HZ Books
华章教育

计 算 机 科 学 丛 书

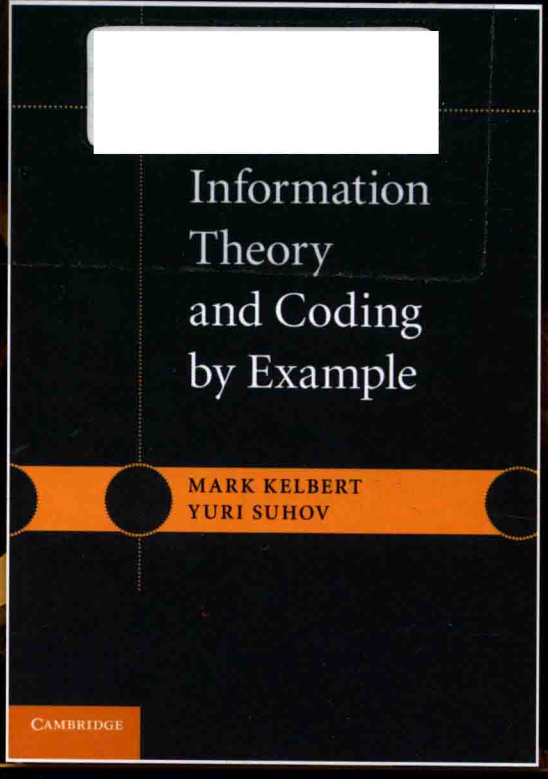
CAMBRIDGE

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

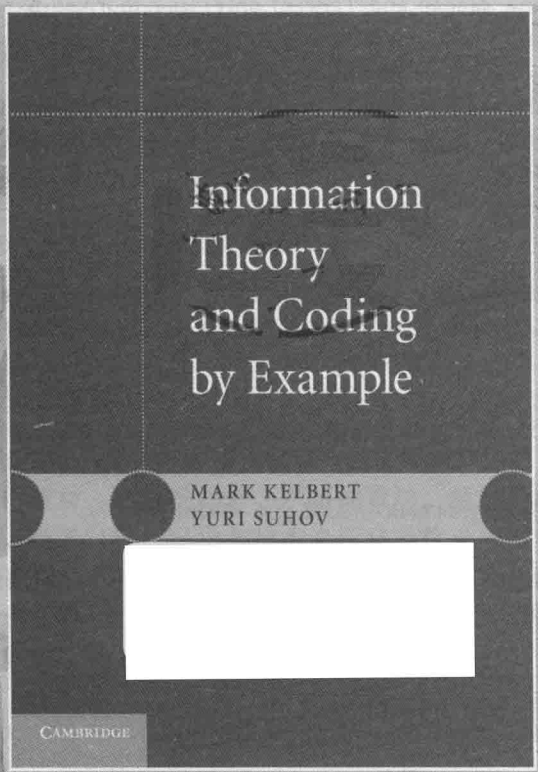
计 算 机 科 学 丛 书

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；高晖等译 . —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－题解 ②信源编码－高等学校－题解 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

本书讲解信息论与编码理论，涵盖概率和代数两个方向。书中素材来自剑桥大学本科生课程“信息论”“编码与密码学”以及几门数学方向的研究生课程。全书最大的特色是例题丰富，并将 Shannon 等科学家的学术历程贯穿其中，在透彻讲解基础知识的同时带领读者逐步探讨深层主题。

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，还能够通过剑桥真题判断自己已达到或期望达到的学习程度。

出版发行：机械工业出版社（北京市西城区百万庄大街 22 号　邮政编码：100037）

责任编辑：曲　熠	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）88378991　88361066　　投稿热线：（010）88379604
购书热线：（010）68326294　88379649　68995259　　读者信箱：hzjsj@hzbook.com

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，还揭示了学术的源变，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：本科三年级的“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教学素材，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。本书的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 Khas'minsky、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 Sagalovich、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[139,141]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虚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[171]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，受压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[35]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[65]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[41]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。在英国，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他理论中

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[35]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[52,36]的影响。我们与 Charles Goldie 长久的友谊以及同 Tom Cover 和睦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(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

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 $\sharp A$ 表示有限集合 A 的势。符号 $\mathbf{1}$ 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：

信源 → 编码器 → 信道 → 译码器 → 信宿

例子 1.1.1 (a) 信源：剑桥大学学院唱诗班。

(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

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

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 $\mathbb{P}(U_j=u)=\lambda(u)$, $j\geqslant 1$ 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,

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

其中 $0\leqslant p(u)\leqslant 1$ 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$q\leqslant m$(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

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

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 s 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 $\boldsymbol{u}^{(n)}=u_1u_2\cdots u_n$ 可以表示为码字的级联

$$f(\boldsymbol{u}^{(n)})=f(u_1)f(u_2)\cdots f(u_n)$$

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，…，s_m，存在一个可译码 f：$I\to J^*$，码字长度是 s_1，…，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，…，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，…，s_m 的码字建立无前缀编码。改写式(1.1.4)

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

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 $n_s\geqslant 0$，可推得

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

或者

$$n_{s-1}\leqslant q^{s-1}-n_1q^{s-2}-\cdots-n_{s-2}q \tag{1.1.6b}$$

重复这个过程可得

$$\begin{aligned} n_{s-2} &\leqslant q^{s-2}-n_1q^{s-3}-\cdots-n_{s-3}q \\ &\vdots \qquad\qquad \vdots \qquad\qquad \vdots \qquad\qquad (1.1.6.\text{s-1}) \\ n_2 &\leqslant q^2-n_1q \\ n_1 &\leqslant q \qquad\qquad (1.1.6.\text{s}) \end{aligned}$$

可以看到实际上对于所有 $i=1$，…，$s-1$，如果满足 $n_{i+1}=0$ 或 n_i 小于不等式的 RHS(根据定义，$n_s\geqslant 1$，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_3 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。)

回顾一下，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

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

（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

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

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

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

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：

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

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 $\mathbb{E}s(X)=4/3<h(X)=\log 3=1.585$。

(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

$$\sum_{x\in I}2^{-s(x)}=\sum_{l\leqslant L}\sum_{x:s(x)=l}2^{-s(x)}=\sum_{l\leqslant L}2^l\times 2^{-l}=L$$

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：

$$h_q\leqslant\min\mathbb{E}S<h_q+1\qquad(1.1.12)$$

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

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

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

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

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_m^H：$I_m \to J_q^*$ 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，下列算法构造了一个二进制树：

(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。

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

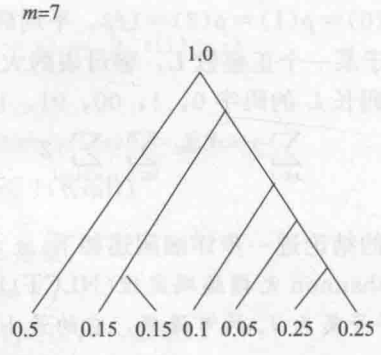


图 1-1

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。

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

$$p(i)\geqslant p(i') \quad \text{意味着} \quad s(i)\leqslant s(i') \tag{1.1.14}$$

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^*：

$$\mathbb{E}S_m^* < \mathbb{E}S_m^H \tag{1.1.15}$$

在此情况下概率分布可以假设为

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

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 $\mathbb{E}S_m^H$ 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 $\mathbb{E}S$ 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 $\mathbb{E}S_{m-1}^H$和 $\mathbb{E}S_{m-1}^*$的贡献与对 $\mathbb{E}S_m^H$ 和 $\mathbb{E}S_m^*$ 的贡献都一样。所以，f_{m-1}^*要优于 f_{m-1}^H：$\mathbb{E}S_{m-1}^*<\mathbb{E}S_{m-1}^H$，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 $p(1)\geqslant p(2)\geqslant\cdots\geqslant p(m)$。在任何二元码中：

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

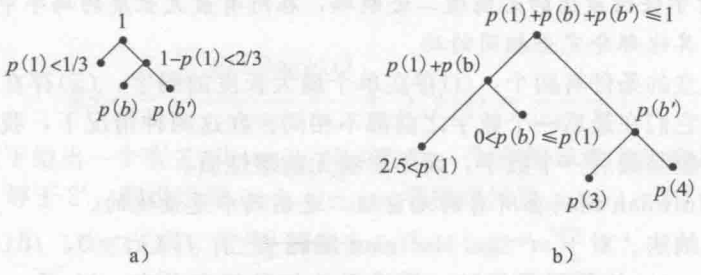
(b) 如果 $p(1)>2/5$，那么字母 1 必须用长度为 1 的码字编码。

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

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

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 $0\leqslant p(b)$，$p(b')\leqslant 1-p(1)$，$p(b)+p(b')=1-p(1)$。



a) b)

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

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

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

12 (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