}|F_n(\omega)|^2\mathrm{d}\omega\right)^{1/2}$$

范数 $\|Be^{ip}\widetilde{f}_n\|$ 在 p 上连续，并且当 $p\to\infty$ 时候，其值趋近 0。这就完成了对情况(a)的分析。

(b) 当 $0<\alpha<\sqrt{\lambda_1}<1$ 时，我们令

$$\tilde{f}=\frac{\sqrt{\alpha^2-\lambda_n}\Psi_1-\sqrt{\lambda_0-\alpha^2}\Psi_n}{\sqrt{\lambda_0-\lambda_n}}$$

当特征根 λ_n 趋近于 0 时，n 很大，我们可以得到 $\tilde{f}\in\mathscr{B}$，$\|\tilde{f}\|=\|B\tilde{f}\|=1$，而通过简单的运算可以证明 $\|D\tilde{f}\|=\alpha$，这包括了情况 $\beta=1$，正如通过选择合适的 $e^{i\theta}\tilde{f}(t)$ 我们可以得到任意的 $0<\beta<1$。

(c)和(d)如果 $\sqrt{\lambda_0}<\alpha<1$，如下所示，我们可以将 $f\in\mathscr{G}(\alpha)$ 分解为

$$f=a_1Df+a_2Bf+g \tag{4.3.72}$$

其中 g 和 Df 与 Bf 都正交。通过将式(4.3.72)的右边分别与 f、Df、Bf 做内积，我们可以得到四个等式：

$$1=a_1\alpha^2+a_2\beta^2+\int g(t)\overline{f}(t)\mathrm{d}t$$

$$\alpha^2=a_1\alpha^2+a_2\int Bf(t)\,\overline{Dg}(t)\mathrm{d}t$$

$$\beta^2=a_1\int Df(t)\,\overline{Bf}(t)\mathrm{d}t-a_2\beta^2$$

$$\int f(t)\overline{g}(t)\mathrm{d}t=\|g\|^2$$

这些等式意味着

[434]

$$\alpha^2+\beta^2-1+\|g\|^2=a_1\int Df(t)\,\overline{Bf}(t)\mathrm{d}t-a_2\int Bf(t)\,\overline{Df}(t)\mathrm{d}t$$

通过消除 $\int g(t)\overline{f}(t)\mathrm{d}t$，$a_1$ 和 a_2，我们发现对于 $\alpha\beta\neq0$，有

$$\beta^2=\frac{1-\alpha^2-\|g\|^2}{\left(\beta^2-\int Bf(t)\,\overline{Df}(t)\mathrm{d}t\right)}\beta^2$$
$$+\left(1-\frac{1-\alpha^2-\|g\|^2}{\alpha^2\left(\beta^2-\int Bf(t)\,\overline{Df}(t)\mathrm{d}t\right)}\int Bf(t)\,\overline{Df}(t)\mathrm{d}t\right)\times\int Df(t)\,\overline{Bf}(t)\mathrm{d}t$$

该式也等价于

$$\beta^2-2\mathrm{Re}\int Df(t)\,\overline{Bf}(t)\mathrm{d}t\leqslant-\alpha^2+\left(1-\frac{1}{\alpha^2\beta^2}\left|\int Df(t)\,\overline{Bf}(t)\mathrm{d}t\right|^2\right) \tag{4.3.73}$$

$$-\|g\|^2\left(1-\frac{1}{\alpha^2\beta^2}\left|\int Df(t)\,\overline{Bf}(t)\mathrm{d}t\right|^2\right) \tag{4.3.74}$$

对于角度 θ，我们可以写出

$$\alpha\beta\cos\theta=\mathrm{Re}\int Df(t)\,\overline{Bf}(t)\mathrm{d}t\leqslant\left|\int Df(t)\,\overline{Bf}(t)\mathrm{d}t\right|\leqslant\alpha\beta$$

将其代入式(4.3.74)中，并完成平方；当且仅当 $g=0$，$\int Df(t)\overline{Bf}(t)$ 是实数时，我们就可以等价地得到

$$(\beta-\alpha\cos\theta)^2\leqslant(1-\alpha^2)\sin^2\theta \tag{4.3.75}$$

因为 $\theta\geqslant\cos^{-1}\sqrt{\lambda_0}$，所以式(4.3.75)意味着

$$\cos^{-1}\alpha+\cos^{-1}\beta\geqslant\cos^{-1}\sqrt{\lambda_0} \tag{4.3.76}$$

满足式(4.3.76)的点(α,β)的迹是在曲线的右上方

$$\cos^{-1}\alpha+\cos^{-1}\beta=\cos^{-1}\sqrt{\lambda_3} \tag{4.3.77}$$

见图(4-6)。

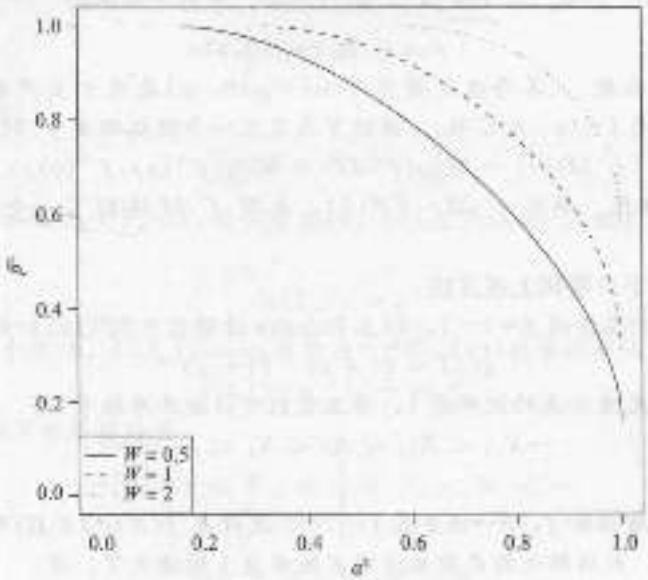


图 4-6

等式(4.3.77)在

$$b_1=\sqrt{\frac{1-\alpha^2}{1-\lambda_3}} \text{ 和 } b_2=\frac{\alpha}{\sqrt{\lambda_3}}-\sqrt{\frac{1-\alpha^2}{1-\lambda_3}}$$

时对于函数 $\tilde{f}=b_1\Psi_0+b_2D\Psi_0$ 成立。

所有 β 的中间值都可以通过 $e^{iX}\tilde{f}$ 再次得到。 □ 435

4.4 空间点过程和网络信息论

为了讨论分布系统的容量以及基于点过程的随机码本，我们需要知道一些背景知识。我们在这里学习的是 $\mathbb{R}^d$ 中的空间点过程，也介绍了一些更先进的点过程模型，给出了一个好的码距离。虽然 PSE II 中的一些材料会很有用，但是读者也可以独立于 PSE II 来阅读这部分。

定义 4.4.1 （参见 PSE II，211 页）令 μ 为 $\mathbb{R}$ 上的一个测度，用值 $\mu(A)$ 表示可测子集 $A\subset\mathbb{R}$ 的测度。假定 μ 是(i)非原子和(ii)σ-局部有限，即(i)对于所有可列集 $A\subset\mathbb{R}$ 都有 $\mu(A)=0$，(ii)存在一个 $\mathbb{R}$ 上的分割 $\mathbb{R}=\bigcup_j J_j$，可以得到两两互不相交区间 J_1，J_2，…使得 $\mu(J_j)<\infty$。如果对于所有在 $\mathbb{R}$ 上互不相交区间 I_1，…，I_n 的集合，值 $M(I_k)$ $(k=1,\cdots,n)$ 是相互独立的，并且每个 $M(I_k)\sim\mathrm{Po}(\mu(I_k))$，那么就可以说随机计数测度 M 定义了一个 Poisson 随机测度(缩写 PRM)，其均值或者说密度、测度为 μ。 436

关于在定义 4.4.1 中介绍的 Poisson 随机测度的存在性和特性，我们将在没有证明的情况下陈述一些事实。

定理 4.4.2 对于在 $\mathbb{R}$ 任何非原子和 σ-局部有限的测度 μ 都存在一个满足定义 4.4.1 的唯一 PRM。如果测度 μ 的形式是 $\mu(\mathrm{d}t)=\lambda\mathrm{d}t$，$\lambda>0$ 是一个常数(称为 μ 的密度)，这个 PRM 是一个 Poisson 过程 $\mathrm{PP}(\lambda)$。如果测度 μ 的形式是 $\mu(\mathrm{d}t)=\lambda(t)\mathrm{d}t$，其中 $\lambda(t)$ 是一个给定的函数，那么这个 PRM 就给出了一个非齐次的 Poisson 过程 $\mathrm{PP}(\lambda(t))$。

定理 4.4.3 （映射定理）令 μ 是在 $\mathbb{R}$ 上非原子和 σ-局部有限的测度，使得对于所有的 $t\geq 0$ 和 $h>0$，区间 $(t, t+h)$ 的测度 $\mu(t, t+h)$ 是正数且有限（也即值 $\mu(t, t+h)\in(0, \infty)$），满足 $\lim_{h\to 0}\mu(0, h)=0$，$\mu(\mathbb{R}_+)=\lim_{u\to\infty}\mu(0, u)=+\infty$。考虑下面函数

$$f: u\in\mathbb{R}_+\mapsto\mu(0, u)$$

并令 f^{-1} 为 f 的反函数。（其存在是因为 $f(u)=\mu(0, u)$ 是关于 u 严格单调的）令 M 为 $\mathrm{PRM}(\mu)$。对于区间 $I=(a, b)\subset\mathbb{R}_+$，通过下式定义一个随机测度 f^*M

$$(f^*M)(I)=M(\mu(f^{-1}I))=M(\mu(f^{-1}(a), f^{-1}(b))) \tag{4.4.1}$$

然后将它在 $\mathbb{R}$ 上延续。那么 $f^*M\sim PP(1)$，也即 f^*M 得到了一个单位率的 Poisson 过程。

我们在两个例子中阐明上述方法。

举例 4.4.4 令一个在区间 $S=(-1, 1)$ 上 Poisson 过程 $\Pi=PP(\lambda(x))$ 的率函数为

$$\lambda(x)=(1+x)^{-3}(1-x)^{-3}$$

证明 Π 在 S 中有无穷个点的概率是 1，并且它们可以按升序标号

$$\cdots X_{-2}<X_{-1}<X_0<X_1<X_2<\cdots$$

其中 $X_0<0<X_1$。

证明存在一个增函数 $f: S\to\mathbb{R}$ 并且 $f(0)=0$ 使得点 $f(X)(X\in\Pi)$ 形成一个 $\mathbb{R}$ 上单位率的 Poisson 过程，并用强大数定理来证明在概率为 1 的情况下，有

$$\lim_{n\to\infty}(2n)^{1/2}(1-X_n)=\frac{1}{2} \tag{4.4.2}$$

437 找到一个随着 $n\to\infty$ 的相应结果。

解答 因为

$$\int_{-1}^{1}\lambda(x)\mathrm{d}x=\infty$$

所以在 $(-1, 1)$ 里有无限多点的概率是 1。另一方面，对于每一个 $\delta>0$，都有

$$\int_{-1+\delta}^{1-\delta}\lambda(x)\mathrm{d}x<\infty$$

使得 $\Pi(-1+\delta, 1-\delta)$ 是有限的且概率为 1。这足以来按升序给 Π 的点唯一标号。令

$$f(x)=\int_0^x\lambda(y)\mathrm{d}y$$

因为 $f: S\to\mathbb{R}$ 是增的，所以 f 将 Π 映射到一个 Poisson 过程中，该过程的均值测度 μ 由下式给出

$$\mu(a, b)=\int_{f^{-1}(a)}^{f^{-1}(b)}\lambda(x)\mathrm{d}x=b-a$$

选择了这个 f，点 $(f(X_n))$ 形成一个 $\mathbb{R}$ 上单位率的 Poisson 过程。强大数定理证明了随着 $n\to\infty$，下式的概率为 1

$$n^{-1}f(X_n)\to 1, \text{且 } n^{-1}f(X_{-n})\to -1$$

现在我们观察到

$$\lambda(x)\sim\frac{1}{4}(1-x)^{-3}\text{ 且 } f(x)\sim\frac{1}{8}(1-x)^{-2},\quad \text{当 } x\to 1\text{ 时}$$

所以，随着 $n\to\infty$，下式概率为 1，

$$n^{-1}\frac{1}{8}(1-X_n)^{-2}\to 1$$

该式等价于式(4.4.2)。类似地

$$\lambda(x) \sim \frac{1}{8}(1+x)^{-2} \text{ 并且随着 } x \to -1, f(x) \sim \frac{1}{8}(1+x)^{-1}$$

意味着，随着 $n \to \infty$，下式概率为 1

$$n^{-1}\frac{1}{8}(1+X_{-n})^{-1} \to 1$$

所以，下式概率为 1

$$\lim_{n\to\infty} n(1+X_{-n}) = \frac{1}{8}$$ □ 438

举例 4.4.5 证明如果 $Y_1 < Y_2 < Y_3 < \cdots$ 是在 $(0, \infty)$ 上 Poisson 过程的点，其率函数为常数 λ，那么

$$\lim_{n\to\infty} Y_n/n = \lambda$$

的概率为 1。令一个在 $(0, 1)$ 上 Poisson 过程 $\Pi = PP(\lambda(x))$ 的率函数为

$$\lambda(x) = x^{-2}(1-x)^{-1}$$

证明 Π 的点可以按下面这样标号

$$\cdots < X_{-2} < X_{-1} < \frac{1}{2} < X_0 < X_1 < \cdots$$

并且

$$\lim_{n\to-\infty} X_n = 0, \lim_{n\to\infty} X_n = 1$$

证明

$$\lim_{n\to\infty} nX_{-n} = 1$$

的概率是 1。随着 $n \to +\infty$，X_n 的极限特性是什么？

解答 第一部分依然可以根据强大数定理来推断。对于第二部分，我们令

$$f(x) = \int_{1/2}^{x} \lambda(\xi)\mathrm{d}\xi$$

并可知 f 将 Π 映射到一个常速率 $(f(0), f(1))$，$f(\Pi) = PP(1)$ 的 PP 上。在我们讨论的情况中，$f(0) = -\infty$，$f(1) = \infty$，所以 $f(\Pi)$ 是一个在上最的 PP，它的点可能被这样标号

$$\cdots < Y_{-2} < Y_{-1} < 0 < Y_0 < Y_1 < \cdots$$

其中

$$\lim_{n\to-\infty} Y_n = -\infty, \lim_{n\to\infty} Y_n = +\infty$$

那么 $X_n = f^{-1}(Y_n)$ 就有了所需的特性。

强大数定理应用到 Y_n，可以得出

$$\lim_{n\to\infty} \frac{f(X_n)}{n} = \lim_{n\to\infty} \frac{Y_n}{n} = 1 \quad \text{a. s.}$$

现在，随着 $x \to 0$，

$$f(x) = \int_x^{1/2} \xi^{-2}(1-\xi)^{-1}\mathrm{d}\xi \sim -\int_x^{1/2} \xi^{-2}\mathrm{d}\xi \sim -x^{-1}$$ 439

意味着

$$\lim_{n\to-\infty} \frac{X_n^{-1}}{n} = 1, \quad \text{即} \lim_{n\to-\infty} nX_{-n} = 1 \quad \text{a. s.}$$

类似地，

$$\lim_{n\to\infty} \frac{f(X_n)}{n} = 1 \quad \text{a. s.}$$

并且随着 $x\to 1$，

$$f(x)\sim\int_{1/2}^{x}(1-\xi)^{-1}\mathrm{d}\xi=-\ln(1-x)$$

这意味着

$$\lim_{n\to\infty}-\frac{\ln(1-X_n)}{n}=1\quad \text{a.s.}$$ □

接下来，我们讨论在一般集合 E 上的 Poisson 随机测度(PRM)，形式上，我们假定 E 被赋予子集的一个 σ 代数 $\mathscr{E}$，和一个测度 μ，该测度给每一个 $A\in\mathscr{E}$ 分配一个值 $\mu(A)$，使得如果 A_1，A_2，…是 $\mathscr{E}$ 中两两互不相交的子集，那么有

$$\mu(\bigcup_n A_n)=\sum_n\mu(A_n)$$

值 $\mu(E)$ 可以是有限的或无限的。我们的目的是定义一个随机计数测度 $M=(M(A),A\in\mathscr{E})$，该测度有以下特征：

(a) 随机变量 $M(A)$ 是非负整数(可能包括正无穷)，并且

$$M(A)\begin{cases}\sim \mathrm{Po}(\lambda\mu(A)), & \mu(A)<\infty\\ =+\infty \text{ 概率为 } 1, & \mu(A)=\infty\end{cases}\tag{4.4.3}$$

(b) 如果 A_1，A_2，…$\in\mathscr{E}$ 是不相交子集，那么

$$M(\bigcup_i A_i)=\sum_i M(A_i)\tag{4.4.4}$$

(c) 如果集合 A_1，A_2，…$\in\mathscr{E}$ 是不相交的，则随机变量 $M(A_1)$，$M(A_2)$，…是独立的。就是说，对于所有不相交集合 A_1，A_2，…$\in\mathscr{E}$ 的有限族和非负整数 k_1，…，k_n

440

$$\mathbb{P}(M(A_i)=k_i,1\leqslant i\leqslant n)=\prod_{1\leqslant i\leqslant n}\mathbb{P}(M(A_i)=k_i)\tag{4.4.5}$$

首先假设 $\mu(E)<\infty$(如果不是，将 E 分解为有限测度的子集)。给定一个随机变量 $M(E)\sim\mathrm{Po}(\lambda\mu(E))$。考虑 E 中 IID 随机点的一个序列 X_1，X_2，…，其中 $X_i\sim\mu/\mu(E)$，独立于 $M(E)$。也就是说对于所有 $n\geqslant 1$ 和集合 A_1，…$A_n\in\mathscr{E}$(不一定不相交)

$$\mathbb{P}(M(E)=n,X_1\in A_1,\cdots,X_n\in A_n)=\mathrm{e}^{-\lambda\mu(E)}\frac{(\lambda\mu(E))^n}{n!}\prod_{i=1}^{n}\frac{\mu(A_i)}{\mu(E)}\tag{4.4.6}$$

并且有条件地

$$\mathbb{P}(X_1\in A_1,\cdots,X_n\in A_n\mid M(E)=n)=\prod_{i=1}^{n}\frac{\mu(A_i)}{\mu(E)}\tag{4.4.7}$$

那么集合

$$M(A)=\sum_{i=1}^{M(E)}\mathbb{1}(X_i\in A),A\in\mathscr{E}\tag{4.4.8}$$

定理 4.4.6 如果 $\mu(E)<\infty$，等式(4.4.8)定义了一个 E 上的随机测度，满足上述(a)-(c)的特性。

举例 4.4.7 令 M 是一个在平面 $\mathbb{R}^2$ 上密度为 λ 的 Poisson 随机测度。圆 $\{x\in\mathbb{R}^2:|x|<r\}$ 在 $\mathbb{R}^2$ 上由 $C(r)$ 表示，其半径为 r，圆心在原点处，并令 R_k 为可以使得 $C(R_k)$ 准确包含 M 的 k 个点的最大半径。(所以 $C(R_k)$ 是一个在圆点处不包含 M 中任何点的最大的圆，$C(R_1)$ 是在圆点处仅包含 M 中一个点的最大的圆，以此类推。)计算 $\mathbb{E}R_2$，$\mathbb{E}R_1$ 和 $\mathbb{E}R_0$。

解答 易知

$$\mathbb{P}(R_0>r)=\mathbb{P}(C(r)\text{ 不包含 } M \text{ 中的点})=\mathrm{e}^{-\lambda\pi r^2},r>0,$$

以及

$$\mathbb{P}(R_1 > r) = \mathbb{P}(C(r)\text{ 至多包含一个 } M \text{ 中的点}) = (1+\lambda\pi r^2)e^{-\lambda\pi r^2}, r > 0.$$

类似地

$$\mathbb{P}(R_2 > r) = \left[1 + \lambda\pi r^2 + \frac{1}{2}(\lambda\pi r^2)^2\right]e^{-\lambda\pi r^2}, r > 0$$ 441

那么

$$\begin{aligned}
\mathbb{E}R_0 &= \int_0^\infty \mathbb{P}(R_0 > r)\mathrm{d}r = \frac{1}{\sqrt{2\pi\lambda}}\int_0^\infty e^{-u^2}\mathrm{d}(\sqrt{2\pi\lambda}r) = \frac{1}{2\sqrt{\lambda}} \\
\mathbb{E}R_1 &= \int_0^\infty \mathbb{P}(R_1 > r)\mathrm{d}r \\
&= \frac{1}{2\sqrt{\lambda}} + \int_0^\infty e^{-\lambda\pi r^2}(\lambda\pi r^2)\mathrm{d}r \\
&= \frac{1}{2\sqrt{\lambda}} + \frac{1}{2\sqrt{2\pi\lambda}}\int_0^\infty (2\pi\lambda r^2)e^{-\pi\lambda r^2}\mathrm{d}(\sqrt{2\pi\lambda}r) \\
&= \frac{3}{4\sqrt{\lambda}} \\
\mathbb{E}R_2 &= \frac{3}{4\sqrt{\lambda}} + \int_0^\infty \frac{(\lambda\pi r^2)^2}{2}e^{-\lambda\pi r^2}\mathrm{d}r \\
&= \frac{3}{4\sqrt{\lambda}} + \frac{1}{8\sqrt{2\pi\lambda}}\int_0^\infty (2\lambda\pi r^2)^2 e^{-\pi\lambda r^2}\mathrm{d}(\sqrt{2\lambda\pi}r) \\
&= \frac{3}{4\sqrt{\lambda}} + \frac{3}{16\sqrt{\lambda}} \\
&= \frac{15}{16\sqrt{\lambda}}
\end{aligned}$$

我们应该使用 PRM(E, μ)表示在相位空间 E 上的 PRM M，其测度为定理 4.4.6 中建立的密度为 μ 的测度。然后，我们将 PRM 的定义扩展到积分和上：对于所有函数：$g: E \to \mathbb{R}$，定义

$$M(g) = \sum_{i=1}^{M(E)} g(X_i) := \int g(y)\mathrm{d}M(y) \tag{4.4.9}$$

在所有点 $X_i \in E$ 上求总和，并且 $M(E)$ 是这些点的数目总和。然后，对于一个常规的 $g: E \to \mathbb{R}$，我们令

$$M(g) = M(g_+) - M(-g_-)$$

对于所有 $a \in (0, \infty)$，约定 $+\infty - a = +\infty$ 以及 $a - \infty = -\infty$。（当 $M(g_+)$ $M(-g_-)$ 都等于 ∞ 时候，值 $M(g)$ 被声明为是未定义的。）

然后

定理 4.4.8 （Campbell 定理）对于所有 $\theta \in \mathbb{R}$ 以及所有函数 $g: E \to \mathbb{R}$ 使得 $e^{\theta g(y)} - 1$ 是 μ 可积的

$$\mathbb{E}e^{\theta M(g)} = \exp\left[\lambda\int_E (e^{\theta g(y)} - 1)\mathrm{d}\mu(y)\right] \tag{4.4.10}$$ 442

证明 记

$$\begin{aligned}
\mathbb{E}e^{\theta M(g)} &= \mathbb{E}[\mathbb{E}(e^{\theta M(g)} \mid M(E))] \\
&= \sum_k \mathbb{P}(M(E) = k)\mathbb{E}\left(\exp\left[\theta\sum_{i=1}^k g(X_i)\right] \middle| M(E) = k\right)
\end{aligned}$$

由于条件独立(4.4.7)

$$\mathbb{E}\left(\exp\left[\theta\sum_{i=1}^{k}g(X_i)\right]\middle| M(E)=k\right)=\prod_{i=1}^{k}\mathbb{E}e^{\theta g(X_i)}=\left(\mathbb{E}e^{\theta g(X_1)}\right)^k$$
$$=\left(\frac{1}{\mu(E)}\int_E e^{\theta g(x)}\mathrm{d}\mu(x)\right)^k$$

以及

$$\mathbb{E}e^{\theta M(g)}=\sum_k e^{-\lambda\mu(E)}\frac{(\lambda\mu(E))^k}{k!}\frac{1}{(\mu(E))^k}\left(\int_E e^{\theta g(x)}\mathrm{d}\mu(x)\right)^k$$
$$=e^{-\lambda\mu(E)}\exp\left[\lambda\int_E e^{\theta g(x)}\mathrm{d}\mu(x)\right]$$
$$=\exp\left[\lambda\int_E(e^{\theta g(x)}-1)\mathrm{d}\mu(x)\right]$$

□

推论 4.4.9 $M(g)$的期望值由下式给出

$$\mathbb{E}M(g)=\lambda\int_E g(y)\mathrm{d}\mu(y)$$

当且仅当等式右边被很好地定义时，它才存在。

例子 4.4.10 假定无线发送端位于 Poisson 过程 Π 的点上，该过程在 $\mathbb{R}^2$ 上且速率为 λ。令 r_i 为发端 i 到中心接收者的距离，并且 r_0 是到发端的最小距离。假定对于 $\alpha>2$，接收信号的功率是 $Y=\sum_{r_i\geq a}\frac{P}{r_i^\alpha}$。那么

$$\mathbb{E}e^{\theta Y}=\exp\left[2\lambda\pi\int_{r_0}^{\infty}(e^{\theta g(r)}-1)r\mathrm{d}r\right] \tag{4.4.11}$$

[443] 其中 $g(r)=\frac{P}{r^\alpha}$，这里 P 是发送功率。

实际应用中有一个很常见的模型，就是所谓的标记点过程，其空间是标记 D。这只是一个在$\mathbb{R}^d\times D$或其子集上的随机测度。在最简单的构建中，我们将需要下面所证明的乘积特性。

定理 4.4.11 (乘积定理)假定给定一个在 $\mathbb{R}$ 上的 Poisson 过程，其常速率为 λ，并记 Y_i 是分布为 v 的 IID。在 M 上按下式 $\mathbb{R}_+\times D$ 定义一个随机测度

$$M(A)=\sum_{i=1}^{\infty}I((T_i,Y_i)\in A),A\subseteq\mathbb{R}\times D \tag{4.4.12}$$

这个测度是在 $\mathbb{R}_+\times D$ 上的一个 PRM，其密度测度为 $\lambda m\times v$，其中 m 是一个 Lebesgue 测度。

证明 首先考虑一个集合 $A\subseteq[0,t)\times D$，其中 $t>0$。那么

$$M(A)=\sum_{i=1}^{N_t}\mathbb{I}((T_i,Y_i)\in A)$$

考虑 MGF $\mathbb{E}e^{\theta M(A)}$ 并用标准条件

$$\mathbb{E}e^{\theta M(A)}=\mathbb{E}[\mathbb{E}(e^{\theta M(A)}\mid N_t)]=\sum_{k=0}^{\infty}\mathbb{P}(N_t=k)\mathbb{E}(e^{\theta M(A)}\mid N_t=k)$$

我们知道 $N_t\sim\mathrm{Po}(\lambda t)$。另外，给定 $N_t=k$，断点 $T_1,\cdots,T_k$ 条件分布 PDF $f_{T_1,\cdots,T_k}(\cdot\mid N_t=k)$ 由式(4.4.7)给出。然后，通过进一步调整，考虑到 $T_1,\cdots,T_k$ 和 Y_i 的独立性，我们有

$$\mathbb{E}(e^{\theta M(A)}\mid N_t=k)$$
$$=\mathbb{E}[\mathbb{E}(e^{\theta M(A)}\mid N_t=k;T_1,\cdots,T_k)]$$

$$-\int_0^t\cdots\int_0^t \mathrm{d}x_k\cdots\mathrm{d}x_1 f_{T_1,\cdots,T_k}(x_1,\cdots,x_k\mid N=k)$$

$$\times\mathbb{E}\Big(\exp\theta\Big(\sum_{i=1}^{k}\mathrm{I}((x_i,Y_i)\in A)\Big)\Big|N_t=k;T_1=x_1,\cdots,T_k=x_k\Big)$$

$$-\frac{1}{t^k}\Big(\int_0^t\int_D e^{\theta\mathrm{I}_A(x,y)}\mathrm{d}v(y)\mathrm{d}x\Big)^k$$

444

那么

$$\mathbb{E}e^{\theta M(A)}=e^{-\lambda t}\sum_{k=0}^{\infty}\frac{(\lambda t)^k}{k!}\frac{1}{t^k}\Big(\int_0^t\int_D e^{\theta\mathrm{I}_A(x,y)}\mathrm{d}v(y)\mathrm{d}x\Big)^k$$

$$=\exp\Big[\lambda\int_0^t\int_D(e^{\theta\mathrm{I}_A(x,y)}-1)\mathrm{d}v(y)\mathrm{d}x\Big]$$

表达式 $e^{\theta\mathrm{I}_A(x,y)}-1$ 对 $(x,y)\in A$ 取值 $e^{\theta}-1$，对 $(x,y)\notin A$ 取值 0。

因此

$$\mathbb{E}e^{\theta M(A)}=\exp\Big[(e^{\theta}-1)\lambda\int_A\mathrm{d}v(y)\mathrm{d}x\Big],\quad \theta\in\mathbb{R} \tag{4.4.13}$$

所以，$M(A)\sim\mathrm{Po}(\lambda m\times v(A))$。 □

另外，如果 $A_1,\cdots,A_n$ 是 $[0,t)\times D$ 互不相交的子集，那么随机变量 $M(A_1),\cdots,M(A_n)$ 是独立的。要明白这个问题，我们首先注意根据定义，M 是可加的：$M(A)=M(A_1)+\cdots+M(A_n)$，其中 $A=A_1\cup\cdots\cup A_n$。根据式(4.4.13)

$$\mathbb{E}e^{\theta M(A)}=\exp\Big[(e^{\theta}-1)\lambda\sum_{i=1}^{n}\int_{A_i}\mathrm{d}v(y)\mathrm{d}x\Big]=\prod_{i=1}^{n}\mathbb{E}e^{\theta M(A_i)},\theta\in\mathbb{R}$$

这意味着独立性。

所以，M 到 $\overline{E}_n=[0,t)\times D$ 的约束是一个 $(\overline{E}_n,\lambda\mathrm{d}m_n\times v)$PRM，其中 $m_n=m|_{[0,m_n)}$，那么，通过扩张性质，M 是一个 $(\mathbb{R}_+\times D,\lambda m\times v)$PRM。

举例 4.4.12 用乘积和 Campcell 定理来解决下面问题。星星按一个密度为 $v(X)(X\in\mathbb{R}^3)$ 的 Poisson 过程 Π 分布在三维空间 $\mathbb{R}^3$ 上。星星的质量是 IID 随机变量；X 处星星的质量 m_X 的 PDF 为 $p(X,\mathrm{d}m)$，圆点处的重力势能为

$$F=\sum_{X\in\Pi}\frac{Gm_X}{|X|}$$

其中 G 是一个常数。求出 MGF $\mathbb{E}e^{\theta F}$。

星系是半径为 R，圆心在原点的一个球。球内部点 x 处的星星密度为 $v(x)=1/|x|$；每个星星的质量都是均值为 M 的指数分布。计算预期的势能，因为星系在原点。令 C 为正常数。求出从原点到最近一颗对势能 F 贡献至少是 C 的星星的距离分布。 445

解答 Campbell 定理指出如果 M 是一个在空间 E 上，密度测度为 v 的随机测度，并且 $a:E\to\mathbb{R}$ 是一个有界可测函数，则

$$\mathbb{E}e^{\theta\Sigma}=\exp\Big(\int_E(e^{\theta a(y)}-1)v(\mathrm{d}y)\Big)$$

其中

$$\Sigma=\int_E a(y)M(\mathrm{d}y)=\sum_{X\in\Pi}a(X)$$

根据乘积定理，对 (X,m_X)（位置，质量）在 $\mathbb{R}^3\times\mathbb{R}_+$ 形成了一个 PRM，其密度测度为 $\mu(\mathrm{d}x\times\mathrm{d}m)=v(x)\mathrm{d}x\rho(x,\mathrm{d}m)$。然后根据 Campbell 定理：

$$\mathbb{E}e^{\theta F}=\exp\left(\int_{\mathbb{R}^3}\int_0^\infty \mu(\mathrm{d}x\times\mathrm{d}m)(e^{\theta Gm/|x|}-1)\right)$$

在原点处预期的势能是 $\mathbb{E}F=\frac{\mathrm{d}\mathbb{E}e^{\theta F}}{\mathrm{d}\theta}|_{\theta=0}$ 并且等于

$$\int_{\mathbb{R}^3}v(x)\mathrm{d}x\int_0^\infty \rho(x,\mathrm{d}m)\frac{Gm}{|x|}=GM\int_{\mathbb{R}^3}\mathrm{d}x\frac{1}{|x|^2}\mathbb{1}(|x|\leqslant R)$$

在球坐标中

$$\int_{\mathbb{R}^3}\mathrm{d}x\frac{1}{|x|^2}\mathbb{1}(|x|\leqslant R)=\int_0^R\mathrm{d}r\frac{1}{r^2}r^2\int\mathrm{d}\vartheta\cos\vartheta\int d\phi=4\pi R$$

可以得到

$$\mathbb{E}F=4\pi GMR$$

最后，令 D 为表示到最近一颗对 F 贡献至少为 C 的星星的距离。那么，根据乘积定理，

$$\mathbb{P}(D\geqslant d)=\mathbb{P}(A\text{ 上无点})=\exp(-\mu(A))$$

这里

446

$$A=\{(x,m)\in\mathbb{R}^3\times\mathbb{R}_+:|x|\leqslant d,\frac{Gm}{|x|}\geqslant C\}$$

并且 $\mu(A)=\int_A\mu(\mathrm{d}x\times\mathrm{d}m)$ 由下式表示

$$\begin{aligned}&\int_0^d\mathrm{d}r\frac{1}{r^2}r^2\int\mathrm{d}\vartheta\cos\vartheta/\mathrm{d}\phi M^{-1}\int_{C/G}^\infty\mathrm{d}m e^{-m/M}\\&=4\pi\int_0^d\mathrm{d}r e^{-G/(GM)}\\&=4\pi\left(\frac{GM}{C}\right)^2\left(1-e^{-C/(GM)}-\frac{Cd}{GM}e^{-C/(GM)}\right)\end{aligned}$$

这决定了在[0，R]上的 D 的分布。 □

在发送端和接收端分布式系统中，比如移动手机无线网络，无线网络中节点对之间容许通信速率取决于它们的随机位置和传输策略。通常，传输是沿着源到目的节点的发送链执行的。所以信息论中就出现了一个新的有趣方向：一些专家甚至创造了“网络信息论”。这方面的研究和概率论有千丝万缕的联系，特别是渗流和空间点过程。就算我们这里完全不提这个领域的快速发展，如今信息论的表述也不能完全地避开网络方面。在这里我们只对网络信息论中的一些话题稍作讨论，有兴趣的读者可以参考文献[48]和其中的参考文献。

例子 4.4.13 假定接收者位于点 y 并且发送端分布在 $\mathbb{R}^2$ 平面中速率为 λ 的 Poisson 过程的点 $x_i\in\Pi$ 上。那么接收信号功率最简单的模型是

$$Y=\sum_{x_i\in\Pi}p\ell(|x_i-y|)\tag{4.4.14}$$

其中 P 是发送信号功率，函数 ℓ 用来描述信号的衰落。在所谓的 Rayleigh 衰落情况中 $\ell(|x|)=e^{-\beta|x|}$ 以及功率衰落中 $\ell(|x|)=|x|^{-\alpha}$，$\alpha>2$。根据 Campbell 定理

$$\phi(\theta)=\mathbb{E}[e^{\theta Y}]=\exp\left(2\lambda\pi\int_0^\infty r(e^{\theta p\ell(r)}-1)\mathrm{d}r\right)\tag{4.4.15}$$

下面描述了一个更加实际的无线网络模型。假定接收端位于点 y_j，$j=1,\cdots,J$，上，并

447

且发送方分布在 $\mathbb{R}^2$ 平面中速率为 λ 的 Poisson 过程的点 $x_i\in\Pi$ 上。假定点 x_i 处的信号 S_i 被按照参数 $\sqrt{P}$ 放大，我们可以将信号记为

$$Y_j=\sum_{x_i\in\Pi}h_{ij}S_i+Z_j,j=1,\cdots,J\tag{4.4.16}$$

这个最简单模型的传输函数是

$$h_{jk}=\sqrt{P}\frac{e^{2\pi r_{jk}/v}}{r_{jk}^{\alpha/2}} \tag{4.4.17}$$

其中 v 是发送波长，$r_{jk}=|y_j-x_k|$。假定噪声随机变量 Z_j 是 IDD $N(0,\sigma_0^2)$。Rayleigh 衰落也可以类似地得出。我们知道当 $J=1$，单个发送端 $K=1$ 时，根据 4.3 中 Nyquist-Shannon 定理，连续时间加性 Gauss 白噪声信道 $Y(t)=X(t)\ell(x,y)+Z(t)$ 在衰减因子为 $\ell(x,y)$、功率限制为 $\int_{-\tau/2}^{\tau/2}X^2(t)dt\leqslant P\tau$、带宽为 W、噪声谱密度 σ_0^2 的情况下的容量是

$$C=W\log\left(1+\frac{P\ell^2(x,y)}{2W\sigma_0^2}\right) \tag{4.4.18}$$

然后，考虑有限的 K 个发送端和 J 个接收端

$$y_j(t)=\sum_{k=1}^{K}\ell(x_k,y_j)x_k(t)+z_j(t),j=1,\cdots,J \tag{4.4.19}$$

其中对于发送端 $k=1,\cdots,K$，功率限制为 P_k。将例子 4.3.5 用在并行信道的容量中，可以证明(参考文献[48])信道容量是

$$C=\sum_{k=1}^{K}W\log\left(1+\frac{P_ks_k^2}{2W\sigma_0^2}\right) \tag{4.4.20}$$

其中 s_k 是矩阵 $L=\ell(|y_j-x_k|)$ 第 k 大的奇异值。然后，我们假定带宽 $W=1$。在平均传输功率满足 $K^{-1}\sum_kP_k\leqslant P$ 时，描述有 K 个发送端和 J 个接收端的容量域是很有趣的。有兴趣的读者可以参考文献[48]，对于容许速率 R_{kj}，下面的容量域可以被证实：

$$\sum_{k=1}^{K}\sum_{j=1}^{J}R_{kj}\leqslant\max_{P_k\geqslant 0:\sum_kP_k\leqslant KP}\sum_{k=1}^{K}\log\left(1+\frac{P_ks_k^2}{2\sigma_0^2}\right) \tag{4.4.21}$$

定理 4.4.14 考虑区域 n(也即大小 $\sqrt{n}$) 的盒子 B_n 中放置 $2n$ 个节点构成了一个任意结构的 S；将它们分为两个集合 S_1 和 S_2，使得 $S_1\cap S_2=\varnothing$，$S_1\cup S_2=S$，$\#S_1=\#S_2=n$。由上面可知，模型(4.4.19)中从发送端 $x_k\in S_1$ 到接收端 $y_j\in S_2$ 的可靠传输速率和 $C_n=\sum_{k=1}^{n}\sum_{j=1}^{n}R_{kj}$ 是有界的，即 448

$$C_n=\sum_{k=1}^{n}\sum_{j=1}^{n}R_{kj}\leqslant\max_{P_k\geqslant 0:\sum_kP_k\leqslant nP}\sum_{k=1}^{n}\log\left(1+\frac{P_ks_k^2}{2\sigma_0^2}\right)$$

其中 s_k 是矩阵 $L=(\ell(x_k,y_j))$ 第 k 大的奇异值，σ_0^2 是噪声谱密度，带宽 $W=1$。

这个结果可以让我们找到随着 $n\to\infty$ 时的渐近容量。在最有趣的 Rayleigh 衰落情况下 $R(n)=C_n/n\sim O\left(\frac{(\log n)^2}{\sqrt{n}}\right)$；在功率 $\alpha>2$ 衰落情况下，$R(n)\sim O\left(\frac{n^{1/\alpha}(\log n)^2}{\sqrt{n}}\right)$；参考文献[48]。

下面我们讨论干扰受限网络。令 Π 为一个在 $\mathbb{R}^2$ 上速率为 λ 的 Poisson 过程。函数 $\ell:\mathbb{R}^2\times\mathbb{R}^2\to\mathbb{R}_+$ 描述在点 y 处发送信号 x 的衰减因子，令 ℓ 为对称函数：$\ell(x,y)=\ell(y,x)$，$x,y\in\mathbb{R}^2$。最常见的例子就是 $\ell(x,y)=Pe^{-\lambda|x-y|}$ 和 $\ell(x,y)=\frac{P}{|x-y|^\alpha}$，$\alpha>2$。在下面假设条件下，我们可以得到一个新的一般理论：

(i) 对于一些 $r_0>0$，$\ell(x,y)=\ell(|x-y|)$，$\int_{r_0}^{\infty}r\ell(r)dr<\infty$ 成立，

(ii) 对于所有的 $x>0$，$\ell(0)>k\sigma_0^2/P$，$\ell(x)\leqslant 1$ 成立，其中 $k>0$ 是所允许的干扰程度，

(iii) ℓ 非零时，是连续且严格减的。

对于每一个点对 $x_1,x_2\in\Pi$，定义信号/噪声比

$$\mathrm{SNR}(x_i \to x_j) = \frac{P\ell^2(x_i, x_j)}{\sigma_0^2 + \gamma \sum_{k \neq i} P\ell^2(x_k, x_j)} \tag{4.4.22}$$

其中 P，σ_0^2，$k>0$ 并且 $0 \leqslant \gamma < \frac{1}{k}$。我们认为如果 $\mathrm{SNR}(x_i \to x_j) \geqslant k$，那么在 x_i 处的发送端可以发送信息给在 x_j 处的接收端。对于任意 $k>0$ 以及 $0<\kappa<1$，令 $A_n(k, \kappa)$ 表示存在一个至少有 Π 中 κn 个点的集合 S_n 使得对于任意两个点 s，$d \in S_n$，都有 $\mathrm{SNR}(s, d) > k$。可以证明(见文献[48])对于所有 $\kappa \in (0, 1)$，存在 $k = k(\kappa)$ 使得

$$\lim_{n \to \infty} \mathbb{P}(A_n(k(\kappa), \kappa)) = 1 \tag{4.4.23}$$

然后，我们认为网络在干扰程度 $k(\kappa)$ 上是超临界的；这意味着，可以与给定发送端(即位于原点 0 的发送端)通过使用中间点重传通信的其他点的总数是以正概率趋向于无穷的。

449 首先，我们注意到任何给定的发送端可能直接连接至多 $1+(\gamma k)^{-1}$ 个接收端。事实上，假设 n_c 个节点连接到节点 x。用 x_i 表示连接到 x 的节点，那么

$$\ell(|x_i - x|) \leqslant \ell(|x_1 - x|), i = 2, \cdots, n_c \tag{4.4.24}$$

因为 x_i 连接到节点 x，我们有

$$\frac{P\ell(|x_i - x|)}{\sigma_0^2 + \gamma \sum_{k=1}^{n_c} P\ell(|x_k - x|)} \geqslant k$$

上式表明

$$\begin{aligned} P\ell(|x_i - x|) &\geqslant k\sigma_0^2 + k\gamma \sum_{k \neq i} P\ell(|x_k - x|) \\ &\geqslant k\sigma_0^2 + k\gamma(n_c - 1)P\ell(|x_1 - x|) + k\gamma \sum_{k=n_c+1} P\ell^2(|x_k - x|) \\ &\geqslant k\gamma(n_c - 1)P\ell(|x_1 - x|) \end{aligned} \tag{4.4.25}$$

通过式(4.4.25)我们可以总结得出 $n_c \leqslant 1 + (k\gamma)^{-1}$。然而，网络会根据式(4.4.23)对参数的一些某些值进行过滤。这表明一个发送端可能通过重传连接到无穷多数目个其他节点上。特别地，模型会过滤 $\gamma=0$ 的情况。高于 Poisson 流 λ_c 的临界过滤值，可以证明对于 $\lambda > \lambda_c$，$\gamma^*(\lambda)$ 的临界值一开始随着 λ 增长，之后因为干扰会变得非常强，导致临界值开始下降。接下来的结果后续结论的证明参考文献[48]。

定理 4.4.15 令 λ_c 为 $\gamma=0$ 情况时的临界节点密度。对于任意节点密度 $\lambda > \lambda_c$，存在 $\gamma^*(\lambda) > 0$ 使得对于 $\gamma \leqslant \gamma^*(\lambda)$ 干扰模型能过滤。对于 $\lambda \to \infty$ 我们有

$$\gamma^*(\lambda) = O(\lambda^{-1}) \tag{4.4.26}$$

另一个有趣的与 $\mathbb{R}^d$ 中的空间点过程理论的连接是在产生随机码本时的点过程实现中的应用。一种选择性(并且更高效)的生成随机码的方法获得(4.1.17)中的值 $C(\alpha)$ 的方法如下。选择一个 $\mathbb{R}^d$ 中的 Poisson 过程 $\Pi^{(N)}$，$\lambda_N = e^{NR_N}$。当 $N \to \infty$ 时 $R_N \to R$，这里 $R < \frac{1}{2} \log \frac{1}{2\pi e \sigma_0^2}$，$\sigma_0^2$ 是信道加性 Gauss 噪声的方差。随机点 $X(i)$ 从过程 $\xi^{(N)}$ 进入码本 $\mathscr{C}_{M,N}$，在欧几里得球 $B^{(N)}(\sqrt{N\alpha})$ 内并且在接下来的"净化"过程中继续存在。确定 $r>0$ 的值(随机码的最小距离)，对于任何 Poisson 过程 $\Pi^{(N)}$ 的点 X_i，生成一个独立同分布随机变量

450 $T_i \sim U([0, 1])$(一个随机标记)。接下来，对于每一个原始 Poisson 过程的点 X_i 都检查球心在 X_i、半径为 r 的球 $B^{(N)}(X_i, r)$，点 X_i 只有其标记 T_i 严格小于其他 $B^{(N)}(X_j, r)$ 内的 $\Pi^{(N)}$ 的点的标记时，这个点才会继续存在。产生的点过程 $\xi^{(N)}$ 是 Matern 过程；它是一

个在最近的文献[1]中被讨论过的更加一般性的结构的例子。

一个长度为 N 码字为 $x^{(N)}$ 的随机码本的主要参数是码字之间距离的诱发分布。在码本通过稳定点过程生成的情况下，可以很方便地引入一个函数 $K(t)$，$\lambda^2 K(t)$ 表示在一个小于距离 t 的单位大小中按顺序排列的不同点组成的对的期望个数。换句话说，$\lambda K(t)$ 是一个过程中的任意点在距离 t 内的进一步的点的期望值。对于 $\mathbb{R}^2$ 中的强度为 λ 的 Poisson 过程，$K(t)=\pi t^2$。在随机码本中，我们对小的或者适中的 t 情况下 $K(t)$ 小得多的模型感兴趣，因此，随机码本在码字之间的小距离出现的情况远小于在 Poisson 过程。可以方便地引入乘积密度

$$\rho(t)=\frac{\lambda^2}{c(t)}\frac{\mathrm{d}K(t)}{\mathrm{d}t} \tag{4.4.27}$$

当 $c(t)$ 依赖于点过程的状态空间。也就是说，在 $\mathbb{R}^1$ 上 $c(t)=2\pi t$，在 $\mathbb{R}^2$ 上 $c(t)=2\pi t^2$，在球上 $c(t)=2\pi\sin t$，以此类推。

某些简便的模型已经被 B. Matérn 介绍过了。这里我们讨论两个 $\mathbb{R}^N$ 上的点过程中直观的模型。第一个是通过简化一个强度为 λ 的 Poisson 过程，删除 $2R$ 之内的所有点，不管这个点是否已经被删除了。对于 $N=2$ 的情况下这个过程的速率为

$$\lambda_{M,1}=\lambda e^{-4\pi\lambda R^2} \tag{4.4.28}$$

对于 $t<2R$ 的密度 $k(t)=0$，并且

$$\rho(t)=\lambda^2 e^{-\lambda U(t)}, t>2R$$

这里

$$U(t)=\mathrm{meas}[B((0,0),2R)\cup B((t,0),2R)] \tag{4.4.29}$$

这里 $B((0,0),2R)$ 是球心为 $(0,0)$ 半径为 $2R$ 的球，$B((t,0),2R)$ 是球心为 $(t,0)$ 半径为 $2R$ 的球，改变 λ 的值使得这个模型的最大速率为 $(4\pi eR^2)^{-1}$，这样就不能表示密集的码。通过三角栅格填充获得了理论界 $(\sqrt{12}R^2)^{-1}$ 的 10%，见文献[1]。

第二个 Matérn 模型是一个所谓的标记点过程的例子。一个速率为 λ 的 Poisson 过程的点是根据分布为 $U([0,1])$ 的独立同分布的随机变量进行独立标记的。如果此过程中有一个彼点，它的距离在 $2R$ 之内并且它的标记值更大，无论这个彼点是否已经被删除，此点均被删除。对于 $N=2$，这个过程的速率为 [461]

$$\lambda_{M,2}=(1-e^{-\lambda c})/c, c=U(0)=4\pi R^2 \tag{4.4.30}$$

对于 $t<2R$ 的产品密度 $\rho(t)=0$，并且

$$\rho(t)=\frac{2U(t)(1-e^{-4\pi R^2\lambda})-2c(1-e^{-\lambda U(t)})}{cU(t)(U(t)-c)}, t>2R \tag{4.4.31}$$

下面是一个等价的定义。给出本原 Poisson 过程的两个点 X 和 Y 在距离 $t=|X-Y|$ 上定义了它们俩在第二个过程被保留的概率。那么对于 $t<2R$ 有 $k(t)=0$，并且

$$k(t)=\frac{2U(t)(1-e^{-4\pi R^2\lambda})-8\pi R^2(1-e^{-\lambda U(t)})}{4\lambda^2\pi R^2 U(t)(U(t)-4\pi R^2)}, t>2R$$

例子 4.4.16 （无线网络中的中断概率）假设一个接收端在原点，发送端按照内径为 r_0 的 Matérn 硬核过程分布，我们假设发送端之间的距离不小于 r_0，覆盖距离为 a。中心接收端收到的功率为

$$X_0=\sum_{j\in J_{r_0,a}}\frac{P}{r_j^\alpha} \tag{4.4.32}$$

这里 $J_{r_0,a}$ 表示干扰发送机的集合，使得 $r_0\leqslant r_j\leqslant a$。令 λ_P 表示稀释后用来产生一个 Matérn

过程的 Poisson 过程的速率。被稀释过程的速率为

$$\lambda = \frac{1-\exp(-\lambda_p \pi r_0^2)}{\pi r_0^2}$$

通过 Campbell 定理我们可以计算 X_0 的 MGF：

$$\phi(\theta) = \mathscr{E}(e^{\theta X_0})$$

452

$$= \exp\left(\lambda_p \pi (a^2 - r_0^2) \int_0^1 q(t)\mathrm{d}t \left[\int_{r_0}^{a} \frac{2r}{(a^2 - r_0^2)} e^{\theta g(r)} \mathrm{d}r - 1\right]\right) \tag{4.4.33}$$

这里 $g(r) = \frac{P}{r^\alpha}$，$q(t) = \exp(-\lambda_p \pi r_0^2 t)$ 是一个标记点 t 的保持概率。因为 $\int_0^1 q(t)\mathrm{d}t = \frac{\lambda}{\lambda_p}$，我们得到

$$\phi(\theta) = \exp\left(\lambda \pi (a^2 - r_0^2) \left[\int_{r_0}^{a} \frac{2r}{(a^2 - r_0^2)} e^{\theta g(r)} \mathrm{d}r - 1\right]\right) \tag{4.4.34}$$

现在我们可以计算干扰信号的所有绝对矩

$$\mu_k = \lambda \pi \int_{r_0}^{a} 2r\,(g(r))^k \mathrm{d}r = \frac{2\lambda\pi}{k\alpha - 2}\left(\frac{P^k}{r_0^{k\alpha-2}} - \frac{P^k}{a^{k\alpha-2}}\right) \tag{4.4.35}$$

工程师说在中心接收机发生中断，即干扰阻止了接收机理解发射机从距离 r_0 处发送来的信息，如果

$$\frac{P/r_0^\alpha}{\sigma_0^2 + \sum_{i \in \Phi_{\setminus 0}} P/r_i^\alpha} \leqslant k$$

这里 σ_0^2 是噪声功率，r_i 是到发射机的距离，k 是成功接收所需的最小 SIR(信噪比)。中断概率的不同近似值是基于式(4.4.35)计算出的时刻。典型地，X_0 的分布接近于 log normal，见文献[113]。

4.5 密码学选例与问题

密码学一般被定义为"关于隐藏信息的实践与研究"，已成为很多编码课程的一部分；在我们的论述中，主要遵循剑桥大学课程"编码与密码学"的传统。我们最低限度保留了理论部分，读者可查阅专门的书籍去了解细节。密码学有着很长并且有时令人着迷的历史，其中数学和其他科学甚至非科学交织在一起。它给无数的小说、半小说、电影和广播带来了灵感，并且热度尚未表现出衰退。

一个产生加密数字序列的通常方式是通过所谓的反馈移位寄存器来实现。我们只讨论二进制的情况，面对的是序列空间 $\mathscr{H}_{n,2} = \{0,1\}^n = \mathbb{F}_2^n$。

定义 4.5.1 一个长度为 d 的(一般)二进制反馈位移寄存器，是一个如下形式的 $\{0,1\}^d \to \{0,1\}^d$ 的映射：

453

$$(x_0, \cdots, x_{d-1}) \to (x_1, \cdots, x_{d-1}, f(x_0, \cdots, x_{d-1}))$$

对于某函数 $f: \{0,1\}^d \to \{0,1\}$(一个反馈函数)。初始字符串 $(x_0, \cdots, x_{d-1})$ 被称为初始填充；它生成一个输出流 $(x_n)_{n\geq 0}$，满足递归式

$$x_{n+d} = f(x_n, \cdots, x_{n+d-1}), \text{对于所有的 } n \geqslant 0 \tag{4.5.1}$$

如果函数 f 是线性的并且 $c_0 = 1$，那么反馈位移寄存器被称为线性的(简称 LFSR)：

$$f(x_0, \cdots, x_{d-1}) = \sum_{i=1}^{d-1} c_i x_i, \quad \text{其中 } c_i = 0,1, c_0 = 1 \tag{4.5.2}$$

这种情况下递归等式是线性的

$$x_{n+d}=\sum_{i=0}^{d-1}c_i x_{n+i}\quad 对于所有的\ n\geqslant 0 \tag{4.5.3}$$

可以很方便写出式(4.5.3)的矩阵形式

$$\boldsymbol{x}_{n+1}^{n+d}=\boldsymbol{V}\boldsymbol{x}_n^{n+d-1} \tag{4.5.4}$$

其中

$$\boldsymbol{V}=\begin{bmatrix}0&1&0&\cdots&0&0\\0&0&1&\cdots&0&0\\\vdots&\vdots&\vdots&\ddots&\vdots&\vdots\\0&0&0&\cdots&0&1\\c_0&c_1&c_2&\cdots&c_{d-2}&c_{d-1}\end{bmatrix},\quad \boldsymbol{x}_n^{n+d-1}=\begin{bmatrix}x_n\\x_{n+1}\\\vdots\\x_{n+d-2}\\x_{n+d-1}\end{bmatrix} \tag{4.5.5}$$

通过第一列的行列式展开，可以看出 $\det\boldsymbol{V}=1\bmod 2$；$(n,1)$的余子式元素 c_0 是矩阵$\boldsymbol{I}_{d-1}$。因此

$$\det\boldsymbol{V}=c_0\det\boldsymbol{I}_{d-1}=c_0=1,\quad 并且矩阵\ \boldsymbol{V}\ 是可逆的 \tag{4.5.6}$$

根据式(4.5.3)

$$C(X)=c_0+c_1X+\cdots+c_{d-1}X^{d-1}+X^d \tag{4.5.7}$$

一个 LFSR 的辅助或者反馈多项式是一个有用的概念，因为通过它可以观察到一般的反馈位移寄存器在初始运行之后会变成周期的。

定理 4.5.2 长度为 d 的一般反馈移位寄存器的输出流(x_n)具有如下性质：存在一个整数 r，$0\leqslant r<2^d$，以及整数 D，$1\leqslant D<2^d-r$，使得对所有的 $k\geqslant r$ 都有 $x_{k+D}=x_k$。 454

证明 在(4.5.1)中，一段序列 $x_M\cdots x_{M+d-1}$ 唯一决定输出流剩余的部分，即$(x_n,\ n\geqslant M+d-1)$。我们看出如果这样一段序列在流中被复制将会出现重复。对于一个只知道后面 d 个数字的字符串有 2^d 中不同的可能性。因此，根据鸽巢原理，存在 $0\leqslant r<R<2^d$ 使得从位置 r 和 R 向前，输出流的长度为 d 的两段将是相同的：$x_{r+j}=x_{R+j}$，$0\leqslant r<R<d$。那么，正如所提到的，对于所有的 $j\geqslant 0$，都有 $x_{r+j}=x_{R+j}$，并且当 $D=R-r$ 时这个结论成立。 □

在线性的情况下(LFSR)，对于丢弃的零字符串，我们可以重复上面的证明。这样我们可以从 2^d 降到 2^d-1。然而，在"适当的意义上"LFSR 是周期的。

定理 4.5.3 LFSR(x_n)是周期的，即对于所有的 n，存在 $D\leqslant 2^d-1$ 使得 $x_{n+D}=x_n$。具有这样性质的最小的 D 被称为 LFSR 的周期。

证明 确实，列向量$\boldsymbol{x}_n^{n+d-1}$，$n\geqslant 0$，与等式$\boldsymbol{x}_{n+1}=\boldsymbol{V}\boldsymbol{x}_n=\boldsymbol{V}^{n+1}\boldsymbol{x}_0$，$n\geqslant 0$ 有关系，其中矩阵 $\boldsymbol{V}$ 在式(4.5.5)中已被定义。我们注意到 $\det\boldsymbol{V}=c_0\neq 0$，因此 $\boldsymbol{V}$ 是可逆的。正如前面所说的，我们可以丢弃零初始填充，对于每个向量$\boldsymbol{x}_n\in\{0,1\}^d$，只有 2^d-1 个非零的可能。因此，正如在定理 4.5.2 中的证明中所讨论的，在初始的 2^d-1 个向量$\boldsymbol{x}_n$，$0\leqslant n\leqslant 2^d-2$ 中，要么是重复，要么是零向量。第二种可能也能再次被丢弃，因为它会导致一个零初始填充。因此，假设第一个重复是对于 j 和 $D+j$：$\boldsymbol{x}_j=\boldsymbol{x}_{j+D}$，即$\boldsymbol{V}^{j+D}\boldsymbol{x}_0=\boldsymbol{V}^j\boldsymbol{x}_0$。如果 $j\neq 0$，我们乘以$\boldsymbol{V}^{-1}$并且得到一个更早的重复。所以：$j=0$，$D\leqslant 2^d-1$ 并且$\boldsymbol{V}^D\boldsymbol{x}_0=\boldsymbol{x}_0$。那么，很明显，$\boldsymbol{x}_{n+D}=\boldsymbol{V}^{n+D}\boldsymbol{x}_0=\boldsymbol{V}^n\boldsymbol{x}_0=\boldsymbol{x}_n$。 □

举例 4.5.4 给出一个一般反馈寄存器的例子，其中输出为 k_j，初始填充为$(k_1,k_2,\cdots,k_N)$，使得

$$(k_n,k_{n+1},\cdots,k_{n+N})\neq(k_1,k_2,\cdots,k_N),\quad n\geqslant 1$$

解答 取 $f:\{0,1\}^2\to\{0,1\}^2$，$f(x_1,x_2)=x_2 1$。初始填充 00 产生 00111111111…。这

里，对于所有的 $n\geq 1$ 都有 $k_{n+1}\neq 0-k_1$。 □

举例 4.5.5 对于线性回归(4.5.3)，令矩阵 $\boldsymbol{V}$ 满足式(4.5.5)的定义。定义并计算 $\boldsymbol{V}$ 的特征值和最小多项式。

455

解答 矩阵 $\boldsymbol{V}$ 的特征多项式为 $h_{\boldsymbol{V}}(X)\in\mathbb{F}_2[X]=X\to\det(X\boldsymbol{I}-\boldsymbol{V})$

$$h_{\boldsymbol{V}}(X)=\det\begin{bmatrix} X & 1 & 0 & \cdots & 0 & 0 \\ 0 & X & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \cdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & X & 1 \\ c_0 & c_1 & c_2 & \cdots & c_{d-2} & (c_{d-1}+X) \end{bmatrix} \tag{4.5.8}$$

(回忆一下，考虑 $\mathbb{F}_2$ 中的元素 1 和 c_i)。从最底下一行开始扩展，多项式 $h_{\boldsymbol{V}}(X)$ 写成一个大小为 $(d-1)\times(d-1)$(余子式)的行列式的线性组合：

$$c_0\det\begin{bmatrix} 1 & 0 & \cdots & 0 & 0 \\ X & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \cdots & \vdots & \vdots \\ 0 & 0 & \cdots & X & 1 \end{bmatrix}+c_1\det\begin{bmatrix} X & 0 & \cdots & 0 & 0 \\ X & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & X & 1 \end{bmatrix}$$

$$+\cdots+c_{d-2}\det\begin{bmatrix} X & 1 & \cdots & 0 & 0 \\ 0 & X & \cdots & 0 & 0 \\ \vdots & \vdots & \cdots & \vdots & \vdots \\ 0 & 0 & \cdots & 0 & 1 \end{bmatrix}$$

$$+(c_{d-1}+X)\det\begin{bmatrix} X & 1 & \cdots & 0 & 0 \\ 0 & X & \cdots & 0 & 0 \\ \vdots & \vdots & \cdots & \vdots & \vdots \\ 0 & 0 & \cdots & 0 & X \end{bmatrix}$$

$$=c_0+c_1X+\cdots+c_{d-2}X^{d-2}+(c_{d-1}+X)X^{d-1}$$

$$=\sum_{0\leq i\leq d-1}c_iX^i+X^d$$

这给出了递归式的特征多项式 $C(X)$。

根据 Cayley-Hamilton 定理

$$h_{\boldsymbol{V}}(\boldsymbol{V})=c_0\boldsymbol{I}+c_1\boldsymbol{V}+\cdots+c_{d-1}\boldsymbol{V}^{d-1}+\boldsymbol{V}^d=\boldsymbol{O}$$

矩阵 $\boldsymbol{V}$ 的最小多项式 $m_{\boldsymbol{V}}(X)$ 是使 $m_{\boldsymbol{V}}(\boldsymbol{V})=\boldsymbol{O}$ 的最小次数多项式。它是 $h_{\boldsymbol{V}}(X)$ 的一个除数，$h_{\boldsymbol{V}}(X)$ 的每个根都是 $m_{\boldsymbol{V}}(X)$ 的根。$h_{\boldsymbol{V}}(X)$ 和 $m_{\boldsymbol{V}}(X)$ 的区别在于重数：$m_{\boldsymbol{V}}(X)$ 的根 μ 的重数等于对应着 μ 的矩阵 $\boldsymbol{V}$ 的 Jordan 单元的最大尺寸。然而对于 $h_{\boldsymbol{V}}(X)$ 则是所有对应着 μ 的矩阵 $\boldsymbol{V}$ 的 Jordan 单元尺寸的总和。

456

为了计算出 $m_{\boldsymbol{V}}(X)$，我们：

(i) 取一组基 $\mathbf{e}_1,\cdots,\mathbf{e}_d$(在 $\mathbb{F}_2^{1\times d}$ 中)，

(ii) 那么对于任意向量 $\mathbf{e}_i$ 我们找到向量 $\mathbf{e}_i,\boldsymbol{V}\mathbf{e}_i,\cdots,\boldsymbol{V}^{d_i}\mathbf{e}_i$，$\boldsymbol{V}^{d_i+1}\mathbf{e}_i$ 是线性相关的最小数 d_i。

(iii)确定对应的线性组合

$$a_i^{(0)}\mathbf{e}_i+a_i^{(1)}\boldsymbol{V}\mathbf{e}_i+\cdots+a_i^{(d_i)}\boldsymbol{V}^{d_i}\mathbf{e}_i+\boldsymbol{V}^{d_i+1}\mathbf{e}_i=0$$

(iv) 进一步，我们构造对应的多项式

$$m_V^{(j)}(X)=\sum_{0\le i\le d_j} a_i^{(j)}X^i+X^{d_j}$$

(v) 然后

$$m_V(X)=\mathrm{lcm}[m_V^{(1)}(X),\cdots,m_V^{(d)}(X)]$$

在我们这种情况下，很方便地可以取

$$\mathbf{e}_j=\begin{bmatrix}0\\ \vdots\\ 1\\ \vdots\\ 0\end{bmatrix}\begin{matrix}\\ \\ \leftarrow j\\ \\ \\ \end{matrix}$$

那么$\mathbf{V}^i\mathbf{e}_1=\mathbf{e}_j$，我们得到 $d_1=d$ 和

$$m_V^{(1)}(X)=\sum_{0\le i\le d-1} c_iX^i+X^d=h_V(X) \qquad \square$$

我们发现递归式的反馈多项式 $C(X)$与矩阵 V 的特征值和最小多项式相同。可以观察到在 $X=0$ 的情况下我们有

$$h_V(0)=C(0)=c_0=1=\det\mathbf{V} \tag{4.5.9}$$

任何多项式都可以通过它的根来确定；我们曾发现这样一个描述可能非常有用。在 LFSR 情况下，下面的例子是很有启发性的。

定理 4.5.6 *考虑式(4.5.3)中的二进制线性递归和式(4.5.7)中相应的辅助多项式。*

(a) 假设 $\mathbb{K}$ 是一个包含 $\mathbb{F}_2$ 的域，使得多项式 $C(X)$在域 $\mathbb{K}$ 中有一个 m 重的根 α。那么对于所有的 $k=0,1,\cdots,m-1$，

$$x_n=A(n,k)\alpha^n, n=0,1,\cdots \tag{4.5.10}$$

[457]

是 $\mathbb{K}$ 域中式(4.5.3)的一个解，其中

$$A(n,k)=\begin{cases}1, & k=0\\ \left[\prod_{1\le l\le k-1}(n-l)_+\right]\bmod 2, & k\ge 1\end{cases} \tag{4.5.11}$$

在这里以及下面的论述中，$(a)_+$表示 $\max[a,0]$。换句话说，序列$x^{(k)}=(x_n)$是 LFSR 的一个辅助多项式为 $C(X)$的输出序列，其中 x_n 如式(4.5.10)所示。

(b) 假设 $\mathbb{K}$ 是一个包含 $\mathbb{F}_2$ 的域，使得多项式 $C(X)$在 $\mathbb{K}$ 中能被分解成线性因子。令 $\alpha_1,\cdots,\alpha_r\in\mathbb{K}$ 为 $C(X)$的不同的根，重数为 $m_1,\cdots,m_r$，其中 $\sum_{1\le i\le r}m_i=d$。那么对于某些 $b_{i,k}\in\mathbb{K}$，式(4.5.3)在 $\mathbb{K}$ 中的一般解为

$$x_n=\sum_{1\le i\le r}\sum_{0\le k\le m_i-1} b_{i,k}A(n,k)\alpha_i^n \tag{4.5.12}$$

换句话说，序列$x_{(i,k)}=(x_n)$张成了辅助多项式 $C(X)$的 LFSR 所有输出流的集合，其中 $x_n=A(n,k)\alpha_i^n$ 和 $A(n,k)$在式(4.5.11)中给出。

证明 (a)如果 $C(X)$有一个 m 重的根 $\alpha\in K$，那么 $C(X)=(X-\alpha)^m\bar{C}(X)$，其中$\bar{C}(X)$是一个次数为 $d-m$ 的多项式(其中的系数来自域 $\mathbb{K}'\subseteq\mathbb{K}$)。那么，对于所有的 $k=0,\cdots,m-1$和 $n\ge d$，多项式

$$D_{k,n}(X)=X^d\frac{\mathrm{d}^k}{\mathrm{d}X^k}[X^{-d}C(X)]$$

(系数取模 2)在 $X=\alpha$(在域 $\mathbb{K}$ 中)处为零；

$$D_{d,n}(a)=\sum_{0\leqslant i\leqslant d-1}c_iA(n-d+i,k)a^{n-d+i}+A(n,k)a^n$$

这样就得到

$$A(n,k)a^n=\sum_{0\leqslant i\leqslant d-1}c_iA(n-d+i,k)a^{n-d+i}$$

因此，关于如式(4.5.10)中所示的 x_n，流 $x^{(k)}=(x_n)$，能求解 $\mathbb{K}$ 中的递归 $x_n=\sum_{1\leqslant i\leqslant d-1}c_ix_{n-d+i}$。这种解的个数为 m，即根 a 的重数。

(b) 首先，我们观察到输出流集合 $(x_n)_{n\geqslant 0}$ 形成一个在 $\mathbb{K}$ 上的线性空间 $\mathcal{W}$(来自 $\mathbb{K}$ 中元素的所有序列的集合)。$\mathcal{W}$ 的维度为 d，因为每个流都被一个种子(初始填充)$x_0x_1\cdots x_{d-1}\in\mathbb{K}^d$ 唯一定义。另一方面，序列 $x^{(i,k)}=(x_n^{(i,k)})$ 的总数为 $d=\sum_{1\leqslant i\leqslant r}m_i$ 中的元素为：

[458]

$$x_n^{(i,k)}=A(n,k)a_i^n,n=0,1,\cdots$$

因此，它足以检测流 $x^{(i,k)}$ 在 $\mathbb{K}$ 上是线性独立的，其中 $i=1,\cdots,r$，$k=0,1\cdots,m_i-1$。 □

在最后，取一个线性组合 $\sum_{1\leqslant i\leqslant r}\sum_{0\leqslant k\leqslant m_i-1}b_{i,k}x^{(i,k)}$ 并且假设其为 0，同时我们也认为对于 $k<0$ 有 $x^{(i,k)}=\mathbf{0}$。可以很方便地引入一个位移运算 $x=(x_n)\to Sx$，其中序列 $Sx=(x_n^1)$ 的元素 $x_n^1=x_{n+1}$，$n=0,1,\cdots$。关键的发现如下。令 I 表示恒等变换，那么对于所有的 $\beta\in\mathbb{K}$，

$$(S-\beta I)x^{(i,k)}=(a_i-\beta)x^{(i,k)}+ka_ix^{(i,k-1)}$$

事实上，序列 $(S-\beta I)x^{(i,k)}$ 的第 n 个元素等于

$$\begin{aligned}&A(n+1,k)a_i^{n+1}-\beta A(k,n)a_i^n\\&=[A(n,k)+kA(n,k-1)]a_i^{n+1}-\beta A(n,k)a_i^n\\&=(a_i-\beta)A(n,k)a_i^n+ka_iA(n,k-1)a_i^n\end{aligned}$$

与上面的等式在序列上相同。在这里我们使用了基本公式

$$A(n+1,k)=A(n,k)+kA(n,k-1)$$

然后通过迭代，我们得到

$$\begin{aligned}(S-\beta_1I)(S-\beta_2I)x^{(i,k)}&=(a_i-\beta_1)(a_i-\beta_2)x^{(i,k)}\\&\quad+ka_i(a_i-\beta_1+a_i-\beta_2)x^{(i,k-1)}+k^2a_i^2x^{(i,k-2)}\\&=(S-\beta_2I)(S-\beta_1I)x^{(i,k)}\end{aligned}$$

等(所有系数运算在 $\mathbb{K}$ 上完成)，特别地，$\beta=a_i$ 时

$$(S-a_iI)^lx^{(i,k)}=\begin{cases}(ka_i)^lx^{(i,k-l)}, & 1\leqslant l\leqslant k\\ \mathbf{0}, & l>k\end{cases}$$

现在考虑将乘积运算 $\prod_{1\leqslant i\leqslant r}(S-a_iI)^{m_i}(S-a_iI)_{m_i-1}$ 的结果应用到我们消除的线性组合 $\sum_{1\leqslant i\leqslant r}\sum_{0\leqslant k\leqslant m_i-1}b_{i,k}x^{(i,k)}$，唯一保留的项来自被加数 $b_{i,m_i-1}x^{(i,m_i-1)}$。这给出了

$$b_{i,m_i-1}\prod_{1\leqslant i\leqslant r}(a_i-a_i)^{m_i}[(m_i-1)a_i]^{m_i-1}x^{(i,0)}=\mathbf{0}$$

因此，$b_{i,m_i-1}=0$。然后，我们应用 $\prod_{1\leqslant i\leqslant r}(S-a_iI)^{m_i}(S-a_iI)^{m_i-2}$，获得 $b_{i,m_i-2}=0$。按照相似的方式继续，我们可以确保每个系数 $b_{i,k}=0$。 □

[459]

当观测到一个数字流为 $(x_n)_{n\geqslant 1}$ 时，观测者也许希望去判断它是否是由一个 LFSR 产生的。这可通过所谓的 Berlekamp-Massey(BM)算法来实现，它可以求解线性方程组系统。如果一个序列 (x_n) 是来自于 LFSR，它的反馈多项式为 $C(X)=\sum_{i=0}^{d-1}c_iX^i+X^d$，那么对于

$n=0,\cdots,d$，递归式 $x_{n+d}=\sum_{i=1}^{d-1}c_i x_{n+i}$ 可以写成向量-矩阵形式 $\boldsymbol{A}_d\boldsymbol{c}_d=\boldsymbol{0}$，其中

$$\boldsymbol{A}_d=\begin{pmatrix} x_0 & x_1 & x_2 & \cdots & x_d \\ x_1 & x_2 & x_3 & \cdots & x_{d+1} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ x_d & x_{d+1} & x_{d+2} & \cdots & x_{2d} \end{pmatrix},\quad \boldsymbol{c}_d=\begin{pmatrix} c_0 \\ c_1 \\ \vdots \\ c_{d-1} \\ 1 \end{pmatrix} \tag{4.5.13}$$

因此，$(d+1)\times(d+1)$维矩阵$\boldsymbol{A}_d$的行列式一定为0，并且$(d+1)$维向量$\boldsymbol{c}_d$一定在零空间$\ker \boldsymbol{A}_d$中。

算法从矩阵$\boldsymbol{A}_r$的检测开始，其中r是一个较小的值(已知是小于d的)：

$$\boldsymbol{A}_r=\begin{pmatrix} x_0 & x_1 & x_2 & \cdots & x_r \\ x_1 & x_2 & x_3 & \cdots & x_{r+1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ x_r & x_{r+1} & x_{r+2} & \cdots & x_{2r} \end{pmatrix}$$

我们计算 $\det \boldsymbol{A}_r$：如果 $\det \boldsymbol{A}_r\neq 0$，我们得出结论 $d\neq r$ 并且令 r 加 1。如果 $\det \boldsymbol{A}_r=0$，那么我们解方程$\boldsymbol{A}_r\boldsymbol{a}_r=0$，即试验 $d=r$

$$\boldsymbol{A}_d\begin{pmatrix} a_0 \\ a_1 \\ \vdots \\ a_{d-1} \\ 1 \end{pmatrix}=0,\quad \boldsymbol{A}_d=\begin{pmatrix} x_0 & x_1 & \cdots & x_d \\ x_1 & x_2 & \cdots & x_{d+1} \\ \vdots & \vdots & \ddots & \vdots \\ x_d & x_{d+1} & \cdots & x_{2d} \end{pmatrix}$$

(比如通过 Gauss 消去法)和针对递归式 $x_{n+d}=\sum_{0\le i\le d-1}a_i x_{n+i}$ 来测试序列(x_n)。如果我们发现不相符，那么如果失败的话，我们选择一个不同的向量$\boldsymbol{c}_r\in\ker\boldsymbol{A}_r$ 或者增加 r。

BM 算法可以用优美的代数形式来说明。给定一个序列(x_n)，考虑一个 X 中的幂级数的形式：$\sum_{j=0}^{\infty}x_jX^j$。$(x_n)$是由反馈多项式为 $C(X)$的 LFSR 产生的，它等价于通过多项式 $A(X)=\sum_{j=0}^{d}a_jX^j$ 除以 $C(X)$来获得上面的级数： 460

$$\sum_{j=0}^{\infty}x_jX^j=\frac{A(X)}{C(X)} \tag{4.5.14}$$

确实，因为 $c_0=1$，$A(X)=C(X)\sum_{j=0}^{\infty}x_jX^j$ 等价于

$$a_n=\sum_{i=0}^{k}c_i x_{n-i},n=1,\cdots \tag{4.5.15}$$

或者

$$x_n=\begin{cases} a_n-\sum_{i=1}^{n-1}c_i x_{n-i}, & n=0,1,\cdots,d \\ -\sum_{i=1}^{d-1}c_i x_{n-i}, & n>d \end{cases} \tag{4.5.16}$$

换句话说，$A(X)$参与解释了初始填充，$C(X)$扮演了反馈多项式的角色。

举例 4.5.7 线性反馈位移寄存器(LFSR)到底是什么？解释 Berlekamp-Massey 方法如何通过输出来恢复一个线性反馈移位寄存器的反馈多项式。描述当我们观测到输出为

$$1\,0\,1\,0\,1\,1\,0\,0\,1\,0\,0\,0\cdots$$

$$0\,1\,0\,1\,1\,1\,1\,0\,0\,0\,1\,0\cdots$$

和

$$1\,1\,0\,0\,1\,0\,1\,1$$

的情况。

解答 初始填充 $x_1\cdots x_{d-1}$ 产生输出流$(x_n)_{n\geq 1}$，它满足递归等式

$$x_{n+d}=\sum_{i=1}^{d-1}c_i x_{n+i},\quad n\geq 0$$

反馈多项式

$$C(X)=c_0+c_1X+\cdots+c_{d-1}X^{d-1}+X^d$$

是特征多项式，因为这个回归等式决定了它的解。我们假设系数 $c_0\neq 0$；否则 x_n 不会对 x_{n+d} 产生影响，寄存器的长度可以视为 $d-1$。

461

Berlekamp-Massey 算法从对矩阵的检查开始

$$\boldsymbol{A}_1=\begin{bmatrix}1&0\\0&1\end{bmatrix},\quad \det\boldsymbol{A}_1\neq 0$$

但是

$$\boldsymbol{A}_2=\begin{bmatrix}1&0&1\\0&1&0\\1&0&1\end{bmatrix},\quad \det\boldsymbol{A}_2=0$$

并且$\boldsymbol{A}_2\begin{bmatrix}c_0\\c_1\\c_2\end{bmatrix}=0$ 有解 $c_2=1$，$c_1=0$。这给出了递归式

$$x_{n+2}=x_n$$

它不符合剩余的数字。因此我们转到$\boldsymbol{A}_3$：

$$\boldsymbol{A}_3=\begin{bmatrix}1&0&1&0\\0&1&0&1\\1&0&1&1\\0&1&1&0\end{bmatrix},\quad \det\boldsymbol{A}_3\neq 0$$

然后到$\boldsymbol{A}_4$：

$$\boldsymbol{A}_4=\begin{bmatrix}1&0&1&0&1\\0&1&0&1&1\\1&0&1&1&0\\0&1&1&0&0\\1&1&0&0&1\end{bmatrix},\quad \det\boldsymbol{A}_4=0$$

等式$\boldsymbol{A}_4\boldsymbol{c}_4=0$ 解得 $\boldsymbol{c}_4=\begin{bmatrix}1\\0\\0\\1\end{bmatrix}$，这样得到

$$x_{n+1}=x_n+x_{n-1}$$

它符合序列其余的部分。在第二个例子中，我们有

$$\det\begin{bmatrix}0&1\\1&0\end{bmatrix}\neq 0,\quad \det\begin{bmatrix}0&1&0\\1&0&1\\0&1&1\end{bmatrix}\neq 0,\quad \det\begin{bmatrix}0&1&0&1\\1&0&1&1\\0&1&1&1\\1&1&1&1\end{bmatrix}\neq 0$$ 452

和

$$\begin{bmatrix}0&1&0&1&1\\1&0&1&1&1\\0&1&1&1&1\\1&1&1&1&0\\1&1&1&0&0\end{bmatrix}\begin{bmatrix}1\\1\\0\\0\\1\end{bmatrix}=0$$

求解为 $d=4$，$x_{n+4}=x_n+x_{n+1}$，线性递归关系可以根据输出序列的每一项自得以满足。那么反馈多项式为 X^4+X+1。

在第三个例子中的递归式是 $x_{n+1}=x_n+x_{n-1}$。 □

LFSR 被用来产生加性流密码。加性流密码在 1917 年由 Gilbert Vernam 发明，这时他是 AT&T Bell Labs 的一名工程师。这里，发送方使用 LFSR(k_n)的输出流，通过(z_n)来将纯文本(p_n)加密，其中

$$z_n=p_n+k_n \bmod 2,\quad n\geqslant 0 \tag{4.5.17}$$

接收方通过如下方式进行解密

$$p_n=z_n+k_n \bmod 2,\quad n\geqslant 0 \tag{4.5.18}$$

但是当然，他必须知道初始填充 $k_0\cdots k_{d-1}$ 和字符串 $c_0\cdots c_{d-1}$。这个流密码的主要缺点是它的周期性。确实，如果生成 LFSR 的周期为 D，那么“攻击者”就足以自制一个长为 $2D$ 的密码文本 $z_0z_1\cdots z_{2D-1}$ 和对应的纯文本 $p_0p_1\cdots p_{2D-1}$。(对于当代密尔塞斯来说并不是无法完成的任务。)如果攻击者侥幸知道了周期 D 的值，那么他只需要利用 $z_0z_1\cdots z_{D-1}$ 和 $p_0p_1\cdots p_{D-1}$ 就可以破译密码，也就是不管整个文本有多长都可以解密。

很明显，短周期的 LFSR 重复使用时，它会更容易被破译。二战时期和随后的冷战时期有着一系列惊人的例子(德国密码破译者成功破译了英国海军的密码，英国和美国的密码破译者成功地破译了德国的密码，美国的“Venona”项目破译了苏联的密码)，这都是由于密集的信息传输引起的。然而，即使是极长的周期也不能保证安全。

就本节这一节目前所讲到的内容而言，可以通过组合几个 LFSR 来增加 LFSR 的周期。

定理 4.5.8 假设流(x_n)是由一个长度为 d_1，周期为 D_1，辅助多项式为 $C_1(X)$的 LFSR 产生的，流(y_n)是由长度为 d_2，周期为 D_2，辅助多项式为 $C_2(X)$的 LFSR 产生的。令 $\alpha_1,\cdots,\alpha_r$ 和 $\beta_1,\cdots,\beta_s$ 分别为在某个域 $\mathbb{K}\supset\mathbb{Z}_2$ 中的 $C_1(X)$和 $C_2(X)$的不同根。令 m_i 为根 α_i 的重数，m_i'为根 β_i 的重数，并且有 $d_1=\sum_{1\leqslant i\leqslant r}m_i$，$d_2=\sum_{1\leqslant i\leqslant s}m_i'$。那么， 453

(a) 流(x_n+y_n)是由辅助多项式为 $\mathrm{lcm}(C_1(X),C_2(X))$的 LFSR 产生的。

(b) 流(x_ny_n)是由辅助多项式为 $C(X)=\prod_{1\leqslant i\leqslant r}\prod_{1\leqslant j\leqslant s}(X-\alpha_i\beta_j)^{m_im_j'}$ 的 LFSR 产生的。

特别地，产生的 LFSR 的周期在两种情况下都是可以被 $\mathrm{lcm}(D_1,D_2)$整除的。

证明 根据定理4.5.6，对于某些 $a_{i,k}$，$b_{j,l}\in\mathbb{K}$，问题中LFSR的输出流(x_n)和(y_n)在域$\mathbb{K}$中的形式如下

$$x_n=\sum_{1\le i\le r_1}\sum_{0\le k\le m_i-1}a_{i,k}A(n,k)\alpha_i^n,\quad y_n=\sum_{1\le j\le r_2}\sum_{0\le l\le m_j'-1}b_{j,l}A(n,l)\beta_j^n \tag{4.5.19}$$

(a) 写出式(4.5.19)的和表达式 x_n+y_n，合并同类项就得到了(a)的证明。

(b) 对于积 x_ny_n 我们有如下表达式

$$\sum_{i,j}\sum_{k,l}a_{i,k}b_{j,l}A(n,k)A(n,l)\,(\alpha_i\beta_j)^n$$

积 $a_{i,k}b_{j,l}A(n,k)A(n,l)$ 可以写成和 $\sum_{k,l;1\le t\le k+l-1}A(n,t)u_t(a_{i,k},b_{j,l})$ 的形式，其中系数 $u_t(a_{i,k},b_{j,l})\in\mathbb{K}$。这给出了 x_ny_n 的如下表达式：

$$\sum_{1\le i\le r_1}\sum_{1\le j\le r_2}\sum_{1\le t\le m_i+m_j'-2}A(n,t)\sum_{k,l;k+l=t+1}u_t(a_{i,k},b_{j,l})\,(\alpha_i\beta_j)^n$$

接下来它可以被写为

$$x_ny_n=\sum_{1\le i\le r_1}\sum_{1\le j\le r_2}\sum_{0\le t\le m_i+m_j'-2}A(n,t)v_{i,j,t}\,(\alpha_i\beta_j)^n$$

对应着(b)中辅助多项式为$C(X)$的LFSR输出流的一般形式。 □

尽管有着严重的不足，LFSR依然在很多情况下可以使用：它们允许简单的加解密而无需"前看"和显示一个在编码、传输、解码中错误的"局部"效应。更一般地，非线性LFSR经常只提供边际利益然而却带来严重的缺点，特别是在解密过程中。

464

一个同样例子引发的错误，
会使事态急速加剧。

William Shakespeare(1564—1616)，
英国编剧与诗人，选自《威尼斯商人》

举例4.5.9 (a) 令(x_n)，(y_n)，(z_n)为三个由LFSR产生的流。令

$$k_n=x_n,\quad 如果\ y_n=z_n$$
$$k_n=y_n,\quad 如果\ y_n\ne z_n$$

试证明k_n也是一个由线性反馈寄存器产生的流。

(b) 一个加密流由长度为d的线性反馈寄存器产生的。证明：当给定长度为$2d$的纯文本和加密文本，我们可以发现加密流。

解答 (a) 对于三个由LFSR产生的流(x_n)，(y_n)，(z_n)，我们令

$$k_n=x_n+(x_n+y_n)(y_n+z_n)(在\ \mathbb{F}_2\ 上)$$

所以应该注意的是由LFSR产生的(逐点的)和与积的流也导致了由LFSR产生的某些流。

(b) 假设纯文本为 $y_1y_2\cdots y_{2d}$，加密后的文本为 $x_1+y_1\ x_2+y_2\cdots x_{2d}+y_{2d}$。那么我们可以恢复出 $x_1\cdots x_{2d}$。我们知道 $c_1\cdots c_d$ 必须满足d个联立线性方程

$$x_{d+j}=\sum_{i=1}^{d}c_ix_{j+i-1},\quad 对于\ j=1,2,\cdots,d$$

通过解上述方程组得到 c_1，c_2，…，c_d，并且因此获得加密流。 □

举例4.5.10 一个长度为4的二进制非线性反馈寄存器有

$$x_{n+1}=x_{n-1}+x_nx_{n-2}+x_{n-3}$$

465 证明状态空间包含长度为1，4，9和2的循环。

解答 有 $2^4=16$ 个初始二进制序列，通过检验，

$$0000 \mapsto 0000 \qquad \text{（长度为 1 的循环）}$$
$$0001 \mapsto 0010 \mapsto 0100 \mapsto 1000 \mapsto 0001 \qquad \text{（长度为 4 的循环）}$$
$$0011 \mapsto 0111 \mapsto 1111 \mapsto 1110 \mapsto 1101$$
$$\mapsto 1011 \mapsto 0110 \mapsto 1100 \mapsto 1001 \mapsto 0011 \qquad \text{（长度为 9 的循环）}$$
$$0101 \mapsto 1010 \mapsto 0101 \qquad \text{（长度为 2 的循环）}$$

所有的 16 个初始域充都在列表中出现，所以分析是完整的。 □

举例 4.5.11 描述一个加法流密码是如何工作的。什么是一次一密？简要解释为什么如果一次一密只使用一次是安全的，反复使用多次就不安全了？使用一次一密将 $x_1x_2x_3x_4x_5x_6y_7$ 编码成 0101011 发送，由于发生错误，重新使用密钥将信息 $y_1x_1x_2x_3x_4x_5x_6$ 编码成 0100010 发送，证明 $x_1x_2x_3x_4x_5x_6$ 是两种可能的信息之一，并且找出这两种可能。

解答 一次一密是一个基于随机密钥的一种密码，这是由 Gilbert Vernam 和 Joseph Mauborgne(二战时期美国 Signal Corps 主席)。这个密码使用了随机数生成器来从长度为 q 的字母表 J 中产生序列 $k_1k_2k_3\cdots$。更准确地说，每个字母均匀地分布在 J 上并且不同字母是不相关的。信息 $m=a_1a_2\cdots a_n$ 被加密为 $c=c_1c_2\cdots c_n$，其中

$$c_i = a_i + k_i (\bmod\ q)$$

为了证明一次一密具有完美的安全性，可写出

$$\begin{aligned}\mathbb{P}(M=m, C=c) &= \mathbb{P}(M=m, K=c-m)\\ &= \mathbb{P}(M=m)\mathbb{P}(K=c-m) = \mathbb{P}(M=m)\frac{1}{q^n}\end{aligned}$$

这里差 $c-m$ 是逐位的并且是模 q 的。因此，条件概率

$$\mathbb{P}(C=c \mid M=m) = \frac{\mathbb{P}(M=m, C=c)}{\mathbb{P}(M=m)} = \frac{1}{q^n}$$

不依赖于 m。因此，M 和 C 是不相关的。

在域 $\mathbb{F}_2$ 中，考虑一个密钥流 $k_1k_2k_3\cdots$。输入的纯文本流 $p_1p_2p_3\cdots$ 被编码为密文流 $c_1c_2c_3\cdots$，其中 $c_i=p_i+k_i$。如果 k_i 是独立同分布的随机数，并且密钥流只使用一次(事实上就是这样)，那么我们就有一次一密。(假设只有发送方和接收方知道密钥流)在这个例子中，我们有 466

$$x_1x_2x_3x_4x_5x_6y_7 \mapsto 0101011$$
$$y_1x_1x_2x_3x_4x_5x_6 \mapsto 0100010$$

假设 $x_1=0$，那么

$$k_1=0, k_1=1, x_2=0, k_2=0, x_3=0, k_3=0, x_4=1, k_4=0$$
$$x_5=1, k_5=0, x_6=1, k_6=1$$

因此

$$k=0100101, \quad x=000111$$

如果 $x_1=1$，每个数字改变，所以

$$k=1011010, \quad x=111000$$

作为另一种选择，令 $x_2=y_1$，$x_7=y_7$。如果第一个密码为 $q_1q_2\cdots$，那么第二个密码为 $p_1p_2\cdots$，并且一次一密为 $k_1, k_2, \cdots$，那么

$$q_i = x_{i+1} + k_i, p_i = x_i + k_i$$

所以

$$x_j + x_{j+1} = q_j + p_j$$

并且

$$x_1 + x_2 = 0, x_2 + x_3 = 0$$
$$x_3 + x_4 = 1, x_4 + x_5 = 0, x_5 + x_6 = 0$$

这会获得

$$x_1 = x_2 = x_3, x_4 = x_5 = x_6, x_4 = x_3 + 1$$

信息为 000111 或者 111000。 □

举例 4.5.12 (a) 令 $\theta: Z \to \{0, 1\}$ 被定义为：如果 n 是奇数那么 $\theta(n)=1$，如果 n 是偶数那么 $\theta(n)=0$。考虑在 $\mathbb{Z}_2$ 中的如下的递归关系：

$$u_{n+3} + u_{n+2} + u_{n+1} + u_n = 0 \tag{4.5.20}$$

式(4.5.20)的通解为 $u_n = A + B\theta(n) + C\theta(n^2)$ 是真的吗？如果为真，证明之。如果不为真，解释为什么并且给出正确的结果。

(b) 在 $\mathbb{Z}_2$ 中解递归关系式 $u_{n+2} + u_n = 1$，约束条件为 $u_0 = 1$，$u_1 = 0$，将解写成 θ 和 n 的表达式。

(c) 四个流 w_n，x_n，y_n，z_n 由线性反馈寄存器产生，如果我们令

$$k_n = \begin{cases} x_n + y_n + z_n & \text{如果 } z_n + w_n = 1 \\ x_n + w_n & \text{如果 } z_n + w_n = 0 \end{cases}$$

[467] 证明 k_n 也是一个由线性反馈寄存器产生的流。

解答 (a) 我们观察到 $\theta(n^2) = \theta(n)$，所以提出的和只包含两个任意常量。现在考虑 $g(n) = \theta(n(n-1)/2)$。那么

$$\begin{aligned} & g(n+3) + g(n+2) + g(n+1) + g(n) \\ &= \theta\left(\frac{(n+3)(n+2)}{2}\right) + \theta\left(\frac{(n+2)(n+1)}{2}\right) \\ &\quad + \theta\left(\frac{(n+1)n}{2}\right) + \theta\left(\frac{n(n-1)}{2}\right) \\ &= q((n+2)2 + n^2) = 0 \end{aligned}$$

并且有 $g(0) = g(1) = 0$，$g(2) = 1$。那么我们将 $n=0$ 和 $n=1$ 代入关系式 $a\theta(n) + b + cg(n) = 0$，并且观察到 $a = b = c = 0$。所以，$\theta(n)$，1，$g(n)$ 是独立的。因此，$A\theta(n) + B + Cg(n)$ 是一个三阶差分等式的通解。

(b) 首先尝试解递归关系式 $u_{n+2} + u_n = 1$，没有附加条件

$$\begin{aligned} g(n) + g(n+2) &= \theta\left(\frac{n(n-1)}{2} + \frac{(n+2)(n+1)}{2}\right) \\ &= \theta\left(\frac{n^2 - n + n^2 + 3n + 2}{2}\right) \\ &= \theta(n^2 + n + 1) = 1 \end{aligned}$$

现在将 $n=0$ 和 $n=1$ 代入关系式 $u_n = A + B\theta(n) + g(n)$，得到 $A = B = 1$。因此，$u_n = 1 + \theta(n) + g(n)$。

(c) 序列 k_n 是由如下的线性寄存器产生的

$$k_n = x_n + w_n + (z_n + w_n)(y_n + z_n + w_n)$$ □

在这一章接下来的部分，我们将要讨论一系列现代密码系统的性质，通常称为公钥密码，主要集中在 RSA 和比特承诺密码系统。

定义 4.5.13 我们给定一个正规的加密系统，如果我们可以识别：

(a) 一个纯文本集合$\mathscr{P}$(按第1章中的说法是源信息)。

(b) 一个密文集合$\mathscr{C}$(按第1章中的说法是码字)。

(c) 一个用于标记编码映射的密钥集合$\mathscr{K}$。

(d) 加密函数(编码映射)的集合$\mathscr{E}$，其中每个函数 E_k 取 $P\in\mathscr{P}\to E_k(P)\in\mathscr{C}$，并且被元素 $k\in\mathscr{K}$ 标记。

(e) 解密函数(译码映射)的集合$\mathscr{D}$，其中每个函数 D_k 取 $C\in\mathscr{C}\to D_k(C)\in\mathscr{P}$，并且被元素 $k\in\mathscr{K}$ 再次标记。 468

使得：

(f) 对于所有的密钥 $e\in\mathscr{K}$，存在一个密钥 $d\in\mathscr{K}$，对于所有纯文本 $P\in\mathscr{P}$ 具有性质 $D_d(E_e(P))=P$。

例子 4.5.14 假设两个人，Bob 和 Alice，准备举行一个双边私人通信。他们想通过使用一个不安全的二进制信道交换他们的密钥，E_A 和 E_B。一个简单的协议如下，Alice 将纯文本 m 加密为 $E_A(m)$ 之后发送给 Bob。他将接收的信息加密为 $E_B(E_A(m))$ 之后返回给 Alice。现在我们做出一个关键性的假设：E_A 和 E_B 对任何纯文本 m' 是可交换的：$E_A\circ E_B(m')=E_B\circ E_A(m')$。这种情况下，Alice 可以解密这个信息为 $D_A(E_A(E_B(m)))=E_B(m)$ 并且发送给 Bob，然后 Bob 计算出 $D_B(E_B(m))=m$。在这个协议下，在通信过程中决不是一个未加密的信息传输。

然而，更进一步的思考证实这完全无解。确实，假设 Alice 使用一次一密 k_A，Bob 使用一次一密 k_B。那么任何单次侦听都无法对发送的信息 m 进行提高。然而，如果所有的三个传输都被侦听，那么就足以通过取和

$$(m+k_A)+(m+k_A+k_B)+(m+k_B)=m$$

来获得纯文本 m。所以需要开发更加复杂的协议；这也是公钥密码系统有用的地方。

另一个通俗的例子是，在股票市场进行交易的投资者和经纪人组成的网络，他们使用开放存取密码系统，例如 RSA。投资者顾虑一个经纪人会在不经过授权的情况下购买股票，并且在亏损的时候声称他有一个委托人的书面请求。确实，经纪人生成一个要求购买股票的命令是很容易的，因为编码的密钥不受版权的限制。另一方面，经纪人可能会担心如果他根据投资者的请求购买了股份，并且股票市场下滑，投资人可能会声称他从来没有指示进行这个交易并且他的编码申请是假的。

然而，可以很轻松地开发一个协议来解决这些担忧。投资人 Alice 在给经纪人 Bob 发送购买股票的请求 p 的同时发送她的"电子签名" $f_Af_A^{-1}(p)$。在接到这个信息之后，Bob 发送一个编码为 $f_Af_B^{-1}(r)$ 的接收凭证 r。如果产生了冲突，那么双方可以提供编码信息和密钥给第三方(称为法庭)。因为除了 Alice 之外没有人可以生成由 $f_Bf_A^{-1}$ 编码的信息并且除了 Bob 没有人可以生成由 $f_Af_B^{-1}$ 编码的信息，这样就没有异议了。这就是比特承诺的主旨。上面提到的 RSA(Rivest-Shamir-Adelman)方案是一个关于公钥密码系统的最好例子。这里，一个接收用户(Bob，可能是一个集合元素)设定 469

$$N=pq,\quad \text{其中 } p \text{ 和 } q \text{ 是两个大的素数，被保密} \tag{4.5.21}$$

数字 N 通常被称为 RSA 模(公开的)。欧拉函数的值为

$$\phi(N)=(p-1)(q-1),\quad \text{被保密}$$

接下来，接收用户选择(或者由密钥中心给出)一个整数 l 使得

$$1<l<\phi(N)\quad \text{和}\quad \gcd(\phi(N),l)=1 \tag{4.5.22}$$

最后，整数 d 可以计算出来(再次通过 Bob 或者他自己)，使得

$$1 < d < \phi(N) \quad 和 \quad ld = 1 \bmod \phi(N) \tag{4.5.23}$$

(d 的值可以通过扩展的 Euclid 算法计算)用于加密的公钥 e_B 是数对(N, l)(列在公共目录中)，发送方(Alice)在与 Bob 通信时，知道 Bob 的纯文本和密文集合为$\mathscr{P}=\mathscr{C}=\{1, \cdots, N-1\}$。然后她将她选择的纯文本 $m=1, \cdots, N-1$ 编码为密文

$$E_{N,l}(m) = c, \quad 其中\ c = m^l \bmod N \tag{4.5.24}$$

Bob 的私钥 d_B 是数对(N, d)(或者是简单的数字 d)，这对于公众来说是保密的，但是 Bob 是知道的。接收方解密密文 c 为

$$D_d(c) = c^d \bmod N \tag{4.5.25}$$

在文献中，l 经常被称为加密，而 d 被称为解密指数。下面的定理 4.5.15 将会确保

$$D_d(c) = m^{ld} = m \bmod N \tag{4.5.26}$$

即密文 c 被正确地解密。更准确地说，

定理 4.5.15 *对于所有的整数 $m=0, \cdots, N-1$，当 l 和 d 满足式(4.5.22)和(4.5.23)，N 如式(4.5.21)所示时，等式(4.5.26)成立。*

证明 根据式(4.5.23)

$$ld = 1 + b(p-1)(q-1)$$

其中 b 是一个整数，那么

470

$$(m^l)^d = m^{ld} = m^{1+b(p-1)(q-1)} = m\,(m^{(p-1)})^{(q-1)b}$$

回想 Euler-Fermat 定理：如果 $\gcd(m, p)=1$，那么 $m^{\phi(p)}=1 \bmod p$。

我们推导出如果 m 不可被 p 整除，那么

$$(m^l)^d = m \bmod p \tag{4.5.27}$$

否则，即如果当 $p|m$ 时，因为 m 和$(m^l)^d$ 等于 0(模 p)，式(4.5.7)依然成立。以此类推

$$(m^l)^d = m \bmod q \tag{4.5.28}$$

根据中国剩余定理(CRT，Chinese remainder theorem)(文献[28]、[114])式(4.5.27)和式(4.5.28)推出式(4.5.26)。 □

例子 4.5.16 假设 Bob 已经选择 $p=29$，$q=31$，$N=899$，并且有 $\phi(N)=840$。在 $\gcd(l, \phi(N))=1$ 的情况下 e 可能的最小值为 $l=11$，之后是 13 紧随其后的是 17，等等。(扩展的)Euclid 算法得到当 $l=13$ 时，$d=517$；当 $l=11$ 时，$d=611$，等等。在第一种情况下，加密密钥 $E_{899,11}$ 为

$$m \mapsto m^{11} \bmod 899, \quad 即\ E_{899,11}(2) = 250$$

密文 250 被解码为

$$D_{611}(250) = 250^{611} \bmod 899 = 2$$

在计算机的辅助下。(即使在 CRT 简化之后还是需要计算机。例如 Mathematica 中的命令为 PowerMod[250, 611, 899]。)

举例 4.5.17 (a) *关于公钥为(N, l)和私钥为$(\phi(N), d)$的 RSA 的密码系统，讨论取(i) $l=2^{32}+1$ 或者(ii) $d=2^{32}+1$ 时可能的优缺点。*

(b) *给定一个(大)数 N，并且我们知道 N 是两个不同素数的积，$N=pq$，但是我们不知道具体的 p 和 q。假设另一个正整数 m 被给定，它是 $\phi(N)$ 的倍数。解释如何找出 p 和 q 的值。*

(c) *描述如何通过 RSA 的方法解决卧承诺问题。*

解答 (a)用 $l=2^{32}+1$ 提供快速加密(你只需要使用重复平方的 33 次乘法)，在 $d=2^{32}+1$ 的情况下可以快速解密信息(但是攻击者可以很容易猜到它)。

471

(b) 接下来，我们证明如果我们知道 $\phi(N)$ 的倍数 m，就可以很“轻松”地分解 N。给定正整数 $y>1$ 和 $M>1$，用 $\mathrm{ord}_M(y)$ 表示与 M 相关的 y 的阶数：

$$\mathrm{ord}_M(y) = \min\{s = 1,2,\cdots; y^s = 1 \bmod M\}$$

假设当 $a\geq 0$，b 为奇数时 $m=2^a b$。令

$$\mathbb{X} = \{x = 1,2,\cdots,N; \mathrm{ord}_p(x^b) \neq \mathrm{ord}_q(x^b)\} \tag{4.5.29}$$

给定 N，l 和 d，我们令 $m=dl-1$。因为 $\varphi(N)\mid dl-1$，我们可以使用下面的引理 4.5.18 来分解 N。我们选择 $x<N$。假设 $\gcd(x, N)=1$，反之搜索已经成功。找到一个非平凡因子的概率为 1/2，所以在 r 随机选择 $x\in\mathbb{X}$ 之后失败的概率为 $1/2^r$。

(c) 比特承诺问题出现在下面的情况：Alice 通过如下方式给 Bob 发送信息。

(i) Bob 无法阅读信息，直到 Alice 发送更进一步的信息。

(ii) Alice 无法改变信息。

一种解决方法为使用电子签名：直到 Alice(随后)透露了她的私钥后 Bob 才能阅读信息。这不违反条件(i)，(ii)并且 Alice 拒绝承认她的作者身份(合法地)是不可能的。 □

引理 4.5.18 (i) 令 $N=pq$，m 如前面的定义，即 $\phi(N)\mid m$，并且定义如式(4.5.29)所示的集合 $\mathbb{X}$。如果 $x\in\mathbb{X}$，那么存在 $0\leq s<a$ 使得 $\gcd(x^{2^s}-1, N)>1$ 是一个 $N=pq$ 的非平凡因子。

(ii) 势 $\#\mathbb{X}\geq\phi(N)/2$。

证明 (i) 令 $y=x^b \bmod N$。Euler-Fermat 定理表明 $x^{\phi(N)}=1 \bmod N$，因此有 $y^m = 1 \bmod N$。那么

$$\mathrm{ord}_p(x^b) \text{ 和 } \mathrm{ord}_q(x^b) \text{ 是 2 的指数}$$

众所周知，$\mathrm{ord}_p(x^b)\neq\mathrm{ord}_q(x^b)$；假设 $\mathrm{ord}_p(x^b)<\mathrm{ord}_q(x^b)$。那么存在 $0\leq t<a$ 使得

$$y^{2^t} = 1 \bmod p, y^{2^t} \neq 1 \bmod q$$

所以，正如所需要的，$\gcd(y^{2^t}-1, N)=p$。

(ii) 根据 CRT，对于 $N\leftrightarrow(p, q)$，有一个双射

$$x \in \{1,\cdots,N\}\leftrightarrow(x \bmod p, x \bmod q) \in \{1,\cdots,p\}\times\{1,\cdots,q\}$$

那么它足以说明如果我们根据 $\mathrm{ord}_p(x^b)$，$x\in\mathbb{X}$ 的值把集合 $\{1, \cdots, p\}$ 分为多个子集，那么每个子集的大小 $\leq(p-1)/2$。我们将列举一个大小为 $(p-1)/2$ 的子集来说明。注意 [472]

$$\phi(N)\mid 2^a b \text{ 表明存在 } \gamma \in \{1,\cdots,p-1\} \quad \text{使得 } \mathrm{ord}_p(\gamma^b) \text{ 是 2 的幂}$$

反过来说，后面的描述意味着

$$\mathrm{ord}_p(\gamma^\delta)\begin{cases} = \mathrm{ord}_p(\gamma^b), & \delta \text{ 为奇数} \\ < \mathrm{ord}_p(\gamma^b), & \delta \text{ 为偶数} \end{cases}$$

因此，$\{\gamma^{\delta b} \bmod p: \delta \text{ 为奇数}\}$ 是所需的子集。 □

我们下一个关于密码的例子是 Rabin，或者叫 Rabin-Williams 密码系统。这里我们再次使用因式分解问题来确保安全性。对于这个系统，与因式分解问题的关系被证明是相互的：已知因式分解问题的解破坏了密码系统，破坏密码系统的能力导致了因式分解。(在 RSA 中并不是如此：无法知道破坏一个 RSA 系统是否能够解因式分解问题。)

在 Rabin 系统中接收用户(Alice)在随机的两个大素数 p 和 q 中选择，其中

$$p = q = 3 \bmod 4 \tag{4.5.30}$$

此外

$$\begin{aligned} &\text{Alice 的公钥为 } N = pq\text{；她的密钥为对 } (p,q) \\ &\text{Alice 的纯文本和密文为数字 } m = 0,1\cdots,N-1 \end{aligned} \tag{4.5.31}$$

并且她的加密准则为 $E_K(m)=c$，其中 $c=m^2 \bmod N$。
为了对一个发送给她的密文 c 进行解密，Alice 计算

$$m_p = c^{(p+1)/4} \bmod p \quad 和 \quad m_q = c^{(q+1)/4} \bmod q \tag{4.5.32}$$

那么

$$\pm m_p = c^{1/2} \bmod p \quad 和 \quad \pm m_q = c^{1/2} \bmod q$$

即 $\pm m_p$ 和 $\pm m_q$ 分别是 c 模 p 和模 q 的平方根。事实上

$$(\pm m_p)^2 = c^{(p+1)/2} = c^{(p-1)/2}c = (\pm m_p)^2 = c \bmod p$$

在最后一步使用了 Euler-Fermat 定理。对于 $\pm m_q$ 的证明与此相似。然后 Alice 通过 Euclid 算法计算整数 $u(p)$ 和 $v(p)$，使得

[473]

$$u(p)p + v(q)q = 1$$

最终，Alice 计算

$$\pm r = \pm[u(p)pm_q + v(q)qm_p] \bmod N$$

和

$$\pm s = \pm[u(p)pm_q - v(q)qm_p] \bmod N$$

有四个 c 模 N 的平方根。纯文本 m 是其中的一个。为了保证她可以识别初始的文本，Alice可能会减少纯文本的空间，只允许具有某些特殊性质（例如初始 32 个和最后 32 个数字互相重复）的纯文本。两个或两个以上的平方根不可能具有这样的性质。然而，这样的方法可能会导致破解密码的难度，因为“简化”后的问题并不总是等同于因式分解。

> 我总是非常钦佩毕德哥拉斯的神秘方法和数字的秘密魔力。
>
> Thomas Browne(1605—1682)，英国作家，
> 写作领域涵盖医学、宗教、科学和旅行

例子 4.5.19 Alice 使用素数 $p=11$ 和 $q=23$，那么 $N=253$，Bob 的加密信息 $m=164$，其中

$$c = m^2 \bmod N = 78$$

Alice 计算 $m_p=1$，$m_q=3$，$u(p)=-2$，$v(q)=1$。那么 Alice 计算

[474]

$$r = -[u(p)pm_q + v(q)qm_p] \bmod N = 210 \text{ 和 } 43$$

$$s = \pm[u(p)pm_q - v(q)qm_p] \bmod N = 164 \text{ 和 } 89$$

并且从解 $164^2 = 78 \bmod 253$ 中找出信息 $m=164$。

我们继续从 Diffie-Hellman 密钥交换方案出发。Diffie 和 Hellman 提出一种协议使一对用户能够通过非安全信道交换密钥。Diffie-Hellman 策略不是一个公钥密码系统，但是它的重要性得到了广泛的认可，因为它形成了 ElGamal 签名密码系统的基础。

Diffie-Hellman 协议与离散对数问题(DLP，discrete logarithm problem)有关：我们有一个素数 p，乘法群为 $\mathbb{F}_p^* \simeq \mathbb{Z}_{p-1}$ 和 $\mathbb{F}_p^*$ 的生成子（即一个 $\mathbb{F}_p^*$ 的本元）为 γ 的域 $\mathbb{F}_p$。那么对于所有的 $b \in \mathbb{F}_p^*$，存在一个唯一的 $a \in \{0, 1, \cdots, p-2\}$，使得

[474]

$$b = \gamma^a \bmod p \tag{4.5.33}$$

那么 a 被称为离散对数，模 p，b 以 γ 为底的对数，某些作者写成 $a=\operatorname{dlog}_\gamma b \bmod p$。计算离散对数被认为是一个困难的问题：没有已知的有效（多项式）算法，尽管没有证据表明这确实是一个非多项式问题。（在一个加性循环群 $\mathbb{Z}/(n\mathbb{Z})$ 中，DLP 变为 $b=\gamma a \bmod n$，并且通过 Euclid 算法求解。）

对于足够多的素数 p，Diffie-Hellman 协议允许 Alice 和 Bob 使用针对 $\mathbb{F}_p$ 的域表建立

一个共同的密钥。这样的话，他们知道一个在各自域中的本原元素 γ。他们同意固定一个大的素数 p 和一个本原元素 $\gamma\in\mathbb{F}_p$。数对 (p,γ) 可能是公开的：Alice 和 Bob 可以通过非安全信道确定 p 和 γ。

接下来，Alice 随机选择 $a\in\{0,1,\cdots,p-2\}$，计算

$$A=\gamma^a \bmod p$$

然后把 A 发送给 Bob，a 被保密。对称地，Bob 随机选择 $b\in\{0,1,\cdots,p-2\}$，计算

$$B=\gamma^b \bmod p$$

然后把 B 发送给 Alice，b 被保密。那么

$$\text{Alice 计算 } B^a \bmod p,\text{Bob 计算 } A^b \bmod p$$

他们的密钥是同样的值

$$K=\gamma^{ab}=B^a=A^b \bmod p$$

攻击者可能要拦截 p，γ，A 和 B，但是都不知道

$$a=\operatorname{dlog}_\gamma A \bmod p \quad 和 \quad b=\operatorname{dlog}_\gamma B \bmod p$$

如果攻击者可以找到离散对数模 p，那么他可以破解密钥：这是唯一已知可以这么做的方法。对于逆问题——如果他可以破解协议就可以解这个离散对数问题——是开放的（这在公钥密码系统中被认为是一个重要的问题）。

然而，像之前讨论的方案，Diffie-Hellman 协议有一个特殊的弱点：对于中间人攻击这种协议是很脆弱的。这里，攻击者利用了这样一个事实：Alice 和 Bob 都不能证明给定的信息实际上来自对方而不是来自第三方。假设攻击者可以截获 Alice 和 Bob 之间的所有信息。假设他可以模仿 Bob 去和 Alice 交换密钥，同时模仿 Alice 去和 Bob 交换密钥。那么就需要使用电子签名来辨别这种伪造。 475

我们用基于电子签名的 ElGamal 密码系统来总结 4.5 节。ElGamal 密码可以被认为是一种 Diffie-Hellman 协议的发展。两种方案都是基于离散对数问题（DLP）的困境，在 ElGamal 系统中，接收用户 Alice 选择一个素数 p 和一个本原元素 $\gamma\in\mathbb{F}_p$。接下来她随机选择一个指数 $a\in\{0,\cdots,p-2\}$，计算

$$A=\gamma^a \bmod p$$

并且声明/广播

$$\text{三元组}(p,\gamma,A),\text{她的公钥}$$

与此同时，她保密

$$\text{指数 } a,\text{她的私钥}$$

Alice 的纯文本集合 $\mathcal{P}$ 为数字 $0,1,\cdots,p-1$。

另一个用户 Bob，期望给 Alice 发送信息并且已知三元组 (p,γ,A)，再次随机选择一个指数 $b\in\{0,1,\cdots,p-2\}$ 并且计算

$$B=\gamma^b \bmod p$$

然后 Bob 让 Alice 知道 B（可以通过广播 B 的值做到）。B 的值扮演着 Bob 的"数字签名"的作用。与此相反，Bob 的指数 b 被保密。

现在，给 Alice 发送信息 $m\in\{0,1,\cdots,p-1\}$，Bob 加密 m 通过数对

$$E_k(m)=(B,c),\quad c=A^b m \bmod p$$

即，Bob 的密文由两部分组成：加密信息 c 和他的数字签名 B。

很显然，值 A 和 B 是 Diffie-Hellman 协议的一部分；在这种意义下，后者可以被认为是 ElGamal 密码的一部分。此外，加密信息 c 是 m 乘上 A^b 的积，A^b 组合了 Alice 的公钥

部分 A 和 Bob 的指数 b。

当 Alice 接收到密文(B, c)时，她使用她的密钥 a。换句话说，她用 B^x 除 c 模 p。一种方便的方法是计算 $x=p-1-a$；因为 $1\leqslant a\leqslant p-2$，值 x 也满足 $1\leqslant x\leqslant p-2$，那么 Alice通过 $B^x c \bmod p$ 解密 c。这样得到初始信息 m，因为

$$B^x c = \gamma^b(p-1-a)A^b m = (\gamma^{p-1})^b(\gamma^a)^{-b}A^b m = A^{-b}A^b m = m \bmod p$$

例子 4.5.20 对于 $p=37$，$\gamma=2$ 和 $a=12$，我们有

475

$$A = \gamma^a \bmod p = 26$$

Alice 的公钥为$(p=37, \gamma=2, A=26)$，她的纯文本为 0，1，…，36，并且私钥为 $a=12$。假设 Bob 已经选了 $b=32$，那么

$$B = 2^{32} \bmod 37 = 4$$

假设 Bob 想发送 $m=31$，他通过如下方式加密 m

$$c = A^b m \bmod p = (26)^{32} m \bmod 37 = 10\times 31 \bmod 37 = 14$$

Alice 解码这个信息为 $2^{12}=7$ 和 $7^{-1}=26 \bmod 37$

$$14\times 2^{(37-12-1)} \bmod 37 = 14\times 7^{24} = 14\times 26 \bmod 37 = 31$$

举例 4.5.21 假设 Alice 想通过 ElGamal 密码系统发送信息“today”给 Bob。描述她如使用如下参数：$p=15485863$，$\gamma=6$ 作为一个本原根模 p，她选择 $b=63$。假设 Bob 的私钥 $a=5$，Bob 如何使用数学程序来发送信息？

解答 Bob 有公钥$(15485863, 6, 7776)$，Alice 也获得了它。她将英文的纯文本通过使用字母顺序转化为等效的数值：19，14，3，0，24。因为$26^5<p<26^6$，她可以将这个纯文本信息表示为一个 5 位的基为 26 的数：

$$m = 19\times 26^4 + 14\times 26^3 + 3\times 26^2 + 0\times 26 + 24 = 8930660$$

现在她计算 $\gamma^b = 6^{63} = 13733130 \bmod 15485863$，那么

$$m\alpha^{\alpha b} = 8930660\times 7776^{63} = 4578170 \bmod 15485863$$

Alice 发送 $c=(13733130, 4578170)$给 Bob。他用他的私钥来计算

$$(\gamma^b)^{p-1-a} = 13733130^{15485863-1-5} = 2620662 \bmod 15485863$$

和

$$(\gamma^b)^{-a} m\gamma^{ab} = 2620662\times 4578170 = 8930660 \bmod 15485863$$

并且将信息变回英文原文。 □

举例 4.5.22 (a) 对于将信息 x 编码为 x^2 并以某个 N 为模，描述 Rabin-Williams 方案。证明如果 N 的选择适当，破解这个码等同于分解两个素数的积。(b) 描述一个公钥为 e，私钥为 d，两个大素数的积为 N 的 RSA 系统。

给出一个简单的例子说明为什么对于同态攻击来说这样的系统是很脆弱的。解释一个有数字签名的系统是如何避免这种攻击的。解释当 e，d 和 N 已知的情况下如何对 N 进行因式分解。

477

解答 (a) 固定两个大素数 p，$q\equiv -1 \bmod 4$，形成一个私钥；广播的公钥为积 $N=pq$。使用的性质如下：

(i) 如果 p 为素数，同余 $a^2\equiv d \bmod p$ 有最多两个解。

(ii) 对于素数 $p\equiv -1 \bmod 4$，即 $p=4k-1$，如果同余 $a^2\equiv c \bmod p$ 有一个解，那么 $a=c^{(p+1)/4} \bmod p$ 是一个解，并且 $a=-c^{(p+1)/4} \bmod p$ 是另一个解。(确实，如果 $c\equiv a^2 \bmod p$，那么根据 Euler-Fermat 定理，$c^{2k}\equiv a^{4k}\equiv a^{(p-1)/2}\equiv a^2 \bmod p$，表明 $c^k\equiv \pm a$。)

这个信息是来自于$\mathscr{M}=\{0, 1, \cdots, N-1\}$中的一个数字 m。加密者(Bob)发送(广播)

$\tilde{m}-m^2 \bmod N$。解密者(Alice)使用性质(ii)来恢复 $m \bmod p$ 的两个可能值和 $m \bmod q$ 的两个可能值。那么 CRT 产生四个可能的 m 的值：三个错的，一个对的。

所以，如果可以对 N 进行因式分解，那么码是可以破解的。相反，假设我们可以破解一个码，那么对于一个总的 u，我们可以找到四个不同的平方根 u_1，u_2，u_3，$u_4 \bmod N$。(CRT 加法性质(i)表明 u 有零个或者四个平方根，除非它是 p 和 q 的倍数。)那么 u_1u^{-1}(通过 Euclid 算法可计算)产生四个平方根 1，-1，ε_1 和 ε_2，$1 \bmod N$，其中

$$\varepsilon_1 = 1 \bmod p, \varepsilon_1 = -1 \bmod q$$

和

$$\varepsilon_2 = -1 \bmod p, \varepsilon_2 = 1 \bmod q$$

通过交换 p 和 q，如果需要的话，我们可以假设我们已经知道了 ε_1。因为 ε_1-1 可以被 p 整除而无法被 q 整除，所以 $\gcd(\varepsilon_1-1, N)=p$；这样的话 p 可以通过 Euclid 算法找到，那么 q 也可以确定。

事实上，可以通过如下方法来做。假设我们可以找到平方根模 N，我们随机选取 x 并且解同余 $x^2=y^2 \bmod N$，当概率为 1/2 时，我们有 $x \not\equiv \pm y \bmod N$，那么 $\gcd(x-y, N)$是 N 的非平凡因子。我们重复这个过程直到我们确定一个因子，在 k 次尝试之后成功的概率为 $1-2^{-k}$。

(b) 为了定义 *RSA* 密码系统，我们随机选择大素数 p 和 q。根据 Fermat's little 定理

$$x^{p-1} = 1 \bmod p, x^{q-1} = 1 \bmod q$$

因此，通过令 $N=pq$ 和 $\lambda(N)=\mathrm{lcm}(p-1, q-1)$，对于所有的与 N 互素的整数 x，我们有

$$x^{\lambda(N)} = 1 \bmod N$$

478

接下来，我们随机选择 e。要么 Euclid 算法可以揭示 e 与 $\lambda(N)$ 不是互素的，要么我们可以使用 Euclid 算法来找到 d 使得

$$de = 1 \bmod \lambda(N)$$

有很高的概率在几次尝试就给出合适的 d 和 e。

我们现在给出公钥 e 的值和 N 的值，但是对私钥 d 保密。给定一个信息 m，满足 $1 \leqslant m \leqslant N-1$，它被编码为整数 c，满足

$$1 \leqslant c \leqslant N-1 \text{ 和 } c = m^e \bmod N$$

除非 m 与 N 不是互素的(一个小概率事件)，否则我们可以通过观察如下式子进行解码

$$m = m^{ed} = c^d \bmod N$$

作为一个同态攻击的例子，假设系统被用于传输一个数字 m(需要支付的钱数)并且某人用 c^2 替换了编码信息 c，那么

$$(c^2)^d = m^{2ed} = m^2$$

信息的接收方相信需要支付的钱数为 m^2。

假设数字签名 $B(m)$也被编码并且传输，其中 B 是一个多对一的没有简单代数性质的函数。那么，上面的攻击者会产生一个信息和数字签名，它们不相对应，并且接收方会知道信息是被篡改过的。

假设 e，d 和 N 是已知的。因为

$$de-1 = 0 (\bmod \lambda(N))$$

并且 $\lambda(N)$是偶数，$de-1$ 是偶数。因此当 b 为奇数并且 $a \geqslant 1$ 时，$de-1=2^a b$。

随机选择 x。令 $z=xb \bmod N$。根据 CRT，当且仅当 z 是 $1 \bmod p$ 和 q 的平方根时，它是 $1 \bmod N=pq$ 的平方根，因为 $\mathbb{F}_2$ 是一个域，

$$x^2 = 1 \bmod p \Leftrightarrow (x-1)(x+1) = 0 \bmod p$$
$$\Leftrightarrow (x-1) = 0 \bmod p \text{ 或 } (x+1) = 0 \bmod p$$

因此 1 有四个平方根 $w \bmod N$，满足 $w=\pm 1 \bmod p$ 和 $w=\pm 1 \bmod q$。换句话说，

$$w = 1 \bmod N, w = -1 \bmod N$$
$$w = w_1 \bmod N, \text{其中 } w_1 = 1 \bmod p \text{ 且 } w_1 = -1 \bmod q$$

或者

479 $$w = w_2 (\bmod N), \text{其中 } w_2 = -1 \bmod p \text{ 且 } w_1 = 1 \bmod q$$

现在 z（1 mod N 的平方根）无法满足 $z=1 \bmod N$。如果 $w=-1 \bmod N$，非常不幸我们还得再试一次。否则我们知道 $z+1$ 与 0 mod N 不同余，但是却可以被 N 的两个素数因子中的一个整除。可以利用 Euclid 算法得到共同因子。在已经找到一个素数因子的情况下，我们可以通过除法或者查找 $z-1$ 发现另一个。

因为 1 的平方根在代数上是无法区分的，所以这种方法失败的概率随着尝试次数的增加趋向于零。 □

4.6 本章附加问题

问题 4.1 (a) 令 $(N_t)_{t\geq 0}$ 为在 $p\in(0,1)$ 上速率为 $\lambda>0$ 的 Poisson 过程。假设 (N_t) 里的每一跳中概率为 p 被算作类型 1，概率为 $1-p$ 被算作类型 2，其中跳与跳之间是独立的，并且与 Poisson 过程也是独立的。令 $M_t^{(1)}$ 为类型 1 跳在时间 t 时的数目，$M_t^{(2)}=N_t-M_t^{(1)}$ 为类型 2 跳在时间 t 时的数目。求过程对 $(M_t^{(1)})_{t\geq 0}$ 和 $(M_t^{(2)})_{t\geq 0}$ 的联合分布。如果我们设定有 m 种类型（并不是两种类型）的概率 $p_1, \cdots, p_m$，$p_1+\cdots+p_m=1$，这会是什么样的情况？

(b) 在速率为 λ 的 Poisson 过程 $(N_t)_{t\geq 0}$ 的跳转时刻，一个人以每次一人的方式收集优惠券，有 m 型优惠券，每次 j 型优惠券以概率 p_j 被收集，并且与之前所收集到的优惠券不相关，与 Poisson 过程也不相关。令 T 为所有的优惠券都被收集的第一时间，证明

$$\mathbb{P}(T<t) = \prod_{j=1}^{m}(1-e^{-p_j\lambda t}) \tag{4.6.1}$$

令 $L=N_T$ 表示优惠券类型的全集被获得时所收集优惠券的总数。证明 $\lambda\mathbb{E}T=\mathbb{E}L$，或相反，推理得 $\mathbb{E}L$ 与 λ 无关。

解答 (a) 部分直接根据 Poisson 过程的定义可得。

(b) 令 T_j 为收集到类型 j 的优惠券的第一时间，那么 $T_j\sim\mathrm{Exp}(p_j\lambda)$，对于不同类型 j 它们是互相独立的。我们有

$$T = \max[T_1, \cdots, T_m]$$

因此

480 $$\mathbb{P}(T<t) = \mathbb{P}(\max[T_1,\cdots,T_m]<t) = \prod_{j=1}^{m}\mathbb{P}(T_j<t) = \prod_{j=1}^{m}(1-e^{-p_j\lambda t})$$

接下来，我们观察随机变量计数初始 Poisson 过程 (N_t) 的跳转，直到收集到优惠券类型的全集时为止。即

$$T = \sum_{i=1}^{L} S_i$$

其中 $S_1, S_2, \cdots$ 是在 (N_t) 中的持续时间，并且有 $S_j\sim\mathrm{Exp}(\lambda)$，对于不同的 j 是互相独立的，那么

$$\mathbb{E}(T\mid L=n) = n\mathbb{E}S_1 = n\lambda^{-1}$$

此外，L 与随机变量 S_1，S_2，…是独立的，因此

$$\mathbb{E}T = \sum_{n\geq 1}\mathbb{P}(L=n)\mathbb{E}(T\mid L=n) = \mathbb{E}S_1\sum_{n\geq 1}n\mathbb{P}(L=n) = \lambda^{-1}\mathbb{E}L$$

但是

$$\begin{aligned}\lambda\mathbb{E}T &= \lambda\int_0^\infty \mathbb{P}(T>t)\mathrm{d}t\\ &= \lambda\int_0^\infty\Big[1-\prod_{j=1}^{\infty}(1-\mathrm{e}^{-\lambda p_j t})\Big]\mathrm{d}t\\ &= \int_0^\infty\Big[1-\prod_{j=1}^{\infty}(1-\mathrm{e}^{-p_j t})\Big]\mathrm{d}t\end{aligned}$$

并且 RHS 与 λ 无关。

相当于，不管前面的收集结果，对于获得类型 j 的优惠券的概率为 p_j，当收集发生在正整数时间 $t=1$，2，…时，L 被定义为收集到优惠券全集所需的收集数目。在这种结构下，λ 不出现，所以均值 $\mathbb{E}L$ 不依赖于 λ（事实上是因为 L 的整体分布）。 □

问题 4.2 排队系统在 PSE II 中进行了详细的讨论。我们经常涉及这个话题是因为它可以在点过程中提供非常丰富的例子。考虑一个 k 个队列的排队系统，每个队列里可以有无限多的顾客，其中对于 $i=1$，…，$k-1$，顾客在离开第 i 个队列的同时到达第 $(i+1)$ 个队列。到达第一个队列符合速率为 λ 的 Poisson 过程。对于所有的 i，第 i 个队列的服务时间与分布 F 都是独立的，而且与其他队列的服务时间也是互相独立的。假设初始时系统是空的，用 $V_i(t)$ 表示在队列 i 中，时间 $t\geq 0$ 时的顾客数目。证明 $V_1(t)$，…，$V_k(t)$ 是互相独立的 Poisson 随机变量。

在 $F(t)=1-\mathrm{e}^{-\mu t}$ 的情况下证明

$$\mathbb{E}V_i(t) = \frac{\lambda}{\mu}\mathbb{P}(N_t\geq i),\ t\geq 0,\ i=1,\cdots,k \tag{4.6.2}$$

其中 $(N_t)_{t\geq 0}$ 是速率为 μ 的 Poisson 过程。 481

假设现在第一个到达队列在时间 T 时停止。确定第 i 个队列在每个时间 $t\geq T$ 时的顾客平均数。

解答 我们将乘积定理应用到随机向量为 $Y_n=(S_n^1,\cdots,S_n^k)$ 的到达 Poisson 过程中，其中 S_n^i 是第 i 个队列中第 n 个顾客的服务时间。那么

$$\begin{aligned}V_i(t) &= \text{第 } i \text{ 个队列在时间 } t \text{ 的顾客数}\\ &= \sum_{n=1}^{\infty}\mathbb{1}(\text{在时间 } J_n \text{ 到达第一个队列的第 } n \text{ 个用户}\\ &\qquad \text{是第 } i \text{ 个队列中在时间 } t \text{ 时离开的用户})\\ &= \sum_{n=1}^{\infty}\mathbb{1}(J_n\geq 0,S_n^1,\cdots,S_n^k\geq 0\\ &\qquad J_n+S_n^1+\cdots+S_n^{i-1}<t<J_n+S_n^1+\cdots+S_n^i)\\ &= \sum_{n=1}^{\infty}\mathbb{1}[(J_n,(S_n^1,\cdots,S_n^k))\in A_i(t)] = M(A_i(t))\end{aligned}$$

这里 $(J_n:n\in\mathbb{N})$ 表示一个速率为 λ 的 Poisson 过程的跳转时间，在 $(0,\infty)\times\boldsymbol{R}_+^k$ 上的测度 M 和 v 定义为

$$M(A) = \sum_{n=1}^{\infty}\mathbb{1}((J_n,Y_n)\in A),\ A\subset(0,\infty)\times\boldsymbol{R}_+^k$$

和

$$\nu((0,t]\times B)=\lambda t\mu(B)$$

乘积定理表明 M 是一个关于紧测度 ν 的在 $(0,\infty)\times\mathbf{R}_+^k$ 上的 Poisson 随机测度。接下来，集合 $A_i(t)\subset(0,\infty)\times\mathbf{R}_+^k$ 定义为

$$\begin{aligned}A_i(t)&=\{(\tau,s^1,\cdots,s^k):0<\tau<t,s^1,\cdots,s^k\geqslant 0\\&\quad 且\ \tau+s^1+\cdots+s^{i-1}\leqslant t<\tau+s^1+\cdots+s^i\}\\&=\{(\tau,s^1,\cdots,s^k):0<\tau<t,s^1,\cdots,s^k\geqslant 0\\&\quad 且\ \sum_{l=1}^{i-1}s^l\leqslant t-\tau<\sum_{l=1}^{i}s^l\}\end{aligned}$$

对于 $i=1,\cdots,k$ 集合 $A_i(t)$ 为逐点不相交的(在 $t-\tau$ 时可以将到随后的部分和 $\sum_{l=1}^{i-1}s^l$ 与 $\sum_{l=1}^{i}s^l$ 之间一次)。所以随机变量 $V_i(t)$ 为独立的 Poisson 随机变量。

482

一个直接的验证是通过联合的 MGF，也就是令 $N_t\sim\mathrm{Po}(\lambda t)$ 为在时间 t 到达第一个队列的数目，那么可以写为

$$\begin{aligned}M_{V_1(t),\cdots,V_k(t)}(\theta_1,\cdots,\theta_k)&=\mathbf{E}\exp(\theta_1V_1(t)+\cdots+\theta_kV_k(t))\\&=\mathbf{E}\Big[\mathbf{E}\exp\Big(\sum_{i=1}^{k}\theta_iV_i(t)\Big|N_t;J_1,\cdots,J_{N_t}\Big)\Big]\end{aligned}$$

反过来，给定 $n=1,2,\cdots$ 和点 $0<\tau_1<\cdots<\tau_n<t$，那么条件期望为

$$\begin{aligned}&\mathbf{E}\exp\Big(\sum_{i=1}^{k}\theta_iV_i(t)\Big|N_t=n;J_1=\tau_1,\cdots,J_n=\tau_n\Big)\\&=\mathbf{E}\exp\Big(\sum_{i=1}^{k}\theta_i\sum_{j=1}^{n}\mathbf{1}[(\tau_j,(S_j^1,\cdots,S_j^k))\in A_i(t)]\Big)\\&=\mathbf{E}\exp\Big(\sum_{j=1}^{n}\sum_{i=1}^{k}\theta_i\mathbf{1}[(\tau_j,(S_j^1,\cdots,S_j^k))\in A_i(t)]\Big)\\&=\prod_{j=1}^{n}\mathbf{E}\exp\Big(\sum_{i=1}^{k}\theta_i\mathbf{1}[(\tau_j,(S_j^1,\cdots,S_j^k))\in A_i(t)]\Big)\end{aligned}$$

接下来对 n 进行求和，对 $\tau_1,\cdots,\tau_n$ 进行积分：

$$\begin{aligned}&\mathbf{E}\Big[\mathbf{E}\exp\Big(\sum_{i=1}^{k}\theta_iV_i(t)\Big|N_t;J_1,\cdots,J_{N_t}\Big)\Big]=\sum_{n=1}^{\infty}\lambda^ne^{-\lambda t}\int_0^t\int_0^{\tau_n}\cdots\int_0^{\tau_2}\\&\quad\times\prod_{j=1}^{n}\mathbf{E}\exp\Big(\sum_{i=1}^{k}\theta_i\mathbf{1}[(\tau_j,(S_j^1,\cdots,S_j^k))\in A_i(t)]\Big)d\tau_1\cdots d\tau_{n-1}d\tau_n\\&=\sum_{n=1}^{\infty}\frac{\lambda^n}{n!}e^{-\lambda t}\Big(\int_0^t\mathbf{E}\exp\Big(\sum_{i=1}^{k}\theta_i\mathbf{1}[(\tau,(S^1,\cdots,S^k))\in A_i(t)]\Big)d\tau\Big)^n\\&=\exp\Big(\lambda\int_0^t\Big[\mathbf{E}\exp\Big(\sum_{i=1}^{k}\theta_i\mathbf{1}[(\tau,(S^1,\cdots,S^k))\in A_i(t)]\Big)-1\Big]d\tau\Big)\\&=\exp\Big[\lambda\int_0^t\sum_{i=1}^{k}\mathbf{P}((\tau,(S^1,\cdots,S^k))\in A_i(t))(e^{\theta_i}-1)d\tau\Big]\\&=\prod_{i=1}^{k}\exp\Big[(e^{\theta_i}-1)\lambda\int_0^t\mathbf{P}\Big(\sum_{l=1}^{i-1}S^l<t-\tau<\sum_{l=1}^{i}S^l\Big)d\tau\Big]\end{aligned}$$

对于一个给定的 MGF，根据随机变量的唯一性，可得

$$V_i(t) \sim \mathrm{Po}\Big(\lambda \int_0^t \mathbb{P}\Big(\sum_{j=1}^{i-1} S^j < t - \tau < \sum_{j=1}^{i} S^j\Big) \mathrm{d}\tau\Big), \quad \text{相互独立}$$ 483

如果 $F(t)=1-\mathrm{e}^{-\mu t}$，那么部分和 S_1，S_1+S_2，…标记了速率为 μ 的 Poisson 过程 $(\widetilde{N}_t)$ 后面的点，在这种情况下，$\mathbb{E}V_i(t)=v(A_i(t))$ 等于

$$\lambda \int_0^t \mathbb{P}\Big(\sum_{j=1}^{i-1} S^j \leqslant t-\tau < \sum_{j=1}^{i} S^j\Big) \mathrm{d}\tau = \lambda \int_0^t \mathbb{P}(\widetilde{N}_{t-\tau} = i-1) \mathrm{d}\tau$$

$$= \lambda \mathbb{E} \int_0^t \mathbf{1}(\widetilde{N}_s = i-1) \mathrm{d}s = \frac{\lambda}{\mu} \mathbb{P}(\widetilde{N}_t \geqslant i)$$

最终，写出在时间 T 关闭入口之后时刻 t 时队列 i 中的顾客数 $V_i(t, T)$，即

$$\mathbb{E}V_i(t,T) = \lambda \int_0^{t} \mathbb{P}(\widetilde{N}_{t-\tau} = i-1) \mathrm{d}\tau = \lambda \mathbb{E} \int_{t-T}^{t} \mathbf{1}(\widetilde{N}_s = i-1) \mathrm{d}s$$

$$= \frac{\lambda}{\mu} [\mathbb{P}(\widetilde{N}_t \geqslant i) - \mathbb{P}(\widetilde{N}_{t-T} \geqslant i)] \qquad \square$$

问题 4.3 超市的用户到达时间形成了速率为 λ 的 Poisson 过程。每个顾客花一个随机长度的时间 S 来挑选要购买的物品，其中 S 对于在时间 t 到达的顾客来说有 PDF$(f(s, t): s \geqslant 0)$。顾客之间的行为互相独立。在收银台要花费时间 $g(S)$ 来结账。超市有一个政策，就是顾客不应该在收银台等待，所以根据需求要有更多的可以使用的收银台。求

(i) 第一个顾客在第二个顾客到来之前的概率。

(ii) 在时间 T 收银台数目的分布。

解答 (i) 如果 J_1 为第一个顾客的到达时间，那么 J_1+S_1 为他进入收银台的时间，$J_1+S_1+g(S_1)$ 为他离开时的时间。令 J_2 为第二个顾客到达的时间。那么 J_1，$J_2-J_1 \sim \mathrm{Exp}(\lambda)$，并且互相独立。

那么

$$\mathbb{P}(S_1 + g(S_1) < J_2 - J_1) = \int_0^\infty \mathrm{d}t_1 \lambda \mathrm{e}^{-\lambda t_1} \int_0^\infty \mathrm{d}t_2 \lambda \mathrm{e}^{-\lambda t_2} \int_0^{t_2} \mathrm{d}s_1 f(s_1, t_1) \mathbf{1}(s_1 + g(s_1) < t_2)$$

$$= \int_0^\infty \mathrm{d}t_1 \lambda \mathrm{e}^{-\lambda t_1} \int_0^\infty \mathrm{d}s_1 f(s_1, t_1) \int_{s_1+g(s_1)}^\infty \mathrm{d}t_2 \lambda \mathrm{e}^{-\lambda t_2}$$

$$= \int_0^\infty \mathrm{d}t_1 \lambda \mathrm{e}^{-\lambda t_1} \int_0^\infty \mathrm{d}s_1 f(s_1, t_1) \mathrm{e}^{-\lambda(s_1+g(s_1))}$$ 484

(ii) 令 $N_T^\dagger$ 表示在时间 T 时使用的收银台数目，根据乘积定理 4.4.11，可知 $N_T^\dagger \sim \mathrm{Po}(\Lambda(T))$，其中

$$\Lambda(T) = \lambda \int_0^T \mathrm{d}u \int_0^\infty \mathrm{d}s f(s,u) \mathbf{1}(u + s < T, u + s + g(s) > T)$$

$$= \lambda \int_0^T \mathrm{d}u \int_0^\infty \mathrm{d}s f(s,u) \mathbf{1}(T - g(s) < u + s < T)$$

事实上，如果在时间 T 时到达顾客的数目是 $N_T^\dagger \sim \mathrm{Po}(\lambda T)$，则可得

$$N_T^\dagger = \sum_{i=1}^{N_T^\dagger} \mathbf{1}(J_i + S_i < T < J_i - S_i + g(S_i))$$

其中 MGF 为

$$\mathbb{E}\exp(\theta N_T^\dagger) = \mathbb{E}[\mathbb{E}(\exp(\theta N_T^\dagger) \mid N_T^\dagger; J_1, \cdots, J_{N_T^\dagger})]$$

$$= \mathrm{e}^{-\lambda T} \sum_{k=0}^\infty \lambda^k \int_0^T \int_0^{t_k} \cdots \int_0^{t_2} \prod_{i=1}^k \mathbb{E}\exp[\theta \mathbf{1}(t_i - S_i < T < t_i - S_i + g(S_i))] \mathrm{d}t_1 \cdots \mathrm{d}t_k$$

$$= e^{-\lambda T}\sum_{k=0}^{\infty}\frac{\lambda^k}{k!}\int_0^T\cdots\int_0^T\prod_{i=1}^{k}\mathbb{E}\exp[\theta\mathbf{1}(t_i+S_i<T<t_i+S_i+g(S_i))]\mathrm{d}t_1\cdots\mathrm{d}t_k$$

$$= e^{-\lambda T}\sum_{k=0}^{\infty}\frac{\lambda^k}{k!}\left(\int_0^T\mathbb{E}\exp[\theta\mathbf{1}(t+S<T<t+S+g(S))]\right)^k$$

$$= \exp\left[\lambda\int_0^T\mathbb{E}(\exp[\theta\mathbf{1}(t+S<T<t+S+g(S))]-1)\mathrm{d}t\right]$$

$$= \exp\left[\lambda(e^\theta-1)\int_0^T\mathbb{P}(t+S<T<t+S+g(S))\mathrm{d}t\right]$$

$$= \exp\left[(e^\theta-1)\lambda\int_0^T\int_0^\infty f(s,u)\mathbf{1}(u+s<T<u+s+g(s))\mathrm{d}s\mathrm{d}u\right]$$

它验证了上述结论。 □

问题 4.4 图书馆的开门时间为上午9点到下午5点。下午5点之后没有学生进入图书馆，但是已经在图书馆里的学生可以待到5点之后。在上午9点到下午5点这段时间，学生到达图书馆的时间满足速率为 λ 的 Poisson 过程。每个学生待在图书馆的总时间 H 是随机的，即 $0\leqslant H\leqslant 8$ 是 PDF 为 h 的随机变量，且 $\mathbb{E}[H]=1$，不同学生待在图书馆的时间是独立同分布的随机变量。

485

(a) 求在下午3点到4点之间离开图书馆学生人数的分布。

(b) 证明在下午3点到4点之间离开图书馆学生的平均人数是 $\mathbb{E}[\min(1,(7-H)_+)]$，其中 w_+ 表示 $\max[w,0]$。

(c) 在关门时仍然在图书馆的学生人数是多少？

解答 图书馆的开门时间是上午9点到下午5点。学生的到达满足 PP(λ)，这个问题等价于 $M/GI/\infty$ 排队。(直到下午5点，不允许任何人再进入的限制是有用的，但涉及更早时间的问题这个限制是不重要的。)

用 J_n 表示第 n 个学生的到达时间。用24小时制表示。

用 H_n 表示第 n 个学生待在图书馆的时间。

同样使用乘积定理4.4.11关于原子 (J_n, Y_n) 在 $(0,8)\times(0,8)$ 上的随机测量，其中 $(J_n, n\in\mathbf{N})$ 是到达时间，$(Y_n, n\in\mathbf{N})$ 是学生待在图书馆的时间。通过 $\mu((0,t)\times B)=\lambda t\mu(B)$，$N(A)=\sum_n \mathbf{1}_{(J_n,H_n)\in A}$ 定义在 $(0,\infty)\times\mathbf{R}_+$ 上的测度。那么 N 是关于强度为 $\nu([0,t]\times[0,y])=\lambda tF(y)$ 的 Poisson 随机测度，其中 $F(y)=\int_0^y h(x)\mathrm{d}x$ ($t=0$ 对应上午9点)。

(a) 现在，在下午3点到4点之间离开图书馆的学生人数满足 Poisson 分布 $\mathrm{Po}(\nu(A))$，其中 $A=\{(r,s): s\in[0,7]$，如果 $s\leqslant 6$，$r\in[0,7]$；如果 $s>6$，$r\in[0,7-s]\}$。可得

$$\nu(A)=\int_0^{\infty}\lambda\mathrm{d}F(r)\int_{(6-r)_-}^{(7-r)_+}\mathrm{d}s=\int_0^{\infty}\lambda[(7-r)_+-(6-r)_-]\mathrm{d}F(r)$$

因此在下午3点到4点之间离开图书馆的学生人数满足速率为 $\lambda\int_0^{\infty}[(7-y)_+-(6-y)_+]\mathrm{d}F(r)$ 的 Poisson 分布。

(b)

486

$$(7-y)_+-(6-y)_+=\begin{cases}0, & y\geqslant 7\\ 7-y, & 6\leqslant y\leqslant 7\\ 1, & y\leqslant 6\end{cases}$$

根据要求，在下午3点到4点之间离开图书馆的学生平均人数为

$$v(A)=\int_0^8\lambda[\min(1,(7-r)_+)]\mathrm{d}F(r)=\lambda E[\min(1,(7-H)_+)]$$

(c) 对于在闭馆时间仍然在图书馆的学生，我们要求 $J+H\geqslant 8$，其中 H 的范围为 $[0,8]$，J 的范围为 $[8-H,8]$。令

$$B=\{(t,x):t\in[0,8],x\in[8-t,8]\}$$

因此有

$$\begin{aligned}v(B)&=\lambda\int_0^8\mathrm{d}t\int_{8-t}^8\mathrm{d}F(x)=\lambda\int_0^8\mathrm{d}F(x)\int_{8-x}^8\mathrm{d}t\\&=\lambda\int_0^8(8-x)\mathrm{d}F(x)=8\lambda\int_0^8\mathrm{d}F(x)-\lambda\int_0^8x\mathrm{d}F(x)\end{aligned}$$

但是因为 $\int_0^8\mathrm{d}F(x)=1$ 和 $\int_0^8x\mathrm{d}F(x)=E[H]=1$，说明 $\lambda E[H]=\lambda$。因此，在闭馆时间留在图书馆的剩余学生人数为 7λ。 □

问题 4.5 (i) 证明 Campbell 定理，即证明如果 M 是在状态空间 E 上的 Poisson 随机测度，其中强度为 μ 和 a，$E\to\mathbf{R}$ 是有界可测函数，则有

$$E[\mathrm{e}^{\theta X}]=\exp\left[\int_E(\mathrm{e}^{\theta a(y)}-1)\mu(\mathrm{d}y)\right]\tag{4.6.3}$$

其中 $X=\int_E a(y)M(\mathrm{d}y)$（假定 $\lambda=\mu(E)<\infty$）。

(ii) 在速率为 λ 的 Poisson 过程的跳转时间 J_1，J_2，…时听到枪声。枪声的初始振幅 A_1，A_2，…~Exp(2)是满足参数为 2 的独立同分布的指数分布，且振幅以速率 α 呈线性衰减。计算在时间 t 处总振幅 X_t 的 MGF：

$$X_t=\sum_n A_n(1-\alpha(t-J_n)_+)\mathbf{1}_{\{J_n\leqslant t\}}$$

如果 $x\geqslant 0$ 则 $x_+=x$，否则为 0。 487

解答 (i)如果存在条件 $M(E)=n$，M 的原子构成了分布为 $\frac{1}{\lambda}\mu$ 的随机样本 Y_1，…，Y_n，因此

$$\begin{aligned}E[\mathrm{e}^{\theta X}\mid M(E)=n]&=E[\mathrm{e}^{\theta\sum_{k=1}^n a(Y_k)}]\\&=\left(\int_E\mathrm{e}^{\theta a(y)}\mu(\mathrm{d}y)/\lambda\right)^n\end{aligned}$$

因此有

$$\begin{aligned}E[\mathrm{e}^{\theta X}]&=\sum_n E[\mathrm{e}^{\theta X}\mid M(E)=n]P(M(E)=n)\\&=\sum_n\left(\int_E\mathrm{e}^{\theta a(y)}\mu(\mathrm{d}y)/\lambda\right)^n\frac{\mathrm{e}^{-\lambda}\lambda^n}{n!}\\&=\exp\left(\int_E(\mathrm{e}^{\theta a(y)}-1)\mu(\mathrm{d}y)\right)\end{aligned}$$

(ii) 固定 t 且令 $E=[0,t]\times\mathbf{R}^+$，则 v 和 M 满足 $v(\mathrm{d}s,\mathrm{d}x)=2\lambda\mathrm{e}^{-2x}\mathrm{d}s\mathrm{d}x$，$M(B)=\sum_n\mathbf{1}_{\{(J_n,A_n)\in B\}}$。根据乘积定理，$M$ 是密度为 v 的 Poisson 随机测度。设 $a_t(s,x)=x(1-\alpha(t-s))_+$，则有 $X_t=\int_E a_t(s,x)M(\mathrm{d}s,\mathrm{d}x)$，因此根据 Campbell 定理，对于 $\theta<2$ 时，有

$$
\begin{aligned}
E[e^{\theta X_t}] &= \exp\left(\int_E (e^{\theta s_t(x,s)}-1)\nu(ds,dx)\right) \\
&= e^{-\lambda t}\exp\left(2\lambda\int_0^t\int_0^\infty e^{-\alpha(t-s)-\alpha(t-s)_+}dxds\right) \\
&= e^{-\lambda t}\exp\left(2\lambda\int_0^t ds\,\frac{1}{2-\theta(1-\alpha(t-s))_+}\right) \\
&= e^{-\min[t,1/\alpha]}\left(\frac{2-\theta+\theta\alpha\min[t,1/\alpha]}{2-\theta}\right)^{\frac{2}{\alpha}}
\end{aligned}
$$

当在 $t>\frac{1}{\alpha}$ 时，拆分积分 $\int_0^t=\int_0^{t-\frac{1}{\alpha}}+\int_{t-\frac{1}{\alpha}}^t$。 □

问题 4.6 在域 S 上要中种植种子，种子是以随机的方式播种意味着它们在 S 上形成了强度为 $\lambda(x,y)$ Poisson 过程。种子长成植物后以农作物的形式被收割，在 (x,y) 处植物重量的均值为 $m(x,y)$，方差为 $v(x,y)$。不同植物的重量是独立随机变量。证明所有植物的总重量 W 是一个随机变量且存在有限的均值

488

$$I_1=\iint_S m(x,y)\lambda(x,y)dxdy$$

和方差

$$I_2=\iint_S\{m(x,y)^2+v(x,y)\}\lambda(x,y)dxdy$$

只要这些积分是有限的。

解答 首先假定

$$\mu=\int_S\lambda(x,y)dxdy$$

是有限的。那么植物的数目 N 是有限的并满足分布 $Po(\mu)$。在条件 N 上，它们的位置可以被看成独立随机变量 (X_n,Y_n)，$n=1,\cdots,N$，并且 S 上的密度为 λ/μ。那么植物的重量是独立的，且有

$$EW=\int_S m(x,y)\lambda(x,y)\mu^{-1}dxdy=\mu^{-1}I_1$$

和

$$EW^2=\int_S[m(x,y)^2+v(x,y)]\lambda(x,y)\mu^{-1}dxdy=\mu^{-1}I_2$$

其中 I_1，I_2 是有限的，因此，根据要求

$$E(W|N)=\sum_{i=1}^N\mu^{-1}I_1=N\mu^{-1}I_1$$

和

$$\mathrm{Var}(W|N)=\sum_{i=1}^N(\mu^{-1}I_2-\mu^{-2}I_1^2)=N(\mu^{-1}I_2-\mu^{-2}I_1^2)$$

那么

$$EW=EN\mu^{-1}I_1=I_1$$

和

$$
\begin{aligned}
\mathrm{Var}W &= E[\mathrm{Var}(W|N)]+\mathrm{Var}[E(W|N)] \\
&= \mu(\mu^{-1}I_2-\mu^{-2}I_1^2)+(\mathrm{Var}N)\mu^{-2}I_1^2=I_2
\end{aligned}
$$

489

如果 $\mu=\infty$，我们将 S 分解成 λ 可积的不相交集合 S_k，那么 $W=\sum_k W_{(k)}$，其中基于 S_k

上的产量 $W_{(i)}$ 是相互独立的，然后利用

$$\mathbb{E}W=\sum_i \mathbb{E}W_{(i)}=\sum_i\int_{S_i} m(x,y)\lambda(x,y)\mathrm{d}x\mathrm{d}y$$
$$=\int_S m(x,y)\lambda(x,y)\mathrm{d}x\mathrm{d}y$$

得到结论。$\mathrm{Var}\,W$ 也可以通过类似方法得到。 □

问题 4.7 在 $\mathbb{R}^2$ 上的不经过原点 O 的一条直线 L 可以由它离原点 O 的垂直距离 $p>0$ 和从 O 到 L 的垂线与 x 轴的角度 $\theta\in[0,2\pi)$ 来定义。详细解释这种直线 L 的 Poisson 过程是什么。

对于 $B\subseteq(0,\infty)\times[0,2\pi)$，直线 L 的一个 Poisson 过程 Π 有均值测度如下：

$$\mu(B)=\int\int_B \mathrm{d}p\mathrm{d}\theta \tag{4.6.4}$$

一个随机可数集合 $\Phi\subset\mathbb{R}^2$ 定义为在 Π 中的直线对的所有交集。证明至少在 Φ 中有一个点在以 O 为圆心，r 为半径的圆中的概率小于

$$1-(1+2\pi r)e^{-2\pi r}$$

Φ 是一个 Poisson 过程吗？

解答 假设 μ 是 $\mathbb{R}^2$ 上不经过原点的直线空间 $\mathscr{L}$ 的一个测度，均值测度 μ 的一个 Poisson 过程是 $\mathscr{L}$ 的一个随机可数子集 Π，使得

(1) $\mathscr{L}$ 的一个可测子集 A 中 Π 的点数 $N(A)$ 的分布为 $\mathrm{Po}(\mu(A))$。

(2) 对于不相交的 $A_1,\cdots,A_n$，$N(A_i)$ 相互独立。

在这个问题中，与原点为 0，半径为 r 的圆 D 相交的直线数目 N 等于直线数目，其中 $p<r$。它是 Poisson 分布，均值为

$$\int_0^r\int_0^{2\pi}\mathrm{d}p\mathrm{d}\theta=2\pi r$$

如果 D 中至少存在 Φ 的一个点，那么 Π 中至少有两条直线与 D 相交，这个概率为

$$\sum_{n\geq 2}\frac{(2\pi r)^n}{n!}e^{-2\pi r}=1-(1+2\pi r)e^{-2\pi r}$$

490

Φ 中一个点位于 D 中的概率严格小于上式，因为可能与 D 相交的两个直线的交点在 D 的外面。

最后，Φ 不是一个 Poisson 过程，因为它有正概率的共线的点。 □

问题 4.8 对于参数为 θ 的随机序列 $(p_1,p_2,\cdots)$，它的 Poisson-Dirichlet 分布的特例出现在 PSE II，定义如下。证明对于 $\phi(0)=0$ 的任何多项式 ϕ，

$$\mathbb{E}\left\{\sum_{n=1}^{\infty}\phi(p_n)\right\}=\theta\int_0^1 f(x)x^{-1}(1-x)^{\theta-1}\mathrm{d}x \tag{4.6.5}$$

这揭示了关于 p_i 的分布的哪些方面？

解答 最简单的引入 Poisson-Dirichlet 分布的方式是说 $p=(p_1,p_2,\cdots)$ 有与 (ξ_n/σ) 相同的分布，其中 $\{\xi_n,n=1,2\cdots\}$ 是在 $(0,\infty)$ 上 Poisson 过程降序排列的点，其中速率为 $\theta x^{-1}e^{-x}$，$\sigma=\sum_{n\geq 1}\xi_n$。由 Campbell 定理可知，$\sigma$ 几乎是有限的，分布为 $\mathrm{Gam}(\theta)$（其中 $\theta>0$ 是任意的），并且独立于向量 $p=(p_1,p_2,\cdots)$，其中

$$p_1\geq p_2\geq\cdots,\ \sum_{n\geq 1}p_n=1,\quad 概率为\ 1$$

在这里 Gam 代表了 Gamma 分布，见附录 PSE I。

为了证明(4.6.5)，我们取 $p_n=\xi_n/\sigma$ 并利用 σ 和 p 独立的事实，对于 $k\geq 1$，

$$\mathbb{E}\Big(\sum_{n\geq 1}\xi_n^k\Big)=\int_0^\infty x^k\theta x^{-1}e^{-x}dx=\theta\Gamma(k)$$

左边等于

$$\mathbb{E}\Big(\sigma^k\sum_{n\geq 1}p_n^k\Big)=\Gamma(\theta+k)\Gamma(\theta)^{-1}\mathbb{E}\Big(\sum_{n\geq 1}p_n^k\Big)$$

所以

$$\mathbb{E}\Big(\sum_{n\geq 1}p_n^k\Big)=\frac{\theta\Gamma(k)\Gamma(\theta)}{\Gamma(k+\theta)}=\theta\int_0^1 x^{k-1}(1-x)^{\theta-1}dx$$

我们看到对于 $\phi(x)=x^k(k>=1)$ 恒等式(4.6.5)成立，所以根据所有 $\phi(0)=0$ 的多项式的线性特性，式(4.6.5)也成立。

491

通过多项式近似阶跃方程说明在区间 $(a,b)(0<a<b<1)$ 中 p_n 的平均数等于

$$\theta\int_a^b x^{-1}(1-x)^{\theta-1}dx$$

如果 $a>1/2$，最多有一个这样的 p_n，使得 p_1 的 PDF 为

$$\theta x^{-1}(1-x)^{\theta-1}\ 在(1/2,1)$$

但在 $(0,1/2)$ 不成立，恒等式(4.6.5)并不能确定 p_1 在这区间内的分布。 □

问题 4.9 在大森林中树木的位置可以建模为在 $\mathbb{R}^2$ 上速率恒为 λ 的一个 Poisson 过程 Π，每棵树产生随机数目的种子，满足均值为 ρ 的 Poisson 分布。每个种子落在地上，均匀分布在圆心为树，半径为 r 的圆内。不同种子相对于它们的母树，一棵给定树的种子数相互之间独立并且与 Π 独立。证明以 Π 为条件，种子构成一个 Poisson 过程 Π^*，其均值测度依赖于 Π。Π^* 的无条件分布是一个 Poisson 过程吗？

解答 通过直接的计算，通过一个在 X 的树上产生的种子构成一个 Poisson 过程，速率为

$$\rho_X(x)=\begin{cases}\rho\pi^{-1}r^{-2}, & |x-X|\leq r\\ 0, & 其他\end{cases}$$

这些独立的 Poisson 过程的叠加得到一个 Poisson 过程，速率为

$$\Lambda_\Pi(x)=\sum_{X\in\Pi}\rho_X(x)$$

显然依赖于 Π。为了不怀疑这种依赖，选择圆周均匀分布这种不实际的假设。选择一个假想的圆圈——在这种情况下 Π 可以从 Λ_Π 的等高线重建。

这里我们第一次遇到双重随机(Cox)过程，即具有随机强度的 Poisson 过程。在有限集合 A 中种子数的均值为

492

$$\mathbb{E}N(A)=\mathbb{E}\mathbb{E}[N(A)|\Pi]=\mathbb{E}\int_A\Lambda_\Pi(x)dx$$

方差为

$$\begin{aligned}\mathrm{Var}N(A)&=\mathbb{E}(\mathrm{Var}[N(A)|\Pi])+\mathrm{Var}(\mathbb{E}[N(A)|\Pi])\\&=\mathbb{E}N(A)+\mathrm{Var}\Big[\int_A\Lambda_\Pi(x)dx\Big]\\&>\mathbb{E}N(A)\end{aligned}$$

所以，Π^* 不是一个 Poisson 过程。 □

问题 4.10 在 $\mathbb{R}^3$ 上单位球内的一个均匀 Poisson 过程 Π，其均值测度是 Lebesgue 测度(体积)，在

$$B=\{(x,y,z)\in\mathbb{R}^3:r^2=x^2+y^2+z^2\leq 1\}$$

证明

$$\Pi_1 = \{r:(x,y,z) \in \Pi\}$$

是一个在[0，1]上的 Poisson 过程，并求它的均值测度。证明

$$\Pi_2 = \{(x/r, y/r, z/r):(x,y,z) \in \Pi\}$$

是一个在 $\boldsymbol{B}$ 边界上的 Poisson 过程，其均值测度是表面积的倍数吗？Π_1 和 Π_2 是独立过程吗？

解答 由映射定理可知，Π_1 是一个 Poisson 过程，在(a, b)上点的平均数目等于$\lambda\times$(半径为 a 和 b 的壳的体积)，即

$$\lambda\left(\frac{4}{3}\pi b^3 - \frac{4}{3}\pi a^3\right)$$

所以，Π_1 的均值测度的 PDF 为

$$4\lambda\pi r^2 (0 < r < 1)$$

类似地，在 $\boldsymbol{A}\subset\partial\boldsymbol{B}$ 中 Π_2 的点的平均数目等于

$$\lambda\times(0\text{ 到 }A\text{ 的圆锥体积}) = \frac{1}{3}\lambda\times(A\text{ 的表面积})$$

最后，Π_1 和 Π_2 不独立，因为它们有相同的点。 □

问题 4.11 Π 的点随机涂上红色或绿色，任何一点涂成红色的概率为 r，$0<r<1$，并且不同点的颜色是相互独立的。证明红色和绿色的点构成独立的 Poisson 过程。 493

解答 如果 $A\subseteq S$，有 $\mu(A)<\infty$，那么可写为

$$N(A) = N_1(A) + N_2(A)$$

其中 N_1 和 N_2 是红点和绿点的数目。给定 $N(A)=n$，$N_1(A)$服从二项式分布 $\mathrm{Bin}(n, r)$。所以

$$\begin{aligned}
&\mathbb{P}(N_1(A) = k, N_2(A) = l) \\
&= \mathbb{P}(N(A) = k+l)\mathbb{P}(N_1(A) = k \mid N(A) = k-l) \\
&= \frac{\mu(A)^{k+l}e^{-\mu(A)}}{(k+l)!}\begin{pmatrix} k+l \\ k \end{pmatrix} r^k(1-r)^l \\
&= \frac{[r\mu(A)]^k e^{-r\mu(A)}}{k!}\frac{[(1-r)\mu(A)]^l e^{-(1-r)\mu(A)}}{l!}
\end{aligned}$$

所以，$N_1(A)$和 $N_2(A)$分别是均值为 $r\mu(A)$和$(1-r)\mu(A)$的独立 Poisson 随机变量。

如果 A_1，A_2，…是不相交的集合，那么数对

$$(N_1(A_1), N_2(A_1)), (N_1(A_2), N_2(A_2)), \cdots$$

是相互独立的，所以

$$(N_1(A_1), N_1(A_2), \cdots) \text{ 和 } (N_2(A_1), N_2(A_2), \cdots)$$

是两个独立随机变量的独立序列。如果 $\mu(A)=\infty$，那么 $N(A)=\infty$几乎处处成立。由于 $r>0$，$1-r>0$，在 A 中几乎有无限多红点和绿点。 □

问题 4.12 一个关于暴雨降落在水平面(设为平面 $\mathbb{R}^2$)的模型将每个雨点建模成三元组(X, T, V)，其中 $X\in\mathbb{R}^2$ 是雨滴中心的水平位置，T 是雨滴落地的时刻，V 是雨滴中水的体积。点(X, T, V)被假设是在 $\mathbb{R}^4$ 上构成一个 Poisson 分布，给定速率为 $\lambda(x, t, v)$，雨滴在平面上构成一个涨圆盘，中心为 X，半径随时间增大，在时间$(T+t)$时半径为一个给定的函数 $r(t, V)$。求在时间 τ，一个点 $\xi\in\mathbb{R}^2$ 是干的概率，并且证明如果这个积分收敛，则在暴雨中总雨量的期望为

$$\int_{\mathbb{R}^4} v\lambda(x,t,v)\mathrm{d}x\mathrm{d}t\mathrm{d}v$$ 494

解答 当且仅当满足 $t<\tau$ 的 Π 中一个点，并且

$$\|X-\xi\|<r(\tau-t,V)$$

成立时，$\xi\in\mathbb{R}^d$ 是湿的。

(因为这种差异涉及 0 概率事件，所以无论不等式是否严格都并不影响结果)。满足这两个不等式的 Π 的点数是 Poisson 分布，均值为：

$$\mu=\int\lambda(x,t,v)\mathbf{1}(t<\tau,\|x-\xi\|<r(\tau-t,v))\mathrm{d}x\mathrm{d}t\mathrm{d}v$$

因此，ξ 是干的概率是 $\mathrm{e}^{-\mu}$(或者如果 $\mu=+\infty$ 时，等于 0)。最后，期望总降水量的公式是

$$\sum_{(x,\tau,V)\in\Pi}V$$

它是 Campbell 定理的一个直接应用。

问题 4.13 设 M 是 $E=\mathbf{R}\times[0,\pi)$ 上具有常数强度 λ 的 Poisson 随机测度。对于 $(x,\theta)\in E$，用 $l(x,\theta)$ 表示 $\mathbf{R}^2$ 中的直线，它是以原点为中心，旋转直线 $\{(x,y):y\in\mathbf{R}\}$ 的角度为 θ 得到的。

考虑线性过程 $L=M\circ l^{-1}$。

(i) 与圆 $D_a=\{z\in\mathbf{R}^2:|z|\leqslant a\}$ 相交的直线数量的分布是什么？

(ii) 从原点到最近直线的距离的分布是什么？

(iii) 从原点到第 k 个最近直线的距离的分布是什么？

解答 (i) 一条直线与圆 $D_a=\{z\in\mathbf{R}^2:|z|\leqslant a\}$ 相交，当且仅当它表示的点 (x,θ) 位于 $(-a,a)\times[0,\pi)$，因而

$$\#\text{ 相交线 }D_a\sim\mathrm{Po}(2a\pi\lambda)$$

(ii) 令 Y 表示原点到最近直线的距离。那么

$$\mathbb{P}(Y\geqslant a)=\mathbb{P}(M((-a,a)\times[0,\pi))=0)=\exp(-2a\lambda\pi)$$

即，$Y\sim\mathrm{Exp}(2\pi\lambda)$。

(iii) 令 $Y_1,Y_2,\cdots$ 表示从原点到最近直线、到第二近直线等等的距离。那么 Y_i 是 $\mathbb{R}_+$ 上 PRMN 的原子。它是通过投影 $(x,\theta)\mapsto|x|$ 从 M 上获得的。通过映射定理，N 是 $\mathbb{R}_+$ 上速率为 $2\lambda\pi$ 的 Poisson 过程。因此，因为 $Y_k=S_1+\cdots+S_k$，其中 $S_i\sim\mathrm{Exp}(2\pi\lambda)$，所以 $Y_k\sim\mathrm{Gam}(k,2\lambda\pi)$，且是相互独立的。 □

496

问题 4.14 一个人想通过噪声信道中传输 M 个等概信号中的一个，第 j 个信息被标量序列 $a_{ji}(i=1,2\cdots,n)$ 编码得到，在传输后，接收信号为 $a_{ji}+\varepsilon_i(i=1,2\cdots,n)$。这里的噪声随机变量 ε_i 是服从独立正态分布 d 的，均值为 0，时变方差为 $\mathrm{Var}\varepsilon_i=v_i$。

在接收端得到一个推理规则：信息依被不正确推理的平均概率有上界

$$\mathbb{P}(\text{error})\leqslant\frac{1}{M}\sum_{1\leqslant j\neq k\leqslant M}\exp(-d_{jk}/8)\tag{4.6.6}$$

这里

$$d_{jk}=\sum_{1\leqslant i\leqslant n}(a_{ji}-a_{ki})^2/v_i$$

假设 $M=2$，传输波形受限于功率约束 $\sum_{1\leqslant i\leqslant n}a_{ji}^2\leqslant K,j=1,2$，这两个波形哪一个能最小化错误概率？(提示：你可以假设界限 $\mathbb{P}(Z\geqslant a)\leqslant\exp(-a^2/2)$ 的有效性，其中 Z 是一个标准 $N(0,1)$ 随机变量。)

解答 如果波形 $A_j=(a_{ji})$ 被发射，令 $f_j=f_n(y|X=a_j)$ 为接收向量 y 的 PDF。那么

$$\mathbb{P}(\text{error})\leqslant\frac{1}{M}\sum_j\sum_{k\neq j}\mathbb{P}(\{y:f_k(y)\geqslant f_j(y)\}|X=A_j))$$

令 V 是对角元素为 v_i 的对角矩阵，在当前这种情况下，

$$f_j = C\exp\left(-\frac{1}{2}\sum_{i=1}^{n}(y_i - a_{ji})^2/v_i\right)$$
$$= C\exp\left(-\frac{1}{2}(Y - A_j)^T V^{-1}(Y - A_j)\right)$$

那么，如果 $X = A_i$ 和 $Y = A_i + \varepsilon$，我们有

$$\log f_k - \log f_i = -\frac{1}{2}(A_i - A_k + \varepsilon)^T V^{-1}(A_i - A_k + \varepsilon) - \frac{1}{2}\varepsilon^T V^{-1}\varepsilon$$
$$= -\frac{1}{2}d_{ik} - (A_i - (A_k)^T V^{-1}\varepsilon)$$
$$= -\frac{1}{2}d_{ik} + \sqrt{d_{ik}}Z$$

496

其中 $Z \sim N(0, 1)$. 因此，根据提示，式(4.6.6)可写为如下所示：

$$\mathbb{P}(f_k \geqslant f_i) = \mathbb{P}(Z > \sqrt{d_{ik}}/2) \leqslant e^{-d_{ik}/8}$$

在 $M=2$ 情况下，我们必须最大化

$$d_{12} = (A_1 - A_2)^T V^{-1}(A_1 - A_2) = \sum_{1 \leqslant i \leqslant n}(a_{1i} - a_{2i})^2/v_i$$

它受限于

$$\sum_i a_{ji}^2 \leqslant K \quad 或 \quad (a)_j^T (a)_j \leqslant K, j = 1, 2$$

根据 Cauchy-Schwarz

$$(A_1 - A_2)^T V^{-1}(A_1 - A_2) \leqslant (\sqrt{A_1^T V^{-1} A_1} + \sqrt{A_2^T V^{-1} A_2})^2 \tag{4.6.7}$$

当 $A_1 = \text{const} A_2$ 时，等号成立，此外，在我们的方案中，V 是对角的。当 $A_j^T A_j = K$，$j = 1, 2, \cdots$ 时，式(4.6.7)被最大化。我们可以得出结论

$$a_{1i} = -a_{2i} = b_i$$

对于 i 来说，只有 b_i 非零时才使得 v_i 是最小的，并且 $\sum_i b_i^2 = K$. □

问题 4.15 随机变量 Y 是一个非负整数，证明 Y 的最大熵，受限于 $\mathbb{E}Y \leqslant M$，它是

$$M\log M + (M+1)\log(M+1)$$

它通过均值为 M 的几何分布获得。

对于非负整数输入值 X，无记忆信道产生输出 Y，写为

$$Y = X + \varepsilon$$

其中 ε 和 X 是相互独立的，$\mathbb{P}(\varepsilon = 1) = p$，$\mathbb{P}(\varepsilon = 1) = 1 - p = q$，输入值 X 受 $\mathbb{E}X \leqslant q$ 的约束。证明假如 $p \leqslant 1/3$，最优输入分布是

$$\mathbb{P}(X = r) = (1+p)^{-1}\left(\frac{1}{2^{r+1}} - \left(\frac{-p}{q}\right)^{r+1}\right), \quad r = 0, 1, 2, \cdots$$

并求信道的容量。

简单地描述如果 $p > 1/3$，确定信道容量的问题。 497

解答 首先，考虑这个问题

$$\text{maximise } h(Y) = -\sum_{y \geqslant 0} p_y \log p_y \log p_y, \quad 约束于 \begin{cases} p_y \geqslant 0 \\ \sum_y p_y = 1 \\ \sum_y y p_y = M \end{cases}$$

使用 Lagrangian 乘子法，解为

$$p_y = (1-\lambda)\lambda^y, y = 0, 1, \cdots, \quad 满足 M = \frac{\lambda}{1-\lambda} 或 \lambda = \frac{M}{M+1}$$

其中最优值是

$$h(Y)=(M+1)\log(M+1)-M\log M$$

接着，对于 $g(m)=(m+1)\log(m+1)-m\log m$，

$$g'(m)=\log(m+1)-\log m>0$$

这意味着当 M 增加时，$h(Y)$ 增加。因此，正如所要求的一样，对于 $EY\leqslant M$，最大值和最优值是相同的。

现在，容量 $C=\sup_{\lambda}[h(Y)-h(Y|X)]=h(Y)-h(z)$，条件 $EX\leqslant q$ 表明 $EY\leqslant q+Ez=q+p-1$。关于 $h(z)=-p\log p-q\log q$，我们考虑到 Y 是几何分布，对于 $M=1$，$\lambda=1/2$，则有

$$C=2\log 2+p\log p-q\log q=\log(4p^pq^q)$$

那么

$$Ez^X=\frac{Ez^Y}{Ez^Z}=\left(\frac{1-\lambda}{1-\lambda z}\right)/(pz+q)=\frac{1}{(2-z)(q-pz)}$$

$$=\frac{(2-z)^{-1}-p(q-pz)^{-1}}{1-p}$$

$$=\frac{1}{1+p}\left(\sum(1/2)^{i+1}z^i+(p/q)\sum(-p/q)^iz^i\right)$$

如果 $p>1/3$，那么 $p/q>1/2$。交替概率变成负数，这意味着给定 Y 的一个最优值，不存在 X 的分布。那么，我们必须最大化

$$-\sum_i p_i\log p_i,\quad 当来于\ p_i=p\pi_{i-1}+qx_i$$

其中 $\pi_i\geqslant 0$，$\sum_i\pi_i=1$，$\sum_i i\pi_i\leqslant q$。 □

问题 4.16 假设由第二编码定理得到信道容量界成立，推导无记忆 Gauss 信道的容量。

[498] 一个信道由 r 个相互独立的无记忆 Gauss 信道组成，在第 i 个信道中噪声的方差为 v_i，$i=1,2,\cdots,n$。对于每一个 i，混合信道受限于一个总功率约束 $E(\sum_i x_i^2)\leqslant p$，其中 x_i 是在时间 i 上第 i 个信道的输入值。请确定混合信道的容量。

解答 第一部分请看 4.3 节。

如果在第 i 个信道上的功率减少到 p_i，我们会有容量

$$C^i=\frac{1}{2}\sum_i\log\left(1+\frac{p_i}{v_i}\right)$$

实际容量通过 $C=\max C^i$，受限于 $p_1,\cdots,p_r\geqslant 0$，$\sum_i p_i=p$ 被给出。因此，我们需要最大化 Lagrangian 乘子

$$\mathscr{L}=\frac{1}{2}\sum_i\log\left(1+\frac{p_i}{v_i}\right)-\lambda\sum_i p_i$$

满足

$$\frac{\partial}{\partial p_i}\mathscr{L}=\frac{1}{2}(v_i-p_i)^{-1}-\lambda,i=1,\cdots,r$$

最大值位于

$$p_i=\max\left(0,\frac{1}{2\lambda}-v_i\right)=\left(\frac{1}{2\lambda}-v_i\right)_+$$

为了调整约束，选择 $\lambda=\lambda^*$，其中 λ^* 由

$$\sum_i \left(\frac{1}{2\lambda^*}-v_i\right)_+ = p$$

决定。

因为 LHS 从 $+\infty$ 到 0 是单调减少的，所以 λ^* 的存在性与唯一性成立，因此

$$C=\frac{1}{2}\sum_i \log\left(\frac{1}{2\lambda^* v_i}\right) \qquad \square$$

问题 4.17 这里我们考虑在给定集合 $\mathcal{A}$(有限，可数，或者不可数) 上取值的随机变量，它们的分布由给定的参考测度 μ 的 PMF 来决定。令 Ψ 是一个实函数，β 是一个实数。证明受限于 $\mathbb{E}\Psi(X)=\beta$ 的熵 $h(X)=-\int f_X(x)\log f_X(x)\mu(\mathrm{d}x)$ 的最大值 $h^{\max}(X)$ 在随机变量 X^* 的 PMF 为

$$f_{X^*}(x)=\frac{1}{\Xi}\exp[-\gamma\Psi(x)] \tag{4.6.8a}$$

499

时是可达的。

其中 $\Xi=\Xi(\gamma)=\int \exp-\gamma\Psi(x)\mu(\mathrm{d}x)$ 为归一化常量，选择 γ 使得

$$\mathbb{E}\Psi(X^*)=\int \frac{\Psi(x)}{\Xi}\exp[-\gamma\Psi(x)]\mu(\mathrm{d}x)=\beta \tag{4.6.8b}$$

假设满足性质 $\int \frac{\Psi(x)}{\Xi}\exp[-\gamma\Psi(x)]\mu(\mathrm{d}x)=\beta$ 的 γ 存在。

另外，证明如果函数 Ψ 是非负的，那么对于任意给定的 $\beta>0$，式(4.6.8a)和式(4.6.8b)的 PMF f_{X^*} 可以使熵 $h(X)$ 在宽泛的约束 $\mathbb{E}\Psi(X)\leqslant\beta$ 下取得最大值。

因此，在下面的情况下，计算受限于 $\mathbb{E}\Psi(X)\leqslant\beta$ 条件，$h(X)$ 的最大值：(i)当 $\mathcal{A}$ 是一个有限集合，μ 是 $\mathcal{A}$ 上的一个正测度(满足 $\mu_i=\mu(\{i\})=1/\mu(\mathcal{A})$，其中 $\mu(\mathcal{A})=\sum_{i\in\mathcal{A}}\mu_i$) 并且 $\Psi(x)=1$，$x\in\mathcal{A}$；(ii)当 $\mathcal{A}$ 是一个任意集合时，μ 是 $\mathcal{A}$ 上的一个正测度，满足 $\mu(\mathcal{A})<\infty$，$\Psi(x)=1$，$x\in\mathcal{A}$；(iii)当 $\mathcal{A}=\mathbb{R}$ 是一个实直线时，μ 是 Lebesgue 测度，$\Psi(x)=|x|$；(iv)当 $\mathcal{A}=\mathbb{R}^d$，μ 是一个 d 维的 Lebesgue 测度，$\Psi(x)=\sum_{i,j\leqslant d}K_{ij}x_ix_j$，其中 $K=(K_{ij})$ 是一个 $d\times d$ 的正定实矩阵。

解答 关于 $\ln f_{X^*}(x)=-\gamma\Psi(x)-\ln\Xi$，我们使用 Gibbs 不等式

$$\begin{aligned}h(X)&=-\int f_X(x)\ln f_X(x)\mu(\mathrm{d}x)\leqslant\int f_X(x)[\gamma\Psi(x)+\ln\Xi]\mu(\mathrm{d}x)\\&=\int f_{X^*}(x)[\gamma\Psi(x)+\ln\Xi]\mu(\mathrm{d}x)=h(X^*)\end{aligned}$$

当且仅当 $X\sim X^*$ 等号成立。第一个证明完成。

如果 $\Psi\geqslant0$，期望值 $\mathbb{E}\Psi(X)\geqslant0$，并且 γ 为满足约束的最小值。 $\square$ 500

参 考 文 献

[1] V. Anantharam, F. Baccelli. A Palm theory approach to error exponents. In *Proceedings of the 2008 IEEE Symposium on Information Theory*, Toronto, pp. 1768–1772, 2008.

[2] J. Adámek. *Foundations of Coding: Theory and Applications of Error-Correcting Codes, with an Introduction to Cryptography and Information Theory*. Chichester: Wiley, 1991.

[3] D. Applebaum. *Probability and Information: An Integrated Approach*. Cambridge: Cambridge University Press, 1996.

[4] R.B. Ash. *Information Theory*. New York: Interscience, 1965.

[5] E.F. Assmus, Jr., J.D. Key. *Designs and their Codes*. Cambridge: Cambridge University Press, 1992.

[6] K.A. Arwini, C.T.J. Dodson. *Information Geometry: Near Randomness and Near Independence*. Lecture notes in mathematics, 1953. Berlin: Springer, 2008.

[7] D. Augot, M. Stepanov. A note on the generalisation of the Guruswami–Sudan list decoding algorithm to Reed–Muller codes. In *Gröbner Bases, Coding, and Cryptography*. RISC Book Series. Springer, Heidelberg, 2009.

[8] R.U. Ayres. *Manufacturing and Human Labor as Information Processes*. Laxenburg: International Institute for Applied System Analysis, 1987.

[9] A.V. Balakrishnan. *Communication Theory* (with contributions by J.W. Carlyle et al.). New York: McGraw-Hill, 1968.

[10] J. Baylis. *Error-Correcting Codes: A Mathematical Introduction*. London: Chapman & Hall, 1998.

[11] A. Betten et al. *Error-Correcting Linear Codes Classification by Isometry and Applications*. Berlin: Springer, 2006.

[12] T. Berger. *Rate Distortion Theory: A Mathematical Basis for Data Compression*. Englewood Cliffs, NJ: Prentice-Hall, 1971.

[13] E.R. Berlekamp. *A Survey of Algebraic Coding Theory*. Wien: Springer, 1972.

[14] E.R. Berlekamp. *Algebraic Coding Theory*. New York: McGraw-Hill, 1968.

[15] J. Berstel, D. Perrin. *Theory of Codes*. Orlando, FL: Academic Press, 1985.

[16] J. Bierbrauer. *Introduction to Coding Theory*. Boca Raton, FL: Chapman & Hall/CRC, 2005.

[17] P. Billingsley. *Ergodic Theory and Information*. New York: Wiley, 1965.

[18] R.E. Blahut. *Principles and Practice of Information Theory*. Reading, MA: Addison-Wesley, 1987.

[19] R.E. Blahut. *Theory and Practice of Error Control Codes*. Reading, MA: Addison-Wesley, 1983. See also *Algebraic Codes for Data Transmission*. Cambridge: Cambridge University Press, 2003.

[20] R.E. Blahut. *Algebraic Codes on Lines, Planes, and Curves*. Cambridge: Cambridge University Press, 2008.

[21] I.F. Blake, R.C. Mullin. *The Mathematical Theory of Coding*. New York: Academic Press, 1975.

[22] I.F. Blake, R.C. Mullin. *An Introduction to Algebraic and Combinatorial Coding Theory*. New York: Academic Press, 1976.

[23] I.F. Blake (ed). *Algebraic Coding Theory: History and Development*. Stroudsburg, PA: Dowden, Hutchinson & Ross, 1973.

[24] N. Blachman. *Noise and its Effect on Communication*. New York: McGraw-Hill, 1966.

[25] R.C. Bose, D.K. Ray-Chaudhuri. On a class of errors, correcting binary group codes. *Information and Control*, **3**(1), 68–79, 1960.

[26] W. Bradley, Y.M. Suhov. The entropy of famous reals: some empirical results. *Random and Computational Dynamics*, **5**, 349–359, 1997.

[27] A.A. Bruen, M.A. Forcinito. *Cryptography, Information Theory, and Error-Correction: A Handbook for the 21st Century*. Hoboken, NJ: Wiley-Interscience, 2005.

[28] J.A. Buchmann. *Introduction to Cryptography*. New York: Springer-Verlag, 2002.

[29] P.J. Cameron, J.H. van Lint. *Designs, Graphs, Codes and their Links*. Cambridge: Cambridge University Press, 1991.

[30] J. Castiñeira Moreira, P.G. Farrell. *Essentials of Error-Control Coding*. Chichester: Wiley, 2006.

[31] W.G. Chambers. *Basics of Communications and Coding*. Oxford: Clarendon, 1985.

[32] G.J. Chaitin. *The Limits of Mathematics: A Course on Information Theory and the Limits of Formal Reasoning*. Singapore: Springer, 1998.

[33] G. Chaitin. *Information-Theoretic Incompleteness*. Singapore: World Scientific, 1992.

[34] G. Chaitin. *Algorithmic Information Theory*. Cambridge: Cambridge University Press, 1987.

[35] F. Conway, J. Siegelman. *Dark Hero of the Information Age: In Search of Norbert Wiener, the Father of Cybernetics*. New York: Basic Books, 2005.

[36] T.M. Cover, J.M. Thomas. *Elements of Information Theory*. New York: Wiley, 2006.

[37] I. Csiszár, J. Körner. *Information Theory: Coding Theorems for Discrete Memoryless Systems*. New York: Academic Press, 1981; Budapest: Akadémiai Kiadó, 1981.

[38] W.B. Davenport, W.L. Root. *Random Signals and Noise*. New York: McGraw Hill, 1958.

[39] A. Dembo, T. M. Cover, J. A. Thomas. Information theoretic inequalities. *IEEE Transactions on Information Theory*, **37**, (6), 1501–1518, 1991.

[40] R.L. Dobrushin. Taking the limit of the argument of entropy and information functions. *Teoriya Veroyatn. Primen.*, **5**, (1), 29–37, 1960; English translation: *Theory of Probability and its Applications*, **5**, 25–32, 1960.

[41] F. Dyson. The Tragic Tale of a Genius. *New York Review of Books*, July 14, 2005.

[42] W. Ebeling. *Lattices and Codes: A Course Partially Based on Lectures by F. Hirzebruch*. Braunschweig/Wiesbaden: Vieweg, 1994.

[43] N. Elkies. Excellent codes from modular curves. *STOC'01: Proceedings of the 33rd Annual Symposium on Theory of Computing* (Hersonissos, Crete, Greece), pp. 200–208, NY: ACM, 2001.

[44] S. Engelberg. *Random Signals and Noise: A Mathematical Introduction*. Boca Raton, FL: CRC/Taylor & Francis, 2007.

[45] R.M. Fano. *Transmission of Information: A Statistical Theory of Communication*. New York: Wiley, 1961.

[46] A. Feinstein. *Foundations of Information Theory*. New York: McGraw-Hill, 1958.

[47] G.D. Forney. *Concatenated Codes*. Cambridge, MA: MIT Press, 1966.

[48] M. Franceschetti, R. Meester. *Random Networks for Communication. From Statistical Physics to Information Science*. Cambridge: Cambridge University Press, 2007.

[49] R. Gallager. *Information Theory and Reliable Communications*. New York: Wiley, 1968.

[50] A. Gofman, M. Kelbert, Un upper bound for Kullback–Leibler divergence with a small number of outliers. *Mathematical Communications*, **18**, (1), 75–78, 2013.

[51] S. Goldman. *Information Theory*. Englewood Cliffs, NJ: Prentice-Hall, 1953.

[52] C.M. Goldie, R.G.E. Pinch. *Communication Theory*. Cambridge: Cambridge University Press, 1991.

[53] O. Goldreich. *Foundations of Cryptography*, Vols 1, 2. Cambridge: Cambridge University Press, 2001, 2004.

[54] V.D. Goppa. *Geometry and Codes*. Dordrecht: Kluwer, 1988.

[55] S. Gravano. *Introduction to Error Control Codes*. Oxford: Oxford University Press, 2001.

[56] R.M. Gray. *Source Coding Theory*. Boston: Kluwer, 1990.

[57] R.M. Gray. *Entropy and Information Theory*. New York: Springer-Verlag, 1990.

[58] R.M. Gray, L.D. Davisson (eds). *Ergodic and Information Theory*. Stroudsburg, CA: Dowden, Hutchinson & Ross, 1977.

[59] V. Guruswami, M. Sudan. Improved decoding of Reed–Solomon codes and algebraic geometry codes. *IEEE Trans. Inform. Theory*, **45**, (6), 1757–1767, 1999.

[60] R.W. Hamming. *Coding and Information Theory*. 2nd ed. Englewood Cliffs, NJ: Prentice-Hall, 1986.

[61] T.S. Han. *Information-Spectrum Methods in Information Theory*. New York: Springer-Verlag, 2002.

[62] D.R. Hankerson, G.A. Harris, P.D. Johnson, Jr. *Introduction to Information Theory and Data Compression*. 2nd ed. Boca Raton, FL: Chapman & Hall/CRC, 2003.

[63] D.R. Hankerson et al. *Coding Theory and Cryptography: The Essentials*. 2nd ed. New York: M. Dekker, 2000. (Earlier version: D. G. Hoffman et al. *Coding Theory: The Essentials*. New York: M. Dekker, 1991.)

[64] W.E. Hartnett. *Foundations of Coding Theory*. Dordrecht: Reidel, 1974.

[65] S.J. Heims. *John von Neumann and Norbert Wiener: From Mathematics to the Technologies of Life and Death*. Cambridge, MA: MIT Press, 1980.

[66] C. Helstrom. *Statistical Theory of Signal Detection*. 2nd ed. Oxford: Pergamon Press, 1968.

[67] C.W. Helstrom. *Elements of Signal Detection and Estimation*. Englewood Cliffs, NJ: Prentice-Hall, 1995.

[68] R. Hill. *A First Course in Coding Theory*. Oxford: Oxford University Press, 1986.

[69] T. Ho, D.S. Lun. *Network Coding: An Introduction*. Cambridge: Cambridge University Press, 2008.

[70] A. Hocquenghem. Codes correcteurs d'erreurs. *Chiffres*, **2**, 147–156, 1959.

[71] W.C. Huffman, V. Pless. *Fundamentals of Error-Correcting Codes*. Cambridge: Cambridge University Press, 2003.

[72] J.F. Humphreys, M.Y. Prest. *Numbers, Groups, and Codes*. 2nd ed. Cambridge: Cambridge University Press, 2004.

[73] S. Ihara. *Information Theory for Continuous Systems*. Singapore: World Scientific, 1993.

[74] F.M. Ingels. *Information and Coding Theory*. Scranton: Intext Educational Publishers, 1971.

[75] I.M. James. *Remarkable Mathematicians. From Euler to von Neumann*. Cambridge: Cambridge University Press, 2009.

[76] E.T. Jaynes. *Papers on Probability, Statistics and Statistical Physics*. Dordrecht: Reidel, 1982.

[77] F. Jelinek. *Probabilistic Information Theory*. New York: McGraw-Hill, 1968.

[78] G.A. Jones, J.M. Jones. *Information and Coding Theory*. London: Springer, 2000.

[79] D.S. Jones. *Elementary Information Theory*. Oxford: Clarendon Press, 1979.

[80] O. Johnson. *Information Theory and the Central Limit Theorem*. London: Imperial College Press, 2004.

[81] J. Justensen. A class of constructive asymptotically good algebraic codes. *IEEE Transactions Information Theory*, **18**(5), 652–656, 1972.

[82] M. Kelbert, Y. Suhov. Continuity of mutual entropy in the large signal-to-noise ratio limit. In *Stochastic Analysis 2010*, pp. 281–299, 2010. Berlin: Springer.

[83] N. Khalatnikov. *Dau, Centaurus and Others*. Moscow: Fizmatlit, 2007.

[84] A.Y. Khintchin. *Mathematical Foundations of Information Theory*. New York: Dover, 1957.

[85] T. Klove. *Codes for Error Detection*. Singapore: World Scientific, 2007.

[86] N. Koblitz. *A Course in Number Theory and Cryptography*. New York: Springer, 1993.

[87] H. Krishna. *Computational Complexity of Bilinear Forms: Algebraic Coding Theory and Applications of Digital Communication Systems*. Lecture notes in control and information sciences, Vol. 94. Berlin: Springer-Verlag, 1987.

[88] S. Kullback. *Information Theory and Statistics*. New York: Wiley, 1959.

[89] S. Kullback, J.C. Keegel, J.H. Kullback. *Topics in Statistical Information Theory*. Berlin: Springer, 1987.

[90] H.J. Landau, H.O. Pollak. Prolate spheroidal wave functions, Fourier analysis and uncertainty, II. *Bell System Technical Journal*, 64–84, 1961.

[91] H.J. Landau, H.O. Pollak. Prolate spheroidal wave functions, Fourier analysis and uncertainty, III. The dimension of the space of essentially time- and band-limited signals. *Bell System Technical Journal*, 1295–1336, 1962.

[92] R. Lidl, H. Niederreiter. *Finite Fields*. Cambridge: Cambridge University Press, 1997.

[93] R. Lidl, G. Pilz. *Applied Abstract Algebra*. 2nd ed. New York: Wiley, 1999.

[94] E.H. Lieb. Proof of entropy conjecture of Wehrl. *Commun. Math. Phys.*, **62**, (1), 35–41, 1978.

[95] S. Lin. *An Introduction to Error-Correcting Codes*. Englewood Cliffs, NJ; London: Prentice-Hall, 1970.

[96] S. Lin, D.J. Costello. *Error Control Coding: Fundamentals and Applications*. Englewood Cliffs, NJ: Prentice-Hall, 1983.

[97] S. Ling, C. Xing. *Coding Theory*. Cambridge: Cambridge University Press, 2004.

[98] J.H. van Lint. *Introduction to Coding Theory*. 3rd ed. Berlin: Springer, 1999.

[99] J.H. van Lint, G. van der Geer. *Introduction to Coding Theory and Algebraic Geometry*. Basel: Birkhäuser, 1988.

[100] J.C.A. van der Lubbe. *Information Theory*. Cambridge: Cambridge University Press, 1997.

[101] R.E. Lewand. *Cryptological Mathematics*. Washington, DC: Mathematical Association of America, 2000.

[102] J.A. Llewellyn. *Information and Coding*. Bromley: Chartwell-Bratt; Lund: Studentlitteratur, 1987.

[103] M. Loève. *Probability Theory*. Princeton, NJ: van Nostrand, 1955.

[104] D.G. Luenberger. *Information Science*. Princeton, NJ: Princeton University Press, 2006.

[105] D.J.C. Mackay. *Information Theory, Inference and Learning Algorithms*. Cambridge: Cambridge University Press, 2003.

[106] H.B. Mann (ed). *Error-Correcting Codes*. New York: Wiley, 1969.

[107] M. Marcus. Dark Hero of the Information Age: In Search of Norbert Wiener, the Father of Cybernetics. *Notices of the AMS* **53**, (5), 574–579, 2005.

[108] A. Marshall, I. Olkin. *Inequalities: Theory of Majorization and its Applications*. New York: Academic Press, 1979.

[109] V.P. Maslov, A.S. Chernyi. On the minimization and maximization of entropy in various disciplines. *Theory Probab. Appl.* **48**, (3), 447–464, 2004.

[110] F.J. MacWilliams, N.J.A. Sloane. *The Theory of Error-Correcting Codes*, Vols I, II. Amsterdam: North-Holland, 1977.

[111] R.J. McEliece. *The Theory of Information and Coding*. Reading, MA: Addison-Wesley, 1977. 2nd ed. Cambridge: Cambridge University Press, 2002.

[112] R. McEliece. *The Theory of Information and Coding*. Student ed. Cambridge: Cambridge University Press, 2004.

[113] A. Menon, R.M. Buecher, J.H. Read. Impact of exclusion region and spreading in spectrum-sharing ad hoc networks. ACM 1-59593-510-X/06/08, 2006.

[114] R.A. Mollin. *RSA and Public-Key Cryptography*. New York: Chapman & Hall, 2003.

[115] R.H. Morelos-Zaragoza. *The Art of Error-Correcting Coding*. 2nd ed. Chichester: Wiley, 2006.

[116] G.L. Mullen, C. Mummert. *Finite Fields and Applications*. Providence, RI: American Mathematical Society, 2007.

[117] A. Myasnikov, V. Shpilrain, A. Ushakov. *Group-Based Cryptography*. Basel: Birkhäuser, 2008.

[118] G. Nebe, E.M. Rains, N.J.A. Sloane. *Self-Dual Codes and Invariant Theory*. New York: Springer, 2006.

[119] H. Niederreiter, C. Xing. *Rational Points on Curves over Finite Fields: Theory and Applications*. Cambridge: Cambridge University Press, 2001.

[120] W.W. Peterson, E.J. Weldon. *Error-Correcting Codes*. 2nd ed. Cambridge, MA: MIT Press, 1972. (Previous ed. W.W. Peterson. *Error-Correcting Codes*. Cambridge, MA: MIT Press, 1961.)

[121] M.S. Pinsker. *Information and Information Stability of Random Variables and Processes*. San Francisco: Holden-Day, 1964.

[122] V. Pless. *Introduction to the Theory of Error-Correcting Codes*. 2nd ed. New York: Wiley, 1989.

[123] V.S. Pless, W.C. Huffman (eds). *Handbook of Coding Theory*, Vols 1, 2. Amsterdam: Elsevier, 1998.

[124] P. Piret. *Convolutional Codes: An Algebraic Approach*. Cambridge, MA: MIT Press, 1988.

[125] O. Pretzel. *Error-Correcting Codes and Finite Fields*. Oxford: Clarendon Press, 1992; Student ed. 1996.

[126] T.R.N. Rao. *Error Coding for Arithmetic Processors*. New York: Academic Press, 1974.

[127] M. Reed, B. Simon. *Methods of Modern Mathematical Physics*, Vol. II. Fourier analysis, self-adjointness. New York: Academic Press, 1975.

[128] A. Rényi. *A Diary on Information Theory*. Chichester: Wiley, 1987; initially published Budapest: Akad'emiai Kiadó, 1984.

[129] F.M. Reza. *An Introduction to Information Theory*. New York: Constable, 1994.

[130] S. Roman. *Coding and Information Theory*. New York: Springer, 1992.

[131] S. Roman. *Field Theory*. 2nd ed. New York: Springer, 2006.

[132] T. Richardson, R. Urbanke. *Modern Coding Theory*. Cambridge: Cambridge University Press, 2008.

[133] R.M. Roth. *Introduction to Coding Theory*. Cambridge: Cambridge University Press, 2006.

[134] B. Ryabko, A. Fionov. *Basics of Contemporary Cryptography for IT Practitioners*. Singapore: World Scientific, 2005.

[135] W.E. Ryan, S. Lin. *Channel Codes: Classical and Modern*. Cambridge: Cambridge University Press, 2009.

[136] T. Schürmann, P. Grassberger. Entropy estimation of symbol sequences. *Chaos*, **6**, (3), 414–427, 1996.

[137] P. Seibt. *Algorithmic Information Theory: Mathematics of Digital Information Processing*. Berlin: Springer, 2006.

[138] C.E. Shannon. A mathematical theory of cryptography. *Bell Lab. Tech. Memo.*, 1945.

[139] C.E. Shannon. A mathematical theory of communication. *Bell System Technical Journal*, **27**, July, October, 379–423, 623–658, 1948.

[140] C.E. Shannon: *Collected Papers*. N.J.A. Sloane, A.D. Wyner (eds). New York: IEEE Press, 1993.

[141] C.E. Shannon, W. Weaver. *The Mathematical Theory of Communication*. Urbana, IL: University of Illinois Press, 1949.

[142] P.C. Shields. *The Ergodic Theory of Discrete Sample Paths*. Providence, RI: American Mathematical Society, 1996.

[143] M.S. Shrikhande, S.S. Sane. *Quasi-Symmetric Designs*. Cambridge: Cambridge University Press, 1991.

[144] S. Simic. Best possible global bounds for Jensen functionals. *Proc. AMS*, **138**, (7), 2457–2462, 2010.

[145] A. Sinkov. *Elementary Cryptanalysis: A Mathematical Approach*. 2nd ed. revised and updated by T. Feil. Washington, DC: Mathematical Association of America, 2009.

[146] D. Slepian, H.O. Pollak. Prolate spheroidal wave functions, Fourier analysis and uncertainty, Vol. I. *Bell System Technical Journal*, 43–64, 1961 .

[147] W. Stallings. *Cryptography and Network Security: Principles and Practice*. 5th ed. Boston, MA: Prentice Hall; London: Pearson Education, 2011.

[148] H. Stichtenoth. *Algebraic Function Fields and Codes*. Berlin: Springer, 1993.

[149] D.R. Stinson. *Cryptography: Theory and Practice*. 2nd ed. Boca Raton, FL; London: Chapman & Hall/CRC, 2002.

[150] D. Stoyan, W.S. Kendall. J. Mecke. *Stochastic Geometry and its Applications*. Berlin: Academie-Verlag, 1987 .

[151] C. Schlegel, L. Perez. *Trellis and Turbo Coding*. New York: Wiley, 2004.

[152] Š. Šujan. *Ergodic Theory, Entropy and Coding Problems of Information Theory*. Praha: Academia, 1983.

[153] P. Sweeney. *Error Control Coding: An Introduction*. New York: Prentice Hall, 1991.

[154] Te Sun Han, K. Kobayashi. *Mathematics of Information and Coding*. Providence, RI: American Mathematical Society, 2002.

[155] T.M. Thompson. *From Error-Correcting Codes through Sphere Packings to Simple Groups*. Washington, DC: Mathematical Association of America, 1983.

[156] R. Togneri, C.J.S. deSilva. *Fundamentals of Information Theory and Coding Design*. Boca Raton, FL: Chapman & Hall/CRC, 2002.

[157] W. Trappe, L.C. Washington. *Introduction to Cryptography: With Coding Theory*. 2nd ed. Upper Saddle River, NJ: Pearson Prentice Hall, 2006.

[158] M.A. Tsfasman, S.G. Vlăduţ. *Algebraic-Geometric Codes*. Dordrecht: Kluwer Academic, 1991.

[159] M. Tsfasman, S. Vlăduţ, T. Zink. Modular curves, Shimura curves and Goppa codes, better than Varshamov–Gilbert bound. *Mathematics Nachrichten*, **109**, 21–28, 1982.

[160] M. Tsfasman, S. Vlăduţ, D. Nogin. *Algebraic Geometric Codes: Basic Notions*. Providence, RI: American Mathematical Society, 2007.

[161] M.J. Usher. *Information Theory for Information Technologists*. London: Macmillan, 1984.

[162] M.J. Usher, C.G. Guy. *Information and Communication for Engineers*. Basingstoke: Macmillan, 1997

[163] I. Vajda. *Theory of Statistical Inference and Information*. Dordrecht: Kluwer, 1989.

[164] S. Verdú. *Multiuser Detection*. New York: Cambridge University Press, 1998.

[165] S. Verdú, D. Guo. A simple proof of the entropy–power inequality. *IEEE Trans. Inform. Theory*, **52**, (5), 2165–2166, 2006.

[166] L.R. Vermani. *Elements of Algebraic Coding Theory*. London: Chapman & Hall, 1996.

[167] B. Vucetic, J. Yuan. *Turbo Codes: Principles and Applications*. Norwell, MA: Kluwer, 2000.

[168] G. Wade. *Coding Techniques: An Introduction to Compression and Error Control*. Basingstoke: Palgrave, 2000.

[169] J.L. Walker. *Codes and Curves*. Providence, RI: American Mathematical Society, 2000.

[170] D. Welsh. *Codes and Cryptography*. Oxford, Oxford University Press, 1988.

[171] N. Wiener. *Cybernetics or Control and Communication in Animal and Machine*. Cambridge, MA: MIT Press, 1948; 2nd ed. 1961, 1962.

[172] J. Wolfowitz. *Coding Theorems of Information Theory*. Berlin: Springer, 1961; 3rd ed: 1978.

[173] A.D. Wyner. The capacity of the band-limited Gaussian channel. *Bell System Technical Journal*, 359–395, 1996 .

[174] A.D. Wyner. The capacity of the product of channels. *Information and Control*, 423–433, 1966.

[175] C. Xing. Nonlinear codes from algebraic curves beating the Tsfasman–Vladut–Zink bound. *IEEE Transactions Information Theory*, **49**, 1653–1657, 2003.

[176] A.M. Yaglom, I.M. Yaglom. *Probability and Information*. Dordrecht, Holland: Reidel, 1983.

[177] R. Yeung. *A First Course in Information Theory*. Boston: Kluwer Academic, 1992; 2nd ed. New York: Kluwer, 2002.

索　引

索引中的页码为英文原书页码，与书中页边标注的页码一致。

A

additive stream cipher(加法流密码)，463
algebra (a commutative ring and a linear space)(代数，即交换环和线性空间的集合)，318
  group algebra(群代数)，317
  polynomial algebra(多项式代数)，214
  σ-algebra(σ代数)，440
algebraic-geometric code(代数几何码)，348
algorithm(算法)，9
  Berlekamp-Massey (BM) decoding algorithm for BCH codes(BCH 码的 Berlekamp-Massey 解码算法)，240
  Berlekamp-Massey (BM) algorithm for solving linear equations(线性方程的 Berlekamp-Massey 算法)，460
  division algorithm for polynomials(多项式的分割算法)，214
  Euclid algorithm for integers(整数的欧氏算法)，473
  extended Euclid algorithm for integers(整数的扩展欧氏算法)，479
  Euclid algorithm for polynomials(多项式的欧氏算法)，242
  Guruswami-Sudan (GS) decoding algorithm for Reed-Solomon codes(Reed-Solomon 码的 Guruswami-Sudan 解码算法)，298
  Huffman encoding algorithm (Huffman 编码算法)，9
Alphabet(符号集)，3
  source alphabet(源符号集)，8
  coder (encoding) alphabet(编码(解码)符号集)，3
  channel input alphabet(信道输入符号集)，60
  channel output alphabet(信道输出符号集)，65
asymptotic equipartition property(渐近均分性)，44
asymptotically good sequence of codes(渐近优序列码)，78
automorphism(自同构)，283

B

band-limited signal(带限信号)，411
bandwidth(带宽)，409
basis(基础)，149，186
BCH (Bose-Ray-Chaudhuri-Hocquenghem) bound, or BCH theorem(BCH 界或 BCH 理论)，237，295
  BCH code(BCH 码)，213
  BCH code in a narrow sense(狭义的 BCH 码)，235
  binary BCH code in a narrow sense(狭义的二进制 BCH 码)，235
Bernoulli source(Bernoulli 信源)，3
bit (a unit of entropy)(比特，熵的单位)，9
bit commitment cryptosystem(比特交付密码)，468
bound(界)
  BCH bound(BCH 界)，237，295
  Elias bound(Elias 界)，177
  Gilbert bound(Gilbert 界)，198
  Gilbert-Varshamov bound (Gilbert-Varshamov 界)，154
  Griesmer bound(Griesmer 界)，197
  Hamming bound(Hamming 界)，150
  Johnson bound(Johnson 界)，177
  linear programming bound(线性规划界)，322
  Plotkin bound(Plotkin 界)，155
Singleton bound(Singleton 界)，154
bar-product(条乘积)，152

C

capacity(容量)，61
  capacity of a discrete channel(离散信道的容量)，61
  capacity of a memoryless Gaussian channel with white noise(无记忆 Gauss 白噪声信道的容量)，374

capacity of a memoryless Gaussian channel with coloured noise(无记忆 Gauss 有色噪声信道的容量), 375
operational channel capacity(可达信道容量), 102
character (as a digit or a letter or a symbol)(符号, 例如数字, 字母或者信号), 53
character (of a homomorphism)(同态征), 313
modular character(模投征), 314
trivial, or principal, character(单位特征或重征), 313
character transform(符号变换), 319
characteristic of a field(域的特征), 269
channel(信道), 60
additive Gaussian channel (AGC, 加性 Gauss 信道), 368
memoryless Gaussian channel (MGC, 无记忆 Gauss 信道), 368
memoryless additive Gaussian channel (MAGC, 无记忆加性 Gauss 信道), 366
memoryless binary channel (MBC, 无记忆二元信道), 60
memoryless binary symmetric channel (MBSC, 无记忆二元对称信道), 60
noiseless channel(有噪信道), 103
channel capacity(信道容量), 61
operational channel capacity(可达信道容量), 102
check matrix, 参见 parity-check matrix
cipher (or a cryptosystem)(校验矩阵密码(或密码系统)), 463
one-time pad cipher(单次加密密码), 466
public-key cipher(公钥密码), 467
ciphertext(密文), 468
code, or encoding(编码), 4
alternant code(交替码), 332
BCH code(BCH 码), 213
binary code(二元码), 10, 95
cardinality of a code (码的基数), 253
cyclic code(循环码), 216
decipherable code(可译码), 14
dimension of a linear code(线性码维度), 149
D error detecting code(D 检错码), 147
dual code(对偶码), 153
equivalent codes(等效码), 190
E error correcting code(E 纠错码), 147
Golay code(Golay 码), 151
Goppa code(Goppa 码), 160, 336
Hamming code(Hamming 码), 199
Huffman code(Huffman 码), 9
information rate of a code(码信息率), 147
Justesen code(Justesen 码), 248, 332
lossless code(无损码), 4
linear code(线性码), 148
maximal distance separating (MDS, 最大距离分离 MDS), 155
parity-check code(奇偶校验码), 149
perfect code(完美码), 151
prefix-free code(无前缀编码), 4
random code(随机码), 58, 372
rank of a linear code(线性码的秩), 184
Reed-Muller (RM) code (Reed-Muller (RM) 码), 203
Reed-Solomon code(Reed-Solomon 码), 256, 291
repetition code(重复码), 149
reversible cyclic code(自反循环码), 230
self-dual code(自对偶码), 201, 227
self-orthogonal(自正交), 227
symplex code(对称码), 194
codebook(码本), 67
random codebook(随机码本), 68
coder, or encoder(编码器), 3
codeword(码字), 4
random codeword(随机码字), 6
coding, 参见 encoding
coloured noise(有色噪声), 374
concave(凹), 19, 32
strictly concave(严格凹), 32
concavity(凹性), 20
conditionally independent(条件独立), 26
conjugacy(共轭性), 281
conjugate(共轭的), 229
convergence almost surely (a.s., 全概率收敛), 131
convergence in probability(概率收敛), 43
convex(凸), 32
strictly convex(严格凸), 104
convexity(凸性), 142
core polynomial of a field(域的核心多项式), 231
coset(陪集), 192
cyclotomic coset(分圆陪集), 285
leader of a coset(陪集的领), 192
cryptosystem (or a cipher)(密码系统或密码), 468
bit commitment cryptosystem(比特承诺密码系统), 468

ElGamal cryptosystem(ElGamal 密码系统), 475
public key cryptosystem(公钥密码), 468
RSA (Rivest-Shamir-Adelman) cryptosystem (RSA 密码系统), 468
Rabin, or Rabin-Williams cryptosystem(Rabin, 或 Rabin-Williams 密码系统), 473
cyclic group(循环群), 231
generator of a cyclic group(循环群生成子)
cyclic shift(循环移位), 216

D

data-processing inequality(数据处理不等性), 80
detailed balance equations (DBE, 详细平衡方程), 56
decoder, or a decoding rule(解码规则), 65
geometric (or minimal distance) decoding rule(几何(或最小距离)解码规则), 163
ideal observer (IO) decoding rule(理想观测者(IO)解码规则), 65
maximum likelihood (ML) decoding rule(最大似然解码规则), 66
joint typicality (JT) decoder(联合典型性译码器), 372
decoding(解码), 167
decoding alternant codes(解交替码), 337
decoding BCH codes(解 BCH 码), 239, 310
decoding cyclic codes(解循环码), 214
decoding Hamming codes(解 Hamming 码), 230
list decoding(列表译码), 192, 405
decoding Reed-Muller codes (解 Reed-Muller 码), 209
decoding Reed-Solomon codes (解 Reed-Muller 码), 292
decoding Reed-Solomon codes by the Guruswami-Sodan algorithm(通过 Guruswami-Sodan 算法解码 Reed-Solomon 码), 269
syndrome decoding(伴随式译码), 193
decrypt function(解密函数), 469
degree of a polynomial(多项式的次数), 208, 214
density of a probability distribution (PDF, 概率分布密度), 86
differential entropy(微分熵), 86
digit(数字), 3
dimension(维度), 149
dimension of a code(码的维度), 149
dimension of a linear representation (线性表度), 314
discrete Fourier transform (FFT, 离散 Fourier 变换), 296
discrete-time Markov chain (DTMC, 离散时间 Markov 链), 1, 3
discrete logarithm(离散算法), 474
distributed system, or a network (of transmitters) ((发送端的)分布式系统或网络), 436
Dirac δ-function(Dirac δ-函数), 318
distance(距离), 20
Kullback-Leibler distance (Kullback-Leibler 距离), 20
Hamming distance(Hamming 距离), 144
minimal distance of a code(码最小距离), 147
distance enumerator polynomial(距离枚举多项式), 322
divisor(除数), 217
greatest common divisor (GCD, 最大公约数), 223
dot product(点积), 153
doubly stochastic (Cox) random process(双重随机(复)过程), 492
electronic signature(电子签名), 469, 475
encoding, or coding vii, (编码), 4
Huffman encoding Huffman(编码), 9
Shannon-Fano encoding (Shannon-Fano 编码), 9
random coding (随机编码), 67
entropy (熵), 7
axiomatic definition of entropy(熵的公理化定义), 36
binary entropy(二进制熵), 7
conditional entropy(条件熵), 20
differential entropy(微分熵), 86

E

entropy of a random variable(随机变量的熵), 18
entropy of a probability distribution(一个概率分布的熵), 18
joint entropy(联合熵), 20
mutual entropy(互信息), 28
entropy power inequality(熵功率不等式), 92
q-ary entropy (q 元熵), 7
entropy rate (熵率), vii, 41
relative entropy(相对熵), 20
encrypt function(加密函数), 468
ergodic random process (stationary) (遍历随机过

程(稳定)), 397
ergodic transformation of a probability space(一个概率空间的遍历变换), 397
error locator(错误定位器), 311
error locator polynomial(错误定位多项式), 239, 311
error-probability(错误概率), 58
extension of a code (码的扩展), 151
parity-check extension(奇偶检验扩展), 151
extension field(扩展域), 261

F

factor (as a divisor) (因子(作为除数)), 39
irreducible factor(不可约因子), 219
prime factor(主要因子), 39
factorization(分解), 239
fading of a signal(信号的衰落), 447
power fading (功率衰落), 447
Rayleigh fading (Rayleigh 衰落), 447
feedback shift register (反馈位移寄存器), 453
linear feedback shift register (LFSR, 线性反馈移位寄存器), 454
feedback polynomial (反馈多项式), 454
field (a commutative ring with inverses) (域(逆交换环)), 146, 230
extension field (扩展域), 261
Galois field Galois(域), 272
finite field (有限域), 194
polynomial field(多项式域), 231
primitive element of a field (一个域的基本元素), 230, 232
splitting field(分裂域), 236, 271
Frobenius map (映射), 283

G

Gaussian channel (Gauss 信道), 366
additive Gaussian channel (AGC, 加性 Gauss 信道), 368
memoryless Gaussian channel (MGC, 无记忆 Gauss 信道), 368
memoryless additive Gaussian channel (MAGC, 无记忆加性 Gauss 信道), 366
Gaussian coloured noise(Gauss 有色噪声), 374
Gaussian white noise(Gauss 白噪声), 368
Gaussian random process (Gauss 随机过程), 369
generating matrix (生成矩阵), 185
generator (of a cyclic code) (生成器(循环码的)), 218
minimal degree generator polynomial (最小生成多项式), 218
generator (of a cyclic group) (生成器(循环群的)), 232
geometric (or minimal distance) decoding rule(几何(或最小距离)解码规则), 163
group (群), 146
group algebra(群代数), 317
commutative, or Abelian, group (交换, 或者阿贝尔, 群), 146
cyclic group(循环群), 231
linear representation of a group (一个群的线性表示), 314
generalized function (一般性函数), 412
greatest common divisor (GCD), 最大公约数), 223

I

ideal observer (IO) decoding rule(理想观测者译码规则), 56
ideal of a ring (环的理想), 217
principal ideal (主要理想), 219
identity (for weight enumerator polynomials) (标识(加权枚举多项式)), 258
abstract MacWilliams identity (抽象 MacWilliams 标识), 315
MacWilliams identity for a linear code (一个线性码的 MacWilliams 标识), 258, 313
independent identically distributed (IID) random variables(独立同分布的随机变量), 1, 3
inequality(不等式), 4
Brunn-Minkovski inequality (Brunn-Minkovski 不等式), 93
Cauchy-Schwarz inequality (Cauchy-Schwarz 不等式), 124
Chebyshev inequality (Chebyshev 不等式), 128
data-processing inequality(数据处理不等式), 80
entropy power inequality(熵功率不等式), 92
Fano inequality (Fano 不等式), 25
generalized Fano inequality (Fano 生成不等式), 27
Gibbs inequality (Gibbs 不等式), 17
Hadamard inequality (Hadamard 不等式), 91
Kraft inequality (Kraft 不等式), 4
Ky-Fan inequality (Ky-Fan 不等式), 91

log sum inequality (log sum 不等式), 103
Markov inequality (Markov 不等式), 408
pooling inequalities (pooling 不等式), 24
information (信息), 2, 18
mutual information, or mutual entropy(互信息, 或互熵), 28
information rate(信息速率), 15
information source (random source) (信源(随机源)), 2, 44
Bernoulli information source (Bernoulli 信源), 3
Markov information source (Markov 信源), 3
information symbols (信息符号), 208
initial fill(初始填充), 454
intensity (of a random measure) (强度(随机测量)), 437
intensity measure(强度测量), 437

J

joint entropy (联合熵), 20
joint input/output distribution (of a channel) (联合输入/输出分布(一个信道的)), 67
joint typicality (JT) decoder(联合典型性(JT)译码器), 372

K

key (as a part of a cipher) (键(作为密码的一部分)), 466
decoding key (a label of a decoding, or decrypting, map) (解码密钥(解码或解密的标签映射)), 469
encoding key (a label of an encoding, or encrypting, map) (编码密钥(编码或加密的标签映射)), 468
random key of a one-pad cipher (one-pad 随机密钥), 456
private key (私钥), 470
public key (公钥), 469
secret key (密钥), 473
Karhunen-Loève decomposition (Karhunen-Loève 分解), 426

L

law of large numbers (大数定理), 34
strong law of large numbers (强大数定理), 438
leader of a coset (陪集的领), 192
least common multiple (lcm) (最小公倍数), 225
lemma (引理)
Borel-Cantelli lemma (Borel Cantel 引理), 418
Nyquist-Shannon-Kotelnikov-Whittaker lemma (Nyquist-Shannon-Kotelnikov-Whittaker 引理), 431
letter (字母), 2
linear code (线性码), 148
linear representation of a group (群的线性表示), 314
space of a linear representation (线性表示的空间域), 314
dimension of a linear representation (线性表示的维度), 314
linear space (线性空间), 146
linear subspace (线性子空间), 148
linear feedback shift register (LFSR, 线性反馈移位寄存器), 454
auxiliary, or feedback, polynomial of an LFSR (LFSR 的附属, 反馈, 多项式), 454

M

Markov chain (Markov 链), 1, 3
discrete-time Markov chain (DTMC, 离散时间 Markov 链), 1, 3
coupled Markov chain (耦合 Markov 链), 50
irreducible and aperiodic Markov chain (不可约和非周期的 Markov 链), 128
kth order Markov chain approximation (k 阶 Markov 链近似), 467
second-order Markov chain (二阶 Markov 链), 131
transition matrix of a Markov chain (Markov 链的转移矩阵), 3
Markov inequality(Markov 不等式), 408
Markov property (Markov 性), 33
strong Markov property (强 Markov 性), 50
Markov source (Markov 源), 3
stationary Markov source (静态 Markov 源), 3
Markov triple (Markov 三元组), 33
Matérn process (with a hard core) (Matern 过程), 451
first model of the Matérn process (Matern 过程的第一模型), 451
second model of the Matérn process (Matern 过程的第二模型), 451

matrix（矩阵），13
covariance matrix（方差矩阵），88
generating matrix（生成矩阵），185
generating check matrix, canonical, or standard, form of（……的生成校验矩阵，规范式，或标准，形式），189
parity check matrix（奇偶校验矩阵），186
parity-check matrix, canonical, or standard, form of（……的奇偶校验矩阵，规范式，或标准，形式），188
parity-check matrix of a Hamming code（Hamming 码的奇偶校验矩阵），191
positive definite matrix（正定矩阵），93
recursion matrix（递归矩阵），174
Töplitz matrix（Toplitz 矩阵），93
transition matrix of a Markov chain（Markov 链的转移矩阵），3
transition matrix, doubly stochastic（转移矩阵，重随机），34
Vandermonde matrix（Vandermonde 矩阵），295
maximum likelihood（ML）decoding rule（最大似然解码准则），66
measure（as a countably additive function of a set）（测度），366
intensity（or mean）measure（强度（或均值）测度），436
non atomic measure（非原子测度），436
Poisson random measure（Poisson 随测度），436
product-measure（乘积测度），371
random measure（随机测度），436
reference measure（标准测度），372
σ-finite（σ-有限），436
Möbius function（Möbius 方程），277
Möbius inversion formula（Möbius 逆公式），278
moment generating function（矩生成方程），442

N

network（网络），参见 distributed system
supercritical network（超临界网络），449
network information theory（网络信息论），436
noise（in a channel）（噪声在信道中），2，79
Gaussian coloured noise（Gauss 有色噪声），374
Gaussian white noise（Gauss 白噪声），368
noiseless channel（无噪信道），103
noisy（or fully noisy）channel（有噪（或全噪声）信道），81

O

one time pad cipher（一次一密密码），465
operational channel capacity（可行信道容量），102
order of an element（元素的阶），267
order of a polynomial（多项式的阶），231
orthogonal（正交），185
ortho-basis（正交基），430
orthogonal complement（正交补），185
orthoprojection（正交投影），375
self-orthogonal（自正交），227
output stream of a register（寄存器的输出流），454

P

parity check code（奇偶校验码），149
parity-check extension（奇偶校验扩展），151
parity-check matrix（奇偶校验矩阵），186
plaintext（明码文本），458
Poisson process（Poisson 过程），436
Poisson random measure（Poisson 随机测度），436
Polynomial（多项式），206
algebra, polynomial（代数，多项式），214
degree of a polynomial（多项式的度），206，214
distance enumerator polynomial（距离列举多项式），322
error locator polynomial（错误定位多项式），239
Goppa polynomial（Goppa 多项式），335
irreducible polynomial（不可约多项式），219
Mattson-Solomon polynomial（Mattson Solomon 多项式），296
minimal polynomial（最小多项式），236
order of a polynomial（多项式的阶），231
reducible polynomial（可约多项式），221
primitive polynomial（原始多项式），230，257
Kravchuk polynomial（Kravchuk 多项式），320
weight enumerator polynomial（权重列举多项式），319，351
probability distribution（概率方程），1
conditional probability（条件概率），1
probability density function（PDF，概率密度方程），86
equiprobable, or uniform, distribution（等概率或均匀分布），3，22
exponential distribution（with exponential densi-

ty)(指数分布(有指数密度)), 89
geometric distribution (几何分布), 21
joint probability (联合概率), 1
multivariate normal distribution (多元正态分布), 88
normal distribution (with univariate normal density)(正态分布(有单变量正态密度)), 89
poisson distribution (Poisson 分布), 101
probability mass function (PMF, 概率质量密度), 366
probability space (概率空间), 397
prolate spheroidal wave function(PSWF, 椭球波函数), 425
protocol of a private communication (私密通信协议), 469
diffie-hellman protocol(diffie-hellman 协议), 474
prefix (前缀), 4
prefix-free code (无前缀编码), 4
product-channel (乘积信道), 404
public-key cipher (公钥算法), 467

Q

quantum mechanics (量子力学), 431

R

random code (随机码), 68, 372
random codebook (随机码本), 68
random codeword (随机码字), 6
poisson random measure (PRM, Poisson 随机测度), 436
random process (随机过程)
gaussian random process (Gauss 随机过程), 369
poisson random process (Poisson 随机过程), 435
stationary random process (平稳随机过程), 397
stationary ergodic random process (平稳各态历经过程), 397
random variable (随机变量), 18
conditionally independent random variables (条件独立随机变量), 26
equiprobable, or uniform, random variable (均匀随机变量), 3, 23
exponential random variable (with exponential density) (指数随机变量), 89
geometric random variable (几何随机变量), 21
independent identically distributed (IID) random variables (独立同分布随机变量), 1, 3
joint probability distribution of random variables (联合概率分布的随机变量), 1
normal random variable (with univariate normal density) (正态随机变量有单一正态密度), 89
poisson random variable (Poisson 随机变量), 101
random vector (随机向量), 20
multivariate normal random vector (多元正态随机向量), 88
rank of a code (码的秩), 184
rank-nullity property (秩零化性质), 186
rate (速率), 15
entropy rate (熵率), 41
information rate of a source (信源的信息速率), 15
reliable encoding (or encodable) rate (可靠编码速率), 15
reliable transmission rate (可靠传输速率), 62
reliable transmission rate with regional constraint (带有区域约束的可靠传输速率), 373
regional constraint for channel capacity (信道容量的区域约束), 367
register (寄存器), 453
feedback shift register (反馈移位寄存器), 453
linear feedback shift register (LFSR, 线性反馈移位寄存器), 454
feedback, or auxiliary, polynomial of an LFSR (线性反馈移位寄存器的反馈多项式), 454
initial fill of register (寄存器的初值填充), 454
output stream of a register (寄存器的输出流), 454
repetition code (重复码), 149
repetition of a code (码的重复), 152
ring (环), 217
ideal of a ring (环的理想), 217
quotient ring (分式环), 274
root of a cyclic code (循环码的根), 233
defining root of a cyclic code (定义循环码的根), 233
root of apolynomial (多项式的根), 228
root of unity (单位根), 228
primitive root of unity (本原单位根), 236

S

sample(取样), 2

signal/noise ratio(SNR，信噪比)，449
sinc function (sinc 函数)，413
size of a code (码的大小)，147
space (空间)，35
hamming space (汉明空间)，144
space(空) (空间)，415
linear space (线性空间)，146
linear subspace (线性子空间)，168
space of a linear representation (空间的线性表示)，514
state space of a Markov chain (Markov 链的状态空间)，35
vector space over a field (域上的向量空间)，269
stream (流)，453
strictly concave (严格凹的)，32
strictly convex (严格凸的)，104
string, or a word(of characters, digits, letters or symbols) ((字符，位，字母符号)的字符串或字)，2
source of information (random) (信源(随机))，2，44
bernoulli source (Bernoulli 源)，3
equiprobable Bernoulli source (等概率 Bernoulli 源)，3
markov source (Markov 源)，3
stationary markov source (平稳 Markov 源)，3
spectral density (频谱密度)，417
stationary (平稳的)，3
stationary markov source (平稳 Markov 源)，3
stationary random process (平稳随机过程)，398
stationary ergodic random process (平稳各态历经随机过程)，398
superer tiesl network(超级界网络)，449
symbol (符号)，2
syndrome (校验子)，192

T

theorem(定理)
Brunn-Minkowski theorem (Brunn-Minkowski 定理)，93
Campbell theorem (Campbell 定理)，442
Cayley-Hamilton theorem (Cayley-Hamilton 定理)，456
Central limit theorem(CLT，中心极限定理)，94
Doob Lévy theorem (Doob-Levy 定理)，439
local De Moivre-Laplace theorem (本地 De Moivre-Laplace 定理)，53
mapping theorem (映射定理)，437
theorem (cont.)(定理)
product theorem(乘积定理)，464
Shannon theorem(Shannon 定理)，8
Shannon's noiseless coding theorem (NLCT，Shannon 无噪编码定理)，8
Shannon's first coding theorem (FCT，Shannon 第一编码定理)，42
Shannon's second coding theorem (SCT)，or noisy coding theorem (NCT，Shannon 第二编码定理，有噪编码定理)，59，162
Shannon's SCT，converse part(Shannon 第二编码定理：逆命题)69
Shannon's SCT，strong converse part，(Shannon 第二编码定理：强逆命题)175
Shannon's SCT，direct part(Shannon 第二编码定理：直接命题)，72，153
Shannon-McMillan-Breiman theorem (Shannon-McMillan Breiman 定理)，397
totient function(欧拉函数)，270
transform(变换)
character transform(字符变换)，319
Fourier transform(Fourier 变换)，296
Fourier transform, discrete(离散 Fourier 变换)，256
Fourier transform in L_2 (L_2 中 Fourier 变换)，413
transmitter(发射机)，443

U

uncertainty principle(测不准原理)，431

V

Vandermonde determinant(Vandermonde 行列式)，237
Vandermonde matrix(Vandermonde 矩阵)，237

W

wedge product(楔形积)，149
weight enumerator polynomial(权重枚举多项式)，318
white noise(白噪声)，368
word, or a string (of characters, digits, letters or symbols)((字符，位，字母，符号)的字，串)，3
weight of a word(码字的重量)，144

信息论与编码理论 剑桥大学真题精解

Information Theory and Coding by Example

三大不同——从科学到科学家

- 概率与代数。这两个方向往往出现在不同的课程和教材中，而本书跨越了不同科研领域的界限，这与作者多年来的研究和教学风格密不可分。
- 剑桥真题解析。多数信息论专著侧重理论分析，而本书包含大量例题，它们有些来自剑桥大学课堂练习，有些则是学位考试真题，并配有详尽解答。
- 科学巨匠之路。跟上Shannon、Markov、Hamming等科学家的学术历程，这里既有划时代论文的光芒，也有学术观点的争鸣，鲜明的态度赋予科学以温度。

两大思维——站在数学的肩膀上

- 跨学科思维。作者曾工作于俄罗斯科学院与剑桥大学，它们都具有跨学科研究的优良传统，这种思维方式不仅滋养着云集其中的智者，也间接塑造了本书的精妙。
- 数学思维。尽管在应用研究中颇有建树，但作者却从不掩饰自己骨子里的数学基因，并且认为在当今世界中，数学思维依然是我们生存和自我完善的重要方式。

作者简介

马克·凯尔伯特（Mark Kelbert）斯旺西大学数学系讲师，曾在位于莫斯科的信息传输问题研究所和数学地理及地震预测研究所工作多年。

尤里·苏霍夫（Yuri Suhov）剑桥大学纯数学和数学统计系荣誉退休教授，主要研究方向为动态系统、统计力学等，曾工作于圣保罗大学和信息传输问题研究所。

CAMBRIDGE
UNIVERSITY PRESS
www.cambridge.org

营销热线：(010) 88379604
客服热线：(010) 88378991 88361066
购书热线：(010) 68326294 88379649 68995259

华章网站：www.hzbook.com
网上购书：www.china-pub.com
数字服务：www.hzmedia.com.cn

ISBN 978-7-111-53352-6

定价：89.00元

嘛屩嚣嚣跨哨喊嚣哪嚣嚣嚣嚣
哨号哨蛐蛐蛐蛐蛐