信息论基础教程习题解答与实验指导

李　梅　李亦农　编著

北京邮电大学出版社
·北京·

内 容 简 介

本书为《信息论基础教程》的配套教材,包括原书各章的习题解答、新增的习题及解答以及上机实验的参考程序等。

本书作为教材配套的实验指导,强调实践环节,编写了 5 个实验的实验指导:信道容量的迭代算法、唯一可译码判决准则、Huffman 编码、LZW 编码和 Shannon 编码,每个实验都给出了实验目的、实验要求、算法和参考程序。

图书在版编目(CIP)数据

信息论基础教程习题解答与实验指导/ 李梅,李亦农编著 .—北京:北京邮电大学出版社,2005

ISBN 7-5635-1163-6

Ⅰ.信... Ⅱ.①李...②李... Ⅲ.信息论—高等学校—教学参考资料 Ⅳ.G201

中国版本图书馆 CIP 数据核字(2005)第 120215 号

书　　名:信息论基础教程习题解答与实验指导
编　　著:李梅　李亦农
责任编辑:李欣一
出 版 者:北京邮电大学出版社(北京市海淀区西土城路 10 号)
　　　　　邮编:100876　电话:62282185　62283578(Fax)
电子信箱:publish@bupt.edu.cn
经　　销:各地新华书店
印　　刷:
印　　数:1—5 000 册
开　　本:787 mm×1 092 mm　1/16　印张:9.5　字数:206 千字
版　　次:2005 年 10 月第 1 版　2005 年 10 月第 1 次印刷

ISBN 7-5635-1163-6/TN·401　定 价:15.00 元

·如有印装质量问题,请与北京邮电大学出版社发行部联系·

前　言

随着信息技术的飞速发展，信息论已经成为越来越多的学校和专业的专业基础课和专业选修课，而同时各个学校都在大量压缩专业课课时，一般通信工程和电子信息专业信息论的授课时数都在34课时左右，因此迫切需要一本内容精简、符合少学时教学要求的教材。基于这样的考虑，我们在借鉴了国内外众多的信息论优秀教材和参考书之后编写了《信息论基础教程》教材。教材以香农的3个编码定理为中心，重点讲述了相关的基本概念、基本原理和基本方法。

本书为《信息论基础教程》的配套教材，包括原书各章的习题解答、新增的习题及解答以及上机实验的参考程序等。

习题是读者深入理解和消化基本理论，锻炼独立解决问题能力的重要环节。本书给出了教材的习题解答，并且补充了一些典型习题，所选配的习题是从国内外大量优秀习题中精选而来的，具有一定的代表性，同时很多题与日常生活密切相关，比较新颖，具有一定的趣味性。大部分习题解答需要综合运用所学知识，能有效地帮助学生巩固、加深对授课内容的理解。本书的解答仅仅作为读者的学习参考，希望读者能通过自己的思考得出不同的解题方法。

信息论是一门理论和实践紧密结合的课程，如果仅仅停留于理论，学生学到的就是死的知识，不能很好地调动学生对课程学习的兴趣，很难体现培养创新思维和动手能力的要求。本书作为教材配套的实验指导，强调实践环节，编写了5个实验的实验指导：信道容量的迭代算法、唯一可译码判决准则、Huffman编码、LZW编码和Shannon编码，每个实验都给出了实验目的、实验要求、算法和参考程序。

第1～4章由李梅编写，第5～7章由李亦农编写，第8章由两人共同完成。

在本书的编写过程中,参阅了国内外一些经典著作(列于本书参考文献中),同时参考了很多国外知名大学信息论基础课程的课后习题及解答,在此向有关作者表示感谢!在本书编写的过程中,得到了陈剑、李洪、徐益民、唐晓晟等同志的大力支持,在此也表示感谢!本套书在成书过程中得到了北京邮电大学出版社周明老师的热心支持,北京邮电大学田宝玉教授对教材提出了宝贵的意见,北京邮电大学信息工程学院的苏泗浠老师给出了部分习题的解答,北京邮电大学电信工程学院2000级陈敏茹同学为部分上机题编制了源程序,在此也一并表示感谢!

本书曾得到地下信息探测技术与仪器教育部重点实验室的资助。

恳请读者对书中的谬误给予指正。

作　者

2005 年 5 月

目　录

第 1 章

绪　　论

《信息论基础教程》的学习要点如下:

自信息

自信息表示随机事件 x_i 发生的不确定性或发生所含有的信息量。自信息量 $I(x_i)$ 定义为

$$I(x_i) = -\log_2 p(x_i) = \log_2 \frac{1}{p(x_i)}$$

式中, $p(x_i)$**为该事件发生的概率。**

互信息

互信息表示已知事件 y_j 后所消除的关于事件 x_i 的不确定性,它等于事件 x_i 本身的不确定性 $I(x_i)$减去已知事件 y_j 后对 x_i 仍然存在的不确定性 $I(x_i|y_j)$。互信息 $I(x_i;y_j)$ 定义为

$$I(x_i;y_j) = I(x_i) - I(x_i|y_j) = \log_2 \frac{p(x_i|y_j)}{p(x_i)}$$

平均自信息

平均自信息表示整个信源(用随机变量 X 表示)的平均不确定性,它等于随机变量 X 的每一个可能取值的自信息 $I(x_i)$的统计平均值。平均自信息量 $H(X)$定义为

$$H(X)=E[I(x_i)]=-\sum_{i=1}^{q}p(x_i)\log_2 p(x_i)$$

式中, q **为随机变量** X **的所有可能取值的个数。**

离散信源的最大熵

离散信源中各消息等概率出现时熵最大,也称为最大离散熵定理:

$$H(p_1,p_2,\cdots,p_n)\le H\left(\frac{1}{n},\frac{1}{n},\cdots,\frac{1}{n}\right)=\log_2 n$$

式中, n **是随机变量** X **的所有可能取值的个数。**

联合熵

联合熵表示二维随机变量 XY **的平均不确定性,它等于联合自信息的统计平均值。联合熵** $H(XY)$**定义为**

$$H(XY)=\sum_{i=1}^{n}\sum_{j=1}^{m}p(x_iy_j)I(x_iy_j)=-\sum_{i=1}^{n}\sum_{j=1}^{m}p(x_iy_j)\log_2 p(x_iy_j)$$

式中, n **和** m **分别为随机变量** X **和** Y **的所有可能取值的个数。**

条件熵

条件熵表示已知随机变量 X **后,对随机变量** Y **仍然存在的平均不确定性。条件熵** $H(Y|X)$**定义为**

$$H(Y|X)=\sum_{i=1}^{n}p(x_i)H(Y|x_i)=-\sum_{i=1}^{n}\sum_{j=1}^{m}p(x_iy_j)\log_2 p(y_j|x_i)$$

式中, $H(Y|x_i)$**表示已知** $X=x_i$ **的条件下,对随机变量** Y **的平均不确定性。**

各类熵之间的关系:

$$H(XY)=H(X)+H(Y|X)=H(Y)+H(X|Y)\le H(X)+H(Y)$$

当 X, Y **统计独立时,有**

$$H(XY)=H(X)+H(Y)$$

平均互信息

平均互信息表示收到一个符号集(用随机变量 Y **表示)后所消除的关于另一个符号集(用随机变量** X **表示)的不确定性,也就是从** Y **所获得的关于** X **的平均信息量。平均互信息** $I(X;Y)$**定义为**

$$I(X;Y)=\sum_{i=1}^{n}\sum_{j=1}^{m}p(x_iy_j)I(x_i;y_j)=\sum_{i=1}^{n}\sum_{j=1}^{m}p(x_iy_j)\log_2\frac{p(x_i|y_j)}{p(x_i)}$$

式中, n **和** m **分别为随机变量** X **和** Y **的所有可能取值的个数。**

平均互信息和各类熵的关系:

$$I(X;Y)=H(X)-H(X|Y)=H(Y)-H(Y|X)=H(X)+H(Y)-H(XY)$$

当 X, Y 统计独立时,有

$$I(X;Y)=0$$

数据处理定理

如果随机变量 X, Y, Z 构成一个马尔可夫链,则有以下关系成立:

$$I(X;Z)\leq I(X;Y),\qquad I(X;Z)\leq I(Y;Z)$$

等号成立的条件是对于任意的 x, y, z,有 $p(x|yz)=p(x|z)$ 和 $p(z|xy)=p(z|x)$。

数据处理定理中不等式 $I(X;Z)\leq I(X;Y)$ 表明从 Z 所得到的关于 X 的信息量小于等于从 Y 所得到的关于 X 的信息量。如果把 $Y\rightarrow Z$ 看作数据处理系统,那么通过数据处理后,虽然可以满足我们的某种具体要求,但是从信息量来看处理后会损失一部分信息,最多保持原来获得的信息,也就是说,对接收到的数据 Y 进行处理后,决不会减少关于 X 的不确定性。

熵率

熵率表示离散多符号信源的平均不确定性,它是信源输出的符号序列中平均每个符号所携带的信息量。

如果当 $N\rightarrow\infty$ 时极限 $\lim\limits_{N\rightarrow\infty} H_N(X)$ 存在,则称之为熵率,或称为极限熵。熵率定义为

$$H_\infty=\lim_{N\rightarrow\infty} H_N(X)$$

式中,$H_N(X)$ 称为平均符号熵,它表示随机变量序列中,对前 N 个随机变量的联合熵的平均:

$$H_N(X)=\frac{1}{N}H(X_1X_2\cdots X_N)$$

离散平稳无记忆信源的熵率

多符号信源中最简单的是离散平稳无记忆信源,它的熵率为

$$H_\infty=H(X)$$

m 阶马尔可夫信源的熵率

离散平稳有记忆信源中比较简单的是记忆长度有限的信源,即马尔可夫信源。如果信源在某时刻发出的符号仅与在此之前发出的 m 个符号有关,则称为 m 阶马尔可夫信源,它的熵率为

$$H_\infty=H(X_{m+1}|X_1X_2\cdots X_m)$$

$H(X_{m+1}|X_1X_2\cdots X_m)$ 常记为 H_{m+1},它表示已知前面 m 个符号的条件下,输出下一个符号的平均不确定性。

对于离散平稳马尔可夫信源,通常将上述符号的不确定性问题转化为齐次遍历的马

尔可夫链的状态转移问题:

$$
\begin{aligned}
H_{m+1} &= H(X_{m+1} \mid X_1 X_2 \cdots X_m) \\
&= E[p(x_{i_{m+1}} \mid x_{i_1} x_{i_2} \cdots x_{i_m})] \\
&= E[p(x_{i_{m+1}} \mid s_i)] \\
&= -\sum_{i=1}^{q^m} \sum_{i_{m+1}=1}^{q} p(s_i)\, p(x_{i_{m+1}} \mid s_i) \log_2 p(x_{i_{m+1}} \mid s_i) \\
&= \sum_i p(s_i) H(X \mid s_i) \\
&= -\sum_i \sum_i p(s_i)\, p(s_j \mid s_i) \log_2 p(s_j \mid s_i)
\end{aligned}
$$

其中, $p(s_i)$是马尔可夫链的状态的平稳分布, $H(X|s_i)$表示信源处于某一状态 s_i 时发出下一个符号的平均不确定性, $p(s_j|s_i)$是状态的一步转移概率。

信源剩余度

信源剩余度定义为

$$
\gamma = 1 - \eta = 1 - \frac{H_\infty}{H_0} = 1 - \frac{H_\infty}{\log_2 q}
$$

其中, η 为熵的相对率, H_0 为离散信源的最大熵, $H_0 - H_\infty$ 越大,信源的剩余度越大。

连续信源的微分熵

$$
h(X) = -\int_{\mathrm{R}} p(x) \log_2 p(x) \,\mathrm{d}x
$$

连续信源的最大熵

对于输出信号幅度受限的连续信源,当满足均匀分布时达到最大熵;对于平均功率受限的连续随机变量,当服从高斯分布时具有最大熵。

熵功率

设某连续信源的微分熵为 $h(X)$,则将与它具有相同熵的高斯信源的平均功率 P 定义为熵功率,即

$$
P = \frac{1}{2\pi \mathrm{e}} \mathrm{e}^{2h(X)}
$$

假定该连续信源的平均功率为 P,则 $P \leq P$。熵功率和信源的平均功率相差越大,说明信源的剩余越大。信源平均功率和熵功率之差$(P - P)$称为连续信源的剩余度。

信道容量

对于给定的信道,即信道转移概率 $p(y_j|x_i)$固定后, $I(X;Y)$是 $p(x_i)$的上凸函数。

因此对于给定的信道,总存在一种信源(某种输入概率分布),使信道平均传输一个符号接收端获得的信息量最大,也就是说对于每个固定信道都有一个最大的信息传输率,这个最大的信息传输率即信道容量,而相应的输入概率分布称为最佳输入分布。信道容量定义为平均互信息对于输入概率分布的最大值:

$$C=\max_{p(x)}\{I(X;Y)\}$$

具有扩展性能的无损信道的信道容量

$$C=\max_{p(x)}\{I(X;Y)\}=\max_{p(x)}H(X)=\log_2 r$$

具有归并性能的无噪信道的信道容量

$$C=\max_{p(x)}\{I(X;Y)\}=\max_{p(x)}H(Y)=\log_2 s$$

具有一一对应关系的无噪无损信道的信道容量

$$C=\max_{p(x)}\{I(X;Y)\}=\max_{p(x)}H(Y)=\max_{p(x)}H(X)=\log_2 s=\log_2 r$$

对称信道的信道容量

$$C=\log_2 s-H(p_1',p_2',\cdots,p_s')$$

式中 $p_1',p_2',\cdots,p_s'$**为信道矩阵中的任一行元素,当输入为等概分布时,达到信道容量。**

准对称信道的信道容量

$$C=\log_2 r-\sum_{k=1}^{n}N_k\log_2 M_k-H(p_1',p_2',\cdots,p_s')$$

式中,假定准对称信道分成 n **个对称的子信道,** N_k **是第** k **个子矩阵中行元素之和,** M_k **是第** k **个子矩阵中列元素之和,当输入为等概分布时,达到信道容量。**

信道容量定理

设有一般离散信道,它有 r **个输入符号,** s **个输出符号。当且仅当存在常数** C **使下式成立时,** $I(X;Y)$ **达到信道容量。**

$$\begin{cases}I(x_i;Y)=C, & p(x_i)\neq 0\\ I(x_i;Y)\leq C, & p(x_i)=0\end{cases}$$

式中 $I(x_i;Y)=\sum_j p(y_j|x_i)\log_2\dfrac{p(y_j|x_i)}{p(y_j)}$**,它表示信道输入** x_i **时,所给出关于输出** Y **的信息量。常数** C **即为所求的信道容量。**

离散平稳无记忆信道的 N 次扩展信道的信道容量

$$C^N=NC$$

独立并联信道的信道容量

当 N **个独立并联信道的信道容量都相同时,有**

$$C_{并}=NC$$

级联信道的信道容量

级联信道的总的信道矩阵等于所级联信道的信道矩阵的乘积。求得级联信道的总的信道矩阵后，级联信道的信道容量就可以用求离散单符号信道的信道容量的方法计算。

波形信道的信道容量

$$C_t = B\log_2\left(1 + \frac{\sigma_X^2}{N_0 B}\right)$$

这就是著名的香农公式，它适用于加性高斯白噪声信道。式中 B 为信道的带宽，N_0 为高斯白噪声的单边功率谱密度。σ_X^2 是输入随机变量 X 的方差，当随机变量 X 的均值为 0 时即为其平均功率。只有当输入信号为功率受限的高斯白噪声信号时，才能达到该信道容量。

码的分类

$$\text{码}\begin{cases}\text{非分组码(树码)}\\ \text{分组码(块码)}\begin{cases}\text{奇异码}\\ \text{非奇异码}\begin{cases}\text{非唯一可译码}\\ \text{唯一可译码}\begin{cases}\text{即时码}\\ \text{非即时码}\end{cases}\end{cases}\end{cases}\end{cases}$$

无失真定长信源编码定理

离散无记忆信源的熵为 $H(S)$，若对信源长为 N 的序列进行定长编码，码符号集中有 r 个码符号，码长为 l，则对于任意小的正数 ε，只要满足 $\frac{l}{N} \geq \frac{H(S)+\varepsilon}{\log_2 r}$，则当 N 足够大时，可实现几乎无失真编码，即译码错误概率为任意小。反之，如果 $\frac{l}{N} \leq \frac{H(S)-2\varepsilon}{\log_2 r}$，则不可能实现几乎无失真编码，当 N 足够大时，译码错误概率为 1。

Kraft 不等式(McMillan 不等式)

设信源符号集为 $S=\{s_1, s_2, \cdots, s_q\}$，码符号集为 $X=\{x_1, x_2, \cdots, x_r\}$，对信源进行编码，得到的码为 $C=\{w_1, w_2, \cdots, w_q\}$，码长分别为 $l_1, l_2, \cdots, l_q$，即时码(唯一可译码)存在的充要条件是 $\sum_{i=1}^{q} r^{-l_i} \leq 1$，这称为 Kraft 不等式(McMillan 不等式)。

无失真变长信源编码定理(香农第一定理)

设离散无记忆信源 S 的信源熵 $H(S)$，它的 N 次扩展信源 $S^N=\{s_1, s_2, \cdots, s_{q^N}\}$，其熵 $H(S^N)$，用码符号 $X=\{x_1, x_2, \cdots, x_r\}$ 对信源 S^N 进行编码，总可以找到一种唯一可译码使单个信源符号所需的平均码长满足：

$$\frac{H(S)}{\log_2 r} \leq \frac{\bar{L}_N}{N} < \frac{H(S)}{\log_2 r} + \frac{1}{N}$$

当 $N\to\infty$**时**,$\frac{L_N}{N}=H_r(S)$, **式中** $L_N=\sum_{i=1}^{q^N}p(s_i)\lambda_i$, λ_i **是扩展信源的信源符号** s_i **所对应的码字长度。**

译码规则

设信道的输入符号集 $X=\{x_i,\ i=1,2,\cdots,r\}$,**输出符号集** $Y=\{y_j,\ j=1,2,\cdots,s\}$,**若对每一个输出符号** y_j **都有一个确定的函数** $F(y_j)$,**使** y_j **对应于唯一的一个输入符号** x_i,**则称这样的函数为译码规则,记为**

$$F(y_j)=x_i \qquad i=1,2,\cdots,r;\quad j=1,2,\cdots,s$$

对于有 r **个输入、**s **个输出的信道而言,输出** y_j **可以对应** r **个输入中的任何一个,所以译码规则共有** r^s **种。**

错误概率

在确定译码规则 $F(y_j)=x_i$ **后,若信道输出端接收到符号** y_j,**则一定译成** x_i,**如果发送端发送的确实就是** x_i,**就是正确译码;反之,若发送端发送的不是** x_i **就认为是错误译码。于是收到符号** y_j **条件下,译码正确概率为**

$$p\ [F(y_j)|y_j]=p(x_i|y_j)$$

而译码错误概率 $p(e|y_j)=1-p(x_i|y_j)=1-p[F(y_j)|y_j]$,**其中** e **表示除了** $F(y_j)=x_i$ **以外的所有符号的集合。**

平均错误概率

译码后的平均错误概率 P_{E} **是译码错误概率** $p(e|y_j)$**对** Y **的统计平均值,即**

$$P_{\mathrm{E}}=E[p(e|y_j)]=\sum_{j=1}^{s}p(y_j)p(e|y_j)$$

它表示平均每接收到一个符号后的译码错误概率。

最大后验概率译码规则

选择译码函数 $F(y_j)=x^*$,**使之满足条件:**

$$p(x^*|y_j)\geq p(x_i|y_j) \qquad x^*\in X$$

则称为最大后验概率译码规则,又称为最小错误概率准则。对于每一个输出符号 y_j, $j=1,2,\cdots,s$ **均译成与之具有最大后验概率的那个输入符号** x^*,**则信道译码的平均错误概率最小。**

极大似然译码规则

选择译码函数 $F(y_j)=x^*$,**使之满足条件:**

$$p(y_j|x^*)\geq p(y_j|x_i) \qquad x^*\in X$$

称为极大似然译码规则。

当输入符号等概时，最大后验概率译码规则和极大似然译码规则是等价的。

费诺不等式

信道疑义度 $H(X|Y)$ 与平均错误概率 P_E 满足以下关系：

$$H(X|Y) \le H(P_E) + P_E \log_2(r-1)$$

费诺不等式表明接收到 Y 后关于 X 的平均不确定性可以分为两部分：第一部分 $H(P_E)$ 是指接收到 Y 后是否产生错误的不确定性；第二部分 $P_E \log_2(r-1)$ 是已知错误 P_E 发生后，判断是哪个输入符号造成错误的最大不确定性，是 $(r-1)$ 个符号的最大可能不确定性与 P_E 的乘积。

汉明距离

长度相同的两个符号序列(码字) x_i 与 y_j 之间的距离是指序列 x_i 和 y_j 对应位置上码元符号不同的位置的个数，称为汉明距离：

$$D(x_i, y_j) = \sum_{k=1}^{n} x_{i_k} \qquad y_{j_k}$$

式中 n 为符号序列的长度。

码的最小距离

码 C 中，任意两个码字的汉明距离的最小值称为该码 C 的最小距离，即

$$D_{\min} = \min\{D(w_i, w_j)\} \quad w_i, w_j \in C \text{ 且 } w_i \ne w_j$$

编码选择码字时，要使码的最小距离越大越好。译码时，则要将接收序列译成与其距离最小的码字，这样得到的平均错误概率最小。

有噪信道编码定理(香农第二定理)

设有一个离散无记忆平稳信道，其信道容量为 C。当待传输的信息率 $R < C$ 时，只要码长 n 足够长，则总存在一种编码可以使译码错误概率任意小。若 $R > C$，则无论 n 取多大，也找不到一种编码，使译码错误概率 P_E 任意小。

失真函数

设离散无记忆信源 X，经过信道传输后信道输出 Y，对于每一对 (x_i, y_j)，指定一个非负的函数 $d(x_i, y_j) \ge 0$ $(i = 1, 2, \cdots, r;\ j = 1, 2, \cdots, s)$，称 $d(x_i, y_j)$ 为单个符号的失真函数或失真度。失真函数表示信源发出一个符号 x_i，而在接收端再现为 y_j 所引起的误差或失真的大小。失真函数通常排列成矩阵形式。

平均失真度

信源的平均失真度表示某个信源通过某个信道传输后失真的大小。信源的平均失真度定义为

$$\bar{D} = E[d(x_i, y_j)] = \sum_{i=1}^{r} \sum_{j=1}^{s} p(x_i y_j)\, d(x_i, y_j) = \sum_{i=1}^{r} \sum_{j=1}^{s} p(x_i)\, p(y_j|x_i)\, d(x_i, y_j)$$

保真度准则

如果要求信源的平均失真度小于所允许的失真 D,即 $D \le D$,称为保真度准则。

D 失真许可的试验信道

凡满足保真度准则($D \le D$)的信道称为 D 失真许可的试验信道。D 失真许可的试验信道的集合 B_D 定义为

$$B_D = \{ p(y_j | x_i): D \le D \} \quad i = 1, 2, \cdots, r; \quad j = 1, 2, \cdots, s$$

信息率失真函数

对于给定的信源,即信源概率分布 $p(x_i)$固定后,$I(X;Y)$是信道转移概率 $p(y_j | x_i)$的下凸函数。因此对于给定的信源,总存在一种信道使 $I(X;Y)$达到最小。在满足保真度准则的 D 失真许可的试验信道集合 B_D 中必然有一个信道 $p(y_j | x_i)$使 $I(X;Y)$达到最小,这个最小值就是信息率失真函数,也称率失真函数。率失真函数 $R(D)$定义为

$$R(D) = \min_{p(y_j | x_i) \in B_D} I(X;Y)$$

$R(D)$是关于 D 的下凸函数,并且在定义域内是严格递减函数。其定义域的下限和上限分别为

$$D_{\min} = \sum_i p(x_i) \min_j d(x_i, y_j)$$

$$D_{\max} = \min_j \sum_i p(x_i) d(x_i, y_j)$$

限失真信源编码定理(香农第三定理)

设 $R(D)$是离散无记忆信源的信息率失真函数并且失真函数为有限值,对于任意的允许失真度 $D \ge 0$ 和任意的正数 $\varepsilon > 0$,当信源序列长度 N 足够长时,一定存在一种编码,其码字个数 $M = e^{N[R(D) + \varepsilon]}$,而编码后的平均失真度 $D \le D + \varepsilon$。反过来说,若码字个数 $M < e^{NR(D)}$,则必然 $D > D$。

第 2 章

信息的度量

2.1　**同时掷 2 颗骰子，事件 A、B、C 分别表示：(A) 仅有一个骰子是 3；(B) 至少有一个骰子是 4；(C) 骰子上点数的总和为偶数。试计算事件 A、B、C 发生后所提供的信息量。**

解：

掷 2 颗骰子，共有 $6^2=36$ 种结果。记投掷结果为 (x, y)，其中 x 和 y 是 2 颗骰子分别掷出的点数，则事件 A 对应于 10 种结果：$(1,3)$，$(2,3)$，$(4,3)$，$(5,3)$，$(6,3)$，$(3,1)$，$(3,2)$，$(3,4)$，$(3,5)$，$(3,6)$；**事件 B、C 分别对应于 11 种结果和 18 种结果。因此它们的概率分别为**

$$p(A)=10/36$$

$$p(B)=11/36$$

$$p(C)=18/36$$

A、B、C 所提供的信息量分别为

$$I(A)=-\log_2 p(A)=1.848\,0\ \text{bit}$$

$$I(B)=-\log_2 p(B)=1.710\,5\ \text{bit}$$

$$I(C)=-\log_2 p(C)=1\ \text{bit}$$

2.2　**设有 n 个球，每个球都以同样的概率 $1/N$ 落入 N 个格子 ($N\geq n$) 的每一个格子中。假定：(A) 某指定的 n 个格子中个各落入一个球；(B) 任何 n 个格子中各落入一**

球。试计算事件 A、B 发生后所提供的信息量。

解:

事件 A、B 发生的概率分别为

$$p(A)=\frac{n!}{N^n}$$

$$p(B)=\frac{C_N^n\cdot n!}{N^n}=\frac{N!}{N^n(N-n)!}$$

事件 A、B 发生提供的信息量分别为

$$I(A)=\log_2\frac{1}{p(A)}=\log_2\frac{N^n}{n!}=n\log_2 N-\log_2 n!$$

$$I(B)=\log_2\frac{1}{p(B)}=\log_2\frac{N^n(N-n)!}{N!}=n\log_2 N+\log_2(N-n)!-\log_2 N!$$

2.3　**一信源有 4 种输出符号 x_i, $i=0,1,2,3$, 且 $p(x_i)=1/4$。设信源向信宿发出 x_3, 但由于传输中的干扰, 接收者收到 x_3 后, 认为其可信度为 0.9。于是信源再次向信宿发送该符号(x_3), 信宿无误收到。问信源在两次发送中发出的信息量各是多少? 信宿在两次接收中得到的信息量又各是多少?**

解:

假设信宿第一次收到的符号为 y, 由于第二次发送无误收到, 因此发、收信息量相等, 均为

$$I(x_3\mid y)=-\log_2 p(x_3\mid y)=-\log_2 0.9=0.15\ \text{bit}$$

第一次发出的信息量为

$$I(x_3)=-\log_2 p(x_3)=-\log_2 0.25=2\ \text{bit}$$

第一次传送的信息量为两次发送信息量之差, 即

$$I(x_3;y)=I(x_3)-I(x_3\mid y)=1.85\ \text{bit}$$

2.4　**用递推性计算熵函数 $H(1/3,1/3,1/6,1/6)$ 的值。**

解:

$$\begin{aligned}H\left(\frac{1}{3},\frac{1}{3},\frac{1}{6},\frac{1}{6}\right)&=H\left(\frac{1}{3},\frac{2}{3}\right)+\frac{2}{3}\times H\left(\frac{1}{2},\frac{1}{4},\frac{1}{4}\right)\\&=H\left(\frac{1}{3},\frac{2}{3}\right)+\frac{2}{3}\times H\left(\frac{1}{2},\frac{1}{2}\right)+\frac{2}{3}\times\frac{1}{2}\times H\left(\frac{1}{2},\frac{1}{2}\right)\\&=H\left(\frac{1}{3},\frac{2}{3}\right)+H\left(\frac{1}{2},\frac{1}{2}\right)\\&=1.915\ \text{比特/符号}\end{aligned}$$

2.5 **一信源有 6 种输出状态,概率分别为**

$p(A)=0.5$, $p(B)=0.25$, $p(C)=0.125$, $p(D)=p(E)=0.05$, $p(F)=0.025$

试计算 $H(X)$。**然后求消息** $ABABBA$ **和** $FDDFDF$ **的信息量(设信源先后发出的符号相互独立),并将之与长度为 6 的消息序列信息量的期望值相比较。**

解:

由信息熵定义,该信源输出的信息熵为

$$\begin{aligned} H(X) &= \sum_{i=1}^{6} p(x_i)\log_2\frac{1}{p(x_i)} \\ &= p(A)\log_2\frac{1}{p(A)} + p(B)\log_2\frac{1}{p(B)} + \cdots + p(F)\log_2\frac{1}{p(F)} \\ &= 0.5\log_2 2 + 0.25\log_2 4 + 0.125\log_2 8 + 2\times 0.05\log_2 20 + 0.025\log_2 40 \\ &= 1.94\ \text{比特/符号} \end{aligned}$$

消息 $ABABBA$ **所含的信息量为**

$$I_1 = 3I(A) + 3I(B) = 3[-\log_2 p(A) - \log_2 p(B)] = 3(\log_2 2 + \log_2 4) = 9\ \text{bit}$$

消息 $FDDFDF$ **所含的信息量为**

$$I_2 = 3I(D) + 3I(F) = 3[-\log_2 p(D) - \log_2 p(F)] = 3(\log_2 20 + \log_2 40) = 28.932\ \text{bit}$$

6 **位长消息序列的信息量期望值为**

$$\overline{I} = 6H(X) = 11.64\ \text{bit}$$

三者比较为 $I_1 < \overline{I} < I_2$。

2.6 **中国国家标准局所规定的二级汉字共 6 763 个。设每字使用的频度相等,求一个汉字所含的信息量。设每个汉字用一个 16×16 的二元点阵显示,试计算显示方阵所能表示的最大信息。显示方阵的利用率是多少?**

解:

由于假定每个汉字的使用频度相同,它们有相同的出现概率,即

$$p(x) = \frac{1}{6\ 763}$$

因此每个汉字所含的信息量为

$$I(x) = -\log_2 p(x) = -\log_2\frac{1}{6\ 763} = 12.7\ \text{bit}$$

每个显示方阵能显示 $2^{16\times 16}=2^{256}$ **种不同的状态,这** 2^{256} **种状态等概出现时信息熵最大,这时每种状态出现的概率为**

$$p(y) = \frac{1}{2^{256}}$$

因此一个显示方阵所能显示的最大信息量是

$$I(y) = -\log_2 p(y) = -\log_2 \frac{1}{2^{256}} = 256 \text{ bit}$$

显示方阵的利用率或显示效率为

$$\eta = \frac{I(x)}{I(y)} = \frac{12.7}{256} = 0.0497$$

2.7　**已知信源发出** a_1 **和** a_2 **两种消息,且** $p(a_1) = p(a_2) = 1/2$。**此消息在二进制对称信道上传输,信道传输特性为** $p(b_1 | a_1) = p(b_2 | a_2) = 1 - \varepsilon$, $p(b_1 | a_2) = p(b_2 | a_1) = \varepsilon$。**求互信息量** $I(a_1; b_1)$**和** $I(a_1; b_2)$。

解:

由 $p(a_1) = p(a_2) = 1/2$, $p(b_1 | a_1) = p(b_2 | a_2) = 1 - \varepsilon$, $p(b_1 | a_2) = p(b_2 | a_1) = \varepsilon$ **可知**

$$p(a_1 b_1) = p(b_1 | a_1) \cdot p(a_1) = \frac{1}{2}(1 - \varepsilon)$$

$$p(a_1 b_2) = p(b_2 | a_1) \cdot p(a_1) = \frac{1}{2}\varepsilon$$

$$p(a_2 b_1) = p(b_1 | a_2) \cdot p(a_2) = \frac{1}{2}\varepsilon$$

$$p(a_2 b_2) = p(b_2 | a_2) \cdot p(a_2) = \frac{1}{2}(1 - \varepsilon)$$

$$p(b_1) = \sum_{i=1}^{2} p(a_i b_1) = \frac{1}{2}$$

$$p(b_2) = \sum_{i=1}^{2} p(a_i b_2) = \frac{1}{2}$$

所以

$$I(a_1; b_1) = \log_2 \frac{p(a_1 | b_1)}{p(a_1)} = \log_2 \frac{p(a_1 b_1)}{p(a_1) p(b_1)} = \log_2 \frac{\frac{1}{2}(1 - \varepsilon)}{\frac{1}{2} \times \frac{1}{2}} = 1 + \log_2 (1 - \varepsilon) \text{ bit}$$

$$I(a_1; b_2) = \log_2 \frac{p(a_1 | b_2)}{p(a_1)} = \log_2 \frac{p(a_1 b_2)}{p(a_1) p(b_2)} = \log_2 \frac{\frac{1}{2}\varepsilon}{\frac{1}{2} \times \frac{1}{2}} = 1 + \log_2 \varepsilon \text{ bit}$$

2.8　**已知二维随机变量** XY **的联合概率分布** $p(x_i y_j)$ **为**: $p(0,0) = p(1,1) = 1/8$, $p(0,1) = p(1,0) = 3/8$, **求** $H(X | Y)$。

解:

由 $p(y_j) = \sum_{i=1}^{2} p(x_i y_j)$,**得**

$$p_Y(0) = p(0,0) + p(1,0) = \frac{1}{2}, \qquad p_Y(1) = p(0,1) + p(1,1) = \frac{1}{2}$$

又由 $p(x_i|y_j)=\dfrac{p(x_iy_j)}{p(y_j)}$**可得**

$$p_{X|Y}(0|0)=\frac{p(0,0)}{p_Y(0)}=\frac{1}{4},\qquad p_{X|Y}(1|1)=\frac{p(1,1)}{p_Y(1)}=\frac{1}{4}$$

$$p_{X|Y}(1|0)=\frac{p(1,0)}{p_Y(0)}=\frac{3}{4},\qquad p_{X|Y}(0|1)=\frac{p(0,1)}{p_Y(1)}=\frac{3}{4}$$

所以

$$\begin{aligned}&H(X|Y)\\&=-\sum p(x_iy_j)\log_2 p(x_i|y_j)\\&=-p(0,0)\log_2 p(0|0)-p(0,1)\log_2 p(0|1)-p(1,0)\log_2 p(1|0)-p(1,1)\log_2 p(1|1)\\&=0.811\text{ 比特/符号}\end{aligned}$$

2.9　X **和** Y **是**$\{0,1,2,3\}$**上的独立、均匀分布的随机变量,求:**

(1) $H(X+Y)$, $H(X-Y)$, $H(X\cdot Y)$

(2) $H(X+Y,X-Y)$, $H(X+Y,X\cdot Y)$

解:

(1) **由于** X **和** Y **相互独立并且都是均匀分布,因此可求出** X **和** Y **之间**+, -, ×**运算的规则如表** 2.1 **所示。**

表 2.1

+	0	1	2	3
0	0	1	2	3
1	1	2	3	4
2	2	3	4	5
3	3	4	5	6

-	0	1	2	3
0	0	-1	-2	-3
1	1	0	-1	-2
2	2	1	0	-1
3	3	2	1	0

×	0	1	2	3
0	0	0	0	0
1	0	1	2	3
2	0	2	4	6
3	0	3	6	9

其中每个元素对应的概率都是 1/16,**由此可得**+, -, ×**运算结果的概率分布如表** 2.2**所示。**

表 2.2

运算	样本空间	概率分布
+	$\{0,1,2,3,4,5,6\}$	$\left\{\frac{1}{16},\frac{2}{16},\frac{3}{16},\frac{4}{16},\frac{3}{16},\frac{2}{16},\frac{1}{16}\right\}$
-	$\{-3,-2,-1,0,1,2,3\}$	$\left\{\frac{1}{16},\frac{2}{16},\frac{3}{16},\frac{4}{16},\frac{3}{16},\frac{2}{16},\frac{1}{16}\right\}$
×	$\{0,1,2,3,4,6,9\}$	$\left\{\frac{7}{16},\frac{1}{16},\frac{2}{16},\frac{2}{16},\frac{1}{16},\frac{2}{16},\frac{1}{16}\right\}$

再根据熵的定义可得

$$H(X+Y) = -\frac{2}{16}\log_2\frac{1}{16} - \frac{4}{16}\log_2\frac{2}{16} - \frac{6}{16}\log_2\frac{3}{16} - \frac{4}{16}\log_2\frac{4}{16}$$

$$= \log_2 16 - \frac{2}{16}\log_2 1 - \frac{4}{16}\log_2 2 - \frac{6}{16}\log_2 3 - \frac{4}{16}\log_2 4$$

$$= 4 - \frac{1}{8}\times 0 - \frac{1}{4}\times 1 - \frac{3}{8}\times 1.585\,0 - \frac{1}{4}\times 2$$

$$= 2.655\,6 \text{ 比特/符号}$$

$$H(X-Y) = H(X+Y) = 2.655\,6 \text{ 比特/符号}$$

$$H(X\cdot Y) = -\frac{7}{16}\log_2\frac{7}{16} - \frac{3}{16}\log_2\frac{1}{16} - \frac{6}{16}\log_2\frac{2}{16}$$

$$= \log_2 16 - \frac{7}{16}\log_2 7 - \frac{3}{16}\log_2 1 - \frac{6}{16}\log_2 2$$

$$= 4 - \frac{7}{16}\times 2.807\,4 - \frac{3}{16}\times 0 - \frac{3}{8}\times 1$$

$$= 2.396\,8 \text{ 比特/符号}$$

因为任何数和 0 相乘都等于 0,因此乘运算损失了信息,导致 $H(X\cdot Y) < H(X+Y)$。

(2) $X+Y$ 和 $X-Y$ 的运算如表 2.3 所示。

表 2.3

+, -	0	1	2	3
0	0,0	1, -1	2, -2	3, -3
1	1,1	2,0	3, -1	4, -2
2	2,2	3,1	4,0	5, -1
3	3,3	4,2	5,1	6,0

其中每个元素各不相同,对应的概率都是 1/16,所以

$$H(X+Y, X-Y) = -16\times\frac{1}{16}\log_2\frac{1}{16} = 4 \text{ 比特/符号}$$

或者说,由于 X 和 Y 也可以从 $X+Y$ 和 $X-Y$ 唯一地得到,所以它们的不确定性相同:

$$H(X+Y, X-Y) = H(XY) = H(X) + H(Y) = \log_2 4 + \log_2 4 = 2 + 2 = 4 \text{ 比特/符号}$$

$X+Y$ 和 $X\cdot Y$ 的运算如表 2.4 所示。

表 2.4

+,×	0	1	2	3
0	0,0	1,0	2,0	3,0
1	1,0	2,1	3,2	4,3
2	2,0	3,2	4,4	5,6
3	3,0	4,3	5,6	6,9

$$H(X+Y, X\cdot Y) = -\frac{4}{16}\log_2\frac{1}{16} - \frac{12}{16}\log_2\frac{2}{16} = \frac{1}{4}\log_2 16 + \frac{3}{4}\log_2 8 = \frac{4}{4} + \frac{9}{4} = 3.25\text{ 比特/符号}$$

因为 16 个元素中有 12 个是成对出现的,因此 X 和 Y 不可以从 $X+Y$ 和 $X\cdot Y$ 唯一地得到,因为方程组 $x+y=a$, $xy=b$ 通常有两个解,所以 $H(X+Y, X\cdot Y) < H(XY)$。

2.10 棒球比赛中大卫和麦克在前面的比赛中打平,最后 3 场与其他选手的比赛结果将最终决定他们的胜、负或平。

(1) 假定最后 3 场他们与其他选手的比赛结果胜负的可能性均为 0.5,把麦克的最终比赛结果{胜、负、平}作为随机变量,计算它的熵;

(2) 假定大卫最后 3 场比赛全部获胜,计算麦克的最终比赛结果的条件熵。

解:

(1) 当最后 3 场比赛麦克胜的次数比大卫多时,麦克最终才能胜,因此

$$\begin{aligned} P(\text{胜}) &= P(\text{麦克胜 3 场})\cdot P(\text{大卫胜少于 3 场}) + P(\text{麦克胜 2 场})\cdot P(\text{大卫胜少于 2 场}) + \\ &\quad P(\text{麦克胜 1 场})\cdot P(\text{大卫胜 0 场}) \\ &= \frac{1}{8}\times\frac{7}{8} + \frac{3}{8}\times\frac{4}{8} + \frac{3}{8}\times\frac{1}{8} \\ &= \frac{22}{64} \end{aligned}$$

同理

$$P(\text{负}) = \frac{22}{64}, \qquad P(\text{平}) = 1 - \frac{22}{64} - \frac{22}{64} = \frac{20}{64}$$

麦克最终比赛结果的熵为

$$\begin{aligned} H\left(\frac{22}{64},\frac{22}{64},\frac{20}{64}\right) &= -\frac{22}{64}\log_2\frac{22}{64} - \frac{22}{64}\log_2\frac{22}{64} - \frac{20}{64}\log_2\frac{20}{64} \\ &= \log_2 64 - 2\times\frac{22}{64}\log_2 22 - \frac{20}{64}\log_2 20 \\ &= 6 - \frac{44}{64}\times 4.4594 - \frac{20}{64}\times 4.3219 \\ &= 6 - 3.0659 - 1.3506 \\ &= 1.5835\text{ 比特/结果} \end{aligned}$$

因为胜、负、平这 3 种结果接近等概,所以该随机变量的熵接近最大熵。

(2) 假定大卫最后 3 场比赛全部获胜,那么麦克也必须全部获胜最后 3 场比赛最终才能得平,否则就是负。麦克 3 场比赛全部获胜的可能性是 $2^{-3}=1/8$,因此在假定大卫最后 3 场比赛全部获胜情况下麦克的最终比赛结果的条件熵是

$$H\left(\frac{1}{8}\right)=3-\frac{7}{8}\log_2 7=0.5436 \text{ 比特/结果}$$

2.11　X, Y, Z 为 3 个随机变量,证明以下不等式成立并指出等号成立的条件。

(1) $H(XY|Z)\geq H(X|Z)$

(2) $I(XY;Z)\geq I(X;Z)$

(3) $H(XYZ)-H(XY)\leq H(XZ)-H(X)$

(4) $I(X;Z|Y)\geq I(Z;Y|X)-I(Z;Y)+I(X;Z)$

解:

(1) $H(XY|Z)=H(X|Z)+H(Y|XZ)\geq H(X|Z)$

当 $H(Y|XZ)=0$,即 Y 是 X 和 Z 的函数时,原式等号成立。

(2) $I(XY;Z)=I(X;Z)+I(Y;Z|X)\geq I(X;Z)$

当 $I(Y;Z|X)=0$,即在给定 X 的情况下 Y 和 Z 统计独立时,原式等号成立。

(3)
$$\begin{aligned}H(XYZ)-H(XY)=H(Z|XY)&=H(Z|X)-I(Z;Y|X)\\&\leq H(Z|X)\\&=H(XZ)-H(X)\end{aligned}$$

当 $I(Z;Y|X)=0$,即在给定 X 的情况下 Y 和 Z 统计独立时,原式等号成立。

(4) 根据互信息的链规则:

$$I(X;Z|Y)+I(Z;Y)=I(XY;Z)=I(Z;Y|X)+I(X;Z)$$

因此

$$I(X;Z|Y)=I(Z;Y|X)-I(Z;Y)+I(X;Z)$$

即要证明的不等式其实是一个等式。

2.12　找出一个概率分布 $\{p_1, p_2, \cdots, p_5\}$,并且 $p_i>0$,使得 $H(p_1, p_2, \cdots, p_5)=2$。

解:

假定 p_1 为最大的概率。根据熵函数的性质,如果 $p_1>\frac{1}{2}$,则熵小于 2;如果 $p_1=\frac{1}{2}$,则只有一种可能:$\left\{\frac{1}{2},\frac{1}{8},\frac{1}{8},\frac{1}{8},\frac{1}{8}\right\}$。如果 $\frac{1}{4}<p_1<\frac{1}{2}$,则有无数个解,其中之一为 $\{0.375, 0.25, 0.25, 0.097859, 0.027141\}$(编程计算得出);如果 $p_1\leq\frac{1}{4}$,则没有解。

2.13 **有两个二元随机变量 X 和 Y,它们的联合概率分布如表 2.5 所列,同时定义另一随机变量 $Z = X \cdot Y$(一般乘积)。试计算:**

表 2.5

Y \ X	0	1
0	$\frac{1}{8}$	$\frac{3}{8}$
1	$\frac{3}{8}$	$\frac{1}{8}$

(1) **熵** $H(X), H(Y), H(Z), H(XZ), H(YZ)$**和** $H(XYZ)$;

(2) **条件熵** $H(X|Y), H(Y|X), H(X|Z), H(Z|X), H(Y|Z), H(Z|Y), H(X|YZ), H(Y|XZ)$**和** $H(Z|XY)$;

(3) **互信息** $I(X;Y), I(X;Z), I(Y;Z), I(X;Y|Z), I(Y;Z|X)$**和** $I(X;Z|Y)$。

解:

(1) **由** X, Y **的联合概率分布可求得** X **和** Y **的概率分布,如表 2.6 所示。**

表 2.6

(a)

X	0	1
$P(X)$	$\frac{1}{2}$	$\frac{1}{2}$

(b)

Y	0	1
$P(Y)$	$\frac{1}{2}$	$\frac{1}{2}$

可得

$$H(X) = H(Y) = -\frac{1}{2}\log_2\frac{1}{2} - \frac{1}{2}\log_2\frac{1}{2} = 1 \text{ 比特/符号}$$

由 $Z = X \cdot Y$ **可得**

$$p_Z(1) = P(X=1, Y=1) = \frac{1}{8}, \qquad p_Z(0) = 1 - p_Z(1) = \frac{7}{8}$$

因此 Z **的概率分布如表 2.7 所示。**

表 2.7

Z	0	1
$P(Z)$	$\frac{7}{8}$	$\frac{1}{8}$

所以

$$H(Z)=-\frac{1}{8}\log_2\frac{1}{8}-\frac{7}{8}\log_2\frac{7}{8}=0.54\ 比特/符号$$

XZ 和 YZ 的联合概率分布如表 2.8 所示。

表 2.8

(a)

XZ	00	10	11
$P(XZ)$	$\frac{1}{2}$	$\frac{3}{8}$	$\frac{1}{8}$

(b)

YZ	00	10	11
$P(YZ)$	$\frac{1}{2}$	$\frac{3}{8}$	$\frac{1}{8}$

所以

$$H(XZ)=H(YZ)=-\frac{1}{2}\log_2\frac{1}{2}-\frac{3}{8}\log_2\frac{3}{8}-\frac{1}{8}\log_2\frac{1}{8}=1.41\ 比特/符号$$

XYZ 的联合概率分布如表 2.9 所示。

表 2.9

XYZ	000	010	100	101
$P(XYZ)$	$\frac{1}{8}$	$\frac{3}{8}$	$\frac{3}{8}$	$\frac{1}{8}$

所以

$$H(XYZ)=2\times\left(-\frac{1}{2}\log_2\frac{1}{2}-\frac{3}{8}\log_2\frac{3}{8}\right)=1.81\ 比特/符号$$

(2) 由 X, Y 的联合概率分布可求出 $H(XY)=1.81$ 比特/符号,再根据各类熵之间的关系可得

$$H(X|Y)=H(XY)-H(Y)=0.81\ 比特/符号$$

同样可得

$$H(Y|X)=0.81\ 比特/符号$$

$$H(X|Z)=H(Y|Z)=0.87\ 比特/符号$$

$$H(Z|X)=H(Z|Y)=0.41\ 比特/符号$$

$$H(X|YZ)=H(Y|XZ)=0.4\ 比特/符号$$

$$H(Z|XY)=H(XYZ)-H(XY)=0\ 比特/符号$$

(3) $$I(X;Y)=H(X)-H(X|Y)=1-0.81=0.19\ 比特/符号$$

同样可得

$$I(X;Z)=I(Y;Z)=0.13\ 比特/符号$$

$$I(X;Y|Z) = H(X|Z) - H(X|YZ) = 0.87 - 0.4 = 0.47 \text{ 比特/符号}$$

同样可得

$$I(X;Z|Y) = I(Y;Z|X) = 0.41 \text{ 比特/符号}$$

2.14 假定 $X_1 \to X_2 \to X_3 \to \cdots \to X_n$ 形成一个马尔可夫链，那么 $p(x_1 x_2 \cdots x_n) = p(x_1)p(x_2|x_1)\cdots p(x_n|x_{n-1})$，请化简 $I(X_1;X_2\cdots X_n)$。

解：

根据马尔可夫链的特性，已知现在，则过去与将来无关，所以

$$\begin{aligned} I(X_1;X_2\cdots X_n) &= H(X_1) - H(X_1|X_2\cdots X_n) \\ &= H(X_1) - H(X_1|X_2) \\ &= I(X_1;X_2) \end{aligned}$$

2.15 给定 X，Y 的联合概率分布，如表 2.10 所示。求：

表 2.10

Y \ X	0	1
0	$\frac{1}{3}$	$\frac{1}{3}$
1	0	$\frac{1}{3}$

(1) $H(X)$，$H(Y)$

(2) $H(X|Y)$，$H(Y|X)$

(3) $H(XY)$

(4) $H(Y) - H(Y|X)$

(5) $I(X;Y)$

解：

由 X，Y 的联合概率分布求出 X，Y 的边缘概率分布，如表 2.11 所示。

表 2.11

(a)

X	0	1
$P(X)$	$\frac{2}{3}$	$\frac{1}{3}$

(b)

Y	0	1
$P(Y)$	$\frac{1}{3}$	$\frac{2}{3}$

(1) $H(X)=\frac{2}{3}\log_2\frac{3}{2}+\frac{1}{3}\log_2 3=0.918=H(Y)$

(2) $H(X|Y)=\frac{1}{3}H(X|Y=0)+\frac{2}{3}H(X|Y=1)=0.667=H(Y|X)$

(3) $H(XY)=3\times\frac{1}{3}\log_2 3=1.585$ 比特/符号

(4) $H(Y)-H(Y|X)=0.251$ 比特/符号

(5) $I(X;Y)=H(Y)-H(Y|X)=0.251$ 比特/符号

2.16　(1) 假定 X 是一个离散随机变量,$g(X)$是 X 的函数,证明: $H[g(X)]\leq H(X)$。

(2) 假定 X 是一个定义在$\{0,1,2,3,4\}$上的等概分布的离散随机变量, $g(X)=\cos\frac{\pi X}{2}$, $f(X)=x^2$,比较它们的熵的大小。

证明:

(1)

$$H[X,g(X)]=H(X)+H[g(X)|X]=H(X)$$

$$H[X,g(X)]=H[g(X)]+H[X|g(X)]\geq H[g(X)]$$

所以 $H[g(X)]\leq H(X)$。当且仅当 $g(\cdot)$是一个一一映射时, $H[X|g(X)]=0$,等号成立。

(2) $H(X)=\log_2 5$ 比特/符号,因为 $f(\cdot)$是一个一一映射,所以 $H[f(X)]=H(X)$。$g(\cdot)$不是一个一一映射,所以 $H[g(X)]<H(X)$。$g(X)$有 3 种取值$\{-1,0,1\}$,对应的概率分别为$\frac{1}{5}$,$\frac{2}{5}$,$\frac{2}{5}$,因此

$$H[g(X)]=\log_2 5-\frac{4}{5}=1.5219 \text{ 比特/符号}$$

2.17　考虑两个发射机和一个接收机之间的平均联合互信息 $I(X_1X_2;Y)$,证明:

(1) $I(X_1X_2;Y)\geq I(X_1;Y)$,也就是用两台发射机比用一台发射机的效果好;

(2) 如果 X_1 和 X_2 相互独立,那么 $I(X_2;Y|X_1)\geq I(X_2;Y)$;

(3) 如果 X_1 和 X_2 相互独立,那么 $I(X_1X_2;Y)\geq I(X_1;Y)+I(X_2;Y)$,也就是同时用两台发射机比分别用两台发射机的效果好。

证明:

(1) $I(X_1X_2;Y)=I(X_1;Y)+I(X_2;Y|X_1)\geq I(X_1;Y)$

(2) 如果 X_1 和 X_2 相互独立,那么

$$
\begin{aligned}
I(X_2;Y|X_1) &= H(X_2|X_1) - H(X_2|YX_1)\\
&= H(X_2) - H(X_2|YX_1)\\
&\geq H(X_2) - H(X_2|Y)\\
&= I(X_2;Y)
\end{aligned}
$$

(3) **如果** X_1 **和** X_2 **相互独立,利用**(2)**的结果,可以得到**:

$$I(X_1X_2;Y) = I(X_1;Y) + I(X_2;Y|X_1) \geq I(X_1;Y) + I(X_2;Y)$$

2 .18　**在一个布袋中有** 3 **枚硬币,分别用** H、T、F **表示**, H **的两面都是正面**, T **的两面都是反面,而** F **是一个一正一反的均匀硬币。随机选择一枚硬币并投掷两次,用** X **表示所选择的硬币**, Y_1, Y_2 **表示两次投掷的结果**, Z **表示两次投掷中出现正面的次数。求**:

(1) $I(X;Y_1)$

(2) $I(X;Z)$

(3) $I(Y_1;Y_2)$

解:

(1) **先求** Y_1 **的条件和非条件概率分布**:

$$P(Y_1 = \mathrm{H}|X = \mathrm{H}) = 1$$

$$P(Y_1 = \mathrm{H}|X = \mathrm{F}) = \frac{1}{2}$$

$$P(Y_1 = \mathrm{H}|X = \mathrm{T}) = 0$$

$$P(Y_1 = \mathrm{H}) = \frac{1}{3}\times\frac{2}{2} + \frac{1}{3}\times\frac{1}{2} + \frac{1}{3}\times\frac{0}{2} = \frac{3}{6} = \frac{1}{2}$$

因此

$$H(Y_1) = 1 \text{ 比特/ 符号},\ H(Y_1|X) = \frac{1}{3}H\left(\frac{1}{2}\right) = \frac{1}{3} \text{ 比特/ 符号}$$

$$I(X;Y_1) = H(Y_1) - H(Y_1|X) = \frac{2}{3} \text{ 比特/ 符号}$$

(2) **先求条件概率分布** $P(Z|X)$**和联合概率分布** $P(XZ)$,**如表** 2 .12 **所示。**

表 2 .12

(a)

$P(Z\|X)$	0	1	2
H	0	0	1
F	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{4}$
T	1	0	0

(b)

$P(XZ)$	0	1	2
H	0	0	$\frac{1}{3}$
F	$\frac{1}{12}$	$\frac{1}{6}$	$\frac{1}{12}$
T	$\frac{1}{3}$	0	0

从表 2.12 可以得到 Z 的边缘概率分布为$\left\{\frac{5}{12},\frac{1}{6},\frac{5}{12}\right\}$,因此

$$H(Z|X)=\frac{1}{3}\times 0+\frac{1}{3}\times\frac{3}{2}+\frac{1}{3}\times 0=\frac{1}{2}=0.500\,0\text{ 比特/符号}$$

$$H(Z)=H\left(\frac{5}{12},\frac{1}{6},\frac{5}{12}\right)=H\left(\frac{1}{6}\right)+\frac{10}{12}H\left(\frac{1}{2}\right)=0.650\,0+0.813\,3=1.483\,3\text{ 比特/符号}$$

$$I(X;Z)=H(Z)-H(Z|X)=1.483\,3-0.500\,0=0.983\,3\text{ 比特/符号}$$

(3) 先求 Y_1, Y_2 的联合概率分布:

$$\underset{P(Y_1Y_2|X=\mathrm{H})}{\frac{1}{3}\begin{bmatrix}1&0\\0&0\end{bmatrix}}+\underset{P(Y_1Y_2|X=\mathrm{F})}{\frac{1}{3}\begin{bmatrix}\frac{1}{4}&\frac{1}{4}\\\frac{1}{4}&\frac{1}{4}\end{bmatrix}}+\underset{P(Y_1Y_2|X=\mathrm{T})}{\frac{1}{3}\begin{bmatrix}0&0\\0&1\end{bmatrix}}=\underset{P(Y_1Y_2)}{\begin{bmatrix}\frac{5}{12}&\frac{1}{12}\\\frac{1}{12}&\frac{5}{12}\end{bmatrix}}$$

因此可以求出:

$$I(Y_1;Y_2)=H(Y_2)-H(Y_2|Y_1)=1-H\left(\frac{1}{6}\right)=1-0.650\,0=0.350\,0\text{ 比特/符号}$$

2.19　猜宝游戏。3 扇门中有一扇门后藏有一袋金子,并且 3 扇门后面藏有金子的可能性相同。如果有人随机打开一扇门并告诉你门后是否藏有金子,他给了你多少关于金子位置的信息量?

解:

假定 $X\in\{1,2,3\}$表示藏金子的位置, $Y\in\{1,2,3\}$表示这个人开的门,显然 X 和 Y 是相互独立的。如果告知这个人开的门后有金子,那么关于 X 的不确定性降为 0,否则,其余两扇门后有金子的可能性均为 1/2,不确定性为 1 bit。

$$\begin{aligned}H(X|Y)&=P(X=Y)H(X|Y,X=Y)+P(X\neq Y)H(X|Y,X\neq Y)\\&=\frac{1}{3}\times 0+\frac{2}{3}\times H\left(\frac{1}{2},\frac{1}{2}\right)\\&=\frac{2}{3}\times 1=\frac{2}{3}\text{ 比特/符号}\end{aligned}$$

因此

$$I(X;Y)=H(X)-H(X|Y)=\log_2 3-\frac{2}{3}=1.585\,0-0.666\,7=0.918\,3\text{ 比特/符号}$$

2.20　一个年轻人研究了当地的天气纪录和气象台的预报纪录后,得到实际天气和预报天气的联合概率分布,如表 2.13 所示。他发现预报只有 12/16 的准确率,而不管三

七二十一都预报明天不下雨的准确率却是 13/ 16。**他把这个想法跟气象台台长说了后，台长却说他错了。请问这是为什么**？

表 2 .13

实际 / 预报	下雨	不下雨
下雨	1/ 8	3/ 16
不下雨	1/ 16	10/ 16

解:

假设 X **表示当地的实际天气情况**, Y **表示气象台预报的天气情况**, Z **表示总是预报不下雨的天气情况。**

$$H(X) = 0.696 \text{ 比特/ 符号}$$

$$I(X;Y) = \sum_{x,y} p(xy)\log_2 \frac{p(xy)}{p(x)p(y)}$$

$$= \frac{1}{8}\log_2 \frac{\frac{1}{8}}{\frac{3}{16}\times\frac{5}{16}} + \frac{3}{16}\log_2 \frac{\frac{3}{16}}{\frac{5}{16}\times\frac{13}{16}} + \frac{1}{16}\log_2 \frac{\frac{1}{16}}{\frac{3}{16}\times\frac{11}{16}} + \frac{10}{16}\log_2 \frac{\frac{10}{16}}{\frac{11}{16}\times\frac{13}{16}}$$

$$= 0.0906 \text{ 比特/ 符号}$$

$I(X;Y) \ll H(X)$, **可见气象台预报的确实不好。**

但是如果总是预报不下雨的话则会更糟,因为 X **和** Z **是相互独立的两个随机变量,即** $I(X;Z) = 0$, **所以**

$$I(X;Y) > I(X;Z), \qquad H(X|Z) > H(X|Y)$$

因此气象台的预报准确率虽然比总是预报不下雨低,但还是传递了一些信息,消除了一些不确定性。

2 .21 **设** $X_1, X_2, \cdots, X_N$ **为一个独立的贝努利随机变量序列,其分布为** $P(X_i = 0) = p$, $P(X_i = 1) = 1 - p$, **求**:

(1) **使** $S_2 = X_1 + X_2$ **的熵** $H(S_2)$ **取得最大值的** p **值**;

(2) **设** $p = \frac{1}{2}$, **求二项式随机变量** $S_n = X_1 + X_2 + \cdots + X_n$ **的熵** $H(S_n)$。

解:

(1) 随机变量 S_2 的概率空间为$\begin{bmatrix} 0 & 1 & 2 \\ p^2 & 2pq & q^2 \end{bmatrix}$,其中 $q = 1 - p$,根据熵的定义:

$$\begin{aligned} H(S_2) &= -[p^2\log_2 p^2 + 2pq\log_2(2pq) + q^2\log_2 q^2] \\ &= -(2p^2\log_2 p + 2pq\log_2 p + 2pq\log_2 q + 2q^2\log_2 q + 2pq\log_2 2) \\ &= -[2p(p\log_2 p + q\log_2 q) + 2q(p\log_2 p + q\log_2 q) + 2pq\log_2 2] \\ &= 2H(p) - 2pq \end{aligned}$$

$H(S_2)$对 p 求导,得

$$\frac{\mathrm{d}}{\mathrm{d}p}H(S_2) = 2\left[\log_2\frac{1-p}{p} - (1-2p)\right]$$

方括号内是一个单调递减函数,可以看出当 $p = \frac{1}{2}$时上式为 0,这时 $H(S_2)$取得最大值为 1.5 比特/ 符号。

(2) $p = \frac{1}{2}$时 S_n 的概率分布为 $P(S_n = k) = 2^{-n}\cdot C_n^k$,其熵为

$$\begin{aligned} H(S_n) &= -\sum_{k=0}^{n} 2^{-n}C_n^k\log_2(2^{-n}C_n^k) \\ &= -\sum_{k=0}^{n} 2^{-n}C_n^k(-n) - 2^{-n}\sum_{k=0}^{n} C_n^k\log_2 C_n^k \\ &= n - 2^{-n}\sum_{k=0}^{n} C_n^k\log_2 C_n^k \end{aligned}$$

2.22　投掷一枚均匀的硬币直到出现两次正面或两次反面。用 X_1, X_2 表示头两次投掷的结果, Y 表示最后一次投掷的结果, N 表示投掷的次数。计算 $H(X_1)$, $H(X_2)$, $H(Y)$, $H(N)$, $I(X_1;Y)$, $I(X_2;Y)$, $I(X_1X_2;Y)$, $I(X_1;N)$, $I(X_2;N)$, $I(X_1X_2;N)$。

解:

由于是独立投掷均匀的硬币,所以

$$H(X_1) = H(X_2) = H(Y) = 1 \text{ 比特/ 符号}$$

如果头两次投掷的结果相同,则 $N = 2$,否则, $N = 3$,所以 N 的取值为 2 或者 3。

$N = 2$ 的概率为

$$p_N(2) = p_{X_1}(0)p_{X_2}(0) + p_{X_1}(1)p_{X_2}(1) = \frac{1}{2}\times\frac{1}{2} + \frac{1}{2}\times\frac{1}{2} = \frac{1}{4} + \frac{1}{4} = \frac{1}{2}$$

因此

$$p_N(3)=\frac{1}{2},\qquad H(N)=1\text{ 比特/符号}$$

如果 X_1 是正面,那么 Y 必定是正面,除非 X_2 和 Y 同时为反面,所以

$$p_{Y|X_1}(0|1)=\frac{1}{4},\qquad p_{Y|X_1}(1|0)=\frac{1}{4}$$

$$H(Y|X_1)=\frac{1}{2}H\left(\frac{1}{4}\right)+\frac{1}{2}H\left(\frac{3}{4}\right)=\frac{1}{2}\left[H\left(\frac{1}{4}\right)+H\left(\frac{1}{4}\right)\right]=H\left(\frac{1}{4}\right)=0.8113\text{ 比特/符号}$$

$$I(X_1;Y)=H(Y)-H(Y|X_1)=1-H\left(\frac{1}{4}\right)=0.1887\text{ 比特/符号}$$

同理

$$I(X_2;Y)=I(X_1;Y)=0.1887\text{ 比特/符号}$$

由 X_1,X_2,Y 的联合概率分布(如表 2.14 所示)可以求 $H(Y|X_1X_2)$。

表 2.14

Y \ X_1X_2	00	01	10	11
0	1	1/2	1/2	0
1	0	1/2	1/2	1

$$H(Y|X_1X_2)=\frac{1}{4}H\left(\frac{1}{2}\right)+\frac{1}{4}H\left(\frac{1}{2}\right)=\frac{1}{4}+\frac{1}{4}=\frac{1}{2}\text{ 比特/符号}$$

所以

$$I(X_1X_2;Y)=H(Y)-H(Y|X_1X_2)=1-\frac{1}{2}=\frac{1}{2}\text{ 比特/符号}$$

投掷的次数 N 与单独的 X_1 和 X_2 无关,因此 $I(X_1;N)=I(X_2;N)=0$ 比特/符号。但是 X_1 和 X_2 都确定以后,N 就可以确定。当且仅当 $X_1=X_2$ 时,$N=2$,因此 $I(X_1X_2;N)=H(N)=1$ 比特/符号。

2.23 判断题

(1) $H(X)>0$;

(2) 若 X 与 Y 独立,则 $H(X)=H(X|Y)$;

(3) $I(X;Y)\geq I(X;Y|Z)$;

(4) 如果 $H(X|YZ)=0$,则要么 $H(X|Y)=0$,要么 $H(X|Z)=0$;

(5) $I(X;Y)\leq H(Y)$;

(6) $H(X|X)=0$;

(7) 若 X 与 Y 独立,则 $H(Y|X)=H(X|Y)$;

(8) $H(X|Y) \geq H(X|YZ)$。

解:

(1) F　当 X 只有一个可能的结果时, $H(X)=0$。

(2) T　若 X 与 Y 相互独立,则 $p(x)=p(x|y)$,因此对于任意 y 值均有 $H(X|Y=y)=H(X)$,所以 $H(X)=H(X|Y)$。

(3) F　给出一个反例。假定 X 和 Y 为相互独立的二元随机变量, $Z = X \quad Y$,则 $I(X;Y)=0$, $I(X;Y|Z)=H(X|Z)-H(X|YZ)=1-0=1$,因此在本例中 $I(X;Y)<I(X;Y|Z)$。

(4) F　给出一个反例。如果用 X 表示棋子所在的位置, Y 表示棋盘的横格, Z 表示棋盘的纵格,棋子所在的位置由横格和纵格共同决定, $H(X|YZ)=0$,但不能说棋子所在的位置要么由横格决定 $H(X|Y)=0$,要么由纵格决定 $H(X|Z)=0$。

(5) T　$I(X;Y)=H(Y)-H(Y|X) \leq H(Y)$,因为 $H(Y|X) \geq 0$。

(6) T　对于任意 x 值有 $H(X|X=x)=0$。

(7) F　若 X 与 Y 独立,则 $H(Y|X)=H(Y)$, $H(X|Y)=H(X)$。

(8) T　增加条件可以减少不确定性。$H(X|Y)-H(X|YZ)=I(X;Z|Y) \geq 0$。

2.24　设随机变量 X 的概率分布为$\left\{\frac{2}{10},\frac{2}{10},\frac{2}{10},\frac{1}{10},\frac{1}{10},\frac{1}{10},\frac{1}{10}\right\}$。随机变量 Y 是 X 的函数,其分布为将 X 的 4 个最小的概率分布合并为一个:$\left\{\frac{2}{10},\frac{2}{10},\frac{2}{10},\frac{4}{10}\right\}$。

(1) 显然 $H(X) \leq \log_2 7$,请解释原因;

(2) 请解释为什么 $H(X) > \log_2 5$?

(3) 计算 $H(X)$, $H(Y)$;

(4) 计算 $H(X|Y)$并解释其结果。

解:

(1) 根据熵的极值性,当随机变量等概分布时,随机变量的熵最大。有 7 个可能取值的随机变量的最大熵为 $\log_2 7$,随机变量 X 不是等概分布,所以 $H(X) \leq \log_2 7$。

(2) 根据熵的递增性, $H(X)=H\left(\frac{2}{10},\frac{2}{10},\frac{2}{10},\frac{2}{10},\frac{2}{10}\right)+\frac{2}{10}H\left(\frac{1}{2}\right)+\frac{2}{10}H\left(\frac{1}{2}\right)>\log_2 5$。

$$
\begin{aligned}
(3)\quad H(X) &= -\sum_x p(x)\log_2 p(x) = -3\times\frac{2}{10}\log_2\frac{2}{10} - 4\times\frac{1}{10}\log_2\frac{1}{10} \\
&= \log_2 10 - \frac{6}{10}\log_2 2 = 3.322 - 0.6 \\
&= 2.722 \text{ 比特/符号}
\end{aligned}
$$

$$H(Y)=-\sum_{y}p(y)\log_2 p(y)=-3\times\frac{2}{10}\log_2\frac{2}{10}-\frac{4}{10}\log_2\frac{4}{10}$$
$$=\log_2 10-\frac{6}{10}\log_2 2-\frac{4}{10}\log_2 4$$
$$=3.322-0.6-0.8$$
$$=1.922 \text{ 比特/符号}$$

(4) 因为随机变量 Y 是 X 的函数,所以

$$H(Y|X)=0 \text{ 比特/符号}$$

$$H(X|Y)=H(XY)-H(Y)=H(X)+H(Y|X)-H(Y)=0.8 \text{ 比特/符号}$$

2.25 已知 $H(Y|X)=0$,求证 " x, $p(x)>0$,只存在一个 y 使得 $p(xy)>0$。

证明:

反证法。

设对某个 x_0,存在两个 Y 的取值 y_1 和 y_2 使得 $p(x_0y_1)>0$ 且 $p(x_0y_2)>0$,则 $p(x_0)=p(x_0y_1)+p(x_0y_2)+\cdots>p(x_0y_1)>0$,同理 $p(x_0)>p(x_0y_2)>0$。因此条件概率 $p(y_1|x_0)$和 $p(y_2|x_0)$都有定义且不为 0 也不为 1,所以

$$H(Y|X)=-\sum_{x}\sum_{y}p(xy)\log_2 p(y|x)>-p(x_0y_1)\log_2 p(y_1|x_0)>0$$

2.26 猜宝游戏。3 扇门中有一扇门后藏有一袋金子,并且 3 扇门后面藏有金子的可能性相同。你选择其中一扇门,主持人会打开后面没藏有金子的另一扇门(如果你选择的门后藏有金子,则主持人会在另两扇门中任意打开一扇门)。主持人给了你多少关于金子位置的信息?

解:

假定 $X\in\{1,2,3\}$表示藏金子的位置, $Y\in\{1,2,3\}$表示你选择的门, Z 表示主持人打开的门,显然 X 和 Y 是相互独立的,但是 Z 和 X、Z 和 Y 都不相互独立,因为 $z\neq x$ 并且 $z\neq y$。对于任意一个 y 值和一个 $z\neq y$,有

$$P(X=y|Z=z)=\frac{P(X=y,Z=z)}{P(Z=z)}$$
$$=\frac{P(X=y)P(Z=z|X=y)}{P(Z=z)}$$
$$=\frac{\frac{1}{3}\times\frac{1}{2}}{\frac{1}{2}}=\frac{1}{3}$$

换句话说,如果主持人打开一扇空门,那么你选择的门后面藏有金子的可能性是 1/3,因此你应该打开另外一扇门。因为对于任意一个 y 值和一个 $z\neq y$,有

$P(X = y \mid Z = z) = 1/3$，所以

$$H(X \mid Z) = H\left(\frac{1}{3}\right) = 0.9183 \text{ 比特/符号}$$

$$I(X;Z) = H(X) - H(X \mid Z) = \log_2 3 - H\left(\frac{1}{3}\right) = 1.5850 - 0.9183 = 0.6667 \text{ 比特/符号}$$

2.27　在一个布袋中有 r 个红球，w 个白球，b 个黑球，从布袋中取 $k \geq 2$ 个球，每次取出球后放回还是每次取出球后不放回的熵 $H(X_i \mid X_{i-1} \cdots X_1)$ 更大？

解：

(1) 每次取球后放回

这种情况下每次取球的条件概率分布都相同。

$$X_i = \begin{cases} \dfrac{r}{r+w+b} & \text{红球} \\ \dfrac{w}{r+w+b} & \text{白球} \\ \dfrac{b}{r+w+b} & \text{黑球} \end{cases}$$

因此

$$H(X_i \mid X_{i-1} \cdots X_1)$$
$$= H(X_i) = \log_2(r+w+b) - \frac{r}{r+w+b}\log_2 r - \frac{w}{r+w+b}\log_2 w - \frac{b}{r+w+b}\log_2 b$$

(2) 每次取球后不放回

每次取球后的无条件概率分布为

$$X_i = \begin{cases} \dfrac{r}{r+w+b} & \text{红球} \\ \dfrac{w}{r+w+b} & \text{白球} \\ \dfrac{b}{r+w+b} & \text{黑球} \end{cases}$$

因此无条件熵为

$$H(X_i) = \log_2(r+w+b) - \frac{r}{r+w+b}\log_2 r - \frac{w}{r+w+b}\log_2 w - \frac{b}{r+w+b}\log_2 b$$

而条件熵为

$$H(X_i \mid X_{i-1} \cdots X_1) < H(X_i)$$

因此不放回的熵更小。

2.28　X，Y_1，Y_2 为二元随机变量，如果 $I(X;Y_1) = 0$ 并且 $I(X;Y_2) = 0$，能不能推出 $I(X;Y_1Y_2) = 0$，如果能请证明，如果不能请给出反例。

解:

上述推理不成立。尽管 Y_1 不能提供关于 X 的信息, Y_2 也不能提供关于 X 的信息,但不能说 Y_1Y_2 两个一起也不能提供关于 X 的信息。因为 $I(X;Y_1Y_2) \neq I(X;Y_1) + I(X;Y_2)$,**而** $I(X;Y_1Y_2) = I(X;Y_1) + I(X;Y_2 \mid Y_1)$。

在概率论中已经知道,3 个随机变量两两独立不等于它们相互独立。 $I(X;Y_1) = 0$ **说明 X 和 Y_1 相互独立,** $I(X;Y_2) = 0$ **说明 X 和 Y_2 相互独立,但不能说 X 和 (Y_1, Y_2) 相互独立。例如, Y_1, Y_2 表示独立地投掷两个均匀硬币的结果,** $X = Y_1 \quad Y_2$,**因此 X 和 Y_1 相互独立, X 和 Y_2 相互独立,但显然 X 和 (Y_1, Y_2) 不相互独立,因为一旦 (Y_1, Y_2) 确定,则 X 也就确定了。**

2.29 **X 是一个几何分布的随机变量,求它的熵。**

解:

因为 X 是一个几何分布的随机变量,所以

$$p_X(k) = p(1-p)^{k-1}, \qquad k = 1,2,3,\cdots$$

根据熵的定义可得

$$\begin{aligned}
H(X) &= -\sum_{k=1}^{\infty} p_X(k) \cdot \log_2[p_X(k)] \\
&= -\sum_{k=1}^{\infty} p(1-p)^{k-1} \cdot \log_2[p(1-p)^{k-1}] \\
&= -p\log_2 p \cdot \sum_{k=1}^{\infty}(1-p)^{k-1} - p\log_2(1-p) \cdot \sum_{k=1}^{\infty}(k-1) \cdot (1-p)^{k-1} \\
&= -p\log_2 p \cdot S_1 - p\log_2(1-p) \cdot S_2
\end{aligned}$$

$$S_1 = \sum_{k=1}^{\infty}(1-p)^{k-1} = \frac{1}{p}$$

$$\begin{aligned}
S_2 &= \sum_{k=1}^{\infty}(k-1) \cdot (1-p)^{k-1} = \sum_{k=2}^{\infty}(k-1) \cdot q^{k-1} = q \cdot \sum_{k=2}^{\infty}(k-1) \cdot q^{k-2} \\
&= q \cdot \frac{\mathrm{d}}{\mathrm{d}q}\left(\sum_{k=2}^{\infty} q^{k-1}\right) = q \cdot \frac{\mathrm{d}}{\mathrm{d}q}\left(\frac{q}{1-q}\right) \\
&= \frac{q}{(1-q)^2} = \frac{1-p}{p^2}
\end{aligned}$$

$$H(X) = -\log_2 p - \frac{1-p}{p}\log_2(1-p)$$

2.30　**人口问题。在某个地区,一对夫妻只允许生一个孩子,可是这里所有的夫妻都希望能生一个男孩传宗接代,因此这里的夫妻都会一直生到生了一个男孩为止,假定生男生女的概率相同,问:**

(1) **这个地区男孩会多于女孩吗?**

(2) **一个家庭孩子的个数用离散随机变量 X 表示,计算 X 的熵。**

解:

(1) **假定一个家庭里有 k 个女孩,1 个男孩,相应的概率是** $0.5^k \cdot 0.5$,**因此女孩的平均数是** $0.5\sum_{k=1}^{\infty} k0.5^k = 1$,**女孩的平均数和男孩的平均数相等。**

(2)
$$H(X) = -\sum_{i=1}^{\infty} 0.5^i \log_2(0.5^i) = 2$$

2.31　**就业问题。假如政府的就业问题顾问在考虑全国的就业问题时,把全体国民的就业情况分为 3 类:全就业(100%就业),部分就业(50%就业),失业(0%就业),分别用概率 $p(E)$,$p(F)$,$p(U)$表示,要使全民的就业率达到 95%,请问:**

(1) $p(E)$**的取值范围;**

(2) **就业情况的熵作为 $p(E)$的函数,画出它在 $p(E)$的取值范围内的曲线;**

(3) **求就业情况熵的最大值。**

解:

(1) **根据题意,可以得到:**

$$p(E) + p(F) + p(U) = 1 \tag{2.1}$$

$$1.0p(E) + 0.5p(F) + 0.0p(U) = 0.95 \tag{2.2}$$

由式(2.2)**可以得到:**

$$p(F) = 1.9 - 2p(E) \tag{2.3}$$

将式(2.3)**代入式**(2.2)**得到:**

$$p(E) = 0.9 + p(U) \tag{2.4}$$

由于 $p(E)$,$p(F)$,$p(U)$的取值必须在 0 到 1 之间,由式(2.3)**和式**(2.4)**可以得到 $p(E)$的取值范围在 0.9 到 0.95 之间。**

(2) **就业情况的熵为**

$$\begin{aligned} H &= p(E)\log_2\left[\frac{1}{p(E)}\right] + p(F)\log_2\left[\frac{1}{p(F)}\right] + p(U)\log_2\left[\frac{1}{p(U)}\right] \\ &= p(E)\log_2\left[\frac{1}{p(E)}\right] + [1.9 - 2p(E)]\log_2\left[\frac{1}{1.9 - 2p(E)}\right] + \\ &\quad [p(E) - 0.9]\log_2\left[\frac{1}{p(E) - 0.9}\right] \end{aligned}$$

它在 $p(E)$**的取值范围内的曲线如图** 2.1 **所示。**

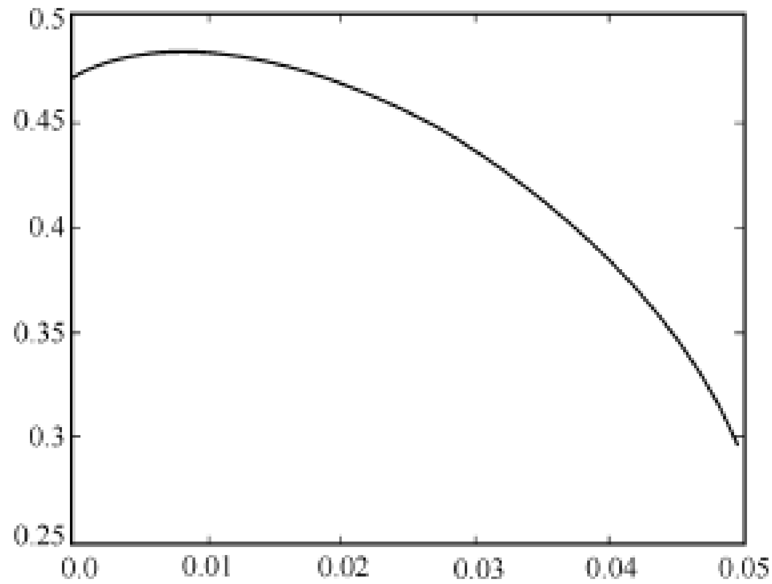


图 2.1

(3) **当** $p(E) = 0.9081$ **时**, $H = 0.4823$ **达到最大值**,**这时** $p(F) = 0.0838$, $p(U) = 0.0081$。

第 3 章

信源及信源熵

3.1　**证明** $\lim\limits_{n\to\infty}\frac{1}{2}H(X_nX_{n-1}\mid X_1\cdots X_{n-2})=H_\infty$。

证明:

$$\begin{aligned}
&\lim_{n\to\infty}\frac{1}{2}H(X_n X_{n-1}\mid X_1\cdots X_{n-2})\\
=&\lim_{n\to\infty}\left[\frac{1}{2}H(X_n\mid X_1\cdots X_{n-1})+\frac{1}{2}H(X_{n-1}\mid X_1\cdots X_{n-2})\right]\\
=&\frac{1}{2}\lim_{n\to\infty}H(X_n\mid X_1\cdots X_{n-1})+\frac{1}{2}\lim_{n\to\infty}H(X_{n-1}\mid X_1\cdots X_{n-2})\\
=&\frac{1}{2}H_\infty+\frac{1}{2}H_\infty=H_\infty
\end{aligned}$$

3.2　**有一无记忆信源的符号集为**$\{0,1\}$,**已知信源的概率空间为**$\begin{bmatrix}X\\P\end{bmatrix}=\begin{bmatrix}0 & 1\\ \frac{1}{4} & \frac{3}{4}\end{bmatrix}$。

(1) **求信源熵**;

(2) **求由** m **个"**0**"和**$(100-m)$**个"**1**"构成的某一特定序列自信息量的表达式**;

(3) **计算由** 100 **个符号构成的符号序列的熵。**

解:

(1) **信源熵为**

$$H(X)=\frac{1}{4}\log_2 4+\frac{3}{4}\log_2\frac{4}{3}=0.8113\ \text{比特/符号}$$

(2) 该特定序列用 A 表示,则

$$I(A)=-\log_2\left(\frac{1}{4}\right)^m\left(\frac{3}{4}\right)^{(100-m)}\approx 41.5+1.585m\ \text{bit}$$

(3) 因为信源是无记忆信源,所以

$$H(X^{100})=100H(X)=81.13\ \text{比特/符号序列}$$

3.3 有一离散无记忆信源,其输出为 $X\in\{0,1,2\}$,相应的概率为 $p_0=\frac{1}{4}$, $p_1=\frac{1}{4}$, $p_2=\frac{1}{2}$,设计两个独立实验去观察它,其结果分别为 $Y_1\in\{0,1\}$, $Y_2\in\{0,1\}$,已知条件概率如表 3.1 所列。

表 3.1

$P(Y_1\mid X)$	0	1
0	1	0
1	0	1
2	1/2	1/2

$P(Y_2\mid X)$	0	1
0	1	0
1	1	0
2	0	1

(1) 求 $I(X;Y_1)$和 $I(X;Y_2)$,并判断哪一个实验好些;

(2) 求 $I(X;Y_1Y_2)$,并计算做 Y_1 和 Y_2 两个实验比做 Y_1 或 Y_2 中的一个实验可多得多少关于 X 的信息;

(3) 求 $I(X;Y_1\mid Y_2)$和 $I(X;Y_2\mid Y_1)$,并解释它们的含义。

解:

(1) $I(X;Y_1)=H(Y_1)-H(Y_1\mid X)$,要求 $H(Y_1)$和 $H(Y_1\mid X)$需要先求 $P(Y_1)$、$P(XY_1)$, $P(Y_1\mid X)$已知。

$I(X;Y_2)=H(Y_2)-H(Y_2\mid X)$,要求 $H(Y_2)$和 $H(Y_2\mid X)$需要先求 $P(Y_2)$、$P(XY_2)$, $P(Y_2\mid X)$已知。

由 $P(XY_1)=P(X)P(Y_1\mid X)$及联合概率分布与边缘概率分布的关系可得 $P(XY_1)$及 $P(Y_1)$,如表 3.2 所示,所以

$$H(Y_1)=\frac{1}{2}\log_2 2+\frac{1}{2}\log_2 2=1\ \text{比特/符号}$$

$$H(Y_1\mid X)=\frac{1}{4}\log_2 1+\frac{1}{4}\log_2 1+\frac{1}{4}\log_2 2+\frac{1}{4}\log_2 2=\frac{1}{2}\ \text{比特/符号}$$

$$I(X;Y_1)=H(Y_1)-H(Y_1|X)=1-\frac{1}{2}=\frac{1}{2}\text{ 比特/符号}$$

同样可求出 $P(XY_2)$**和** $P(Y_2)$,**如表** 3.3 **所示**。

表 3.2

$P(XY_1)$ X \ Y_1	0	1
0	$\frac{1}{4}$	0
1	0	$\frac{1}{4}$
2	$\frac{1}{4}$	$\frac{1}{4}$
$P(Y_1)$	$\frac{1}{2}$	$\frac{1}{2}$

表 3.3

$P(XY_2)$ X \ Y_2	0	1
0	$\frac{1}{4}$	0
1	$\frac{1}{4}$	0
2	0	$\frac{1}{2}$
$P(Y_2)$	$\frac{1}{2}$	$\frac{1}{2}$

所以

$$H(Y_2)=\frac{1}{2}\log_2 2+\frac{1}{2}\log_2 2=1\text{ 比特/符号}$$

$$H(Y_2|X)=\frac{1}{4}\log_2 1+\frac{1}{4}\log_2 1+\frac{1}{2}\log_2 1=0\text{ 比特/符号}$$

$$I(X;Y_2)=H(Y_2)-H(Y_2|X)=1\text{ 比特/符号}$$

因此第二个实验好些。

(2) $I(X;Y_1Y_2)=H(Y_1Y_2)-H(Y_1Y_2|X)$,**因此要求出** $P(Y_1Y_2)$、$P(Y_1Y_2|X)$ **和** $P(XY_1Y_2)$。**由于** Y_1、Y_2 **是相互独立的实验**,**所以** $P(Y_1Y_2|X)=P(Y_1|X)P(Y_2|X)$。

$$\left.\begin{array}{l}P(Y_1|X)\\ P(Y_2|X)\end{array}\right\}\quad P(Y_1Y_2|X)\quad P(XY_1Y_2)\quad P(Y_1Y_2)\quad (\text{见表 3.4 和 3.5})$$

表 3.4

$P(Y_1Y_2\|X)$ X \ Y_1Y_2	00	01	10	11
0	1	0	0	0
1	0	0	1	0
2	0	$\frac{1}{2}$	0	$\frac{1}{2}$

表 3 .5

$P(XY_1Y_2)$ \ Y_1Y_2 / X	00	01	10	11
0	$\frac{1}{4}$	0	0	0
1	0	0	$\frac{1}{4}$	0
2	0	$\frac{1}{4}$	0	$\frac{1}{4}$
$P(Y_1Y_2)$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$

$$H(Y_1Y_2) = \frac{1}{4}\log_2 4 + \frac{1}{4}\log_2 4 + \frac{1}{4}\log_2 4 + \frac{1}{4}\log_2 4 = 2 \text{ 比特/ 符号}$$

$$H(Y_1Y_2 \mid X) = \frac{1}{4}\log_2 1 + \frac{1}{4}\log_2 1 + \frac{1}{4}\log_2 2 + \frac{1}{4}\log_2 2 = \frac{1}{2} \text{ 比特/ 符号}$$

$$I(X;Y_1Y_2) = H(Y_1Y_2) - H(Y_1Y_2 \mid X) = 2 - \frac{1}{2} = \frac{3}{2} \text{ 比特/ 符号}$$

可以看出:做 Y_1 和 Y_2 两个实验比做 Y_1 一个实验可多得到的信息为

$$I(X;Y_1Y_2) - I(X;Y_1) = \frac{3}{2} - \frac{1}{2} = 1 \text{ 比特/ 符号}$$

做 Y_1 和 Y_2 两个实验比做 Y_2 一个实验可多得到的信息为

$$I(X;Y_1Y_2) - I(X;Y_2) = \frac{3}{2} - 1 = \frac{1}{2} \text{比特/ 符号}$$

(3) $I(X;Y_1 \mid Y_2) = I(X;Y_1Y_2) - I(X;Y_2) = \frac{3}{2} - 1 = \frac{1}{2}$ **比特/ 符号,它表示做完 Y_2 实验以后,从 Y_1 实验可得到关于 X 的信息量。**

$I(X;Y_2 \mid Y_1) = I(X;Y_1Y_2) - I(X;Y_1) = \frac{3}{2} - \frac{1}{2} = 1$ **比特/ 符号,它表示做完 Y_1 实验以后,从 Y_2 实验可得到关于 X 的信息量。**

3 .4 **某信源符号集的概率分布和对应的二进制代码如表 3 .6 所示。**

表 3 .6

信源符号	u_0	u_1	u_2	u_3
概率	1/ 2	1/ 4	1/ 8	1/ 8
代码	0	10	110	111

(1) 求信源符号熵;

(2) 求平均每个消息符号所需要的二进制码元的个数或平均代码长度,进而用这一结果求码序列中的二进制码元的熵;

(3) 当消息是由符号序列组成时,各符号之间若相互独立,求其对应的二进码序列中出现"0"和"1"的无条件概率 $p(0)$和 $p(1)$,以及相邻码元间的条件概率 $p(0|0)$、$p(1|0)$、$p(0|1)$和 $p(1|1)$。

解:

(1) 信源熵为

$$H(U) = \frac{1}{2}\log_2 2 + \frac{1}{4}\log_2 4 + \frac{1}{4}\log_2 8 = \frac{7}{4}\text{比特/ 符号}$$

(2) 设平均代码长度为 L,则

$$L = \frac{1}{2}\times 1 + \frac{1}{4}\times 2 + \frac{1}{8}\times 3 + \frac{1}{8}\times 3 = \frac{7}{4}\text{二进制码元/ 符号}$$

二进制码元的熵为

$$\frac{H(U)}{L} = 1\text{ 比特/ 二进制码元}$$

(3) 由于符号间相互独立,因此

$$p(0) = \frac{\frac{1}{2}+\frac{1}{4}+\frac{1}{8}}{L} = \frac{1}{2}, \qquad p(1) = 1 - p(0) = \frac{1}{2}$$

为求相邻码元间的条件概率,先求相邻码元间的联合概率:

$$p(1,1) = \frac{1}{L}\left[\frac{1}{8} + \frac{1}{8}\times 2 + \frac{1}{8}\times\left(\frac{1}{4}+\frac{1}{8}+\frac{1}{8}\right)\right] = \frac{1}{4}$$

所以

$$p(1|1) = \frac{p(1,1)}{p(1)} = \frac{1}{2}$$

$$p(0|1) = 1 - p(1|1) = \frac{1}{2}$$

同理

$$p(0,0) = \frac{1}{L}\left(\frac{1}{2}\times\frac{1}{2} + \frac{1}{4}\times\frac{1}{2} + \frac{1}{8}\times\frac{1}{2}\right) = \frac{1}{4}$$

$$p(0|0) = \frac{p(0,0)}{p(0)} = \frac{1}{2}$$

$$p(1|0) = 1 - p(0|0) = \frac{1}{2}$$

3.5　二次扩展信源的熵为 $H(X^2)$,而一阶马尔可夫信源的熵为 $H(X_2|X_1)$。试比较两者的大小,并说明原因。

解:

$$H(X^2) = 2H(X) > H(X) > H(X_2 \mid X_1)$$

二次扩展信源的熵是一个联合熵,其值应该大于单符号信源熵,而马尔可夫信源的熵是一个条件熵,其值小于单符号信源熵。马尔可夫信源符号间的依赖关系提供了额外的信息量,从而减小了信源的不确定性。

3.6 一个马尔可夫过程的基本符号为 0,1,2,这 3 个符号等概率出现,并且具有相同的转移概率。

(1) 画出一阶马尔可夫过程的状态图,并求稳定状态下的一阶马尔可夫信源熵 H_1 和信源剩余度;

(2) 画出二阶马尔可夫过程的状态图,并求稳定状态下二阶马尔可夫信源熵 H_2 和信源剩余度。

解:

(1) 一阶马尔可夫过程的状态转移图如图 3.1 所示。

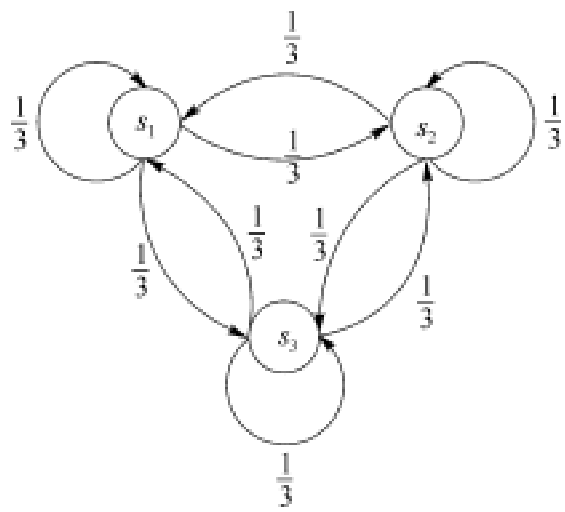


图 3.1

一阶马尔可夫过程共有 3 种状态,每个状态转移到其他状态的概率均为$\frac{1}{3}$,设状态的平稳分布为 $W = (W_1, W_2, W_3)$,根据

$$\begin{cases} W_1 = \frac{1}{3}W_1 + \frac{1}{3}W_2 + \frac{1}{3}W_3 \\ W_2 = \frac{1}{3}W_1 + \frac{1}{3}W_2 + \frac{1}{3}W_3 \\ W_3 = \frac{1}{3}W_1 + \frac{1}{3}W_2 + \frac{1}{3}W_3 \\ W_1 + W_2 + W_3 = 1 \end{cases}$$

可得 $W = (1/3, 1/3, 1/3)$,3 种状态等概率分布。

一阶马尔可夫信源熵为

$$H_2 = 3 \times \frac{1}{3} \times H\left(\frac{1}{3}, \frac{1}{3}, \frac{1}{3}\right) = 1.585 \text{ 比特/符号}$$

信源剩余度为

$$\gamma = 1 - \frac{H_2}{H_0} = 1 - \frac{H_2}{\log_2 3} = 0$$

(2) 二阶马尔可夫信源有 9 种状态(状态转移图略),同样列方程组求得状态的平稳分布为

$$W = \left(\frac{1}{9}, \frac{1}{9}, \frac{1}{9}, \frac{1}{9}, \frac{1}{9}, \frac{1}{9}, \frac{1}{9}, \frac{1}{9}, \frac{1}{9}\right)$$

二阶马尔可夫信源熵为

$$H_3 = 9 \times \frac{1}{9} \log_2 3 = 1.585 \text{ 比特/符号}$$

信源剩余度为

$$\gamma = 1 - \frac{H_3}{H_0} = 1 - \frac{H_3}{\log_2 3} = 0$$

由于在上述两种情况下,3 个符号均为等概率分布,所以信源剩余度都等于 0。

3.7　一阶马尔可夫信源的状态转移图如图 3.2 所示,信源 X 的符号集为{0,1,2}。

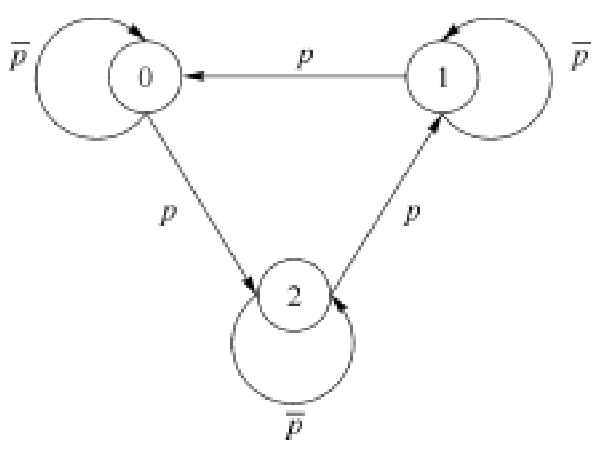


图 3.2

(1) 求平稳后的信源的概率分布;

(2) 求信源熵 H_∞;

(3) 求当 $p = 0$ 或 $p = 1$ 时信源的熵,并说明其理由。

解:

(1) 设状态的平稳分布为 $W = (W_0, W_1, W_2)$,根据

$$\begin{cases} W_0 = \bar{p} W_0 + p W_1 \\ W_1 = \bar{p} W_1 + p W_2 \\ W_2 = p W_0 + \bar{p} W_2 \\ W_0 + W_1 + W_2 = 1 \end{cases}$$

解得稳态分布为:当 $p\neq 0$ 时, $W=\left(\frac{1}{3},\frac{1}{3},\frac{1}{3}\right)$;当 $p=0$ 时, W 为任意的概率分布。

(2) 当 $p\neq 0$ 时, $H_\infty = H_2 = 3\times\frac{1}{3}\times H(p) = H(p)$;当 $p=0$ 时, $H_\infty=0$。

(3) 当 $p=0$ 或 $p=1$ 时, $H_\infty=0$,整个信源为一个确定信源,所以其熵为零。

3.8 有一个二元无记忆信源,其发 0 的概率为 p,而 $p\approx 1$,所以在发出的二元序列中经常出现的是那些一串为 0 的序列,称高概率序列。对于这样的信源我们可以用另一新信源来代替,新信源中只包含这些高概率序列。这时新信源 $S_n=\{s_1, s_2, s_3, \cdots, s_n, s_{n+1}\}$,共有 $n+1$ 个符号,它与高概率的二元序列的对应关系如下:

二元序列:1,01,001,…, 00…01(共 $n-1$ 个 0),00…000(共 n 个 0);

新信源符号: $s_1, s_2, s_3, \cdots, s_n, s_{n+1}$。

(1) 求 $H(S_n)$;

(2) 当 $n\to\infty$ 时,求信源的熵 $H(S)=\lim\limits_{n\to\infty}H(S_n)$。

解:

(1) 新信源的概率空间为

$$\begin{bmatrix} S_n \\ P(S_n) \end{bmatrix} = \begin{bmatrix} s_1 & s_2 & s_3 & \cdots & s_n & s_{n+1} \\ 1-p & p(1-p) & p^2(1-p) & \cdots & p^{n-1}(1-p) & p^n \end{bmatrix}$$

$$\begin{aligned}
H(S_n) &= -\sum_{i=1}^{n+1} p(s_i)\log_2 p(s_i) = -\sum_{i=1}^{n} p^{i-1}(1-p)\log_2[p^{i-1}(1-p)] - p^n\log_2 p^n \\
&= -\sum_{i=1}^{n} p^{i-1}(1-p)\log_2 p^{i-1} - \sum_{i=1}^{n} p^{i-1}(1-p)\log_2(1-p) - p^n\log_2 p^n \\
&= -\sum_{i=1}^{n}(i-1)p^{i-2}p(1-p)\log_2 p - \sum_{i=1}^{n} p^{i-1}(1-p)\log_2(1-p) - np^n\log_2 p \\
&= \frac{np^{n-1}-np^n-1+p^n}{(1-p)^2}p(1-p)\log_2 p - \frac{1-p^n}{1-p}(1-p)\log_2(1-p) - np^{n-1}p\log_2 p \\
&= -\frac{1-p^n}{1-p}(1-p)\log_2(1-p) - \frac{1-p^n}{1-p}p\log_2 p \\
&= \frac{1-p^n}{1-p}H(p)
\end{aligned}$$

(2) 当 $n\to\infty$ 时,信源的熵为

$$H(S)=\lim_{n\to\infty}H(S_n)=\frac{1}{1-p}H(p)$$

3.9 给定状态转移概率矩阵, $P=\begin{pmatrix} 1-p & p \\ 1 & 0 \end{pmatrix}$,求:

(1) 此两状态马尔可夫链的熵率 H_∞;

(2) 此熵率的极大值及相应的 p;

(3) 在达到最大熵率的情况下,求出每一个 n 长序列的概率。

解:

(1) 此两状态马尔可夫链的状态转移图如图 3.3 所示。

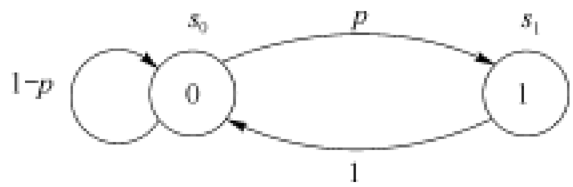


图 3.3

可求出稳态分布为 $W = \left[\frac{1}{1+p}, \frac{p}{1+p}\right]$,即 $p(s_0) = \frac{1}{1+p}$, $p(s_1) = \frac{p}{1+p}$,所以

$$H_\infty = \frac{H(p)}{1+p} \tag{3.1}$$

(2) 将式(3.1)对 p 求导得

$$\frac{\mathrm{d}}{\mathrm{d}p}H_\infty = \frac{\log_2 e}{(1+p)^2} \cdot \ln\frac{(1-p)^2}{p}$$

所以当 $p = (1-p)^2$,即 $p = \frac{3-\sqrt{5}}{2} = 0.3820$ 时熵率达到极值:

$$\max H_\infty = 0.6942 \text{ 比特/符号}$$

(3) 根据链规则,一个允许的符号序列$(x_1 x_2 \cdots x_n)$的概率为

$$\begin{aligned} p(x_1 x_2 \cdots x_n) &= p(x_1)\,p(x_2 \mid x_1)\cdots p(x_n \mid x_{n-1}) \\ &= p(x_1)(1-p)^{n_{00}}\,p^{n_{01}}\,1^{n_{10}} \end{aligned}$$

其中 n_{00}, n_{01}和 n_{10}分别是 0→0,0→1 和 1→0 转移的次数, $n_{00} + n_{01} + n_{10} = n - 1$。又因为每个符号 1 后面肯定是一个 0,所以根据序列的第一个和最后一个符号不同,有 $n_{01} = n_{10} + \delta$,其中 $\delta \in \{-1, 0, 1\}$。

由(2)得 $p = (1-p)^2$,所以有

$$\begin{aligned} p(x_1, x_2, \cdots, x_n) &= p(x_1)(1-p)^{n_{00}}\,p^{n_{01}}\,1^{n_{10}} = p(x_1)(1-p)^{n_{00}+2n_{01}} \\ &= p(x_1)(1-p)^{n_{00}+n_{01}+n_{10}+\delta} \\ &= p(x_1)(1-p)^{n-1+\delta} \end{aligned}$$

由上述结论可知:一个允许的符号序列的概率仅仅由此序列的第一个和最后一个符号确定,如表 3.7 所示。

3.10　在一个 3×3 的国际象棋棋盘上,分别计算“王”、“车”、“左象”、“右象”和“后”随机行走的熵率。

解:

给棋盘编号如表 3.8 所示。

表 3.7

x_1	x_n	n_{01}	$p(x_1\cdots x_n)$
0	0	n_{10}	$p(s_0)(1-p)^{n-1}$
0	1	$n_{10}+1$	$p(s_0)(1-p)^{n}$
1	0	$n_{10}-1$	$p(s_1)(1-p)^{n-2}$
1	1	n_{10}	$p(s_1)(1-p)^{n-1}$

表 3.8

1	2	3
4	5	6
7	8	9

(1) 由于“王”不能停在当前格上,必须走一步,所以这 9 个状态的稳态分布为

$$W_i = p(s_i) = E_i / E$$

其中 E_i 是从第 i 格出发能够到达的格子数, $E = \sum_i E_i$。通过简单的计算可得:

$$W_1 = W_3 = W_7 = W_9 = \frac{3}{40},\quad W_2 = W_4 = W_6 = W_8 = \frac{5}{40},\quad W_5 = \frac{8}{40}$$

再根据“随机行走”的意义可得

$$H(X|s_i) = \begin{cases} \log_2 3 & i = 1,3,7,9 \\ \log_2 5 & i = 2,4,6,8 \\ \log_2 8 & i = 5 \end{cases}$$

因此最终结果为

$$H(\text{王}) = 4\times\frac{3}{40}\times\log_2 3 + 4\times\frac{5}{40}\times\log_2 5 + \frac{8}{40}\times\log_2 8 = 2.2365\ \text{比特/步}$$

(2)“车”不管在哪个格子,它都有 4 个走向,例如它在 1 号格子,它可以去 2、3、4、7 号格子,因此状态的稳态分布为均匀分布: $W_i = p(s_i) = \frac{1}{9}$, $i = 1,\cdots,9$,车随机行走的熵率为

$$H(\text{车}) = 9\times\frac{1}{9}\times\left(4\times\frac{1}{4}\times\log_2 4\right) = 2\ \text{比特/步}$$

同样可得

$H(\text{左象}) = 1$ 比特/步; $H(\text{右象}) = 1.3333$ 比特/步; $H(\text{后}) = 2.6443$ 比特/步

3.11 图 3.4 是一张有 4 个节点的随机行走图,从任何一个节点走到下一个节点的概率都相等。

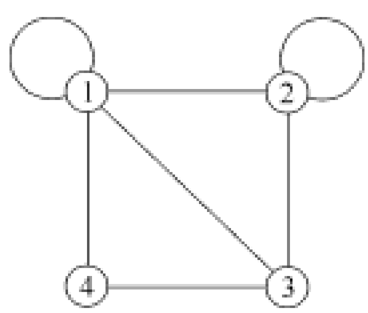


图 3.4

(1) **求随机行走的稳态分布;**

(2) **求随机行走的熵率。**

解:

(1)**随机行走的转移概率矩阵为** $P=\begin{bmatrix}\frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4}\\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & 0\\ \frac{1}{3} & \frac{1}{3} & 0 & \frac{1}{3}\\ \frac{1}{2} & 0 & \frac{1}{2} & 0\end{bmatrix}$。**由于从任何一个节点走到下一个节点的概率都相等,可以初步推测每个节点的稳态分布与和它相连的转移线的数目成正比,即**

$$(W_1, W_2, W_3, W_4)=k(4,3,3,2)\quad (W_1, W_2, W_3, W_4)=\left(\frac{4}{12},\frac{3}{12},\frac{3}{12},\frac{2}{12}\right)$$

通过验证,$\left(\frac{4}{12},\frac{3}{12},\frac{3}{12},\frac{2}{12}\right)P=\left(\frac{4}{12},\frac{3}{12},\frac{3}{12},\frac{2}{12}\right)$,**证明我们的推测是正确的。这里连接不同节点的转移线都是双向的。**

(2)
$$\begin{aligned}H_\infty &= \sum_{i=1}^{4} W_i H(X_2 \mid X_1=i)\\ &= \frac{4}{12}\log_2 4+\frac{3}{12}\log_2 3+\frac{3}{12}\log_2 3+\frac{2}{12}\log_2 2\\ &= \frac{2}{3}+\frac{1}{2}\log_2 3+\frac{1}{6}=\frac{5}{6}+0.792\,5\\ &= 1.625\,8\ \text{比特/步}\end{aligned}$$

3.12　**求具有如下概率密度函数的随机变量的熵。**

(1) **指数分布** $f(x)=\lambda e^{-\lambda x}$, $x\geq 0$;

(2) $f(x)=\frac{1}{2}\lambda e^{-\lambda|x|}$;

(3) **单边高斯分布** $f(x)=\frac{2}{\sqrt{2\pi\sigma^2}}e^{-x^2/2\sigma^2}$, $x\geq 0$。

解:

(1)
$$\begin{aligned}h(X) &= -\int_0^\infty \lambda e^{-\lambda x}\ln(\lambda e^{-\lambda x})\,dx\\ &= -\int_0^\infty \lambda e^{-\lambda x}(\ln\lambda-\lambda x)\,dx\\ &= -\ln\lambda+\lambda(1/\lambda)\\ &= 1-\ln\lambda\ \text{奈特/样值}\end{aligned}$$

(2)
$$h(X) = -\int_{-\infty}^{\infty} \frac{1}{2}\lambda e^{-\lambda|x|}\ln\left[\frac{1}{2}\lambda e^{-\lambda|x|}\right]dx$$
$$= -\int_{-\infty}^{\infty} \frac{1}{2}\lambda e^{-\lambda|x|}\left[\ln\frac{1}{2} + \ln\lambda - \lambda|x|\right]dx$$
$$= -\ln\frac{1}{2} - \ln\lambda + 1$$
$$= \ln\frac{2e}{\lambda}\ \text{奈特/样值}$$

(3) 设 $\phi(x) = \frac{1}{\sqrt{2\pi\sigma^2}}e^{-x^2/2\sigma^2}$表示高斯密度函数,它的微分熵为$\frac{1}{2}\log_2 2\pi e\sigma^2$。单边高斯分布 $f(x) = \frac{2}{\sqrt{2\pi\sigma^2}}e^{-x^2/2\sigma^2} = 2\phi(x)$, $x\geq 0$ 的微分熵为

$$h(X) = -\int_0^{\infty} f(x)\log_2 f(x)dx$$
$$= -\int_0^{\infty} 2\phi(x)\log_2[2\phi(x)]dx$$
$$= -\int_0^{\infty} 2\phi(x)\log_2 2\,dx - \int_0^{\infty} 2\phi(x)\log_2\phi(x)dx$$
$$= -\log_2 2 - \int_{-\infty}^{\infty}\phi(x)\log_2\phi(x)dx \quad (\text{因为 } \phi(-x) = \phi(x))$$
$$= \frac{1}{2}\log_2(2\pi e\sigma^2) - \log_2 2\ \text{比特/样值}$$

3.13 连续随机变量 X 和 Y 的联合概率密度为
$$p(xy) = \frac{1}{2\pi\sqrt{SN}}\exp\left\{-\frac{1}{2N}\left[x^2\left(1+\frac{N}{S}\right) - 2xy + y^2\right]\right\}$$
试求 $h(X)$, $h(Y)$, $h(Y|X)$和 $I(X;Y)$。

解:
$$p(x) = \int_{-\infty}^{+\infty} p(xy)dy = \int_{-\infty}^{+\infty}\frac{1}{\sqrt{2\pi S}}e^{-\frac{x^2}{2S}}\cdot\frac{1}{\sqrt{2\pi N}}e^{-\frac{1}{2N}(y-x)^2}dy = \frac{1}{\sqrt{2\pi S}}e^{-\frac{x^2}{2S}}$$
$$p(y|x) = \frac{p(xy)}{p(x)} = \frac{1}{\sqrt{2\pi N}}e^{-\frac{1}{2N}(y-x)^2}$$
$$h(X) = -\int_{-\infty}^{\infty} p(x)\ln p(x)dx = -\int_{-\infty}^{\infty} p(x)\ln\left[\frac{1}{\sqrt{2\pi S}}e^{-\frac{x^2}{2S}}\right]dx$$
$$= -\int_{-\infty}^{\infty} p(x)\ln\left[\frac{1}{\sqrt{2\pi S}}\right]dx - \int_{-\infty}^{\infty} p(x)\ln e^{-\frac{x^2}{2S}}dx$$
$$= \ln\sqrt{2\pi S} + \int_{-\infty}^{\infty} p(x)\frac{x^2}{2S}dx$$
$$= \ln\sqrt{2\pi S} + \frac{1}{2S}\cdot S = \ln\sqrt{2\pi e S}$$

$$h(Y \mid X) = -\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} p(xy)\ln p(y \mid x)\,\mathrm{d}x\mathrm{d}y$$

$$= -\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} p(xy)\left[\ln\frac{1}{\sqrt{2\pi N}} - \frac{1}{2N}(y - x)^2\right]\mathrm{d}x\mathrm{d}y$$

$$= \ln\sqrt{2\pi N} + \frac{1}{2N}\int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi S}}\mathrm{e}^{-\frac{x^2}{2S}}\int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi N}}\mathrm{e}^{-\frac{1}{2N}(y-x)^2}(y - x)^2\,\mathrm{d}y\mathrm{d}x$$

$$= \ln\sqrt{2\pi N} + \frac{1}{2} = \ln\sqrt{2\pi\mathrm{e}N}$$

$$h(XY) = -\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} p(xy)\ln p(xy)\,\mathrm{d}x\mathrm{d}y$$

$$= -\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi S}}\mathrm{e}^{-\frac{x^2}{2S}}\cdot\frac{1}{\sqrt{2\pi N}}\mathrm{e}^{-\frac{1}{2N}(y-x)^2}\left[\ln\frac{1}{2\pi\sqrt{SN}} - \frac{x^2}{2S} - \frac{1}{2N}(y - x)^2\right]\mathrm{d}x\mathrm{d}y$$

$$= \ln(2\pi\sqrt{SN}) + \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\frac{x^2}{2S}\cdot\frac{1}{\sqrt{2\pi S}}\mathrm{e}^{-\frac{x^2}{2S}}\cdot\frac{1}{\sqrt{2\pi N}}\mathrm{e}^{-\frac{1}{2N}(y-x)^2}\mathrm{d}x\mathrm{d}y +$$

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\frac{1}{2N}(y - x)^2\frac{1}{\sqrt{2\pi S}}\mathrm{e}^{-\frac{x^2}{2S}}\cdot\frac{1}{\sqrt{2\pi N}}\mathrm{e}^{-\frac{1}{2N}(y-x)^2}\mathrm{d}x\mathrm{d}y$$

$$= \ln(2\pi\sqrt{SN}) + \frac{1}{2} + \frac{1}{2} = \ln(2\pi\mathrm{e}\sqrt{SN})$$

$$h(Y) = h(XY) - h(X) = \ln\sqrt{2\pi\mathrm{e}(S + N)}$$

$$I(X;Y) = h(Y) - h(Y|X) = \frac{1}{2}\ln\frac{S + N}{N}$$

3.14　**一信源产生的时不变波形信号(即信号统计特性不随时间而变)的带宽** $W = 4$ kHz,**幅度分布为**

$$p(x) = \mathrm{e}^{-x} \qquad x \geq 0$$

设在信号幅度 0～2 区间按量化单位 $\Delta = 0.5$ 做量化,试求该信源的信息输出率。

解:

该信源的绝对熵

$$H(X) = h(X) - \lim_{\Delta \to 0}\ln\Delta$$

由于本题中 $\Delta = 0.5$,并不趋于 0,所以

$$H(X) \approx h(X) - \ln\Delta \approx -\int_a^h p(x)\ln p(x)\,\mathrm{d}x - \ln\Delta$$

$$\approx \int_0^2 x\mathrm{e}^{-x}\,\mathrm{d}x - \ln\Delta$$

$= 1 - 3\mathrm{e}^{-2} + \ln 2$ **奈特/ 样值**

按照奈奎斯特定理,对该波形信号的抽样率至少为 $2\times4\times10^3$ 次/秒。信源的输出信息率为

$$H_t = nH(X) = 2\times4\times10^3\times(1-3e^{-2}+\ln 2)\ \text{奈特/秒}$$

3.15 随机变量 X 和 Y 的联合概率密度函数在曲线 $y=\frac{1}{\sqrt{2\pi}}e^{-x^2/2}$ 和 X 轴所组成的区域内均匀分布,如图 3.5 所示。

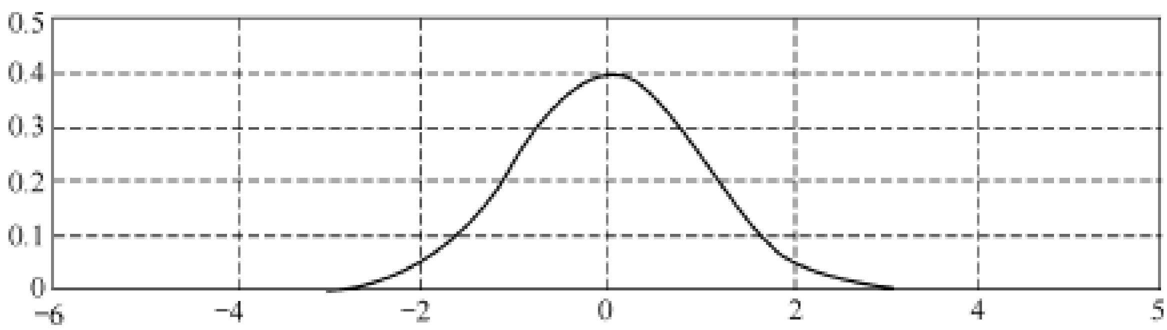


图 3.5

(1) 求 $h(XY)$;

(2) 求 $h(X)$;

(3) Y 的概率密度函数为 $f(y)=2\sqrt{-2\ln(y\sqrt{2\pi})}$, $0<y\leq1/\sqrt{2\pi}$, 证明:

$$-\frac{1}{2}\ln 2\pi-\frac{1}{2}<h(Y)<-\frac{1}{2}\ln 2\pi$$

解:

(1) 这是一个高斯随机变量的概率密度函数图,所以曲线下面的面积为 1,因此 X 和 Y 在一个面积为 1 的区域 S 内均匀分布,所以

$$h(XY) = \iint_S 1\ \ln 1\ \mathrm{d}x\mathrm{d}y = \iint_S 1\cdot 0\ \mathrm{d}x\mathrm{d}y = 0$$

(2) X 的边沿概率密度函数为 $f(x)=\int_0^y 1\ \mathrm{d}y=\frac{1}{\sqrt{2\pi}}e^{-x^2/2}$,也就是说,$X$ 是均值为 0、方差为 1 的高斯随机变量,所以

$$h(X)=\frac{1}{2}\ln 2\pi e$$

(3) Y 的取值范围为 $\left[0,\frac{1}{\sqrt{2\pi}}\right]$,并且它不是均匀分布,它的微分熵必然小于在这个取值范围内均匀分布的随机变量,所以

$$h(Y)<\ln\frac{1}{\sqrt{2\pi}}=-\ln(\sqrt{2\pi})=-\frac{1}{2}\ln 2\pi$$

由 $0 = h(XY) = h(X) + h(Y|X) \quad h(Y|X) = -h(X) = -\frac{1}{2}\ln 2\pi - \frac{1}{2}$，又因为 X 和 Y 不是相互独立的随机变量，所以

$$h(Y) > h(Y|X) = -\frac{1}{2}\ln 2\pi - \frac{1}{2}$$

综合前两步的结果得

$$-\frac{1}{2}\ln 2\pi - \frac{1}{2} < h(Y) < -\frac{1}{2}\ln 2\pi$$

3.16　给定状态转移概率矩阵，$P = \begin{bmatrix} 1-\alpha & \alpha \\ \beta & 1-\beta \end{bmatrix}$，求：此二状态马尔可夫信源的熵率 H_∞。

解：

先求出该马尔可夫信源的状态平稳分布：

$$\begin{cases} W_1 = (1-\alpha)W_1 + \beta W_2 \\ W_2 = \alpha W_1 + (1-\beta)W_2 \end{cases}$$

解得平稳分布为

$$W = \left[\frac{\beta}{\alpha+\beta}, \frac{\beta}{\alpha+\beta}\right]$$

其熵率为

$$H_\infty = H_2 = \frac{\beta}{\alpha+\beta}H(\alpha) + \frac{\alpha}{\alpha+\beta}H(\beta)$$

3.17　布袋中有手感完全相同的 3 个红球和 3 个蓝球，每次从中随机取出一个球，取出后不放回布袋。用 X_i 表示第 i 次取出的球的颜色，$i = 1, 2, \cdots, 6$，求：

(1) $H(X_1)$；

(2) $H(X_2)$；

(3) $H(X_2|X_1)$；

(4) 随着 k 的增加，$H(X_k|X_1 \cdots X_{k-1})$ 是增加还是减少？请解释。

(所有的答案用 $H(p)$ 的形式表示)

解：

(1) P(第一个球是红球) = P(第一个球是蓝球) $= \frac{1}{2}$，因此

$$H(X_1) = (1/2)\log_2 2 + (1/2)\log_2 2 = \log_2 2 = 1$$

(2) P(第二个球是红球) = P(第二个球是蓝球) = $\frac{1}{2}$,因此

$$H(X_2) = (1/2)\log_2 2 + (1/2)\log_2 2 = \log_2 2 = 1$$

同理

$$H(X_3) = H(X_4) = H(X_5) = H(X_6) = 1 \text{ 比特/符号}$$

(3)
$$\begin{aligned} & H(X_2 \mid X_1) \\ = & P(\text{第一个球是红球}) \cdot H(X_2 \mid \text{第一个球是红球}) + P(\text{第一个球是蓝球}) \cdot \\ & H(X_2 \mid \text{第一个球是蓝球}) \\ = & \frac{1}{2} H\left(\frac{2}{5}\right) + \frac{1}{2} H\left(\frac{2}{5}\right) \\ = & H\left(\frac{2}{5}\right) = 0.971\,0 \text{ 比特/符号} \end{aligned}$$

(4) 随着 k 的增加,$H(X_k \mid X_1 \cdots X_{k-1})$会减少,因为知道以前的结果会降低这次结果的不确定性,以前的结果知道得越多,这次结果的不确定性越小。$H(X_6 \mid X_1 \cdots X_5) = 0$。

3.18 已知一个二元一阶马尔可夫信源的状态转移概率矩阵为 $P = \begin{bmatrix} 0.9 & 0.1 \\ 0.2 & 0.8 \end{bmatrix}$。

(1) 求此马尔可夫信源的熵率;

(2) 求符号序列 1000011 的概率(根据平稳分布确定第一个符号的概率);

(3) 计算分布函数 $F(x) = P[(X_1 X_2 \cdots) < x]$当 $x = 1000011$ 时的值。

解:

(1) 先求出该马尔可夫信源的平稳分布为 $W = \left(\frac{2}{3}, \frac{1}{3}\right)$,其熵率为

$$H_\infty = H_2 = \frac{2}{3} H(0.1) + \frac{1}{3} H(0.2) = \frac{2}{3} \times 0.469\,0 + \frac{1}{3} \times 0.721\,9 = 0.553\,3 \text{ 比特/符号}$$

(2) 根据平稳分布确定第一个符号的概率为 $P(X_1 = 1) = 1/3$,因此

$$\begin{aligned} p(1000011) &= p(1)\,p(0|1)\,p(0|0)\,p(0|0)\,p(0|0)\,p(1|0)\,p(1|1) \\ &= \frac{1}{3} \times 0.2 \times 0.9 \times 0.9 \times 0.9 \times 0.1 \times 0.8 \\ &= 3.888 \times 10^{-3} \end{aligned}$$

(3)
$$\begin{aligned} F(1000011) &= p(0) + p(100000) + p(1000010) \\ &= 0.666\,667 + 0.043\,740 + 0.000\,972 \\ &= 0.711\,379 \end{aligned}$$

所以

$$P[(X_1X_2\cdots)<1000011]=0.711\ 379$$

3.19　盒子里有两枚偏畸硬币，硬币 1 正面向上的概率为 p，硬币 2 正面向上的概率为 $1-p$，$0<p<0.5$。随机取一枚硬币并且连续投掷。用 $Z\in\{1,2\}$表示所选择的硬币，$X_1,X_2,X_3,\cdots$表示每次投掷的结果(正面或反面)。

(1) $X_1,X_2,X_3,\cdots$是否为平稳过程？是否为马尔可夫过程？

(2) 求 $H(X_1X_2\cdots X_n|Z)$；

(3) 求 $I(X_1;X_2|Z)$；

(4) 求 $H(X_1X_2)$；

(5) 求 $I(X_1;X_2)$和 $I(X_3;X_{729})$；

(6) 求熵率 $H_\infty=\lim\limits_{n\to\infty}\dfrac{1}{n}H(X_1X_2\cdots X_n)$；

(7) 求熵率 $\lim\limits_{n\to\infty}H(Z|X_1X_2\cdots X_n)$。

解：

(1) 该随机过程是平稳过程，因为一旦选择了一枚硬币，在投掷的过程中它的偏畸性不会改变。

该随机过程不是一个有限阶的马尔可夫过程，因为投掷次数越多，则随机过程给出关于 Z 的信息越多，因此预测下一次投掷的结果越准确。如果用 Y_i 来表示二维随机变量(Z,X_i)，则 $Y_1,Y_2,Y_3,\cdots$是一个一阶马尔可夫过程，而 $X_1,X_2,X_3,\cdots$可以看作是这个马尔可夫链的函数。

(2) 一旦选择了一枚硬币，则每次投掷的结果是相互独立的，即

$$H(X_1X_2\cdots X_n|Z)=H(X_1|Z)+H(X_2|Z)+\cdots+H(X_n|Z)=nH(p)=nH(1-p)$$

(3) $I(X_1;X_2|Z)=0$，因为给定 Z 以后 X_1 和 X_2 相互独立。

(4) 我们可以直接计算 $H(X_1X_2)$或利用 $H(X_1X_2)=H(X_1)+H(X_2|X_1)$。$X_1,X_2$的联合概率分布如表 3.9 所示。

表 3.9

X_1 \ X_2	0	1
0	$\frac{1}{2}(p^2+q^2)$	$\frac{1}{2}(2pq)$
1	$\frac{1}{2}(2pq)$	$\frac{1}{2}(p^2+q^2)$

为了简单起见,**令** $q=1-p$,**所以**

$$H(X_1X_2)=-2pq\log_2 pq-(p^2+q^2)\log_2\frac{1}{2}(p^2+q^2)$$
$$=1-2pq\log_2 2pq-(p^2+q^2)\log_2(p^2+q^2)$$
$$=1+H(p^2+q^2)$$

(5) **任意**(X_i,X_j)**的概率分布均和**(X_1,X_2)**的概率分布相同**,**因此**

$$I(X_1;X_2)=I(X_3;X_{729})$$

而

$$I(X_1;X_2)=H(X_1)+H(X_2)-H(X_1X_2)$$
$$=2-[1+H(p^2+q^2)]$$
$$=1-H(p^2+q^2)$$

(6) $$H(X_1X_2\cdots X_n)\geq H(X_1X_2\cdots X_n|Z)=nH(p)$$

$$H(X_1X_2\cdots X_n)\leq H(X_1X_2\cdots X_nZ)=H(X_1X_2\cdots X_n|Z)+H(Z)=nH(p)+1$$

根据上面两个不等式可以得到熵率:

$$H_\infty=\lim_{n\to\infty}\frac{1}{n}H(X_1X_2\cdots X_n)=H(p)$$

这个结果和经验是一致的,**因为对于很大的** n **值**,**我们几乎可以断定这枚硬币的偏畸性**(p **或** $1-p$),**因此** $H_\infty=\lim_{n\to\infty}\frac{1}{n}H(X_1X_2\cdots X_n)=H(p)$。

(7) **我们猜测** $\lim_{n\to\infty}H(Z|X_1X_2\cdots X_n)=0$,**因为随着** n **的增大**,**几乎可以断定这枚硬币的偏畸性**。**下面来验证这个猜测**。

令 Y_n **表示前** $2n-1$ **次的投掷结果**:

$$Y_n=\begin{cases}1 & \text{如果 } X_1+\cdots+X_{2n-1}\geq n\\ 0 & \text{其他}\end{cases}$$

根据弱大数定理,**有**

$$\lim_{n\to\infty}P(Y_n=1|Z=z)=\begin{cases}0 & z=0,\ p<1/2\\ 1 & z=1\end{cases}$$

由于所选择的硬币是随机选取的,**所以** $P(Y_n=1)=\frac{1}{2}$,**所以**

$$I(Z;X_1X_2\cdots X_n)\geq I(Z;Y_n)=H(Y_n)-H(Y_n|Z)=1-H(Y_n|Z)$$

当 $n\to\infty$**时**, $H(Y_n|Z)\to 0$, $I(Z;X_1X_2\cdots X_n)\to 1$, $H(Z|X_1X_2\cdots X_n)=H(Z)-$

$I(Z; X_1 X_2 \cdots X_n) = 1 - I(Z; X_1 X_2 \cdots X_n) \to 0$。

3.20　**一个二元一阶马尔可夫信源的状态转移图如图 3.6 所示。**

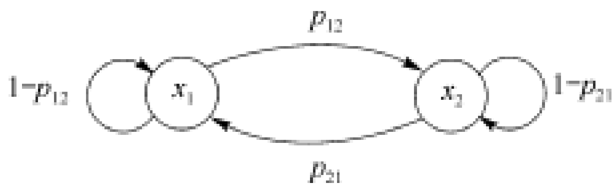


图 3.6

计算当 $p_{12} = 0.2$, $p_{21} = 0.3$ **时该马尔可夫信源的熵率,并求具有同样的符号概率分布的离散无记忆信源的熵。**

解:

该马尔可夫信源的状态转移概率矩阵为 $P = \begin{bmatrix} 0.8 & 0.2 \\ 0.3 & 0.7 \end{bmatrix}$,**可以求得它的状态的平稳分布**: $p(x_1) = 0.6$, $p(x_2) = 0.4$,**因此**

$$H_\infty = H_2 = 0.6(-0.8\log_2 0.8 - 0.2\log_2 0.2) + 0.4(-0.3\log_2 0.3 - 0.7\log_2 0.7)$$

$$= 0.7857 \text{ 比特/ 符号}$$

一阶马尔可夫信源的符号概率分布等于状态的平稳分布,所以具有同样的符号概率分布的离散无记忆信源(DMS)**的熵**:

$$H_{\text{DMS}} = -0.6\log_2 0.6 - 0.4\log_2 0.4 = 0.971$$

显然 $H_2 < H_{\text{DMS}}$。

3.21　**求具有如下概率密度函数的连续随机变量的微分熵**($\lambda > 0$):

$$f_X(x) = \begin{cases} (x+\lambda)/\lambda^2 & -\lambda \leq x \leq 0 \\ (-x+\lambda)/\lambda^2 & 0 < x \leq \lambda \\ 0 & \text{其他} \end{cases}$$

解:

根据微分熵的定义可得

$$h(X) = -\int_{-\infty}^{\infty} f_X(x)\ln f_X(x)\,\mathrm{d}x$$

$$= -\int_{-\lambda}^{0} \frac{x+\lambda}{\lambda^2}\ln\left(\frac{x+\lambda}{\lambda^2}\right)\mathrm{d}x - \int_{0}^{\lambda} \frac{-x+\lambda}{\lambda^2}\ln\left(\frac{-x+\lambda}{\lambda^2}\right)\mathrm{d}x$$

$$= -\ln\left(\frac{1}{\lambda^2}\right)\left[\int_{-\lambda}^{0} \frac{x+\lambda}{\lambda^2}\mathrm{d}x + \int_{0}^{\lambda} \frac{-x+\lambda}{\lambda^2}\mathrm{d}x\right] -$$

$$\int_{-\lambda}^{0}\frac{x+\lambda}{\lambda^2}\ln(x+\lambda)\mathrm{d}x-\int_{0}^{\lambda}\frac{-x+\lambda}{\lambda^2}\ln(-x+\lambda)\mathrm{d}x$$

$$=\ln\lambda^2-\frac{2}{\lambda^2}\int_0^{\lambda}z\ln z\mathrm{d}z\quad(z=\pm x+\lambda)$$

$$=\ln\lambda^2-\frac{2}{\lambda^2}\left[\frac{z^2\ln z}{2}-\frac{z^2}{4}\right]\Bigg|_0^{\lambda}$$

$$=\ln\lambda^2-\ln\lambda+\frac{1}{2}=\ln\lambda+\frac{1}{2}\text{ 奈特/样值}$$

3.22 X 是 n 维连续型随机矢量，$Y=AX$ 是 X 的线性变换，并且 A 是一个 $n\times n$ 的非奇异矩阵，证明 $h(Y)=\log_2|\det(A)|+h(X)$。

证明：

把 $|\det(A)|$ 简记为 $|A|$，因为 Y 是通过线性变换 $Y=AX$ 得到，所以它的密度函数为

$$g(y)=\frac{1}{|A|}f(A^{-1}y)$$

它的微分熵为

$$h(Y)=-\int g(y)\log_2 g(y)\mathrm{d}y$$

$$=-\int\frac{1}{|A|}f(A^{-1}y)[\log_2 f(A^{-1}y)-\log_2|A|]\mathrm{d}y$$

$$=-\int\frac{1}{|A|}f(x)[\log_2 f(x)-\log_2|A|]|A|\mathrm{d}x$$

$$=h(X)+\log_2|A|$$

3.23 设以 8 000 样值/秒的速率抽样一语音信号，并以 $M=2^8=256$ 级对抽样均匀量化，设抽样值取各量化值的概率相等，且抽样间相互统计独立。求：

(1) 每抽样的信息熵；

(2) 求信源的信息输出率。

解：

(1) 每抽样有 256 个量化取值，且各量化值的概率相等，因此每抽样的信息熵为

$$H(X)=-\sum_{i=1}^{256}p_i\log_2 p_i=-\sum_{i=1}^{256}\frac{1}{256}\log_2\frac{1}{256}=\log_2 256=8\text{ 比特/符号}$$

(2) 信源的信息输出率为

$$R=rH(X)=8\,000\times 8=6.4\times 10^4\text{ bps}$$

第 4 章

信道及信道容量

4 .1　**设一个二元信道如图 4 .1 所示,其输入概率空间为** $\begin{bmatrix} X \\ P \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ 0.2 & 0.8 \end{bmatrix}$,**试计算** $I(X=0; Y=1)$, $I(X=1; Y)$ **和** $I(X; Y)$。

解:

(1) $$I(X=0; Y=1) = \log_2 \frac{P(Y=1 \mid X=0)}{P(Y=1)} = \log_2 \frac{0.3}{0.2\times 0.3 + 0.8\times 0.9}$$
$$= -1.38\ \text{bit}$$

(2) $$I(X=1; Y)$$
$$= \sum_y p(y \mid X=1)\log_2 \frac{p(y \mid X=1)}{p(y)}$$
$$= 0.1\times\log_2 \frac{0.1}{0.2\times 0.7 + 0.8\times 0.1} + 0.9\times\log_2 \frac{0.9}{0.2\times 0.3 + 0.8\times 0.9}$$
$$= 0.1\times\log_2 \frac{10}{22} + 0.9\times\log_2 \frac{90}{78} = -0.114 + 0.186 = 0.072\ \text{比特/ 符号}$$

(3) $$I(X; Y)$$
$$= \sum_{x,y} p(xy)\log_2 \frac{p(y \mid x)}{p(y)} = \sum_{x,y} p(x)p(y \mid x)\log_2 \frac{p(y \mid x)}{\sum_u p(u)p(y \mid u)}$$
$$= 0.2\times 0.7\times\log_2 \frac{0.7}{0.2\times 0.7 + 0.8\times 0.1} + 0.2\times 0.3\times\log_2 \frac{0.3}{0.2\times 0.3 + 0.8\times 0.9} +$$

$$0.8\times 0.1\times \log_2 \frac{0.1}{0.2\times 0.7+0.8\times 0.1}+$$

$$0.8\times 0.9\times \log_2 \frac{0.9}{0.2\times 0.3+0.8\times 0.9}$$

$= 0.209$ 比特/符号

4.2　二元删除信道有两个输入:0,1 和 3 个输出:0,1,E,其中 E 表示可检出但无法纠正的错误。信道前向转移概率是

$$p(0|0)=1-\alpha \qquad p(E|0)=\alpha \qquad p(1|0)=0$$
$$p(0|1)=0 \qquad p(E|1)=\alpha \qquad p(1|1)=1-\alpha$$

求信道容量 C。

解:

信道如图 4.2 所示。

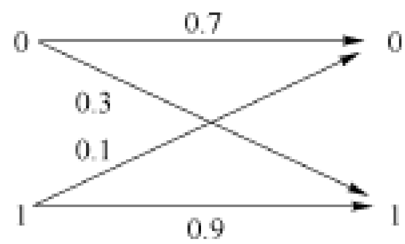


图 4.1

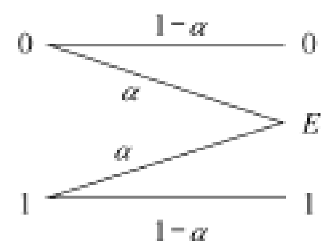


图 4.2

其转移矩阵为

$$\begin{bmatrix} 1-\alpha & \alpha & 0 \\ 0 & \alpha & 1-\alpha \end{bmatrix}$$

这是一个准对称信道,信道容量为

$$C=\log_2 2-H(\alpha)-(1-\alpha)\log_2(1-\alpha)-\alpha\log_2(2\alpha)=1-\alpha \text{ 比特/符号}$$

4.3　设某二进制数字传输系统接收判决器的输入信号电平、噪声密度分布、及判决电平如图 4.3 所示。试求:(1) 信道模型;(2) 平均互信息;(3) 信道容量。

解:

(1) 信道模型如图 4.4 所示。

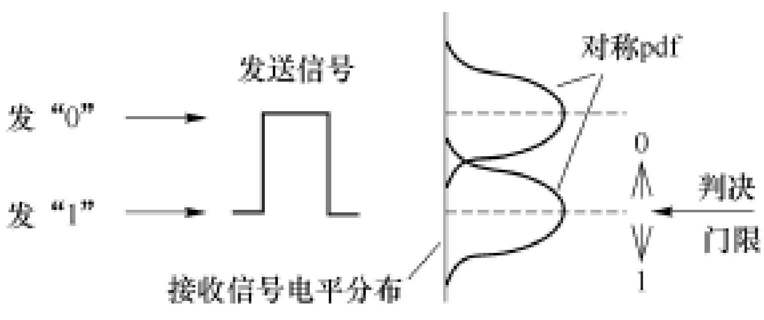


图 4.3

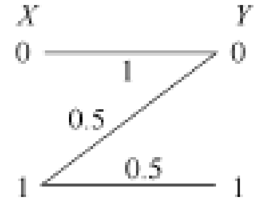


图 4.4

(2) **平均互信息为**

$$I(X;Y)=\sum_{x,y}p(x)p(y\mid x)\log_2\frac{p(y\mid x)}{p(y)}$$

$$=p_X(0)p_{Y|X}(0\mid 0)\log_2\frac{p_{Y|X}(0\mid 0)}{p_Y(0)}+p_X(0)P_{Y|X}(1\mid 0)\log_2\frac{p_{Y|X}(1\mid 0)}{p_Y(1)}+$$

$$p_X(1)p_{Y|X}(0\mid 1)\log_2\frac{p_{Y|X}(0\mid 1)}{p_Y(0)}+p_X(1)p_{Y|X}(1\mid 1)\log_2\frac{p_{Y|X}(1\mid 1)}{p_Y(1)} \tag{4.1}$$

令 $p_X(0)=p$, $p_X(1)=1-p$,**则** Y **的概率分布为**

$$p_Y(0)=p+\frac{1}{2}(1-p)=\frac{1}{2}(1+p)$$

$$p_Y(1)=0+\frac{1}{2}(1-p)=\frac{1}{2}(1-p)$$

将 X、Y **的概率分布代入式**(4.1)**得**

$$I(X;Y)$$

$$=p\times1\times\log_2\frac{1}{\frac{1}{2}\cdot(1+p)}+0+(1-p)\frac{1}{2}\log_2\frac{\frac{1}{2}}{\frac{1}{2}\cdot(1+p)}+(1-p)\frac{1}{2}\log_2\frac{\frac{1}{2}}{\frac{1}{2}\cdot(1-p)}$$

$$=p\log_2\frac{2}{1+p}+\frac{1}{2}(1-p)\log_2\frac{1}{(1+p)(1-p)}$$

$$=p-p\log_2(1+p)-\frac{1}{2}(1-p)\log_2(1-p^2) \tag{4.2}$$

(3) **为求信道容量,对平均互信息求驻点**:

$$\frac{\mathrm{d}I(X;Y)}{\mathrm{d}p}$$

$$=1-\log_2(1+p)-p\cdot(\log_2 \mathrm{e})\cdot\frac{1}{1+p}+\frac{1}{2}\log_2(1-p^2)-\frac{1}{2}(1-p)(\log_2\mathrm{e})\cdot\frac{-2p}{1-p^2}$$

$$=1-\log_2(1+p)-\frac{p}{1+p}\log_2\mathrm{e}+\frac{1}{2}\log_2(1-p^2)+\frac{p}{1+p}\log_2\mathrm{e}$$

$$=1+\log_2\frac{\sqrt{1-p^2}}{1+p}$$

令$\frac{\mathrm{d}I(X;Y)}{\mathrm{d}p}=0$,**由** $1+\log_2\frac{\sqrt{1-p^2}}{1+p}=0$　$\frac{1+p}{\sqrt{1-p^2}}=2$,**即** $p=\frac{3}{5}$**时,互信息达最大值。把** $p=\frac{3}{5}$**代入式**(4.2),**得信道容量为** $C=0.32$ **比特/符号。**

4.4　**设有扰离散信道的输入端是以等概率出现的** A, B, C, D 4 **个字母。该信道的正确传输概率为** 1/2,**错误传输概率平均分布在其他** 3 **个字母上。验证在该信道上每个**

字母传输的平均信息量为 0.208 **比特。**

解法一:

信源概率分布为

$$\begin{bmatrix} X \\ P \end{bmatrix} = \begin{bmatrix} A & B & C & D \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \end{bmatrix}$$

信道的传递概率矩阵为

$$P = \begin{bmatrix} 1/2 & 1/6 & 1/6 & 1/6 \\ 1/6 & 1/2 & 1/6 & 1/6 \\ 1/6 & 1/6 & 1/2 & 1/6 \\ 1/6 & 1/6 & 1/6 & 1/2 \end{bmatrix}$$

这是强对称信道,且输入为等概分布,这时平均互信息达到信道容量:

$$C = \log_2 r - p\log_2(r-1) - H(p) = \log_2 4 - H(1/2, 1/6, 1/6, 1/6)$$
$$= 2 - [H(1/2) + 1/2 \times \log_2 3] = 0.208 \text{ 比特/符号}$$

解法二:

$$\left.\begin{array}{l} p(x_i) \\ p(y_j \mid x_i) \end{array}\right\} \quad \left.\begin{array}{l} p(x_i y_j) \\ p(y_j) \end{array}\right\} \quad I(X; Y)$$

得

$$I(X; Y) = \sum_{i,j} p(x_i y_j)\log_2 \frac{p(y_j \mid x_i)}{p(y_j)} \approx 0.208 \text{ 比特/符号}$$

4.5 Z **信道及它的输入、输出如图** 4.5 **所示。**

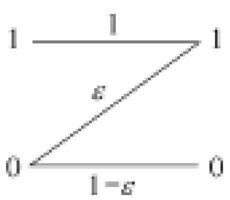


图 4.5

$$p(y \mid x) = \begin{bmatrix} 1 & 0 \\ \varepsilon & 1-\varepsilon \end{bmatrix} \quad x, y \in \{0, 1\}$$

(1) **求最佳输入分布;**

(2) **求** $\varepsilon = 1/2$ **时的信道容量;**

(3) **求当** $\varepsilon \to 0$ **和** $\varepsilon \to 1$ **时的最佳输入分布值。**

解:

(1) **把** $1-\varepsilon$ **记为** δ,**并令** $P(X=1) = p$ **则** $P(Y=1) = p(1-\varepsilon) = p\delta$。

平均互信息表示为 p **的函数:**

$$I(X; Y) = H(Y) - H(Y \mid X) = H(p\delta) - pH(\varepsilon) = H(p\delta) - pH(\delta)$$

对平均互信息求驻点,它的极值即为信道容量。

$$\frac{\mathrm{d}I}{\mathrm{d}p} = \delta\log_2\frac{1-p\delta}{p\delta} - H(\delta) = \delta\log_2\frac{1-p\delta}{p\delta} + \delta\log_2\delta + (1-\delta)\log_2(1-\delta) = 0$$

整理得到:

$$\log_2 \frac{p\delta}{1-p\delta} = \log_2 \delta + \frac{1-\delta}{\delta}\log_2(1-\delta) = \log_2 \delta(1-\delta)^{\frac{1-\delta}{\delta}}$$

去掉对数符号,得

$$\frac{p\delta}{1-p\delta} = \delta(1-\delta)^{\frac{1-\delta}{\delta}} \qquad p = \frac{(1-\delta)^{\frac{1-\delta}{\delta}}}{1+\delta(1-\delta)^{\frac{1-\delta}{\delta}}} = \frac{\varepsilon^{\frac{\varepsilon}{1-\varepsilon}}}{1+(1-\varepsilon)\varepsilon^{\frac{\varepsilon}{1-\varepsilon}}} \tag{4.3}$$

最佳输入分布为

$$P(X=0) = 1 - \frac{\varepsilon^{\frac{\varepsilon}{1-\varepsilon}}}{1+(1-\varepsilon)\varepsilon^{\frac{\varepsilon}{1-\varepsilon}}}, \qquad P(X=1) = \frac{\varepsilon^{\frac{\varepsilon}{1-\varepsilon}}}{1+(1-\varepsilon)\varepsilon^{\frac{\varepsilon}{1-\varepsilon}}}$$

(2) $\varepsilon = 1/2$ **时,有**

$$p = \frac{\left(\frac{1}{2}\right)^{\frac{\frac{1}{2}}{1-\frac{1}{2}}}}{1+\left(1-\frac{1}{2}\right)\times\left(\frac{1}{2}\right)^{\frac{\frac{1}{2}}{1-\frac{1}{2}}}} = \frac{\frac{1}{2}}{1+\frac{1}{4}} = \frac{2}{5}$$

所以

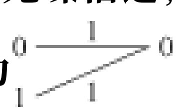
$$C = H(Y) - H(Y|X) = H(1/5) - 2/5 = 0.7219 - 0.4 = 0.3219 \text{ 比特/符号}$$

(3) **记** $\gamma = \varepsilon^{\frac{\varepsilon}{1-\varepsilon}}$**,则**

$$\lim_{\varepsilon\to 0}\gamma = \lim_{\varepsilon\to 0}\varepsilon^{\frac{\varepsilon}{1-\varepsilon}} = \lim_{\varepsilon\to 0}\varepsilon^{\varepsilon} = 1$$

$$\lim_{\varepsilon\to 0} p = \lim_{\varepsilon\to 0}\frac{\gamma}{1+(1-\varepsilon)\gamma} = \frac{1}{1+1} = \frac{1}{2}$$

所以 $\varepsilon\to 0$ **时** Z **信道趋于一个无噪信道,当输入等概时达到信道容量。**

当 $\varepsilon\to 1$ **时,原信道退化为****,此时** $C=0$**,不存在最佳输入分布,即任意的输入分布均可使信道达到信道容量。这从**(1)**求最佳输入分布的过程也可以得到解释。当** $\varepsilon\to 1$ **时,** $\delta\to 0$**,这时** $\frac{\mathrm{d}I}{\mathrm{d}p}=0$ **成为恒等式,任意的** p **均可使该等式成立。**

4.6　**如图 4.6 所示把** n **个二元对称信道串接起来,每个二元对称信道的错误传递概率为** p**。证明这** n **个串接信道可以等效于一个二元对称信道,其错误传递概率为** $\frac{1}{2}[1-(1-2p)^n]$**,并证明** $\lim_{n\to\infty} I(X_0;X_n)=0$**,设** $p\neq 0$ **或 1。**

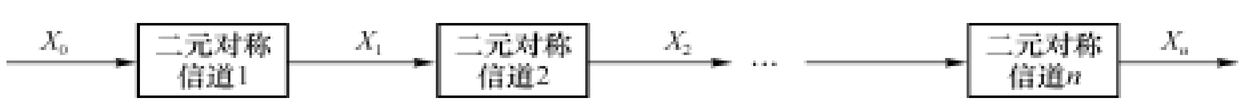


图 4.6

解:

$n=2$ 时,级联信道转移概率矩阵为

$$\begin{bmatrix} 1-p & p \\ p & 1-p \end{bmatrix} \cdot \begin{bmatrix} 1-p & p \\ p & 1-p \end{bmatrix} = \begin{bmatrix} 1-2p+2p^2 & 2p-2p^2 \\ 2p-2p^2 & 1-2p+2p^2 \end{bmatrix}$$

其错误传递概率为 $2p-2p^2=\frac{1}{2}[1-(1-2p)^2]$,故 $n=2$ 时结论成立。

设 $n=k$ 时结论成立,则错误传递概率为 $\frac{1}{2}[1-(1-2p)^k]$。

当 $n=k+1$ 时,级联信道转移概率矩阵为

$$\begin{bmatrix} \frac{1}{2}[1+(1-2p)^k] & \frac{1}{2}[1-(1-2p)^k] \\ \frac{1}{2}[1-(1-2p)^k] & \frac{1}{2}[1+(1-2p)^k] \end{bmatrix} \cdot \begin{bmatrix} 1-p & p \\ p & 1-p \end{bmatrix} = \begin{bmatrix} \frac{1}{2}[1+(1-2p)^{k+1}] & \frac{1}{2}[1-(1-2p)^{k+1}] \\ \frac{1}{2}[1-(1-2p)^{k+1}] & \frac{1}{2}[1+(1-2p)^{k+1}] \end{bmatrix}$$

由数学归纳法可知结论成立,即 n 个二元对称信道的级联信道等效于一个二元对称信道,其错误传递概率为 $\frac{1}{2}[1-(1-2p)^n]$。

当 $n\to\infty$ 时,$\lim\limits_{n\to\infty}\frac{1}{2}[1-(1-2p)^n]=\frac{1}{2}$,所以 $\lim\limits_{n\to\infty} I(X_0;X_n)=1-H\left(\frac{1}{2}\right)=0$。

4.7 试求出准对称信道的信道容量的一般表达式。

解:

设输入符号个数为 r,输出符号个数为 s,对于行对称信道,有

$$\begin{aligned} & I(X;Y) \\ &= H(Y)-H(Y|X) \\ &= p(y_1)\log_2\frac{1}{p(y_1)}+p(y_2)\log_2\frac{1}{p(y_2)}+\cdots+p(y_s)\log_2\frac{1}{p(y_s)}-H(p_1',p_2',\cdots,p_s') \end{aligned}$$

由于准对称信道当输入等概分布 $p(x_i)=\frac{1}{r}$, $i=1,2,\cdots,r$ 时达到信道容量,这时

$$p(y_j)=\sum_i p(y_j\mid x_i)p(x_i)=\frac{1}{r}\sum_i p(y_j\mid x_i)$$

所以

$$\begin{aligned} C = {} & \frac{1}{r}\sum_i p(y_1\mid x_i)\log_2\frac{1}{\frac{1}{r}\sum_i p(y_1\mid x_i)}+\frac{1}{r}\sum_i p(y_2\mid x_i)\log_2\frac{1}{\frac{1}{r}\sum_i p(y_2\mid x_i)}+ \\ & \cdots+\frac{1}{r}\sum_i p(y_s\mid x_i)\log_2\frac{1}{\frac{1}{r}\sum_i p(y_s\mid x_i)}-H(p_1',p_2',\cdots,p_s') \\ = {} & \frac{1}{r}\sum_i p(y_1'\mid x_i)\log_2 r+\frac{1}{r}\sum_i p(y_2\mid x_i)\log_2 r+\cdots+\frac{1}{r}\sum_i p(y_s\mid x_i)\log_2 r - \end{aligned}$$

$$\left[\frac{1}{r}\sum_i p(y_1 \mid x_i)\log_2\sum_i p(y_1 \mid x_i) + \frac{1}{r}\sum_i p(y_2 \mid x_i)\log_2\sum_i p(y_2 \mid x_i) + \cdots + \frac{1}{r}\sum_i p(y_s \mid x_i)\log_2\sum_i p(y_s \mid x_i)\right] - H(p_1', p_2', \cdots, p_s')\Bigg]$$

$$= \frac{1}{r}\left[\sum_i p(y_1 \mid x_i) + \cdots + \sum_i p(y_s \mid x_i)\right]\cdot\log_2 r - \left[\frac{1}{r}\sum_i p(y_1 \mid x_i)\log_2\sum_i p(y_1 \mid x_i) + \frac{1}{r}\sum_i p(y_2 \mid x_i)\log_2\sum_i p(y_2 \mid x_i) + \cdots + \frac{1}{r}\sum_i p(y_s \mid x_i)\log_2\sum_i p(y_s \mid x_i)\right] - H(p_1', p_2', \cdots, p_s')\Bigg] \tag{4.4}$$

因为准对称信道可以按列分成若干个对称的子矩阵 $Q_k(k=1,2,\cdots,n)$,在子矩阵中每一行、每一列的元素都相同,把式(4.4)括号内的求和分成在不同的子矩阵中求和。

$\sum_i p(y_j \mid x_i)$ 表示每一列元素之和,在每一个子矩阵中 $\sum_i p(y_j \mid x_i)$ 都相等,因此在同一子矩阵中求和项可以进行合并,合并后对数符号前的值就等于子矩阵中行元素之和,所以

$$C = \log_2 r - \sum_{k=1}^{n} N_k \log_2 M_k - H(p_1', p_2', \cdots, p_s')$$

其中 N_k 表示子矩阵中行元素之和,M_k 表示子矩阵中列元素之和。

4.8　**试画出三元对称信道在理想(无噪声)和强噪声(输出不依赖于输入)情况下的信道模型,设信道输入等概分布。**

解:

(1) **三元对称理想(无噪声)信道的信道模型如图 4.7 所示。**

(2) **三元对称强噪声信道的信道模型如图 4.8 所示。**

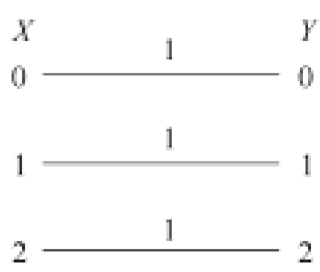


图4.7

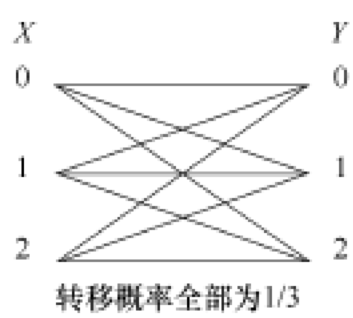


图4.8

4.9　**串联信道如图 4.9 所示,求总的信道矩阵。**

解:

由图 4.9 可知信道 Ⅰ、Ⅱ 的信道矩阵分别为

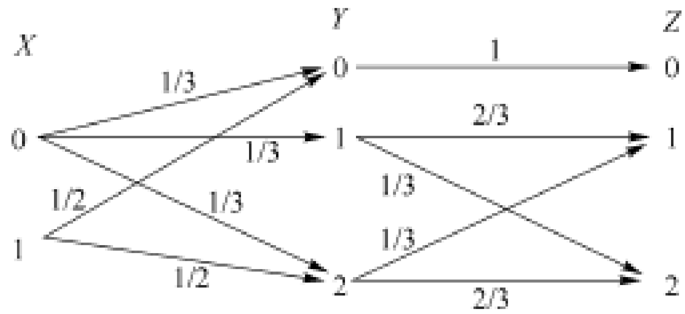


图 4 .9

$$P_{\mathrm{I}}=\begin{bmatrix}\frac{1}{3} & \frac{1}{3} & \frac{1}{3}\\ \frac{1}{2} & 0 & \frac{1}{2}\end{bmatrix},\qquad P_{\mathrm{II}}=\begin{bmatrix}1 & 0 & 0\\ 0 & \frac{2}{3} & \frac{1}{3}\\ 0 & \frac{1}{3} & \frac{2}{3}\end{bmatrix}$$

它们串联后构成一个马尔可夫链,根据马氏链的性质,串联后总的信道矩阵为

$$P=P_{\mathrm{I}}\,P_{\mathrm{II}}=\begin{bmatrix}\frac{1}{3} & \frac{1}{3} & \frac{1}{3}\\ \frac{1}{2} & 0 & \frac{1}{2}\end{bmatrix}\begin{bmatrix}1 & 0 & 0\\ 0 & \frac{2}{3} & \frac{1}{3}\\ 0 & \frac{1}{3} & \frac{2}{3}\end{bmatrix}=\begin{bmatrix}\frac{1}{3} & \frac{1}{3} & \frac{1}{3}\\ \frac{1}{2} & \frac{1}{6} & \frac{1}{3}\end{bmatrix}$$

总的级联信道可以等效为图4 .10。

4 .10　设一时间离散、幅度连续的无记忆信道的输入是一个零均值、方差为 E 的高斯随机变量,信道噪声为加性高斯噪声,方差为 $\sigma^2=1\ \mu\mathrm{W}$,信道的符号传输速率为 $r=8\,000$符号/ 秒。如令一路电话通过该信道,电话机产生的信息率为 64 kbps,求输入信号功率 E 的最小值。

解:

由香农公式,该信道的信道容量为

$$C_t=\frac{r}{2}\log_2\left(1+\frac{E}{\sigma^2}\right)=\frac{8\,000}{2}\log_2(1+E)$$

为使电话机产生的 64 kbps 数据正确通过信道,必须

$$C_t\geq 64\times 10^3\ \mathrm{bps}$$

得

$$E\geq 1\times(2^{16}-1)=65\,535\ \mu\mathrm{W}$$

所以输入信号功率应不小于 65 .5 mW。

4 .11　连续随机变量 X 和 Y 的联合概率密度函数在由$\frac{1}{a}|x|+\frac{1}{b}|y|\leq 1$ 确定的菱形内均匀分布,如图 4 .11 所示。

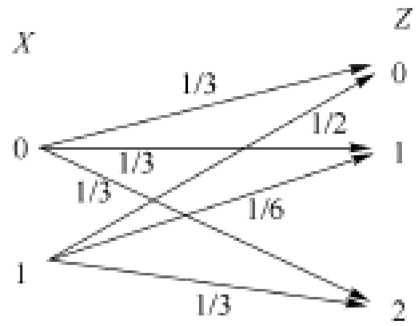


图 4 .10

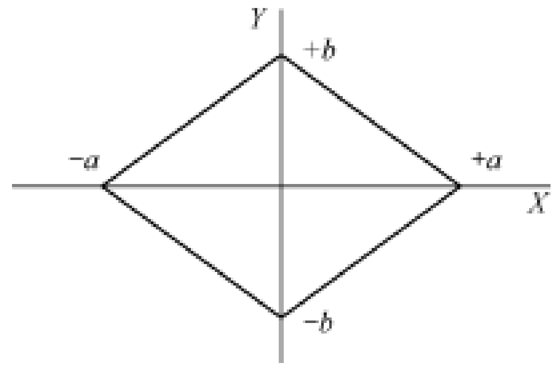


图 4 .11

(1) **求** $I(X;Y)$;

(2) **解释为什么** $I(X;Y)$**与** a **和** b **无关。**

解:

(1) $$I(X;Y)=h(Y)-h(Y|X)=h(X)+h(Y)-h(XY)\quad h(XY)$$

$$=-\frac{1}{2ab}\ln\frac{1}{2ab}\mathrm{d}x\mathrm{d}y=\ln 2ab\text{ 奈特/ 样值}$$

$$h(X)=\int_{-a}^{+a}\left(\frac{1}{a}-\frac{|x|}{a^2}\right)\log_2\left(\frac{1}{a}-\frac{|x|}{a^2}\right)\mathrm{d}x$$

$$=\ln a-\int_{-1}^{+1}(1-|x|)\ln(1-|x|)\mathrm{d}x$$

$$=\frac{1}{2}+\ln a\text{ 奈特/ 样值}$$

同理

$$h(Y)=\frac{1}{2}+\ln b\text{ 奈特/ 样值}$$

所以

$$I(X;Y)=h(X)+h(Y)-h(XY)=1-\ln 2\text{ 奈特/ 样值}$$

(2) a **和** b **由度量** X **和** Y **的单位决定,改变这个度量单位,不影响互信息的大小。用数学的语言来说,假定** X_1, Y_1 **的联合概率密度函数在由** $|x|+|y|\leq 1$ **确定的菱形内均匀分布,那么** $X=aX_1$, $Y=bY_1$**。根据推导,可以得到**:

$$h(XY)=h(X_1Y_1)+\log_2(ab)$$

$$h(X)=h(X_1)+\log_2 a,\qquad h(Y)=h(Y_1)+\log_2 b$$

$$I(X;Y)=h(X)+h(Y)-h(XY)$$

$$=h(X_1)+\log_2 a+h(Y_1)+\log_2 b-[h(X_1Y_1)+\log_2(ab)]$$

$$=h(X_1)+h(Y_1)-h(X_1Y_1)$$

所以 $I(X;Y)$**与** a **和** b **无关。**

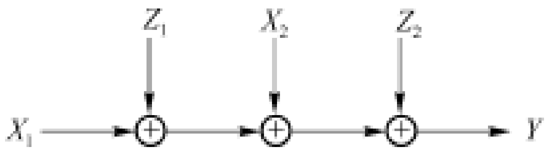


图 4 .12

4 .12 **高斯加性信道,输入信号 X_1, X_2,噪声信号 Z_1, Z_2,输出信号 $Y = X_1 + Z_1 + X_2 + Z_2$,如图 4 .12 所示。输入和噪声均为相互独立的零均值的高斯随机变量,功率分别为 P_1, P_2 和 N_1, N_2。**

(1) **求** $I(X_1; Y_1)$**和** $I(X_2; Y)$;

(2) **求** $I(X_1 X_2; Y)$;

(3) **当输入信号的总功率受限** $P_1 + P_2 \leq P$ **时,求** $I(X_1; Y) + I(X_2; Y)$**的最大值。**

解:

(1) $Y - X_1 = X_2 + Z_1 + Z_2$ **是 3 个独立高斯随机变量的和,因此**

$$h(Y | X_1) = h(Y - X_1) = \frac{1}{2}\log_2[2\pi e(P_2 + N_1 + N_2)]$$

$$h(Y) = \frac{1}{2}\log_2[2\pi e(P_1 + P_2 + N_1 + N_2)]$$

$$I(X_1; Y) = h(Y) - h(Y | X_1) = \frac{1}{2}\log_2\left(\frac{P_1 + P_2 + N_1 + N_2}{P_2 + N_1 + N_2}\right)$$

$$= \frac{1}{2}\log_2\left(1 + \frac{P_1}{P_2 + N_1 + N_2}\right)$$

$$I(X_2; Y) = \frac{1}{2}\log_2\left(1 + \frac{P_2}{P_1 + N_1 + N_2}\right)$$

(2) **当给定** X_1, X_2 **的值** x_1, x_2 **以后**, Y **的条件概率密度是均值为** $x_1 + x_2$、**方差为** $N_1 + N_2$ **的正态分布。**

$$h(Y | X_1 X_2) = h(Y - X_1 - X_2) = \frac{1}{2}\log_2[2\pi e(N_1 + N_2)]$$

$$I(X_1 X_2; Y) = h(Y) - h(Y | X_1 X_2) = \frac{1}{2}\log_2\left(\frac{P_1 + P_2 + N_1 + N_2}{N_1 + N_2}\right) = \frac{1}{2}\log_2\left(1 + \frac{P_1 + P_2}{N_1 + N_2}\right)$$

(3) $I(X_1; Y)$**和** $I(X_2; Y)$**可以写成**:

$$I(X_1; Y) = \frac{1}{2}\log_2\left(\frac{P}{P_2 + N_1 + N_2}\right), \qquad I(X_2; Y) = \frac{1}{2}\log_2\left(\frac{P}{P_1 + N_1 + N_2}\right)$$

其中 $P = P_1 + P_2 + N_1 + N_2 = \mathrm{Var}(Y)$**是输出总功率,因此**

$$I(X_1; Y) + I(X_2; Y) = \frac{1}{2}\log_2\left(\frac{P}{P_2 + N_1 + N_2}\right) + \frac{1}{2}\log_2\left(\frac{P}{P_1 + N_1 + N_2}\right)$$

$$= \frac{1}{2}\log_2\left[\frac{P^2}{(P_2 + N_1 + N_2)(P_1 + N_1 + N_2)}\right]$$

要使 $I(X_1;Y)+I(X_2;Y)$得到最大值,$(P_2+N_1+N_2)(P_1+N_1+N_2)$的值需最小。因为这两个因子的和是个常数,所以当这两个因子相等时,$(P_2+N_1+N_2)(P_1+N_1+N_2)$最大,而这两个因子相差越大时,$(P_2+N_1+N_2)(P_1+N_1+N_2)$越小。即当 $P_1=P$或 $P_2=P$ 时,$I(X_1;Y)+I(X_2;Y)$得到最大值。

4.13　一个无记忆信道输入为离散随机变量 X,噪声 Z 在区间$[-a,+a]$上均匀分布,因此输出 $Y=X+Z$ 是一个连续随机变量。

(1) 当 $X\in\{-1,+1\}$并且等概分布时,求 $I(X;Y)$(用 a 表示);

(2) 当 $X\in\{-1,0,+1\}$, $a=1/2$ 时,求最佳输入分布。

解:

(1) 分两种情况讨论。

对于 $X=-1$,输出 Y 的最大值为 $-1+a$,而对于 $X=+1$,输出 Y 的最小值为 $+1-a$,

$$-1+a<+1-a \qquad 2a<2 \qquad a<1$$

当 $a<1$ 时,不同的输入信号经信道传输后在输出端不会重叠,由输出 Y 可以确定输入 X,所以 $I(X;Y)=H(X)=1$ 比特/样值。

当 $a>1$ 时,对应不同的输入信号的输出 Y 值会发生重叠,重叠区域为 $I_0=[+1-a,-1+a]$,重叠区间长度为$(-1+a)-(1-a)=2a-2$。如果 Y 落入这个区间则不能提供任何关于 X 的信息,因为这时关于不同输入信号的条件概率是相等的:

$$P(X=+1|Y=y)=\frac{P_X(+1)p(y|X=+1)}{P_X(-1)p(y|X=-1)+P_X(+1)p(y|X=+1)}=\frac{1}{2}$$

所以当 $y\in I_0$ 时,$H(X|Y=y)=1$ 比特/符号。而当 $y<+1-a$ 或 $y>-1+a$ 时,则由输出 Y 可以确定输入 X,所以

$$H(X|Y=y)=0$$

$$H(X|Y)=P(Y\in I_0)\times 1+P(Y\notin I_0)\times 0=\frac{2a-2}{2a}=1-\frac{1}{a}\text{ 比特/样值}$$

因此当 $X\in\{-1,+1\}$并且等概分布时,有

$$I(X;Y)=H(X)-H(X|Y)=1-(1-\frac{1}{a})=\frac{1}{a}\text{ 比特/样值}$$

(2) 当 $X\in\{-1,0,+1\}$, $a=1/2$ 时,对应不同的输入信号的输出 Y 值正好能区分开,所以 $I(X;Y)=H(X)$。当 $H(X)$最大时,也就是 X 为等概分布时平均互信息得到最大值:

$$I(X;Y)=\log_2 3=1.585\text{ 比特/样值}$$

4.14 **设某一信号的信息输出率为 5.6 kbps，噪声功率谱为 $N = 5\times 10^{-6}$ mW/Hz，在带宽 $B = 4$ kHz 的高斯信道中传输。试求无差错传输需要的最小输入功率 P 是多少？**

解：

由香农公式，该信道的信道容量为

$$C_t = B\log_2\left(1+\frac{P}{NB}\right) = 4\times 10^3\log_2\left(1+\frac{P}{5\times 10^{-9}\times 4\times 10^3}\right)$$

为使 5.6 kbps 的信号无差错传输，必须

$$C_t \geq 5.6\times 10^3 \text{ bps}$$

得

$$P \geq 3.28\times 10^{-5} \text{ mW}$$

即输入信号功率应不小于 3.28×10^{-5} mW。

4.15 **判断图 4.13 中各信道是否对称，如对称，求出其信道容量。**

解：

(a) 图，准对称信道。

$$P = \begin{bmatrix} 1-\varepsilon-\delta & \delta & \varepsilon \\ \varepsilon & \delta & 1-\varepsilon-\delta \end{bmatrix}$$

$$C = \log_2 2 - H(1-\varepsilon-\delta, \delta, \varepsilon) - (1-\delta)\log_2(1-\delta) - \delta\log_2(2\delta)$$

(b) 图，对称信道。

$$P = \begin{bmatrix} 1-\varepsilon & \varepsilon & 0 \\ 0 & 1-\varepsilon & \varepsilon \\ \varepsilon & 0 & 1-\varepsilon \end{bmatrix}$$

$$C = \log_2 3 - H(1-\varepsilon, \varepsilon)$$

(c) 图，不对称信道。

$$P = \begin{bmatrix} 1-\varepsilon & \varepsilon & 0 \\ \varepsilon & 1-\varepsilon & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

(d) 图，准对称信道。

$$P = \begin{bmatrix} p & p & 0 & p \\ 0 & p & p & p \\ p & 0 & p & p \end{bmatrix}$$

$$C = \log_2 3 - H(p, p, p) - 2p\log_2(2p) - p\log_2(3p) = 0.39 \text{ 比特/符号}$$

(e) 图,不对称信道。

$$P = \begin{bmatrix} 1 & 0 & 0 \\ \frac{1}{2} & \frac{1}{4} & \frac{1}{4} \\ 0 & \frac{1}{2} & \frac{1}{2} \end{bmatrix}$$

(f) 图,对称信道。

$$P = \begin{bmatrix} \alpha & \beta & \alpha & \beta \\ \beta & \alpha & \beta & \alpha \end{bmatrix}$$

$$C = \log_2 4 - H(\alpha, \beta, \alpha, \beta) = 1 - H(p) \text{ 比特/符号}$$

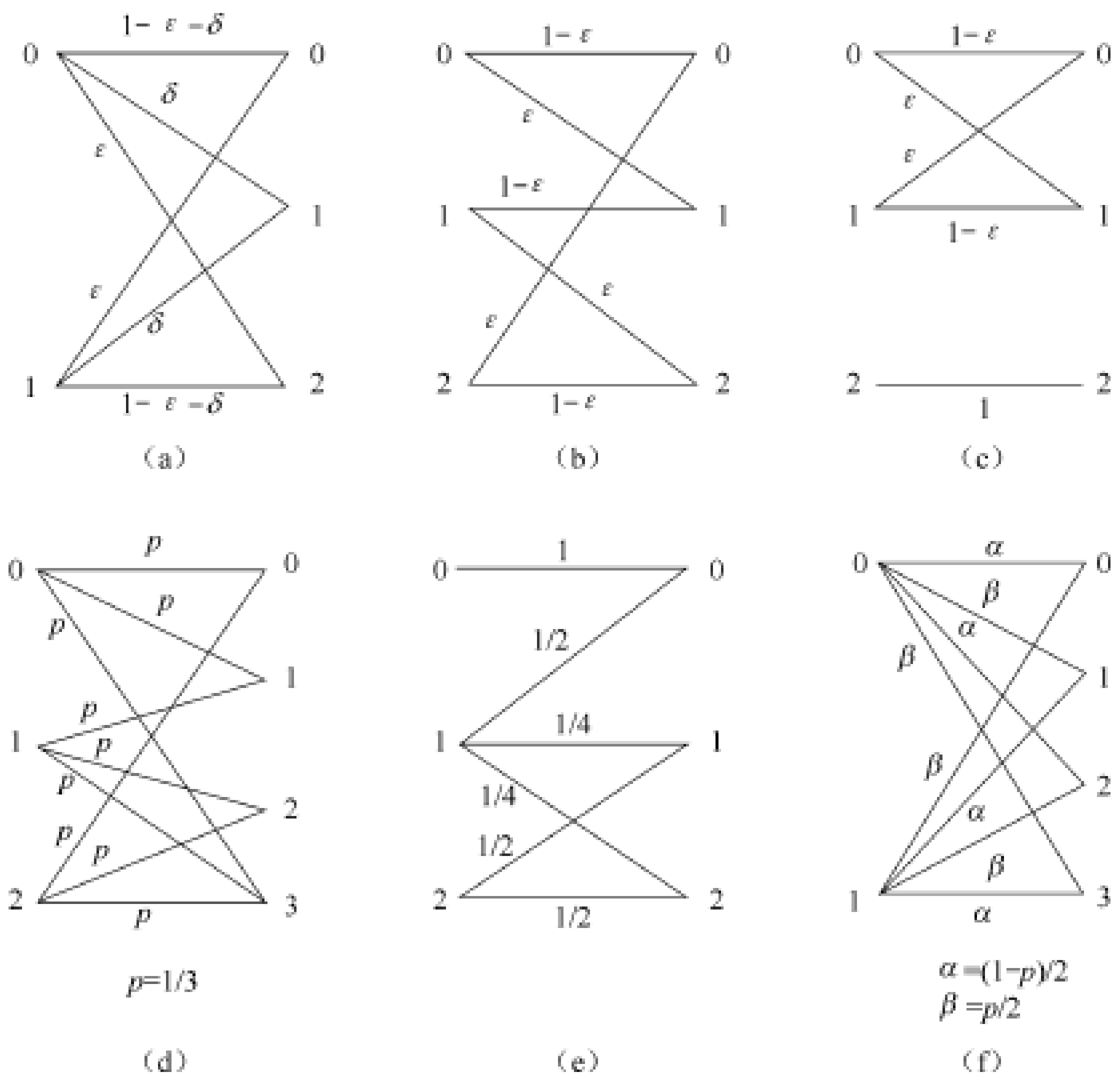


图 4.13

4.16　一个快餐店只提供汉堡包和牛排,当顾客进店以后只需向厨房喊一声“B”或“Z”就表示他点的是汉堡包或牛排,不过通常 8%的概率厨师可能会听错。一般来说进店的顾客 90%会点汉堡包,10%会点牛排。问:

(1) **这个信道的信道容量**;

(2) **每次顾客点菜时提供的信息**;

(3) **在这个信道可不可以正确地传递顾客点菜的信息**?

解:

(1) **这是一个错误传递概率为** 0.08 **的二元对称信道,它的信道容量为**

$$C = 1 - 0.08\log_2 \frac{1}{0.08} - 0.92\log_2 \frac{1}{0.92} = 0.5978 \text{ 比特/样值}$$

(2) **每次顾客点菜时提供的信息为**

$$H(X) = 0.1\log_2 \frac{1}{0.1} + 0.9\log_2 \frac{1}{0.9} = 0.469 \text{ 比特/样值}$$

(3) **由于需要传递的信息量小于信道容量** $H(X) < C$,**因此在这个信道可以正确地传递顾客点菜的信息。**

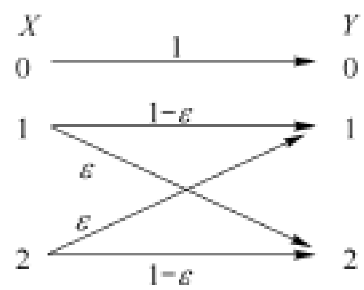


图 4.14

4.17 **求图 4.14 中信道的信道容量及其最佳的输入概率分布,并求当** $\varepsilon = 0$ **和** $1/2$ **时的信道容量值。**

解:

信道转移矩阵为 $P = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1-\varepsilon & \varepsilon \\ 0 & \varepsilon & 1-\varepsilon \end{bmatrix}$。**由于输入为** 1 **和** 2 **时信道的转移概率对称分布,所以可以设信源的概率分布为**

$$\begin{bmatrix} X \\ P \end{bmatrix} = \begin{bmatrix} 0 & 1 & 2 \\ 1-2p & p & p \end{bmatrix}$$

由此可以得到输出分布为

$$\begin{bmatrix} X \\ P \end{bmatrix} = \begin{bmatrix} 0 & 1 & 2 \\ 1-2p & p & p \end{bmatrix}$$

$$\begin{aligned} I(X;Y) &= H(Y) - H(Y \mid X) = H(Y) - \sum_{i=1}^{3} p(x_i) H(Y \mid x_i) \\ &= H(Y) - [p(x_1) H(1,0,0) + p(x_2) H(1-\varepsilon, \varepsilon) + p(x_2) H(1-\varepsilon, \varepsilon)] \\ &= H(Y) - 2pH(\varepsilon) = -(1-2p)\log_2(1-2p) - p\log_2 p - p\log_2 p - 2pH(\varepsilon) \end{aligned}$$

因为 $C = \max_p I(X;Y)$,**所以令**

$$\frac{I}{p} = -2H(\varepsilon) + 2\log_2(1-2p) - 2\log_2 p = 0$$

解得

$$p = \frac{1}{2 + 2^{H(\varepsilon)}}$$

这时

$$\begin{bmatrix} Y \\ P \end{bmatrix} = \begin{bmatrix} 0 & 1 & 2 \\ 1-2p & p & p \end{bmatrix} = \begin{bmatrix} 0 & 1 & 2 \\ 2^{H(\varepsilon)} \cdot [2+2^{H(\varepsilon)}]^{-1} & [2+2^{H(\varepsilon)}]^{-1} & [2+2^{H(\varepsilon)}]^{-1} \end{bmatrix}$$

$$C = \max I(X;Y) = \frac{2^{H(\varepsilon)}}{2+2^{H(\varepsilon)}}\log_2\left[\frac{2+2^{H(\varepsilon)}}{2^{H(\varepsilon)}}\right] + \frac{2}{2+2^{H(\varepsilon)}}\log_2[2+2^{H(\varepsilon)}] - \frac{2}{2+2^{H(\varepsilon)}}H(\varepsilon)$$

当 $\varepsilon = 0$ **时**, $H(\varepsilon) = 0$, **所以**

$$C = \log_2 3 \approx 1.585 \text{ 比特/符号}$$

当 $\varepsilon = 1/2$ **时**, $H(\varepsilon) = 1$, **所以**

$$C = \frac{1}{2}\log_2 2 + \frac{1}{2}\log_2 4 - \frac{1}{2}\times 1 = 1 \text{ 比特/符号}$$

4.18　**积信道。有两个离散无记忆信道**$\{X_1, P(Y_1|X_1), Y_1\}$**和**$\{X_2, P(Y_2|X_2), Y_2\}$，**信道容量分别为** C_1 **和** C_2**。两个信道同时分别输入** X_1 **和** X_2，**输出** Y_1 **和** Y_2，**这两个信道组成一个新的信道。求它的信道容量。**

解:

由于这两个离散无记忆信道相互独立,因此

$$P(Y_1Y_2|X_1X_2) = P(Y_1|X_1)P(Y_2|X_2)$$

$$\begin{aligned} I(X_1X_2;Y_1Y_2) &= H(Y_1Y_2) - H(Y_1Y_2|X_1X_2) \\ &= H(Y_1Y_2) - H(Y_1|X_1X_2) - H(Y_2|X_1X_2Y_1) \\ &= H(Y_1Y_2) - H(Y_1|X_1) - H(Y_2|X_2) \\ &\leq H(Y_1) + H(Y_2) - H(Y_1|X_1) - H(Y_2|X_2) \\ &= I(X_1;Y_1) + I(X_2;Y_2) \end{aligned}$$

当 X_1 **和** X_2 **相互独立时**, Y_1 **和** Y_2 **相互独立,因此**

$$\begin{aligned} C &= \max_{p(x_1x_2)} I(X_1X_2;Y_1Y_2) \\ &\leq \max_{p(x_1)} I(X_1;Y_1) + \max_{p(x_2)} I(X_2;Y_2) \\ &= C_1 + C_2 \end{aligned}$$

当 $p(x_1x_2) = p^*(x_1)p^*(x_2)$**时等号成立**, $p^*(x_1)$**和** $p^*(x_2)$**是使这两个信道分别达到信道容量** C_1 **和** C_2 **的最佳输入分布。**

4.19 **和信道。有两个离散无记忆信道** $\{X_1, P(Y_1 | X_1), Y_1\}$ **和** $\{X_2, P(Y_2 | X_2), Y_2\}$,**信道容量分别为** C_1 **和** C_2。**这两个信道的输入输出符号集各不相同,并且假定每次只有一个信道有输入,试证明:**

(1) $2^C = 2^{C_1} + 2^{C_2}$

(2) **如果** $C_1 > C_2$,**那么一直用信道 1 是不是效率更高?**

$$X\left\{\begin{array}{l} X_1 \rightarrow \boxed{P_1(Y_1 | X_1)} \rightarrow Y_1 \\ \\ X_2 \rightarrow \boxed{P_2(Y_2 | X_2)} \rightarrow Y_2 \end{array}\right\} Y$$

证明:

(1) **假定输入**

$$X = \begin{cases} X_1 & \text{概率为 } \alpha \\ X_2 & \text{概率为}(1-\alpha) \end{cases}$$

$$\theta(X) = \begin{cases} 1 & X = X_1 \\ 2 & X = X_2 \end{cases}$$

由于输出符号集各不相同,所以 θ **也是** Y **的函数,因为**

$$I(X; Y\theta) = I(X;\theta) + I(X; Y|\theta) = I(X; Y) + I(X;\theta| Y)$$

由于 θ **是** Y **的函数,所以** $I(X;\theta| Y) = 0$,**因此**

$$\begin{aligned} I(X; Y) &= I(X;\theta) + I(X; Y|\theta) \\ &= \mathbf{H}(\theta) - H(\theta| X) + \alpha I(X; Y|\theta = 1) + (1-\alpha) I(X; Y|\theta = 2) \\ &= H(\alpha) + \alpha I(X_1; Y_1) + (1-\alpha) I(X_2; Y_2) \end{aligned}$$

$$C = \sup_{\alpha}\{H(\alpha) + \alpha C_1 + (1-\alpha) C_2\}$$

这是一个关于 α **的严格上凸函数,当** $\alpha = 2^{C_1}/(2^{C_1} + 2^{C_2})$ **时最大值存在,最大值为**

$$C = \log_2(2^{C_1} + 2^{C_2})$$

(2) **从** $I(X; Y) = H(\alpha) + \alpha I(X_1; Y_1) + (1-\alpha) I(X_2; Y_2)$ **可以看出,如果以** α **的概率使用信道 1,以** $(1-\alpha)$ **的概率使用信道 2,则不仅得到了** $\alpha I(X_1; Y_1) + (1-\alpha) I(X_2; Y_2)$ **的信息量,而且还有另外的** $H(\alpha)$,**这是因为在选择信道 1 或信道 2 时也传递了信息。**

4.20 **有记忆信道的信道容量高于无记忆信道的信道容量。考虑一个二元对称信道** $Y_i = X_i \quad Z_i$, **表示模 2 加**, $X_i, Y_i \in \{0,1\}$。**假定** $Z_1, Z_2 \cdots, Z_n$ **有相同的边缘概率分布** $P(Z_i = 1) = p = 1 - P(Z_i = 0)$,**但是并不相互独立,但是** Z^n **与输入** X^n **相互独立。如果记** $C = 1 - H(p, 1-p)$,**证明:**

$$\max_{p(x_1 x_2 \cdots x_n)} I(X_1 X_2 \cdots X_n; Y_1 Y_2 \cdots Y_n) \geq nC$$

证明:

根据题意,有

$$Z_i = \begin{cases} 1 & \text{概率为 } p \\ 0 & \text{概率为}(1-p) \end{cases}$$

由于 $Y_i = X_i \quad Z_i$ **而且** $Z_1, Z_2 \cdots Z_n$ **并不相互独立,所以**

$$\begin{aligned} I(X_1 \cdots X_n; Y_1 \cdots Y_n) &= H(X_1 \cdots X_n) - H(X_1 \cdots X_n | Y_1 \cdots Y_n) \\ &= H(X_1 \cdots X_n) - H(Z_1 \cdots Z_n | Y_1 \cdots Y_n) \\ &\geq H(X_1 \cdots X_n) - H(Z_1 \cdots Z_n) \\ &\geq H(X_1 \cdots X_n) - \sum_{i=1}^{n} H(Z_i) \end{aligned}$$

这个离散有记忆信道的信道容量记为 $C^{(n)}$**,当**$\{X_i\}$**为独立同分布且为** $p = \frac{1}{2}$**的贝努利分布时,有**

$$C^{(n)} \geq H(X_1 \cdots X_n) - \sum_{i=1}^{n} H(Z_i) = n - nH(p) = n[1 - H(p)] = nC$$

从直觉上来说,由于噪声样值之间的相关性减小了有效噪声,使得有记忆信道的信道容量高于无记忆信道的信道容量。

4.21　**时变信道。离散无记忆时变信道** $P(Y_1 \cdots Y_n | X_1 \cdots X_n) = \prod_{i=1}^{n} P_i(Y_i | X_i)$**,求它的信道容量** $\max\limits_{p(x_1 x_2 \cdots x_n)} I(X_1 X_2 \cdots X_n; Y_1 Y_2 \cdots Y_n)$。

解:

由于是无记忆信道,所以

$$\begin{aligned} I(X^n; Y^n) &= H(Y^n) - H(Y^n | X^n) \\ &= H(Y^n) - \sum_{i=1}^{n} H(Y_i | Y_1 \cdots Y_{i-1} X^n) \\ &= H(Y^n) - \sum_{i=1}^{n} H(Y_i | X_i) \\ &\leq \sum_{i=1}^{n} H(Y_i) - \sum_{i=1}^{n} H(Y_i | X_i) \\ &= \sum_{i=1}^{n} I(X_i; Y_i) \leq \sum_{i=1}^{n} C_i \end{aligned}$$

其中 C_i **是第** i **时刻的信道容量。当** $X_1, \cdots, X_n$ **相互独立并且均使对应时刻的信道达到信道容量时,等号成立。这时**

$$\max_{p(x_1 x_2 \cdots x_n)} I(X_1 X_2 \cdots X_n; Y_1 Y_2 \cdots Y_n) = \sum_{i=1}^{n} \max_{p(x_i)} I(X_i; Y_i) = \sum_{i=1}^{n} C_i$$

4.22　**假定** C **为有** N **个输入,** M **个输出的离散无记忆信道的信道容量,证明** $C \leq \min\{\log_2 M, \log_2 N\}$。

证明:

因为 $H(X|Y)\geq 0$, $H(Y|X)\geq 0$,所以

$$C=\max_{p(x)} I(X;Y)=\max_{p(x)}[H(Y)-H(Y|X)]=\max_{p(x)}[H(X)-H(X|Y)]$$
$$\leq \min\{\max[H(Y)],\max[H(X)]\}$$

当 X 和 Y 为等概分布时 $H(X)$ 和 $H(Y)$ 达到最大值,因此

$$C\leq \min\{\log_2 M,\log_2 N\}$$

4.23 证明: n 个 $P(Y=1|X=0)=\varepsilon$ 的 Z 信道(如图 4.15 所示)级联相当于一个 $P(Y=1|X=0)=1-(1-\varepsilon)^n$ 的 Z 信道。

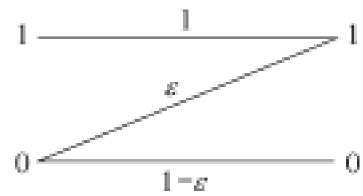


图 4.15

证明:

n 个 Z 信道级联以后的等效信道,当输入为 1 时,输出必定为 1,而当输入为 0 时,输出为 1 的概率为

$$P(Y=1|X=0)=\varepsilon+(1-\varepsilon)\varepsilon+\cdots+(1-\varepsilon)^{n-1}\varepsilon=\varepsilon\sum_{i=0}^{n-1}(1-\varepsilon)^n$$

$$=\varepsilon\frac{1-(1-\varepsilon)^n}{1-(1-\varepsilon)}=1-(1-\varepsilon)^n$$

当 $n\to\infty$时, $P(Y=1|X=0)=1$,因此输出必定为 1。

4.24 一个信道的信道矩阵为

$$\begin{array}{c} \\ 0 \\ 1 \\ 2 \end{array}\begin{array}{c} \begin{array}{ccc} 0 & 1 & 2 \end{array} \\ \begin{bmatrix} 3/4 & 1/4 & 0 \\ 1/3 & 1/3 & 1/3 \\ 0 & 1/4 & 3/4 \end{bmatrix} \end{array}$$

(1) 求信道容量;

(2) 若它的输入概率分布为

$$P(X=0)=1/2-p$$
$$P(X=1)=2p$$
$$P(X=2)=1/2-p$$

画出 $I(X=0;Y)$, $I(X=1;Y)$和 $I(X=2;Y)$作为 p 的函数的函数曲线, $0\leq p\leq 1/2$。

你能从这些曲线得到关于这个信道的信道容量和最佳输入分布的什么结论？

解:

(1) **因为** $X=0$ **和** $X=2$ **时信道的对称性以及** $X=1$ **时的不可靠传输,我们假定最佳输入概率分布为**

$$P(X=0)=1/2;\quad P(X=1)=0;\quad P(X=2)=1/2$$

这时

$$P(Y=0)=3/8;\quad P(Y=1)=1/4;\quad P(Y=2)=3/8$$

因为能够满足信道容量定理:

$$I(X=0;Y)=(3/4)\log_2(2)+(1/4)\log_2(1)=3/4^{>}\lambda$$

$$I(X=1;Y)=(1/3)\log_2(8/9)+(1/3)\log_2(4/3)+(1/3)\log_2(8/9)=0.0251\leq\lambda$$

$$I(X=2;Y)=(1/4)\log_2(1)+(3/4)\log_2(2)=3/4=\lambda$$

所以假定的输入分布确实是最佳输入分布,这时信道容量为

$$C=\lambda=3/4$$

(2) **根据假定的输入概率分布可得**

$$P(Y=0)=3/8+(2/3-3/4)p$$

$$P(Y=1)=1/4+(2/3-1/2)p$$

$$P(Y=2)=3/8+(2/3-3/4)p$$

因此

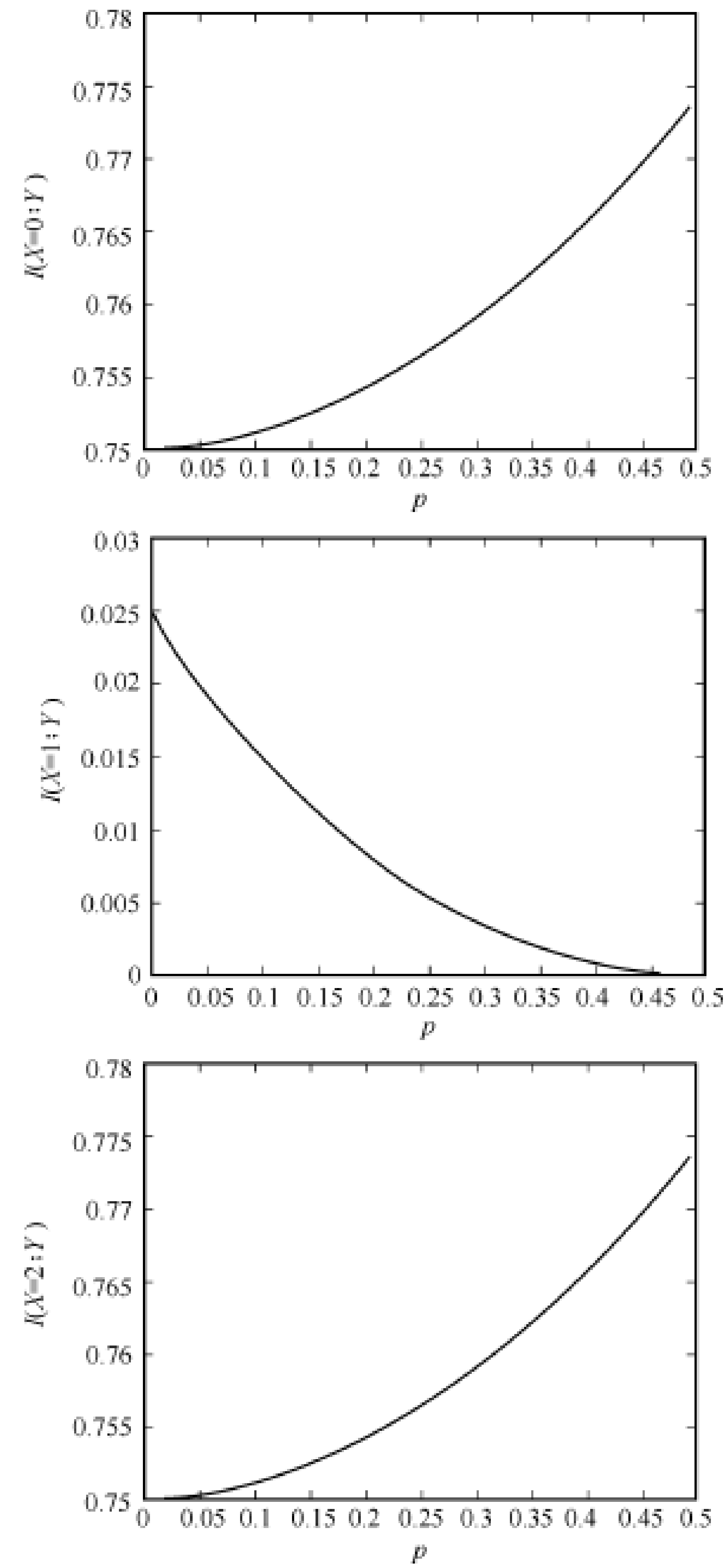
$$I(X=0;Y)=\frac{3}{4}\log_2\left[\frac{3/4}{3/8+(2/3-3/4)p}\right]+\frac{1}{4}\log_2\left[\frac{1/4}{1/4+(2/3-1/2)p}\right]$$

$$I(X=1;Y)=\frac{1}{3}\log_2\left[\frac{1/3}{3/8+(2/3-3/4)p}\right]+\frac{1}{3}\log_2\left[\frac{1/3}{1/4+(2/3-1/2)p}\right]+\frac{1}{3}\log_2\left[\frac{1/3}{3/8+(2/3-3/4)p}\right]$$

$$I(X=2;Y)=\frac{3}{4}\log_2\left[\frac{3/4}{3/8+(2/3-3/4)p}\right]+\frac{1}{4}\log_2\left[\frac{1/4}{1/4+(2/3-1/2)p}\right]$$

$I(X=0;Y)$, $I(X=1;Y)$**和** $I(X=2;Y)$**作为** p **的函数,它们的函数曲线如图 4.16 所示。**

从图中可以看出,当 $0\leq p\leq 1/2$ **时**, $I(X=1;Y)\leq I(X=0;Y)=I(X=2;Y)$,**所以要满足信道容量定理就必须** $P(X=1)=0$,**而只要** p **大于** 0, $P(X=1)$**就大于** 0,**所以必须** $p=0$,**因此输入概率分布为**



0.78
0.775
0.77
0.765
0.76
0.755
0.75
I(X=0:Y)
0 0.05 0.1 0.15 0.2 0.25 0.3 0.35 0.4 0.45 0.5
p
0.03
0.025
0.02
0.015
0.01
0.005
0
I(X=1:Y)
0 0.05 0.1 0.15 0.2 0.25 0.3 0.35 0.4 0.45 0.5
p
0.78
0.775
0.77
0.765
0.76
0.755
0.75
I(X=2:Y)
0 0.05 0.1 0.15 0.2 0.25 0.3 0.35 0.4 0.45 0.5
p

图 4 .16

$$P(X=0)=1/2$$
$$P(X=1)=0$$
$$P(X=2)=1/2$$

这时 $C=3/4$ 比特/符号,和第一问得到的结果一致。

4.25 假定(X, Y, Z)是一个多维高斯分布,并且 $X\to Y\to Z$ 组成一个马尔可夫链,X 和 Y、Y 和 Z 的相关系数分别为 ρ_1 和 ρ_2,求 $I(X;Z)$。

解:

首先,不失一般性我们假设 X、Y 和 Z 的均值为 0,因为这样假设不会影响平均互信息以及给定 Y 后 X 和 Z 的相互独立的关系。同样我们假设 X、Y 和 Z 的方差为 1(这可能会影响微分熵的结果但不会影响平均互信息),记 $R=\begin{bmatrix}1 & \rho_{xz}\\ \rho_{xz} & 1\end{bmatrix}$为 X 和 Z 的协方差矩阵,因此

$$\begin{aligned}I(X;Z) &= h(X)+h(Z)-h(XZ)\\ &=\frac{1}{2}\log_2(2\pi e)+\frac{1}{2}\log_2(2\pi e)-\frac{1}{2}\log_2\left[(2\pi e)^2|R|\right]\\ &=-\frac{1}{2}\log_2(1-\rho_{xz}^2)\end{aligned}$$

因为给定 Y 后 X 和 Z 相互独立,所以

$$\begin{aligned}\rho_{xz} &= E[XZ]=E[E(XZ|Y)]=E[E(X|Y)\cdot E(Z|Y)]\\ &=E[\rho_1 Y\cdot\rho_2 Y]\\ &=\rho_1\rho_2\end{aligned}$$

因此

$$I(X;Z)=-\frac{1}{2}\log_2(1-\rho_1^2\rho_2^2)$$

第 5 章

无失真信源编码

5.1　设 S 为一离散无记忆信源，其符号集合为$\{0,1\}$，概率分布为 $p(0)=0.995$，$p(1)=0.005$。令信源符号序列的长度为 $n=100$，假定对所有只包含 3 个以下符号"1"的序列编制长度为 k 的非奇异二进制码。求：

(1) 信源的熵 $H(S)$及其冗余度；

(2) k 的最小值应该为多少？试比较 k/n 和 $H(S)$；

(3) 信源产生的序列没有码字与其对应的概率。

解：

(1) 根据熵的定义，有

$$\begin{aligned}H(S) &= -p(0)\log_2 p(0) - p(1)\log_2 p(1)\\ &= -0.995\log_2 0.995 - 0.005\log_2 0.005\\ &= 0.0454\text{ 比特/符号}\end{aligned}$$

信源冗余度为

$$\gamma = 1 - \frac{H(S)}{H_0(S)} = 1 - \frac{0.0454}{\log_2 2} = 1 - 0.0454 = 0.9546$$

(2) 只包含 3 个以下符号"1"的长度为 $n = 100$ 的序列的个数为

$$C_{100}^{0} + C_{100}^{1} + C_{100}^{2} = 1 + 100 + \frac{100\,!}{98\,!2\,!} = 1 + 100 + 4\,950 = 5\,051$$

长度为 k 的码符号序列的个数为 2^k 个,要想编制非奇异码,必须满足:

$$2^k \geq 5\,051$$

即

$$k \geq 12.302\,4$$

因此, k 的最小值应该为 13。

$$k/n = 13/100 = 0.13$$

$$H(S) = 0.045\,4 \text{ 比特/样值}$$

比较的结果是 k/n 依然大于 $H(S)$,这说明此信源还存在压缩的余地。

(3) 当信源产生的长度为 100 的序列中包含 2 个以上的符号"1"时,没有相应的码字与此序列对应。

$$\begin{aligned} P_e &= 1 - P_r\{\text{只包含 3 个以下符号"1"的长度为 } n = 100 \text{ 的序列}\} \\ &= 1 - p(0)^{100} C_{100}^{0} - p(0)^{99} p(1) C_{100}^{1} - p(0)^{98} p(2) C_{100}^{2} \\ &= 1 - 0.605\,8 - 0.005\,2 - 0.001\,4 \\ &= 0.387\,6 \end{aligned}$$

5.2 用 Fibonacci 数列的前 8 个非零元素构成相应的概率分布为 $P_{\text{Fib}}^{8} = \left\{\frac{13}{34}, \frac{8}{34}, \frac{5}{34}, \frac{3}{34}, \frac{2}{34}, \frac{1}{34}, \frac{1}{34}, \frac{1}{34}\right\}$,对于这样一个概率分布存在着许多拥有不同码长分布的最优编码方案。

(1) 为 P_{Fib}^{8} 构造一个 Huffman 编码并求其平均码长;

(2) 求出最大码长最小的码的码长分布;

(3) 求出最大码长最大的码的码长分布;

(4) 请问一共有多少种码长分布不同的最优编码?

解:

(1) Huffman 编码的过程如图 5.1 所示。编码结果如表 5.1 所示。

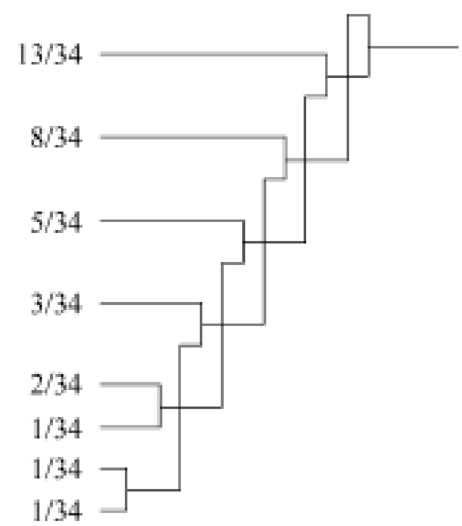


图 5.1 P_{Fib}^{8} 的 Huffman 编码过程

表 5 .1

信源符号	概率分布	码字
s_1	13/ 34	10
s_2	8/ 34	00
s_3	5/ 34	110
s_4	3/ 34	010
s_5	2/ 34	1110
s_6	1/ 34	1111
s_7	1/ 34	0110
s_8	1/ 34	0111

平均码长为

$$\overline{L}=\frac{1}{34}[2\times(13+8)+3\times(5+3)+4\times(2+1+1+1)]=\frac{86}{34}==2\frac{21-3}{34}=2+\frac{F_8-3}{F_9}$$

(2) 深度最小的 Huffman 编码树即对应最大码长最小的码。构造这样的编码树只需在每次缩减信源时尽可能地合并初始的信源符号。按照这个原则得到的 Huffman 编码就是上一步的结果,所以最小的最大码长为 4。

(3) 同理,深度最大的 Huffman 编码树对应最大码长最大的码。在编码时就应该将缩减之后的信源符号排在尽可能靠后的位置,由此得到的编码的码长分布为{1,2,3,4,5,6,7,7},因此最大码的最大码长为 7。

(4) 共有 13 种码长分布不同的最优码。

容易验证 P^2_{Fib}有两种不同的码长分布,P^3_{Fib}有 3 种不同的码长分布,…。通过归纳法可以证明 P^n_{Fib}有 F_n 种不同的码长分配方案。

5 .3 设 X_1, X_2, X_3 为独立的二进制随机变量,并且有 $P_r\{X_1=1\}=\frac{1}{2}$, $P_r\{X_2=1\}=\frac{1}{3}$, $P_r\{X_3=1\}=\frac{1}{4}$,请给出联合随机变量(X_1, X_2, X_3)的 Huffman 编码,并求其平均码长。

解:

根据已知条件可以得到联合随机变量(X_1, X_2, X_3)的概率分布,如表 5 .2 所示。

表 5 .2

$x_1x_2x_3$	$p(x_1x_2x_3)$	$x_1x_2x_3$	$p(x_1x_2x_3)$
000	1/ 4	100	1/ 4
001	1/ 12	101	1/ 12
010	1/ 8	110	1/ 8
011	1/ 24	111	1/ 24

然后根据 Huffman 编码算法得到编码结果如表 5 .3 所示。

表 5 .3

$x_1\ x_2\ x_3$	码　字	$x_1\ x_2\ x_3$	码　字
000	10	100	00
001	1100	101	1101
010	111	110	010
011	0110	111	0111

平均码长为

$$\overline{L} = \frac{1}{24}[2\times(6+6)+3\times(3+3)+4\times(2+2+1+1)] = 2.75$$

5 .4　下面以码字集合的形式给出 5 种不同的编码,第一个码的码符号集合为$\{x, y, z\}$,其他4 个码都是二进制码。

$\{xx, xz, y, zz, xyz\}$

$\{000, 10, 00, 11\}$

$\{100, 101, 0, 11\}$

$\{01, 100, 011, 00, 111, 1010, 1011, 1101\}$

$\{01, 111, 011, 00, 010, 110\}$

对于上面列出的 5 种编码,分别回答下述问题:

(1) 此码的码长分布是否满足 Kraft-McMillan 不等式?

(2) 此码是否是即时码?如果不是,请给出反例。

(3) 此码是否是唯一可译码?如果不是,请给出反例。

解:

1 . $\{xx, xz, y, zz, xyz\}$

(1) 此码满足 Kraft-McMillan 不等式:

$$\frac{1}{3^2}+\frac{1}{3^2}+\frac{1}{3^1}+\frac{1}{3^2}+\frac{1}{3^3} = \frac{3}{27}+\frac{3}{27}+\frac{9}{27}+\frac{3}{27}+\frac{1}{27} = \frac{19}{27}\leq 1$$

(2) 根据码树图可知,此码是即时码;

(3) 由于此码是即时码,所以它也是唯一可译码。

2 . $\{000, 10, 00, 11\}$

(1) 此码满足 Kraft-McMillan 不等式:

$$\frac{1}{2^3}+\frac{1}{2^2}+\frac{1}{2^2}+\frac{1}{2^2} = \frac{1}{8}+\frac{2}{8}+\frac{2}{8}+\frac{2}{8} = \frac{7}{8}\leq 1$$

(2) **此码不是即时码**,**因为码字** 00 **是码字** 000 **的前缀**;

(3) **此码不是唯一可译码**,**因为码符号序列** 000000 **可以译码为** 00,00,00,**也可以译码为** 000, 000。

3 . {100,101,0,11}

(1) **此码满足** Kraft-McMillan **不等式**:

$$\frac{1}{2^3}+\frac{1}{2^3}+\frac{1}{2^1}+\frac{1}{2^2}=\frac{1}{8}+\frac{1}{8}+\frac{4}{8}+\frac{2}{8}=\frac{8}{8}\leq 1$$

(2) **根据码树图可知**,**此码是即时码**;

(3) **由于此码是即时码**,**所以它也是唯一可译码**。

4 . {01,100,011,00,111,1010,1011,1101}

(1) **此码不满足** Kraft-McMillan **不等式**:

$$\frac{1}{2^2}+\frac{1}{2^3}+\frac{1}{2^3}+\frac{1}{2^2}+\frac{1}{2^3}+\frac{1}{2^4}+\frac{1}{2^4}+\frac{1}{2^4}=\frac{4}{16}+\frac{2}{16}+\frac{2}{16}+\frac{4}{16}+\frac{2}{16}+\frac{1}{16}+\frac{1}{16}+\frac{1}{16}=\frac{17}{16}>1$$

(2) **因为不满足** Kraft-McMillan **不等式**,**所以此码不是即时码**;

(3) **因为不满足** Kraft-McMillan **不等式**,**所以此码不是唯一可译码**。

5 . {01,111,011,00,010,110}

(1) **此码满足** Kraft-McMillan **不等式**:

$$\frac{1}{2^2}+\frac{1}{2^3}+\frac{1}{2^3}+\frac{1}{2^2}+\frac{1}{2^3}+\frac{1}{2^3}=\frac{2}{8}+\frac{1}{8}+\frac{1}{8}+\frac{2}{8}+\frac{1}{8}+\frac{1}{8}=\frac{8}{8}\leq 1$$

(2) **此码不是即时码**,**因为码字** 01 **是码字** 011 **的前缀**;

(3) **此码是唯一可译码**。**根据唯一可译码判决准则**,**我们构造以下的尾随后缀集序列**:

$$C=\{01,111,011,00,010,110\}$$

$$F_0=\{1,0\}$$

$$F_1=\{11,10,1,0\}$$

$$F_2=\{11,10,1,0\}$$

5.5 **请问下述编码中哪些不可能是任何概率分布对应的** Huffman **编码**?

(1) {0, 10, 11};

(2) {00, 01, 10, 110};

(3) {01, 10}。

解:

(1) {0, 10, 11}**是概率分布**{1/ 2, 1/ 4, 1/ 4}**的** Huffman **编码**。

(2) {00, 01, 10, 110}**可以被改进为**{00, 01, 10, 11}**而不丧失其即时可译性**,**因此**

原码不是最佳码,因此也就不是一个 Huffman 编码。

另外,此码中最长的码长只有一个码字,这也和 Huffman 编码的特性不符。

(3) {01, 10}可以被改进为{0, 1}而不丧失其即时可译性,因此原码不是最佳码,也就不是一个 Huffman 编码。

5.6　设一信源有 2^k 种不同的符号,其中 k 为任意正整数。对此信源进行二进制 Huffman 编码。假设此信源的分布概率满足 $p_i / p_j < 2, \forall i, j \in \{1, 2, \cdots, 2^k\}$。试证明此 Huffman 编码中所有的码长都为 k。

解:

将此信源按分布概率的降序排好序,假设排序结果为

$$p_1 \geq p_2 \geq \cdots \geq p_i \geq \cdots \geq p_{2^k-1} \geq p_{2^k} \tag{5.1}$$

按照 Huffman 编码过程,下一步应该是合并分布概率最小的两个信源符号以得到缩减信源:$p'_{2^k-1} = p_{2^k-1} + p_{2^k}$。

根据已知条件 $p_i / p_j < 2, \forall i, j \in \{1, 2, \cdots 2^k\}$,有

$$p'_{2^k-1} = p_{2^k-1} + p_{2^k} > \frac{1}{2} p_1 + \frac{1}{2} p_1 = p_1 \tag{5.2}$$

即经过一次缩减得到的符号 s'_{2^k-1} 应该在第二次排序时排在最前面,所以第二次缩减信源时合并的是 p_{2^k-2} 和 p_{2^k-3}。

经过类似的分析可知,前 2^{k-1} 次合并处理的都是原信源的 2^k 个符号,得到的是 2^{k-1} 个新的信源符号,其分布概率为

$$p_i' = p_{2i-1} + p_{2i}$$

容易证明,这 2^{k-1} 个新的分布概率仍然满足条件 $p_i' / p_j' < 2, \forall i, j \in \{1, 2, \cdots, 2^{k-1}\}$,即

$$\frac{p_i'}{p_j'} = \frac{p_{2i-1} + p_{2i}}{p_{2j-1} + p_{2j}} < \frac{2 p_{2i-1}}{2 p_{2j}} < 2 \tag{5.3}$$

所以上面的分析同样适用于这个新信源。

依此类推,整个 Huffman 编码得到的将是一个完全树,即每个码字的码长都一样:$l_i = k, \forall i \in \{1, 2, \cdots, 2^k\}$。

5.7　请找出一个唯一可译码,既不满足前缀条件也不满足后缀条件。

解:

码{0, 010}是唯一可译码,但是它既不满足前缀条件也不满足后缀条件。换句话说,在这个码中,某些码字是另外一些码字的前缀,并且某些码字是另外一些码字的后缀。

5.8　前缀码又称为即时码是因为在译码时产生的译码延时为零。变长码的译码延时定义为在译码时所需查看的下一个码字的码符号个数的最大值。例如,码 $C = \{0, 01\}$ 的译码延时为 1。请给出一个译码延时为 3 的变长码。

解：

码{0，0001}的译码延时为 3。对于接收序列 0000 来说，第一个码字 0 在后 3 个码符号收到之前无法译出。

5.9 等概信源。

(1) 试证明对于一个有 n 个符号的等概信源，其最佳前缀码的各个码长之间最多相差 1 个码符号；

(2) 变长码的冗余度定义为 $L - H$。设一个随机变量 S 有 n 个等概的输出，其中 $2^m \leq n < 2^{m+1}$，对此随机变量进行二进制的变长编码，得到的码的冗余度为 $L - \log_2 n$，试问 n 取何值时冗余度最大？当 $n\to\infty$时码的冗余度的极限值是多少？(提示：0.0861)

解：

(1) 如果等概分布的码字的码长之差大于或等于 2，则此码不是最佳码。这是因为最长的码字有至少一个兄弟节点，否则就可以将此码字的最后一个码符号删除从而改进此码的平均码长。既然最大的码长至少有两个码字与之对应，那么根据图 5.2，可以对此编码进行调整以获得更小的平均码长。

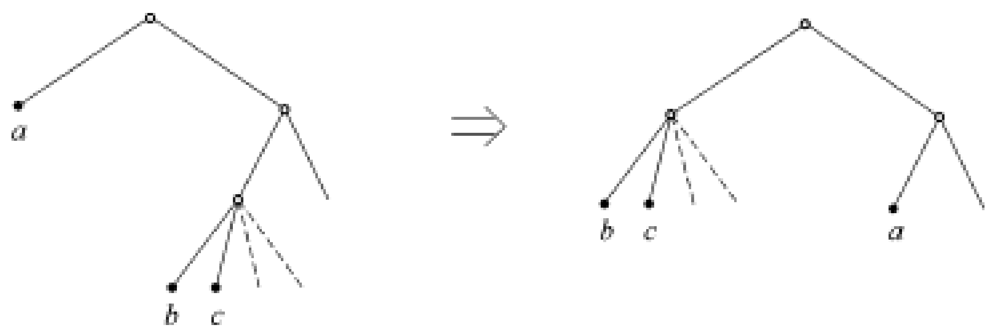


图 5.2

(2) 通过前面的讨论可知，有 n 个等概输出的信源对应的最佳码的最小码长为 $\lfloor \log_2 n \rfloor$，最大码长为 $\lceil \log_2 n \rceil$，令 k 为 n 与下一个最小的 2 的幂次的差，即

$$k = n - 2^{\lfloor \log_2 n \rfloor} = n - 2^m \tag{5.4}$$

其中 $m = \lfloor \log_2 n \rfloor$，因此最佳码有 $2k$ 个码字的码长为 $m+1$，$n-2k$ 个码字的码长为 m，平均码长为

$$\overline{L} = \frac{1}{n}[2k(m+1) + (n-2k)m] = \frac{1}{n}(nm + 2k) = \lfloor \log_2 n \rfloor + \frac{2k}{n} \tag{5.5}$$

因为 $n = 2^m + k$，所以冗余度 $r = L - H$ 为

$$\begin{aligned} r &= L - \log_2 n = \lfloor \log_2 n \rfloor + \frac{2k}{n} - \log_2 n = m + \frac{2k}{n} - \log_2 n(2^m + k) \\ &= m + \frac{2k}{2^m + k} - \frac{\ln(2^m + k)}{\ln 2} \end{aligned} \tag{5.6}$$

k 的取值在 0 和 2^m 之间，要使得 r 达到极大值，则必须有 $\partial r / \partial k = 0$，即

$$\frac{r}{k}=\frac{2(2^m+k)-2k}{(2^m+k)^2}-\frac{1}{(2^m+k)\ln 2}=\frac{2^{m+1}\ln 2-(2^m+k)}{(2^m+k)^2\ln 2}=0 \tag{5.7}$$

使得 r 达到极大值的 k 为 $k^*=(2\ln 2-1)\cdot 2^m=\xi 2^m$,其中 $\xi\approx 0.3863$。

因为冗余度 r 是 k 的凸函数,所以存在一个唯一的最大值,达到最大值的 k 为最接近 $\xi 2^m$ 的两个整数之一。最大冗余度为

$$\begin{aligned}
r^* &\approx m+\frac{2k^*}{2^m+k^*}-\frac{\ln(2^m+k^*)}{\ln 2}\\
&= m+\frac{(2\xi)2^m}{(1+\xi)2^m-\log_2[(1+\xi)2^m]}\\
&=\frac{2\xi}{1+\xi}-\log_2(1+\xi)\\
&=\frac{2(2\ln 2-1)}{1+(2\ln 2-1)}-\log_2[1+(2\ln 2-1)]\\
&=2-\frac{1}{\ln 2}-\log_2(2\ln 2)\\
&=\log_2 4-\log_2 e-\log_2(2\ln 2)\\
&=\log_2\frac{4}{2e\ln 2}\\
&=\log_2\left(\frac{2}{e}\log_2 e\right)=\log_2(1.0615)=0.0861=\sigma
\end{aligned} \tag{5.8}$$

当 $m\to\infty$时,最大冗余度 r^* 趋向于 σ,如图 5-3 所示。

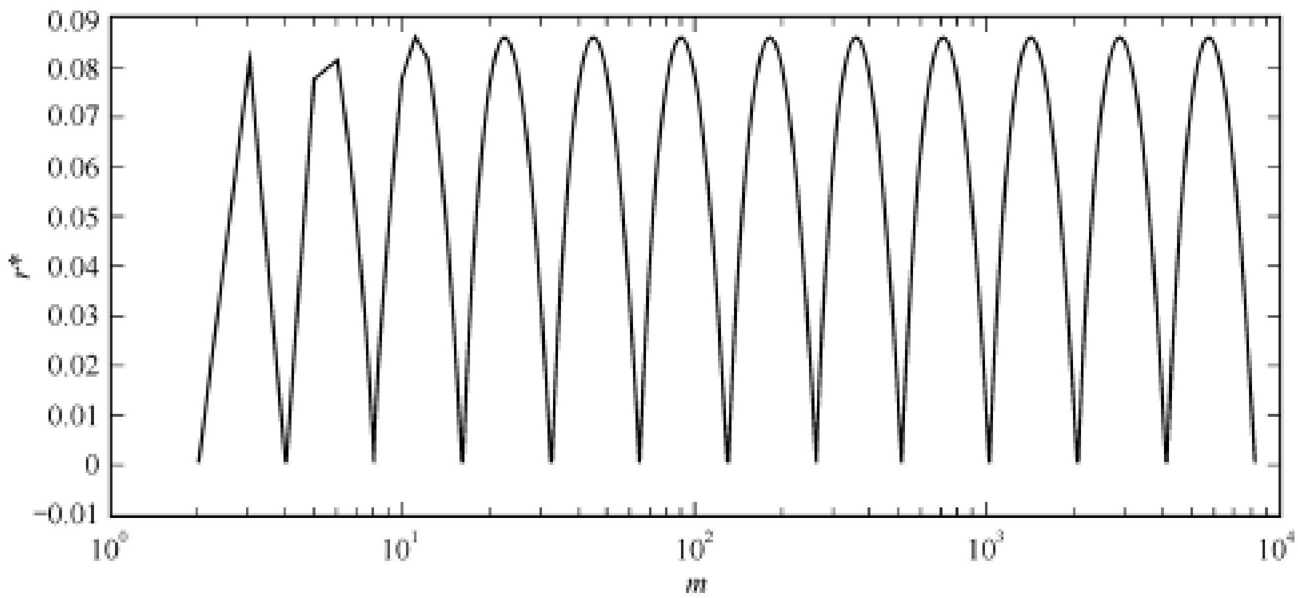


图 5-3

5.10　令符号表 $D=\{1,2\}$的符号代价为 1,2。

(1) 对于一个有 8 个等概输出的信源,找出一个在 D 上的最佳前缀码;

(2) 对于分布$\left\{\frac{1}{n},\cdots,\frac{1}{n}\right\}$,求相应最佳码的平均码字代价的近似值。

解:

(1) 直观地看,对于等概信源来说,其紧致码可以通过使得各码字的代价尽可能相等来获得。我们可以将 Huffman 编码思想加以扩展,排序的依据不再是概率分布,而是代价。由此可知,如果已知码字个数为 $n-1$ 的均匀分布的最小代价编码,则将代价最小的那个码字分裂为两个码字就可以得到码字个数为 n 的最小代价码。表 5.4 列出了符号个数为 2～8 的均匀分布对应的最佳码。

表 5.4

n	总代价	编码 $C^{(n)}$
2	$1\times1+1\times2$	1,2
3	$2\times2+1\times3$	11,12,2
4	$2\times1+2\times3+1\times4$	11,12,21,22
5	$3\times3+2\times4$	111,112,12,21,22
6	$2\times3+3\times4+1\times5$	111,112,12,211,212,22
7	$1\times3+4\times4+2\times5$	111,112,121,122,211,212,22
8	$5\times4+3\times5$	1111,1112,112,121,122,211,212,22

8 符号等概信源的最小代价码的平均代价为 $4\frac{3}{8}$。

(2) 在表 5.4 中,每个码所包含的码字的代价之差最多为 2,这是因为新的码字是通过将代价最小的码字分裂得到的。

通过观察我们发现当 $n=2,3,5,8$ 时每个码中码字的代价之差最多为 1。利用数学归纳法可以很容易地证明这一推理。同时还可以利用数学归纳法证明,当 $n=F_k$ 时,码 $C^{(n)}$ 中有 F_{k-1} 个码字的代价为 $k-1$,另外 F_{k-2} 个码字的代价为 k。大致的思路为:$n'=F_{k+1}$ 时的码是通过分裂 F_{k-1} 个代价为 $k-1$ 的节点得到的,因此会有 F_{k-1} 个新节点的代价为 $k-1+2=k+1$,以及另外 F_{k-1} 个新节点的代价为 $k-1+1=k$,最终 $C^{(n')}$ 拥有 $F_{k-1}+F_{k-2}=F_k$ 个代价为 k 的节点以及 F_{k-1} 个代价为 $k+1$ 的节点。

当 $n=F_k$ 时,平均码代价为

$$\begin{aligned}\overline{L} &= \frac{1}{F_k}[F_{k-1}\cdot(k-1)+F_{k-2}\cdot k] \\ &= \frac{1}{F_k}[(F_{k-1}+F_{k-2})\cdot(k-1)+F_{k-2}] \\ &= \frac{1}{F_k}[F_k\cdot(k-1)+F_{k-2}] \\ &= (k-1)+\frac{F_{k-2}}{F_k} \qquad (5.9)\end{aligned}$$

因为 $F_k \approx c\ ^{k}$,其中 $\ = \frac{1}{2}(1+\sqrt{5}) = 1.618\,03$,为黄金分割比数,所以 $F_{k-2}/\ F_k$ 收敛于 $\ ^{-2} = 0.382\,0$。

当 n 不是 Fibonacci 数时,码 $C^{(n)}$ 的平均代价为 $\log\ n + a$,其中 a 为常量。其求解过程留给读者自行揣摩。

5.11　证明一个码 C 是唯一可译码当且仅当其 k 次扩展

$$C^k(x_1, x_2, \cdots, x_k) = C(x_1)C(x_2)\cdots C(x_k)$$

对任意的 $k \geq 1$ 都是从 X^k 到 D^* 的一一映射。

解:

我们利用反证法。假设对于某些 k 来说,C^k 不是从 X^k 到 D^* 的一一映射,则需证明 C 不是唯一可译码。

根据假设,必然存在两个不同的序列$(x_1, \cdots, x_k)$和$(y_1, \cdots, y_k)$满足:

$$C^k(x_1, \cdots, x_k) = C(x_1)\cdots C(x_k) = C(y_1)\cdots C(y_k) = C^k(y_1, \cdots, y_k) \qquad (5.10)$$

反过来,如果 C 不是唯一可译码,那么根据唯一可译码的定义,必然存在两个不同的序列$(x_1, \cdots, x_i)$和$(y_1, \cdots, y_j)$满足:

$$C(x_1)C(x_2)\cdots C(x_i) = C(y_1)C(y_2)\cdots C(y_j) \qquad (5.11)$$

将这两个序列按照不同的顺序串接起来,得到:

$$C(x_1)\cdots C(x_i)C(y_1)\cdots C(y_j) = C(y_1)\cdots C(y_j)C(x_1)\cdots C(x_i) \qquad (5.12)$$

这表明对于 $k = i + j$ 来说 C^k 不是 X^k 到 D^* 的一一映射。

5.12　香农的信源编码定理告诉我们,一个随机变量 X 的最佳码的平均码长满足 $L < H(X) + 1$,这只是一个最佳的上界,而事实上在很多情况下编码的效率要差得多。例如 X 是一个二进制的随机变量,$p_X(1) = \varepsilon$,则当 $\varepsilon \to 0$ 时 $H(X) = H(\varepsilon) \to 0$,而此时 $L = 1$。请找出一个随机变量,其熵为 2,同时其相应最佳码的平均码长为 3。

解:

如果某个最佳码的平均码长满足 $L \approx H + 1$,则必然有一个输出符号的概率接近 1,其对应的码字的码长也为 1。我们可以假设另外还有 2^m 个输出符号,它们的出现概率的和为 pn 1。为了使得信源的熵最大,这 2^m 个输出符号应该是等概的,因此分布为 $\{1 - p, p2^{-m}, \cdots, p2^{-m}\}$的信源的熵为

$$H = -(1-p)\log_2(1-p) - \sum_{i=1}^{2^m} p2^{-m}\log_2 p2^{-m} = H(p) + p\log_2 2^{-m} = H(p) + pm \qquad (5.13)$$

选择一个很小的 p 使得 $H(p)$n 1,例如 $p = 2^{-10}$。为了让信源的熵等于 2,即

$$pm = 2^{-10}m \approx 2 \qquad m = 2 \times 2^{10} = 2\,048 \qquad 2^m = 2^{2\,048}$$

因此这个信源共有 $1+2^{2\,048}$ **个输出符号,分布为**$\{1-2^{-10},2^{-2\,058},\cdots,2^{-2\,058}\}$。

其相应的最佳码的平均码长为

$$L=(1-2^{-10})\times 1+2^{-10}\times(1+2\,048)=1+2=3$$

$$H=H(2^{-10})+2^{-10}\cdot\log_2(2^{2\,048})=0.011\,2+2^{-10}\times 2^{11}=2.011\,2$$

这个人为拼凑出来的信源主要说明了 $H+1$ **只是最佳码平均码长** L **的上限,对于大多数合理的信源分布来说,** L **与** H **将会非常接近。**

5.13 **对于分布**$\left\{\frac{1}{5\,050},\frac{2}{5\,050},\cdots,\frac{100}{5\,050}\right\}$,**试求其熵和相应的二进制编码的最小平均码长。**

解:

此信源的熵为

$$-\sum_{i=1}^{100}\frac{i}{5\,050}\log_2\frac{i}{5\,050}=\log_2 5\,050-\frac{1}{5\,050}\sum_{i=1}^{100}i\log_2 i=6.372\,2\ \text{比特/符号}$$

利用后面的 Huffman **编码程序,可以得到这个有** 100 **个输出符号的信源的** Huffman **编码。其平均码长为** $\overline{L}=6.395\,6$。**码方差最小的** Huffman **编码共有** 46 **个码字码长为** 6, 27 **个码字码长为** 7, 13 **个码字码长为** 8, 8 **个码字码长为** 9, 3 **个码字码长为** 10, 2 **个码字码长为** 11, 1 **个码字码长为** 12。

5.14 **考虑这样一个信源分布**$\left\{\frac{1}{3},\frac{1}{3},\frac{1}{4},\frac{1}{12}\right\}$。

(1) **为此信源构造一个** Huffman **码;**

(2) **为此信源找出两组不同的最佳码长方案;**

(3) **用实例说明在最佳码中某些码字的码长将会大于相应信源符号对应的** Shannon **编码的码字长度** $l_i=\left\lceil\log_2\frac{1}{p_i}\right\rceil$。

解:

(1) **此信源的** Huffman **编码如表5.5所示。**

表 5.5

符号	概率	码字	码长
s_1	1/3	1	1
s_2	1/3	00	2
s_3	1/4	010	3
s_4	1/12	011	3

(2) **如果将合并后的缩减信源符号排得尽可能靠前,则此时的** Huffman **编码如表5.6所示。**

表 5.6

符号	概率	码字	码长
s_1	1/3	10	2
s_2	1/3	11	2
s_3	1/4	00	2
s_4	1/12	01	2

这里的两种编码的码长分布都满足 Kraft 不等式,并且具有相同的平均码长 2,它们都是最佳码。

(3) 在第一种编码中 s_3 对应的码长为 3,它大于按照 Shannon 公式计算出的码长 $\lceil \log_2 4 \rceil = 2$。也就是说某个特定信源符号对应的 Huffman 码字的码长有可能大于 Shannon 码字的码长,但是从平均码长的角度来看,Huffman 编码的平均码长不会比 Shannon 编码的平均码长大。

5.15　枚举编码。

(1) 求序列 101010001001 在由长度为 12、权重为 5 的二进制序列构成的字典中的索引号;

(2) 假设一二进制信源,其输出中不可能出现两个相连的"1",而序列 101010001001 为此信源的一个输出序列,求此序列对应的字典索引号。(提示:长度为 n 的允许序列有 F_{n+2} 个,其中 $F_0 = 0$, $F_1 = 1$ 且 $F_n = F_{n-1} + F_{n-2}$, $\forall n \geq 2$)

解:

(1) 令 $N(x)$ 表示以 x 作为前缀的、包含 5 个 1 的 12 元组的个数,则序列 101010001001 的索引为

$$\begin{aligned} & N(0) + N(100) + N(10100) + N(101010000) + N(101010001000) \\ &= C_{11}^5 + C_9^4 + C_7^3 + C_3^2 + C_0^1 \\ &= 462 + 126 + 35 + 3 + 0 = 626 \end{aligned}$$

长度为 12、权重为 5 的二进制序列共有 $C_{12}^5 = 792$ 个,因此最大的索引号为 791,它对应着序列 111110000000。

(2) 在这一问中,令 $N(x)$ 表示以 x 作为前缀的、不包含两个连续的 1 的 12 元组的个数。如果 x 的长度为 k 并且以 0 结尾,则将任意长度为 $12 - k$ 且不包含连续两个 1 的序列追加在 x 的后面将得到一个长度为 12 且不包含两个连续的 1 的序列。由提示可知,长度为 n 的不包含两个连续的 1 的序列的个数为 Fibonacci 数 F_{n+2}。前 14 个 Fibonacci 数为:0,1,1,2,3,5,8,13,21,34,55,89,144,233,因此序列 101010001001 的索引为

$$\begin{aligned} & N(0) + N(100) + N(10100) + N(101010000) + N(101010001000) \\ &= F_{13} + F_{11} + F_9 + F_5 + F_2 = 233 + 89 + 34 + 8 + 1 = 365 \end{aligned}$$

共有 $F_{14} = 377$ **个长度为** 12 **且不包含两个连续的** 1 **的序列,最大的索引号为** 376,**它对应着序列** 101010101010。

5.16 **一离散信源的符号表为**$\{a, b, c, d, e\}$,**而** $x = daadcadbea$ **为此信源的观察序列。假设此信源为具体分布未知的独立同分布随机过程,**

(1) **通过观察序列,可以得到信源概率分布函数的一个估计,根据这个估计求信源的熵。**

(2) **根据估计出来的信源概率分布函数构造一个** Huffman **码,计算平均码长并指出对序列** x **编码所需的比特数与平均码长的关系。**

(3) **对序列** x **的前** 4 **个符号进行自适应的** Huffman **编码。初始的估计分布为** $\left\{\frac{1}{5},\frac{1}{5},\frac{1}{5},\frac{1}{5},\frac{1}{5}\right\}$,**此后每收到一个符号,就用当前的码本对其进行编码,然后根据此符号更新估计分布和码本。例如,当收到第一个符号** d **后,更新的概率分布为** $\left\{\frac{1}{6},\frac{1}{6},\frac{1}{6},\frac{2}{6},\frac{1}{6}\right\}$,**更新的码本为**{00,010,011,10,11}。

解:

(1) **此时信源概率分布的估计为**{0.4,0.1,0.1,0.3,0.1},**相应的熵为**

$$H = -0.4\log_2 0.4 - 3\times 0.1\log_2 0.1 - 0.3\log_2 0.3 = 2.0464 \text{ 比特/符号}$$

(2) **由上一问的概率分布,根据** Huffman **编码算法可得到编码结果,如表** 5.7 **所示。**

表 5.7

符号	估计概率分布	码字	码长
a	0.4	1	1
b	0.1	001	3
c	0.1	0000	4
d	0.3	01	2
e	0.1	0001	4

平均码长为

$$\overline{L} = 0.4\times 1 + 0.3\times 2 + 0.1\times 3 + 2\times 0.1\times 4 = 2.1$$

如果用表 5.7 **中的码对序列** $x = daadcadbea$ **进行编码,则总共需要的比特数为**

$$2_d + 1_a + 1_a + 2_d + 4_c + 1_a + 2_d + 3_b + 4_e + 1_a = 21$$

而此序列的自信息为 $10\times H = 20.464$ bit。

(3) **因为对于同一分布可以得到多个不同的** Huffman **编码,所以在给定概率分布的情况下编码器和解码器应该协商决定采用哪种** Huffman **编码。在本问中我们规定在概**

率相同的情况下按符号的字典顺序排列,并且缩减信源符号尽可能的排在后面。按照这种约定我们可以得到前 4 个符号的自适应 Huffman 编码方案,如表 5.8 所示。

表 5.8

当前分布	当前编码	符号	码字
{1/5,1/5,1/5,1/5,1/5}	{01,000,001,10,11}	d	10
{1/6,1/6,1/6,2/6,1/6}	{010,011,000,1,001}	a	010
{2/7,1/7,1/7,2/7,1/7}	{00,11,100,01,101}	a	00
{3/8,1/8,1/8,2/8,1/8}	{1,001,0000,01,0001}	d	01

由此可得序列 x 中前 4 个符号的自适应 Huffman 编码为 100100001。

5.17　一个不合格的骰子,6 个面的出现概率为$\left\{\frac{1}{3},\frac{1}{6},\frac{1}{6},\frac{1}{6},\frac{1}{12},\frac{1}{12}\right\}$。

(1) 求此骰子的熵。($\log_2 3 = 1.58496$)

(2) 令 $A_0^{(n)}$ 是长度为 n 的精确典型列集合:

$$A_0^{(n)} = \{(x_1,\cdots,x_n): p(x_1,\cdots,x_n) = 2^{-nH(X)}\} \tag{5.14}$$

当 $n = 12$ 时求上述骰子对应的 $A_0^{(n)}$ 中元素的个数。

(3) 求此骰子对应的输出序列 x 的累积分布函数的取值范围$[F(x), F(x)+p(x))$。其中, x 的前 4 个输出符号为$(x_1, x_2, x_3, x_4) = (1,2,3,4)$。(注:此问中 $F(x) = P(X < x)$)

解:

(1) 此骰子的熵为

$$-\frac{1}{3}\log_2\frac{1}{3} - 3\times\frac{1}{6}\log_2\frac{1}{6} - 2\times\frac{1}{12}\log_2\frac{1}{12} = \log_2 3 + \frac{5}{6} = 2.4183 \text{ 比特/符号}$$

(2) 精确典型列的概率为

$$2^{-nH(X)} = 2^{-12(\log_2 3+\frac{5}{6})} = 3^{-12}\times 2^{-10}$$

记长度为 12 的序列为 $x = (x_1,\cdots, x_{12})$,令 n_i 表示序列 x 中符号 i 出现的次数。序列 x 的概率为

$$\left(\frac{1}{3}\right)^{n_1}\left(\frac{1}{6}\right)^{n_2+n_3+n_4}\left(\frac{1}{12}\right)^{n_5+n_6} = 3^{-(n_1+n_2+n_3+n_4+n_5+n_6)}\cdot 2^{-(n_2+n_3+n_4+2n_5+2n_6)}$$

令 $r = n_1$, $s = n_2+n_3+n_4$ 以及 $t = n_5+n_6$ 分别为概率为$\frac{1}{3},\frac{1}{6},\frac{1}{12}$的符号出现的次数。

对于一个精确典型列来说,有 $r+s+t = 12$ 以及 $s+2t = 10$,由这两个等式再加上约束条件(r, s, t 都为非负整数)即可得到 5 组解。对于任意一组给定的 r, s, t,12 个输出的排列个数如表 5.9 所示。

$$C_{12}^{t} C_{12-t}^{r} = C_{12}^{t} C_{12-t}^{t+2}$$

表 5.9

t	s	r	$C_{12}^{t} C_{12-t}^{t+2} 2^{t} 3^{10-2t}$
0	2	10	1×6×59 049 = 3 897 234
1	3	8	12×165×13 122 = 25 981 560
2	4	6	66×210×5 832 = 40 415 760
3	5	4	220×126×648 = 17 962 560
4	6	2	495×28×144 = 1 995 840
5	7	0	792×1×32 = 25 344

对于概率为$\frac{1}{12}$的输出,有两个不同的骰子点数,对于概率为$\frac{1}{6}$的输出,有 3 个不同的骰子点数。因此,长度为 12 的精确典型列的个数为

$$\left| A_0^{12} \right| = \sum_{i=1}^{5} C_{12}^{t} C_{12-t}^{t+2} 2^{t} 3^{10-2t}$$

表中最右边一列的和即为结果。长度为 12 的精确典型列的个数为 90 278 298。这个值比 AEP 定理的上界 $2^{nH} = 3^{12} 2^{10} = 544\ 195\ 584$ 要小。

(3) 根据累积概率分布函数的定义,有

$$F(x_1, \cdots, x_n) = \sum_{i=1}^{n} \sum_{u < x_i} p(u \mid x_1, \cdots, x_{i-1}) p(x_1, \cdots, x_{i-1})$$

$$p(x_1, \cdots, x_n) = \prod_{i=1}^{n} p(x_i \mid x_1, \cdots, x_{i-1})$$

由此可知:

$$\begin{aligned}
F(x) &= F(1,2,3,4) = P_r\{(X_1, X_2, X_3, X_4) < (1,2,3,4)\} \\
&= p(1,1) + p(1,2,1) + p(1,2,2) + p(1,2,3,1) + p(1,2,3,2) + p(1,2,3,3) \\
&= \frac{1}{3}\times\frac{1}{3} + \frac{1}{3}\times\frac{1}{6}\times\frac{1}{3} + \frac{1}{3}\times\frac{1}{6}\times\frac{1}{6} + \frac{1}{3}\times\frac{1}{6}\times\frac{1}{6}\times\frac{1}{3} + \frac{1}{3}\times\frac{1}{6}\times\frac{1}{6}\times\frac{1}{6} + \\
&\quad \frac{1}{3}\times\frac{1}{6}\times\frac{1}{6}\times\frac{1}{6} \\
&= \frac{1}{9} + \frac{1}{54} + \frac{1}{108} + \frac{1}{324} + \frac{1}{648} + \frac{1}{648} = \frac{94}{648} = \frac{47}{324}
\end{aligned}$$

$$p(x) = p(1,2,3,4) = \frac{1}{3}\times\frac{1}{6}\times\frac{1}{6}\times\frac{1}{6} = \frac{1}{648}$$

对于序列 x 来说,依此得到的结果如表 5.10 所示。

表 5.10

x_i	$F(x_i)$	$p(x_i)$	$F(x_1,\cdots,x_i)$	$p(x_1,\cdots,x_i)$
-	-	-	0	1
1	0	1/3	0	1/3
2	1/3	1/6	1/9	1/18
3	1/2	1/6	5/36	1/108
4	4/6	1/6	47/324	1/648

结论:以$(1,2,3,4,\cdots)$开头的任意序列的累积概率函数落在区间$\left[\frac{47}{324},\frac{95}{648}\right]$中。

5.18　Kraft 不等式。

(1) 当 $r=2$ 时,试问无限长的即时码 $l_1=1$; $l_2=2,\cdots$; $l_k=k,\cdots$是否满足 Kraft 不等式?

(2) 将上问推广到任意 r 的情况。

解:

(1) 此时 Kraft 不等式为

$$\sum_{i=1}^{\infty} 2^{-i} = 1 \tag{5.15}$$

满足小于等于 1 的条件。

(2) 当 r 为任意大于 2 的整数时,有

$$\sum_{i=1}^{\infty} r^{-i} = \frac{1}{r-1} \leq 1, \quad r > 2 \tag{5.16}$$

因此任意大于 2 的 r 都使得 Kraft 不等式成立。

5.19　试证明:在即时码 $W_1=0$, $W_2=10$, $W_3=11$ 中一个码符号的差错有可能传播到无限远处。(提示:考虑 $W_1(W_3)^n W_2$)

证明:

考虑码符号序列 $W_1(W_3)^n W_2$: $0\underbrace{1111\cdots11}_{2n\text{个}}10$ 当第一个比特发生翻转时,码符号序列将成为$\underbrace{1111\cdots11}_{2(n+1)\text{个}}0$,译码结果为$(W_3)^{n+1}W_1$,即一个码符号的差错被传播到了任意长度的码字序列之后。

5.20　(发酸的酒) 某人得到了 5 瓶酒,已知其中有且仅有一瓶酒坏了(尝起来发酸)。仅凭肉眼观察,发现坏酒的概率分布 p_i 为$(p_1, p_2, p_3, p_4, p_5)=\left(\frac{1}{3},\frac{1}{4},\frac{1}{6},\frac{1}{6},\frac{1}{12}\right)$。通过品尝则可以正确地找出坏酒。

假设此人一瓶一瓶地依次品尝,并且选择一种品尝顺序使得确定出坏酒所必须的品尝次数的期望值最小。请问:

(1) 所需的品尝次数的期望值是多少?

(2) 首先品尝的应该是哪一瓶?

然后此人改变了策略,每次不再品尝单独的一瓶酒,而是将数瓶酒混合起来一起品尝,直到找到坏酒为止。

(3) 在这种方式下所需的品尝次数的期望值是多少?

(4) 首先品尝的应该是哪几瓶酒的混合?答案是唯一的吗?如果是唯一的,请解释为什么;如果不是唯一的,请给出另外一种方案。(提示:品尝数瓶酒的混合等效于品尝其他几瓶酒的混合,因此这不算两种方案。另外,有两瓶酒坏掉的概率相等,都是 1/6,所以交换这两瓶酒不算新的方案)

解:

(1) 我们将品尝次数看作码字的码长,由于只能依次品尝每一瓶酒,所以对于这个题目而言码长只能是 1,2,3,4,4 这样的分布,所需的平均品尝次数就是平均码长,要想使得平均码长最小,则应该将较短的码长安排给概率较大的符号,因此所需的最小平均品尝次数为

$$\sum_{i=1}^{5} p_i l_i = \frac{1}{3}\times 1 + \frac{1}{4}\times 2 + \frac{1}{6}\times 3 + \frac{1}{6}\times 4 + \frac{1}{12}\times 4 = \frac{7}{3} = 2.33$$

(2) 应该首先品尝概率为 1/3 的那瓶酒。

(3) 这个问题可以转化为求取此概率分布的 Huffman 编码,如图 5.4 所示。

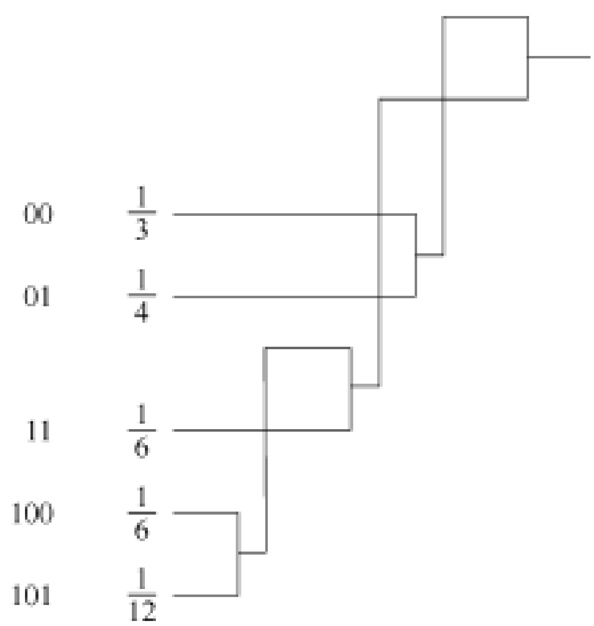


图 5.4

品尝是首先将第一比特为 0 的酒混合起来品尝,分出好坏后再按照第二比特来混合,直到找到坏酒,因此平均品尝次数为

$$\sum_{i=1}^{5} p_i l_i = \frac{1}{3}\times 2 + \frac{1}{4}\times 2 + \frac{1}{6}\times 2 + \frac{1}{6}\times 3 + \frac{1}{12}\times 3 = \frac{27}{12} = 2.25$$

(4) 根据上述编码,首先应该品尝第一和第二瓶酒的混合,或者品尝第三、四、五瓶酒的混合。由于 Huffman 编码不是唯一的,所以还存在另外的方案,其对应的编码如下:

概率	1/3	1/4	1/6	1/6	1/12
码字	00	10	11	010	011

在这种方案的指导下,首先应该品尝第一、四、五瓶酒的混合。

5.21　考虑一个 3 符号的离散信源,其概率分布为: $P_r\{X=a\}=1/2$, $P_r\{X=b\}=1/4$, $P_r\{X=c\}=1/4$。请给出其长度为 n 的精确典型列的数目,不允许使用递归表达式。(提示:最终结果依赖于 n 的奇偶性)

解:

一个长度为 n 的精确典型列 $x=(x_1, x_2, \cdots, x_n)$ 意味着:

$$\left|\frac{1}{n}\log_2\frac{1}{p(x)} - H\right| = 0 \tag{5.17}$$

等价的有

$$\log_2\frac{1}{p(x)} = \sum_{i=1}^{n}\log_2\frac{1}{p(x_i)} = nH \tag{5.18}$$

在本题中 $H=3/2$,并且:

$$\log_2\frac{1}{p(x_i)} = \begin{cases} 1 & x_i = a \\ 2 & x_i = b \text{ 或 } x_i = c \end{cases} \tag{5.19}$$

令 $n(a)$, $n(b)$, $n(c)$分别表示长度为 n 的精确典型列 x 中符号 a, b, c 的个数,则根据式(5.18)和(5.19)有

$$\begin{cases} n(a)+n(b)+n(c)=n \\ n(a)+2n(b)+2n(c)=\dfrac{3}{2}n \end{cases} \tag{5.20}$$

解之可得

$$\begin{cases} n(a)=\dfrac{1}{2}n \\ n(b)+n(c)=\dfrac{1}{2}n \end{cases} \tag{5.21}$$

因此,当 n 为奇数时,不存在长度为 n 的精确典型列。当 n 为偶数时,精确典型列的由 $n/2$ 个符号 a 和一共 $n/2$ 个符号 b 和 c 构成,其个数为在 n 个位置上安排 $n/2$ 个符号 a,并且在剩下的 $n/2$ 个位置上安排符号 b 或 c:

$$C_n^{n/2}\cdot 2^{n/2} \tag{5.22}$$

5.22 **一个离散无记忆信源,其样本空间为**$\{W, B\}$,**符号** W **出现的概率为** 0.99, B **出现的概率为** 0.01。

(1) **对此信源的二次扩展,求出信源符号序列的概率分布,找出相应的** Huffman **编码并求平均码长。**

(2) **对此信源的三次扩展重复上一问。**

(3) **计算信源的单符号熵并与上两问中的单符号平均码长进行比较。**

(4) **要想使得单符号平均码长** $\overline{L}_n/n$ **只比单符号信源熵大出** 10%,**请问扩展次数** n **应该是多少?**

解:

(1) **二次扩展信源的符号序列、概率分布及码字如表** 5.11 **所示。**

表 5.11

符号序列	概　率	码　字
WW	$0.99\times0.99=0.9801$	0
WB	$0.99\times0.01=0.0099$	11
BW	$0.01\times0.99=0.0099$	100
BB	$0.01\times0.01=0.0001$	101

平均码长为

$$\overline{L}_2=0.9801\times1+0.0099\times2+0.0099\times3+0.0001\times3=1.0299$$

(2) **三次扩展信源的符号序列、概率分布及码字如表** 5.12 **所示。**

表 5.12

符号序列	概　率	码　字
WWW	$0.99\times0.99\times0.99=0.970299$	0
WWB	$0.99\times0.99\times0.01=0.009801$	100
WBW	$0.99\times0.01\times0.99=0.009801$	101
BWW	$0.01\times0.99\times0.99=0.009801$	110
WBB	$0.99\times0.01\times0.01=0.000099$	11100
BWB	$0.01\times0.99\times0.01=0.000099$	11101
BBW	$0.01\times0.01\times0.99=0.000099$	11110
BBB	$0.01\times0.01\times0.01=0.000001$	11111

平均码长为

$$\overline{L}_3=0.970299\times1+0.009801\times(3+3+3)+0.000099\times(5+5+5)+0.000001\times5=1.06$$

(3) 信源的单符号熵为

$$H(S)=0.99\log_2\frac{1}{0.99}+0.01\log_2\frac{1}{0.01}=0.080\,179\ \text{比特/符号}$$

二次扩展信源和三次扩展信源的单符号平均码长分别为

$$\frac{\overline{L_2}}{2}=0.514\,95,\quad \frac{\overline{L_3}}{3}=0.353\,33$$

都远大于信源的单符号熵。

(4) 单符号信源熵的 110%是 $1.1\times0.080\,179=0.088\,19$,而 $\overline{L_n}$ 至少为 1,即单符号平均码长至少为 $1/n$,因此,根据题意有

$$\frac{1}{n}\leq 0.088\,19\qquad n\geq 11.339\qquad n=12$$

5.23　假设某个信源的模型为一个独立同分布的随机变量序列 $X_1,X_2,\cdots$,序列中的每个随机变量 X_i 都是有 q 个元素$\{0,1,\cdots,q-1\}$的样本空间上的均匀分布。

(1) 令 n 和 k 为正整数,请问 2^k 个最有可能的长度为 n 的信源符号序列的概率和 $Q(n,k)$是多少?(注:如果 2^k 超出了长度为 n 的信源符号序列的总个数,则 $Q(n,k)$被定义为 1,即所有长度为 n 的信源符号序列的概率和)

(2) 利用上一问的结果,对所有 $R>0$ 计算极限(不许利用 Shannon 编码定理):

$$\lim_{n\to\infty}Q(n,\lceil Rn\rceil)$$

其中$\lceil x\rceil$表示不小于 x 的最小整数。

(3) 现在令 $q=3$,请找出最小的 n 和 k,使得 $Q(n,k)=1$ 以及 $k/n\leq1.60$。若要求 $k/n\leq1.50$,结果又如何?

解:

(1) 既然每个信源符号序列中的每个符号都是从同一个均匀分布中抽取的,所以每个长度为 n 的信源符号序列的概率都相同。因此,由任意 2^k 个信源符号序列构成的子集都可认为是“最有可能的 2^k 个长度为 n 的信源符号序列”,其概率和为

$$Q(n,k)=\frac{2^k}{q^n}$$

(2) 由上一问可得

$$\lim_{n\to\infty}Q(n,\lceil Rn\rceil)=\lim_{n\to\infty}\frac{2^{\lceil Rn\rceil}}{q^n}=\lim_{n\to\infty}\frac{2^{Rn}}{q^n}=\lim_{n\to\infty}\left(\frac{2^R}{q}\right)^n=\begin{cases}0 & q>2^R\\ 1 & q=2^R\\ +\infty & q<2^R\end{cases}$$

或者,如果限定 Q 的取值不能大于 1,则有

$$\lim_{n\to\infty}Q(n,\lceil Rn\rceil)=\begin{cases}0 & q>2^R\\ 1 & q\leq2^R\end{cases}$$

(3) $Q(n,k)=1$ 意味着 $2^k \geq q^n$,将 $q=3$ 代入得

$$\frac{2^k}{3^n} \geq 1$$

$$\frac{3^{k \cdot \log_3 2}}{3^n} \geq 1$$

$$\frac{k}{n} \geq \log_2 3 \approx 1.585$$

再加上 n, k 都是整数,所以有如下列表:

$$(k, n, k/n) = (2, 1, 2.0)$$
$$(k, n, k/n) = (4, 2, 2.0)$$
$$(k, n, k/n) = (5, 3, 1.666\,67)$$
$$(k, n, k/n) = (7, 4, 1.75)$$
$$(k, n, k/n) = (8, 5, 1.6)$$
$$(k, n, k/n) = (16, 10, 1.6)$$
$$(k, n, k/n) = (24, 15, 1.6)$$
$$(k, n, k/n) = (27, 17, 1.588\,24)$$
$$(k, n, k/n) = (32, 20, 1.6)$$
$$(k, n, k/n) = (35, 22, 1.590\,91)$$
$$\cdots$$

从列表中可知,使得 $k/n \leq 1.6$ 的最小的一对(k, n)是(8,5),而要使 $k/n \leq 1.59$,则(k, n)必须为(27,17)。

5.24 令 S 为一离散无记忆信源,其样本空间为 A,熵为 H。令 $R < H$,设 L_n 为 A^n 的一个子集,并且满足条件:

$$|L_n| \leq 2^{Rn}, \quad n = 1, 2, \cdots$$

求证:

$$\lim_{n \to \infty} P_r\{L_n\} = 0$$

证明:

已知有结论:如果 x 是一个 ε-典型列,则

$$2^{-n(H+\varepsilon)} \leq p(x) \leq 2^{-n(H-\varepsilon)}$$

下面将 $P_r\{L_n\}$改写为 L_n 与 ε-典型列 $T_n(\varepsilon)$的交集的形式:

$$L_n = \{L_n \cap T_n(\boldsymbol{\varepsilon})\} \cup \{L_n \cap \overline{T_n}(\boldsymbol{\varepsilon})\}$$

其中 $\overline{T_n}(\varepsilon)$为 ε-典型列的补集。

所以有

$$P_r\{L_n\} = P_r\{L_n \cap T_n(\boldsymbol{\varepsilon})\} + P_r\{L_n \cap \overline{T_n}(\boldsymbol{\varepsilon})\}$$

当 $n\to\infty$**时,因为** $P_r\{\overline{T_n}(\varepsilon)\}\to 0$,**所以上式的第二项趋向于零,因此可得**

$$\lim_{n\to\infty} P_r\{L_n\} = \lim_{n\to\infty} P_r\{L_n \cap T_n(\varepsilon)\} = \lim_{n\to\infty}\sum_{x\in L_n\cap T_n(\varepsilon)} p(x) \le \lim_{n\to\infty}|L_n|2^{-n(H-\varepsilon)}$$

$$\le \lim_{n\to\infty} 2^{Rn}2^{-n(H-\varepsilon)}$$

$$= \lim_{n\to\infty} 2^{-n(H-R-\varepsilon)} \xrightarrow{R<H-\varepsilon} 0$$

5.25　**已知离散无记忆信源如下,试求**:

$$\begin{bmatrix} S \\ P \end{bmatrix} = \begin{bmatrix} s_1 & s_2 & s_3 & s_4 & s_5 & s_6 & s_7 \\ 0.20 & 0.19 & 0.18 & 0.17 & 0.15 & 0.10 & 0.01 \end{bmatrix}$$

(1) **信源符号熵** $H(S)$;

(2) **相应的二元** Huffman **编码及其编码效率**;

(3) **相应的三元** Huffman **编码及其编码效率**;

(4) **若要求** $p_E \le 10^{-3}$,**采用定长二元码要求达到第**(2)**问中的编码效率,至少需要多少信源符号一起编码才能实现**?

解:

(1) **信源符号熵为**

$$H(S) = -(0.2\log_2 0.2 + 0.19\log_2 0.19 + 0.18\log_2 0.18 + 0.17\log_2 0.17 +$$
$$0.15\log_2 0.15 + 0.10\log_2 0.10 + 0.01\log_2 0.01)$$
$$= 2.61 \text{ 比特/符号}$$

(2) **二元** Huffman **编码如表** 5.13 **所示**。

表 5.13

信源符号	s_1	s_2	s_3	s_4	s_5	s_6	s_7
概　率	0.20	0.19	0.18	0.17	0.15	0.10	0.01
码　字	10	11	000	001	010	0110	0111

平均码长为 $\overline{L} = 2.72$,**编码效率为** $\eta = \dfrac{H(S)}{\overline{L}} = 96\%$。

(3) **三元** Huffman **编码如表** 5.14 **所示**。

表 5.14

信源符号	s_1	s_2	s_3	s_4	s_5	s_6	s_7
概率	0.20	0.19	0.18	0.17	0.15	0.10	0.01
码字	2	00	01	02	10	11	12

平均码长为 $\overline{L}=1.8$,**编码效率为** $\eta=\frac{H(S)}{\overline{L}}\log_3 2=92\%$。

(4) **信源自信息量的方差为**

$$\sigma^2=E[I(s_i)^2]-[H(S)]^2=0.241$$

按编码效率为 96%,**即**

$$\frac{H(S)}{H(S)+\varepsilon}=\eta=0.96 \qquad \varepsilon\approx 0.109$$

因此有

$$p_E\leq\frac{\sigma^2}{L\varepsilon^2} \qquad L\geq\frac{\sigma^2}{\varepsilon^2 p_E}\approx 2.03\times10^4$$

即需要约 2.03×10^4 **个信源符号一起编码。**

第 6 章

有噪信道编码

6.1　**考虑将 9 次重复码 R_9 看成是先后进行两次 R_3 编码，即先进行一次 R_3 编码，然后对每个码字的 3 个码符号分别再进行一次 R_3 编码，并将其称之为"R_3^2"。由此可得另外一种 R_9 的译码方法：先对每组 3 个符号进行 R_3 译码，然后再对 3 个译码结果进行一次 R_3 译码。**

计算这种译码方法的错误概率，并与 R_9 的择多译码对应的错误概率进行比较。

解：

假设信道的错误概率为 p，则 R_3^2 的错误概率为

$$p_E(R_3^2) \approx 3[p_E(R_3)]^2 = 3(3p^2)^2 + \cdots = 27p^4 + \cdots$$

而 R_9 的择多译码的错误概率是受 5 比特翻转控制的：

$$p_E(R_9) \approx C_9^5 p^5 (1-p)^4 \approx 126p^5 + \cdots$$

由此可知 R_3^2 译码过程是次优的，因为 4 比特的翻转就可以导致译码错误。但是它也有节省计算资源的优点，因为在 R_3^2 译码过程中只需要缓存 3 个码符号，并且只需要检查 3 个码符号就可以得到译码结果，而在 R_9 的情况下需要缓存 9 个码符号，并且要检查全部的 9 个码符号才能得到译码结果。

6.2　**试证明线性分组码的最小 Hamming 距离等于其非零码字的最小 Hamming 重量。**

证明：

根据线性分组码的封闭性可知，任意两个码字的和仍为码中的码字。根据码字之间

的距离的定义,两个码字和的非零符号的个数即为它们之间的距离,而两个码字和的非零符号的个数又是新码字的重量。所以,线性分组码的最小距离必为它的非零码字的最小重量。

6.3 **假设一个**(n,k)**二进制线性分组码的生成矩阵** G **不包含全零行,例如**(7,4) Hamming **码的生成矩阵为**

$$G=\begin{bmatrix} I_{4\times 4} \\ P_{3\times 4} \end{bmatrix}=\begin{bmatrix} 1&0&0&0\\ 0&1&0&0\\ 0&0&1&0\\ 0&0&0&1\\ 0&1&1&1\\ 1&0&1&1\\ 1&1&0&1 \end{bmatrix}$$

令 C **为以所有** 2^k **个码字为列矢量构成的** $n\times 2^k$ **矩阵**:

$$C = G\times\begin{bmatrix} 0&0&0&0&0&0&0&0&1&1&1&1&1&1&1&1\\ 0&0&0&0&1&1&1&1&0&0&0&0&1&1&1&1\\ 0&0&1&1&0&0&1&1&0&0&1&1&0&0&1&1\\ 0&1&0&1&0&1&0&1&0&1&0&1&0&1&0&1 \end{bmatrix}$$

$$=\begin{bmatrix} 0&0&0&0&0&0&0&0&1&1&1&1&1&1&1&1\\ 0&0&0&0&1&1&1&1&0&0&0&0&1&1&1&1\\ 0&0&1&1&0&0&1&1&0&0&1&1&0&0&1&1\\ 0&1&0&1&0&1&0&1&0&1&0&1&0&1&0&1\\ 0&1&1&0&1&0&0&1&0&1&1&0&1&0&0&1\\ 0&1&1&0&0&1&1&0&1&0&0&1&1&0&0&1\\ 0&1&0&1&1&0&1&0&1&0&1&0&0&1&0&1 \end{bmatrix}$$

试证明对于任意的(n,k)**二进制线性分组码,其** C **矩阵中的每一行正好有** 2^{k-1} **个** 0 **和** 2^{k-1} **个** 1。(**提示**: C **的第**(i,j)**个元素与** G **的第** i **行的关系是什么**?**当** j **从** 0 **变到** 2^{k-1}**时上述关系会产生什么结果**?)

证明:

令 c_{ij}**表示矩阵** C **的第** i **行、第** j **列的元素**, g_{ij}**表示矩阵** G **的第** i **行、第** j **列的元素**, j_l **表示整数** j **的二进制表示的从左边数起第** l **个比特,则有**

$$c_{ij} = \sum_{l=1}^{k} g_{il} j_l \qquad (6.1)$$

即 c_{ij}**是** 0 **还是** 1 **将由矩阵** G **的第** i **行为** 1 **的元素对应位置上的** j_l **的取值是否为** 1 **来决定。当** j **从** 0 **变到** 2^k-1 **时,因为** j **对应的二进制形式取遍了长度为** k **的所有二进制序列,所以** j **对应的二进制形式中从左到右每个位置上的** 0 **和** 1 **出现的次数都相等。**

6.4　**求证:一个(n,k)二进制线性分组码的最小距离 $d_{\min}$满足下述不等式:**

$$d_{\min} \leq \frac{n \cdot 2^{k-1}}{2^k - 1}$$

证明:

我们用反证法证明此题的结论:假设 $d_{\min} > \dfrac{n \cdot 2^{k-1}}{2^k - 1}$,**又因为总共有** $2^k - 1$ **个非零码字,所以所有码字中码符号 1 的总个数** N_1 **将满足:**

$$N_1 \geq (2^k - 1) \times d_{\min} > n \cdot 2^{k-1}$$

又由 6.3 题可知,一个(n,k)二进制线性分组码的所有码字构成的 $n \times 2^k$ **矩阵中共有** $n \cdot 2^{k-1}$**个码符号 1,与上式矛盾,所以得证。**

6.5　**假设想要设计一个(n,k)码,且要求其最小码距至少为 $2t+1$,其中 k 为信息比特的个数,$(n-k)$为附加的冗余比特数,我们当然希望$(n-k)$越小越好。**

(1) **试证明附加的冗余比特数$(n-k)$满足下述不等式:**

$$n - k \geq \log_2(1 + C_n^1 + C_n^2 + \cdots + C_n^t)$$

(2) **试证明 Hamming 码满足上述不等式。**

证明:

(1) (n,k)**码有** 2^k **个码字,如果最小码距为** $2t+1$,**则以每个码字为中心,半径为** t **的球互不相交。每个这样的球内的码字个数为**

$$N_i = 1 + C_n^1 + C_n^2 + \cdots + C_n^t \tag{6.2}$$

而这样的 2^k **个球内的总码字数将小于等于所有可能的码字数:**

$$2^k \cdot (1 + C_n^1 + C_n^2 + \cdots + C_n^t) \leq 2^n \tag{6.3}$$

即

$$n - k \geq \log_2(1 + C_n^1 + C_n^2 + \cdots + C_n^t) \tag{6.4}$$

(2) **对于 Hamming 码而言,** $n = 2^m - 1$, $k = 2^m - m - 1$, $d_{\min} = 3$, $t = 1$, **代入式(6.4)可得**

$$n - k = m = \log_2(1 + C_n^1) = \log_2(1 + n) = \log_2 2^m \tag{6.5}$$

6.6　**证明:对于任何线性分组码,码字的重量或全部为偶数,或奇数重量的码字数为偶数。**

证明:

线性分组码全体码字构成一个线性空间,所有码字之和为零矢量,是一个偶重码。另,偶重码与偶重码之和为偶重码,偶重码与奇重码之和为奇重码,奇重码与奇重码之和为偶重码。因此原命题得证。

6.7 求:(1) $d_{\min}=3$ 且至多只有 3 位校验码元的二进制码的码长;

(2) $d_{\min}=5$ 且至多只有 8 位校验码元的二进制码的码长。

解:

(1) (7,4) Hamming 码正好满足条件,其码长为 7。

(2) 码长 $n=15$ 的 BCH 编码满足条件,其 $d_{\min}=5$ 且只有 8 位校验码元。

6.8 假设信道编码为 Hamming(7,4)码,请将下述接收序列解码:

(1) $r=1101011$

(2) $r=0110110$

(3) $r=0100111$

(4) $r=1111111$

解:

(1) Hamming(7,4)码的校验矩阵为

$$H=\begin{bmatrix}1&1&1&0&1&0&0\\0&1&1&1&0&1&0\\1&0&1&1&0&0&1\end{bmatrix}$$

由此可得其伴随式为

$$s=rH^{\mathrm{T}}=\begin{bmatrix}0&1&1\end{bmatrix}$$

而 Hamming(7,4)码的伴随式与翻转比特之间的关系如表 6.1 所示。

表 6.1

伴随式 s	000	001	010	011	100	101	110	111
翻转的比特	-	r_7	r_6	r_4	r_5	r_1	r_2	r_3

由此可知翻转的比特是第 4 个 r_4,所以正确的接收矢量应该是 1100011,译码结果应该是 1100。

(2) 根据上面的推理同样可以得到 0110110 的译码结果应该是 0100。

(3) 0100111 的译码结果应该是 0100。

(4) 1111111 的译码结果应该是 1111。

6.9 Hamming(7,4)码可以纠正 1 比特的差错。请问是否存在一个(14,8)码能够纠正两个比特的差错?

解:

首先假设构造了一个码长为 14 的线性分组码,并且假设采用伴随式译码。

对于 $N=14$ 个发送比特来说,最多两个比特的错误模式共有

$$C_{14}^{2}+C_{14}^{1}+C_{14}^{0}=91+14+1=106$$

由于采用伴随式译码,所以每个不同的错误模式必须指定一个不同的伴随式,而对于

(14,8)**码来说,伴随式是长度为** $m = n - k = 14 - 8 = 6$ **的矢量,共有** $2^m = 2^6 = 64$ **个。**

因为 $64 < 106$,**所以可以得到结论:不存在能够纠正两个比特错误的**(14,8)**码。**

6.10　**下述问题针对**(15,11) Hamming **码:**

(1) **写出此码的校验矩阵** H;

(2) **写出此码的生成矩阵** G;

(3) **写出此码的一个译码函数** $g:\{0,1\}^{15} \to \{0,1\}^{11}$,**要求它能纠正所有的** 1 **比特错误;(提示:用有效码字** x^{15} **与接收序列** y^{15} **之间的** Hamming **距离来构造** g)

(4) **如果将此码放在一个错误概率为** $p = 0.01$ **的** BSC **上传输,请计算一个** 11 **比特的消息发生译码错误的概率。**

解:

(1) (n, k)Hamming **码的校验矩阵有**$(n-k)$**行和** n **列。构造方法是将由**$(n-k)$**个码符号组成的非零列矢量按照二进制数由小到大的顺序从左到右依次排列。**

$$H = \begin{bmatrix} 0&0&0&0&0&0&0&1&1&1&1&1&1&1&1 \\ 0&0&0&1&1&1&1&0&0&0&0&1&1&1&1 \\ 0&1&1&0&0&1&1&0&0&1&1&0&0&1&1 \\ 1&0&1&0&1&0&1&0&1&0&1&0&1&0&1 \end{bmatrix}$$

$$\xrightarrow{\text{列置换}} \begin{bmatrix} 0&1&1&1&0&1&1&0&0&1&1&1&0&0&0 \\ 1&0&1&1&0&0&1&1&1&0&1&0&1&0&0 \\ 1&1&0&1&1&0&0&1&0&1&1&0&0&1&0 \\ 1&1&1&0&1&1&0&0&1&0&1&0&0&0&1 \end{bmatrix}$$

$$= [\, P^{\mathrm{T}} \quad I_4 \,]$$

(2) **由上一问可知,生成矩阵为**

$$G = [\, I_{11} \quad P \,] = \begin{bmatrix} 1&0&0&0&0&0&0&0&0&0&0&0&1&1&1 \\ 0&1&0&0&0&0&0&0&0&0&0&1&0&1&1 \\ 0&0&1&0&0&0&0&0&0&0&0&1&1&0&1 \\ 0&0&0&1&0&0&0&0&0&0&0&1&1&1&0 \\ 0&0&0&0&1&0&0&0&0&0&0&0&0&1&1 \\ 0&0&0&0&0&1&0&0&0&0&0&1&0&0&1 \\ 0&0&0&0&0&0&1&0&0&0&0&1&1&0&0 \\ 0&0&0&0&0&0&0&1&0&0&0&0&1&1&0 \\ 0&0&0&0&0&0&0&0&1&0&0&0&1&0&1 \\ 0&0&0&0&0&0&0&0&0&1&0&1&0&1&0 \\ 0&0&0&0&0&0&0&0&0&0&1&1&1&1&1 \end{bmatrix}$$

(3) **译码规则为:当接收序列** y **与有效码字** x_i, $i = 1, 2, \cdots, 2^{11}$ **之间的距离** $d(x, y) \leq 1$ **时,将** y **译码为** x_i。

(4) 译码错误概率为

$P_E = 1 - P_r\{$正确译码$\} = 1 - P_r\{11$ 位码符号均未翻转$\} - P_r\{$仅有一位码符号发生翻转$\}$

$= 1 - C_{11}^{0}(1-p)^{11} - C_{11}^{1}p(1-p)^{10} = 0.0052$

6.11 证明:如果两个错误图样 e_1, e_2 的和是一个有效的码字,那么它们具有相同的伴随式。

证明:

(n, k)线性分组码的一致校验矩阵 H 是一个$(n-k)\times n$ 的矩阵。一个码符号序列 $x = (x_1, x_2, \cdots x_n)$是码字当且仅当:

$$Hx^{T} = 0$$

由已知条件可知:

$$H(e_1 + e_2)^{T} = 0$$

所以有

$$He_1^{T} = -He_2^{T} = He_2^{T}$$

即两者的伴随式相同。

6.12 设(7,4)循环码的生成多项式为 $g(x) = x^3 + x + 1$,当接收码字为 0010011 时,试问接收码字是否有错。

解:

接收码字的码多项式为

$$R(x) = x^4 + x + 1$$

$$\frac{R(x)}{g(x)} = \frac{x^4 + x + 1}{x^3 + x + 1} \neq 0 \quad \bmod\ g(x)$$

所以接收码字中存在错误。

6.13 选用一个最短的生成多项式设计一个(6,2)循环码。

(1) 计算该码的生成矩阵(系统形式);

(2) 找出所有可能的码字;

(3) 该码能纠多少差错?

解:

(1) 因为 $x^6 + 1 = (x+1)^2(x^2 + x + 1)^2$,所以可以得到两个生成多项式:

$$g_1(x) = (x+1)^2(x^2 + x + 1)^2 \tag{6.6}$$

$$g_2(x) = (x^2 + x + 1)^2 \tag{6.7}$$

我们选择 $g_2(x)$为生成多项式,其生成矩阵为

$$G = \begin{bmatrix} g_2(x) \\ xg_2(x) \end{bmatrix} = \begin{bmatrix} 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 & 1 \end{bmatrix} \tag{6.8}$$

(2) 所有码字为:(000000) (101010) (010101) (111111)。

(3) **最小码距为** $d_{\min}=3$,**所以能纠 1 个比特的错误。**

6.14　**试证明:当且仅当多项式系数之和等于零(模 2)时,** $x+1$ **才是它的一个因子。**

证明:

对于任何一个实系数多项式 $f(x)$,**考虑它被** $x-a$ **除所得的余数:**

$$f(x)=(x-a)g(x)+r \tag{6.9}$$

将 $x=a$ **代入式**(6.9),**得**

$$f(a)=r \tag{6.10}$$

式(6.10)**表明** $x-a$ **除以** $f(x)$ **的余数为** $f(a)$,**因此** $x+1$ **为** $f(x)$ **的因子当且仅当其余数** $f(-1)=0$。

对于整系数多项式,不妨设 $f(x)=a_0+a_1x+a_2x^2+\cdots+a_nx^n$,**因此有**

$$f(-1)=a_0-a_1+a_2+\cdots+(-1)^na_n=0 \tag{6.11}$$

在式(6.11)**两侧同时加上** $2a_1+2a_3+\cdots$,**有**

$$a_0+a_1+a_2+\cdots+a_n=2a_1+2a_3+\cdots=0 \bmod 2 \tag{6.12}$$

6.15　**如果已知某一次数下多项式的个数以及低次素多项式之积的个数,试推算出该次数的素多项式的个数。**

解:

n **次多项式的总个数为** 2^n,**假设次数为** n **的素多项式的个数为** $A(n)$,**次数为** n **的合多项式的个数为** $B(n)$。**前提条件为已知** 2^n **和** $\sum_{i=1}^{n-1}B(i)$,**需要求取的是** $A(n)$。

显然有

$$A(n)+B(n)=2^n \tag{6.13}$$

将任何一个合多项式分解为两个因子的形式,只能有下述几种情形:两个素因子、一个素因子和一个合因子、两个合因子,所以有

$$B(n)=\sum_{i=1}^{n/2}A(i)A(n-i)+\sum_{i=1}^{n-2}A(i)B(n-i)+\sum_{i=2}^{n/2}B(i)B(n-i) \tag{6.14}$$

式(6.14)**中后两项的求和限比较奇怪,这是因为** $n=1$ **的两个多项式都是素多项式。**

经过简单的分析可知: $A(1)=2$, $A(2)=1$, $A(3)=2$。

由初始值和式(6.13)、**式**(6.13)**可以进行递推,从而得到** $A(n)$。

6.16　**设有一码以素多项式** x^3+x^2+1 **为模,试找出此码的校验矩阵。**

解:

以素多项式 x^3+x^2+1 **为模即意味着以素多项式** x^3+x^2+1 **为生成多项式。**

因为 $1+x^7=(1+x)(1+x^2+x^3)(1+x+x^3)$,**所以此码相应的校验多项式为** $(1+x)(1+x+x^3)$,**相应的校验矩阵为**

$$H=\begin{bmatrix}0&0&1&0&1&1&1\\0&1&0&1&1&1&0\\1&0&1&1&1&0&0\end{bmatrix}$$

6.17 **已知(2,1)卷积码的一致校验方程为** $p_i = m_i + m_{i-1}$,**求校验矩阵和生成矩阵。**

解:

已知:

$$p_i = m_i + m_{i-1}$$

令编码器的输入序列为

$$m = (m_1,\ m_2,\ m_3, \cdots)$$

则编码器的输出序列为

$$C = (m_1 p_1, m_2 p_2, m_3 p_3, \cdots) = (m_1 m_1, m_2(m_1 + m_2), m_3(m_2 + m_3), \cdots)$$

由 m **可以利用下式来生成** C:

$$C = mG = (m_1,\ m_2,\ m_3, \cdots)\begin{pmatrix} 11 & 01 & 00 & 00 & \cdots \\ 00 & 11 & 01 & 00 & \cdots \\ 00 & 00 & 11 & 01 & \cdots \\ \cdots & \cdots & \cdots & \cdots & \ddots \end{pmatrix}$$

生成矩阵可以写成下列分块矩阵的形式:

$$G = \begin{pmatrix} I & P_1 & 0 & P_2 & 0 & 0 & 0 & 0 & \cdots \\ & & I & P_1 & 0 & P_2 & 0 & 0 & \cdots \\ & & & & I & P_1 & 0 & P_2 & \cdots \\ & & & & \cdots & \cdots & \cdots & \cdots & \ddots \end{pmatrix}$$

式中 I **为** $k\times k = 1\times 1$ **阶单位阵**,0**为** $k\times k = 1\times 1$ **阶全** 0 **方阵**, P_i **为** $k\times(n - k) = 1\times 1$ **阶方阵。**

$$P_1 = (1) \qquad P_2 = (1)$$

可以看到生成矩阵每一行都相同,只不过每行是上一行向右移动 2 **列。**

一致校验矩阵为

$$H = \begin{pmatrix} P_1^{\mathrm{T}} & I & \cdots & \cdots & \cdots & \cdots & \cdots \\ P_2^{\mathrm{T}} & 0 & P_1^{\mathrm{T}} & I & \cdots & \cdots & \cdots \\ 0 & 0 & P_2^{\mathrm{T}} & 0 & P_1^{\mathrm{T}} & I & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \ddots \end{pmatrix}$$

6.18 **请在** Hamming (7,4)**码中找出那些使得校验子为全零的噪声矢量。这样的噪声矢量共有多少个**?

解:

共有 15 **个这样的噪声矢量,每个噪声矢量与一个非零码字对应。**

6.19 **对于** N **次重复码** R_N,

(1) **请证明**:**当** N **为奇数时其译码错误概率** P_{E} **为**

$$P_E = \sum_{n=(N+1)/2}^{N} C_N^n p^n (1-p)^{N-n}$$

(2) 应用 Stirling 公式($x! \approx x^x \cdot e^{-x}$)写出 N 次重复码 R_N 的译码错误概率的近似表达式。

(3) 假设信道的错误概率为 $p = 0.1$,请问需要重复多少次才能将总的译码错误概率降到 10^{-15}之下。

解:

(1) 所谓发生了译码错误是指有一半以上的比特发生了翻转,因此译码错误概率为所有超过一半的比特发生了翻转的序列的概率的和:

$$P_E = \sum_{n=(N+1)/2}^{N} C_N^n p^n (1-p)^{N-n} \tag{6.15}$$

(2) 根据 Stirling 公式,有

$$\log_2 C_N^n \approx = \left[(N-n)\ln \frac{N}{N-n} + n\ln \frac{N}{n} \right] \Big/ \ln 2 = NH\left(\frac{n}{N}\right) \tag{6.16}$$

所以有

$$C_N^n \approx 2^{NH(n/N)} \tag{6.17}$$

代入式(6.15)可得

$$P_E \approx 2^N [p(1-p)]^{N/2} = [4p(1-p)]^{N/2} \tag{6.18}$$

(3) 由式(6.18)可得

$$N \approx \frac{2\log_2 P_E}{\log_2 4p(1-p)} \tag{6.19}$$

将 $P_E = 10^{-15}$和 $p = 0.1$ 代入式(6.19),有

$$N \approx \frac{2\log_2 10^{-15}}{\log_2 (4\times 0.1 \times 0.9)} = 68$$

由于 Stirling 公式存在误差,所以求出来的 N 偏大。

6.20　课本的第 6.1.2 小节(134 页)介绍了 3 次重复编码以及这种编码在一个错误概率为 p 的二进制对称信道上传输时的译码方法,最终的译码错误概率为 $3p^2 - 2p^3$,信息传输率为 $R = 1/3$。针对 4 次重复编码,其信息传输率为 $R = 1/4$,请问:

(1) 其最优的译码方案是什么?

(2) 其译码错误概率是多少?

(3) 请将它与 3 次重复编码进行比较。

解:

(1) 在每个接收序列中,如果某种比特的个数比较多,则将接收序列译码为这种比特即可。当接收序列中两种比特的个数相等时,只能进行随机的猜测,这时译码正确的概率只有 1/2。

(2) 译码错误概率为

$$P_{\mathrm{E}}=C_4^4 p^4+C_3^4 p^3(1-p)+\frac{1}{2} C_2^4 p^2(1-p)^2=3p^2-2p^3$$

(3) 两者在译码错误概率上完全一样,但是4次重复编码的信息传输率更低一些。

6.21 令 X, Y 分别为某二进制对称信道的输入和输出随机变量,此信道的错误概率为 $p=0.1$。

(1) 当 $P_{\mathrm{r}}\{X=0\}=P_{\mathrm{r}}\{X=1\}=1/2$ 时根据 Bayes 规则计算 $P_{\mathrm{r}}\{X=1 \mid Y=1\}$;

(2) 当 $P_{\mathrm{r}}\{X=1\}=0.01$ 时,重新计算第(1)问;

(3) 在第(1)问的基础上如何对接收符号 $Y=1$ 进行译码才能使得译码错误概率最小?

(4) 在第(2)问的基础上如何对接收符号 $Y=1$ 进行译码才能使得译码错误概率最小?

解:

(1) 因为 $p_X(0)=p_X(1)=1/2$,所以 $p_Y(0)=p_Y(1)=1/2$,因此有

$$p_{X|Y}(1|1)=\frac{p_{XY}(11)}{p_Y(1)}=\frac{p_X(1)\,p_{Y|X}(1|1)}{p_Y(1)}=\frac{1/2\times(1-0.1)}{1/2}=0.9$$

(2) 因为 $p_X(1)=0.01$,所以 $p_X(0)=0.99$,因此有

$$p_{X|Y}(1|1)==\frac{p_X(1)\,p_{Y|X}(1|1)}{p_Y(1)}=\frac{0.01\times0.9}{0.01\times0.9+0.99\times0.1}=0.0833$$

(3) 我们将 $Y=1$ 译码为 $\hat{X}=1$ 将使得译码错误概率为 0.1。

(4) 由于输入不再是均匀分布,所以此时的译码规则为:无论收到的 Y 是什么,一律都译码为 $\hat{X}=0$,这样才能使得译码错误概率最小。

第 7 章

限失真信源编码

7.1　**试证明对离散信源，$R(D=0)=H(X)$的充要条件是失真矩阵 D 的每行中至少有一个 0，而在每列中至多有一个 0。**

证明：

必要性：$D=0 \quad \sum_{ij} p(x_i)\,p(y_j|x_i)\,d(x_i,y_j)=0$，又因为 $p(x_i)>0$，$p(y_j|x_i)\geq 0$，$d(x_i,y_j)\geq 0$，所以

$$D=0 \quad " \; i,j,\; p(x_i)\,p(y_j|x_i)\,d(x_i,y_j)=0 \qquad \begin{cases} p(y_j|x_i)\neq 0 \text{ 时},\; d(x_i,y_j)=0 \\ p(y_j|x_i)=0 \text{ 时},\; d(x_i,y_j)>0 \end{cases}$$

若$(d(x_i,y_j))$矩阵中某行没有零，则对应的 $p(y_j|x_i)$全为零，必然存在失真，与 $D=0$矛盾。

若$(d(x_i,y_j))$矩阵中某列有超过两个以上的零，则表示有不止一个 i 使得 $p(y_j|x_i)\neq 0$，必然存在失真，与 $D=0$ 矛盾。

充分性：当 $D=0$ 即无失真时，i,j 一一对应，即 $p(y_j|x_i)=\delta_{ij}=\begin{cases}1, & i=j\\ 0, & i\neq j\end{cases}$。即对应的$(d(x_i,y_j))$矩阵中某行必然至少有一个 0，而在每列中至多有一个 0。

易知 $R(D)\big|_{D=0}=\sum_{ij} p(x_i)\,\delta_{ij}\log_2\dfrac{\delta_{ij}}{p(x_i)}=\sum_i p(x_i)\log_2\dfrac{1}{p(x_i)}=H(X)$

7.2　**利用 $R(D)$的性质，画出一般 $R(D)$的曲线，并说明其物理意义？试问为什么**

$R(D)$是非负且非增的？

解：

由 $R(D_{\min}) = H(X)$和 $R(D_{\max}) = 0$ 可以定出 $R(D)$曲线的两个端点。再根据 $R(D)$在其定义域内非负、下凸和严格递减，可以得到曲线图如图 7.1（$D_{\min} \neq 0$）和图 7.2（$D_{\min} = 0$）所示。因为 $R(D)$的物理意义为满足失真度不大于 D 的最小信息速率，而信息速率一定是非负的，所以整个曲线非负。

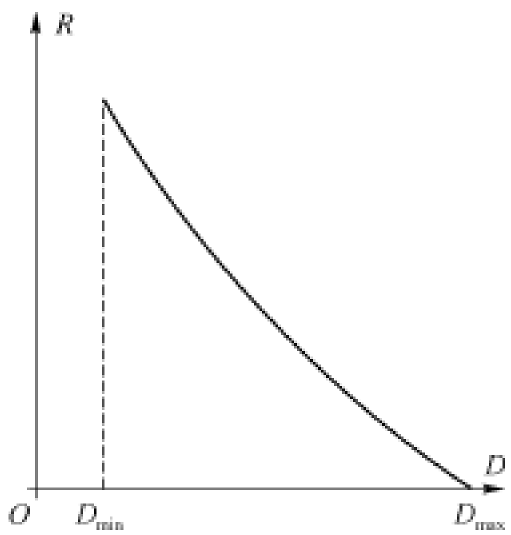


图 7.1

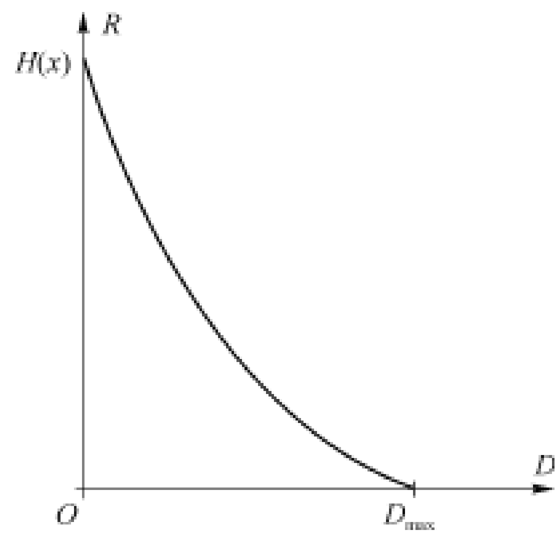


图 7.2

允许的失真门限越大，信息速率就可以降得越低，所以曲线是严格递减的。

7.3 一随机变量 X 的符号集为一个无限可数的集合 χ，假设失真测度定义为 Hamming失真。

(1) 试证明 $R(D) \geq H(X) - \log_2 |\chi| - 1) - H(D)$, " $0 \leq D \leq D_{\max}$；

(2) 试证明当 X 在 χ 上为均匀分布的情况下，上问中的下限是严格的。

证明：

(1) 在 Hamming 失真的情况下，$D = p_E$。

根据信息率失真函数的定义，有

$$R(D) = \min_{p(y_j | x_i)} \{ I(X;Y), (p(y_j | x_i)) \in B_D \} \geq I(X;Y)$$

又根据 Fano 不等式：

$$H(Y|X) \leq H(p_E) + p_E \log_2(|\chi| - 1)$$

所以有

$$R(D) \geq I(X;Y) = H(X) - H(Y|\chi) \geq H(X) - \log_2(|\chi| - 1) - H(D)$$

(2) 当 X 在 χ 上为均匀分布的情况下，由对称性可知输出分布也是均匀分布。在这种情况下要想达到上式的下界，只能选择实验信道为理想的一一对应信道（因为 X 的样本空间为无穷可数集），此时有

$$R(D) = I(X;Y) = H(X) > H(X) - \log_2(|\chi| - 1) - H(D)$$

7.4　令 $\chi=\{0,1\}=\hat{\chi}$,设失真矩阵为

$$d(x,\hat{x})=\begin{bmatrix}1 & 2\\ 3 & 2\end{bmatrix}$$

对于一个等概分布的 Bernoulli 随机变量,求 $R(D)\in(0,1)$对应的定义域$(D_{\min},D_{\max})$。

解:

根据失真矩阵,有

$$D_{\min}=1.5$$

$$D_{\max}=\min_{p(y_j)}\sum_j p(y_j)\sum_i p(x_i)d_{ij}=\min_j D_j$$

而

$$\begin{cases}D_0=2\\ D_1=2\end{cases}$$

所以 $D_{\max}=2$。

7.5　令 X 为等概分布的 Bernoulli 随机变量,相应的失真测度定义为

$$\chi=\{0,1\},\quad d(x,\hat{x})=\begin{bmatrix}0 & 1 & \infty\\ \infty & 1 & 0\end{bmatrix},\quad \hat{\chi}=\{0,e,1\}$$

(1) 请找出使得 $R(D)$非平凡的$(D_{\min},D_{\max})$;

(2) 计算 $R(D)$。

解:

(1) 失真矩阵中的无穷大元素 $d(0,1)=d(1,0)=\infty$将使得 $p(0,1)=p(1,0)=0$。根据对称性可知 $x,\hat{x}$之间的条件分布如下,如图 7.3 所示。$D_{\min}=0$, $D_{\max}=1$。

$$p(\hat{x}\mid x)=\begin{bmatrix}1-\alpha & \alpha & 0\\ 0 & \alpha & 1-\alpha\end{bmatrix}$$

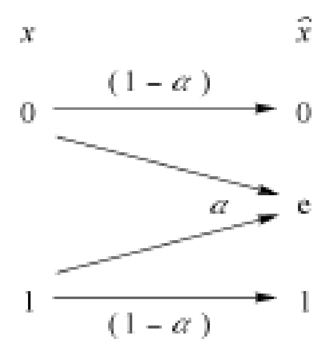


图 7.3

(2) 从上述条件概率很容易得知平均失真为 $D=\alpha$以及 $I(x;\hat{x})=1-\alpha$,因此有

$$R(D)=\begin{cases}1-D & 0\le D\le 1\\ 0 & D>1\end{cases}$$

因此可得 $D_{\min} = 0$ 以及 $D_{\max} = 1$。

7.6 考虑一个离散信源 $X \in \chi = \{1, 2, \cdots, m\}$，其分布为 $p_1, p_2, \cdots, p_m$，失真测度定义为 $d(i, j)$，令 $R(D)$ 为此信源和相应失真测度对应的信息率失真函数。通过 $d'(i, j) = d(i, j) - w_i$ 生成一个新的失真测度定义，相应的信息率失真函数变为 $R'(D)$。

(1) 试证明 $R'(D) = R(D + w')$，其中 $w' = \sum p_i w_i$；

(2) 根据上述结论说明可以通过归一化失真使得：对任意的 $x \in \chi$，至少存在一个符号 $\hat{x}$ 使得 $d(x, \hat{x}) = 0$。

证明：

(1) $d'(i, j) = d(i, j) - w_i$ 意味着 x_i 对应的平均失真减少了 w_i，所以总的平均失真将减少：

$$D - D' = \sum_i p_i w_i = w'$$

前后两个信息率失真函数相当于在 D 轴的方向上平移了 w' 的距离，因此 $R'(D) = R(D + w')$。

(2) 假设对于所有的 $\hat{x}$ 都有 $d(x, \hat{x}) \neq 0$，则选择矩阵 $(d(x, \hat{x}))$ 中最小的那个元素，令其为 w，然后对所有的 x_i 都进行下面的变换：

$$d'(x_i, \hat{x}_j) = d(x_i, \hat{x}_j) - w$$

这样得到的新的失真矩阵就会满足至少存在一个符号 $\hat{x}$ 使得 $d(x, \hat{x}) = 0$。

7.7 三元信源的概率分布为 $\{0.4, 0.4, 0.2\}$，失真函数为

$$d(i, j) = \begin{cases} 0 & i = j \\ 1 & i \neq j \end{cases}$$

试求：

(1) 信息率失真函数 $R(D)$；

(2) 若此信源用容量为 1 比特/符号和 0.1 比特/符号的信道传输，请分别计算出最小误码率 P_E。

解：

(1) 失真矩阵为

$$(d_{ij}) = \begin{bmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{bmatrix}$$

信源分布为

$$\begin{bmatrix} X \\ P \end{bmatrix} = \begin{bmatrix} 0 & 1 & 2 \\ 0.4 & 0.4 & 0.2 \end{bmatrix}$$

因为

$$\begin{cases} D_0 = \sum_i p(x_i)d(x_i, y_0) = 0.6 \\ D_1 = 0.6 \\ D_2 = 0.8 \end{cases}$$

所以有

$$D_{\min} = 0$$

$$D_{\max} = \min_{p(y_j)} \sum_j p(y_j) \sum_i p(x_i)\ d(x_i, y_j) = \min_j D_j = 0.6$$

由参量表达式 $\sum_j p(x_i)\lambda_i e^{sd_{ij}} = 1$ **可得下列方程组**:

$$\begin{cases} 0.4\lambda_0 + 0.4\lambda_1 e^s + 0.2\lambda_2 e^s = 1 \\ 0.4\lambda_0 e^s + 0.4\lambda_1 + 0.2\lambda_2 e^s = 1 \\ 0.4\lambda_0 e^s + 0.4\lambda_1 e^s + 0.2\lambda_2 = 1 \end{cases} \qquad \begin{cases} \lambda_0 = \dfrac{2.5}{1+2e^s} \\ \lambda_1 = \dfrac{2.5}{1+2e^s} \\ \lambda_2 = \dfrac{5}{1+2e^s} \end{cases}$$

又根据 $\sum_j p(y_j)e^{sd_{ij}} = \dfrac{1}{\lambda_i}$ **可得**

$$\begin{cases} p(y_0) + p(y_1)e^s + p(y_2)e^s = 0.4(1+2e^s) \\ p(y_0)e^s + p(y_1) + p(y_2)e^s = 0.4(1+2e^s) \\ p(y_0)e^s + p(y_1)e^s + p(y_2) = 0.2(1+2e^s) \end{cases} \qquad \begin{cases} p(y_0) = \dfrac{0.2(2-e^s)}{1-e^s} \\ p(y_1) = \dfrac{0.2(2-e^s)}{1-e^s} \\ p(y_2) = \dfrac{0.2(1-3e^s)}{1-e^s} \qquad e^s \leq \dfrac{1}{3} \end{cases}$$

以及

$$D(s) = \sum_i \sum_j p(x_i)p(y_j)\lambda_i d_{ij} e^{sd_{ij}} = \frac{2e^s}{1+2e^s} \qquad e^s = \frac{D}{2(1-D)} \leq \frac{1}{3} \qquad D \leq \frac{2}{5}$$

$$\begin{aligned} R(s) &= sD(s) + \sum_i p(x_i)\log_2 \lambda_i \\ &= s\frac{2e^s}{1+2e^s} + 2 \times 0.4\log_2 \frac{2.5}{1+2e^s} + 0.2\log_2 \frac{5}{1+2e^s} \\ &= D\log_2 \frac{D}{2(1-D)} + \log_2 5 - 0.8\log_2 2 + \log_2(1-D) \\ &= \log_2 5 - (D+0.8)\log_2 2 - H(D, 1-D) \quad (D \leq 0.4) \end{aligned}$$

若 $D > 0.4$,**则** $e^s > \dfrac{1}{3}$,**代入得** $p(y_2) < 0$。**所以,令** $p(y_2) = 0$,**则** $p(y_0) = p(y_1) = \dfrac{1}{2}$。**根据** $\sum_j p(y_j)e^{sd_{ij}} = \dfrac{1}{\lambda_i}$ **可得** $\lambda_0 = \lambda_1 = \dfrac{2}{1+e^s}$, $\lambda_2 = \dfrac{1}{e^s}$,**以及**

$$D(s) = \frac{0.2 + e^s}{1+e^s} \qquad e^s = \frac{D - 0.2}{1 - D}$$

$$R(s) = sD + 0.8\log_2 \frac{2}{1+e^s} + 0.2\frac{1}{e^s}$$
$$= (D-0.2)\log_2(D-0.2) + (1-D)\log_2(1-D) - 0.8\log_2 0.4 \quad (0.4 < D \leq 0.6)$$

几个关键点的函数值为

$$R(0) = 1.522, \quad R(0.4) = 0.15, \quad R(0.6) = 0$$

(2) **根据上一问得到的** $R(D)$ **函数,在已知** $R(D) = C = 1$ **时,可求得** $D = p_E = 0.089$;**在已知** $R(D) = C = 0.1$ **时,可求得** $D = p_E = 0.436$。

7.8 **一个不对称的失真测度** $d(x, \hat{x})$ **定义为**

$$d(x, \hat{x}) = \begin{bmatrix} 0 & \infty \\ 1 & 0 \end{bmatrix}, \qquad x = 0, 1; \quad \hat{x} = 0, 1$$

即:**不允许用** 1 **来表示** 0。**设随机变量** X **的概率分布为** $p_X(0) = p_X(1) = 1/2$,**并令** $R(D)$ **表示基于** $d(x, \hat{x})$ **的随机变量** X **的信息率失真函数**。

(1) $R(0)$ **的值是多少**?

(2) **使得** $R(D_0) = 0$ **的最小** D_0 **是多少**?

(3) **在满足条件** $E[d(X, \hat{X})] \leq D$ **的前提下,找出使得** $I(X; \hat{X})$ **最小并且满足** $D < D_0$ **的** $p(\hat{x} | x)$。

(4) **给出** $0 \leq D \leq D_0$ **之上的信息率失真函数** $R(D)$。

解:

(1) **为了使得失真为** 0,**必须有** $P(X = \hat{X}) = 1$,**即** $I(X; \hat{X}) = 1$,**所以有** $R(0) = 1$。

(2) **平均失真为**

$$E[d(x, \hat{x})] = p_{X,\hat{X}}(1,0) \times 1 + p_{X,\hat{X}}(0,1) \times \infty$$

如果失真是有限的,则 $p_{X,\hat{X}}(0,1) = 0$ **并且** $E[d(x, \hat{x})] = p_{X,\hat{X}}(1,0)$。**如果信息率为** 0,**则** $\hat{X}$ **与** X **统计独立,即** $p_{\hat{X}}(1) = p_{X,\hat{X}}(0,1) = p_X(0) = 0$。**因此,使得信息率为** 0 **的平均失真为** $E[d(x, \hat{x})] = p_{X,\hat{X}}(1,0) = 1/2$,**即** $D_0 = 1/2$ **是使得信息率为** 0 **的最小失真**。

(3) **根据第**(2)**问的结果,有** $E[d(x, \hat{x})] = p_{X,\hat{X}}(1,0)$。**对于任意的** $D < D_0 = 1/2$,**最小平均互信息在下述联合概率分布和条件概率分布时达到**:

$$p(\hat{x}, x) = \begin{bmatrix} 1/2 & 0 \\ D & 1/2 - D \end{bmatrix} \qquad p(\hat{x} | x) = \begin{bmatrix} 1 & 0 \\ 2D & 1-2D \end{bmatrix}$$

其中,条件概率分布 $p(\hat{x} | x)$ **的第一行是为了满足** $d(0,1) = \infty$。

(4) **第**(3)**问中联合概率分布对应的平均互信息量为**

$$R(D) = I(X; \hat{X}) = H(\hat{X}) - H(X | \hat{X}) = H\left[\frac{1}{2} - D\right] - \frac{1}{2}(2D)$$
$$= H\left[\frac{1}{2}(1-2D)\right] - \frac{1}{2}H(1-2D)$$

$R(D)$的另外一种表达式为

$$R(D)=H(\hat{X})-H(X|\hat{X})=1-\left(\frac{1}{2}+D\right)H\left(\frac{1/2}{1/2+D}\right)$$

$$=1-\frac{1}{2}(1+2D)H\left(\frac{1}{1+2D}\right)$$

$$=1+\frac{1}{2}\log_2\frac{1}{1+2D}+D\log_2\frac{2D}{1+2D}$$

7.9　信源 X 遵循$\{1,2,\cdots,m\}$上的均匀分布。Hamming 失真定义为

$$d(x,\hat{x})=\begin{cases}0 & x=\hat{x}\\ 1 & x\neq\hat{x}\end{cases}$$

请求出此信源的信息率失真函数。

解:

因为平均失真 $D=P_r\{X\neq\hat{X}\}$,根据 Fano 不等式,有

$$H(X|\hat{X})\leq H(D)+D\log_2(m-1)$$

所以有

$$I(X;\hat{X})=H(X)-H(X|\hat{X})\geq\log_2 m-H(D)-D\log_2(m-1) \tag{7.1}$$

我们可以通过选择 $p(\hat{x})$为均匀分布,并选择条件概率分布为

$$p(x|\hat{x})=\begin{cases}1-D & x=\hat{x}\\ D/(m-1) & x\neq\hat{x}\end{cases}$$

来达到式(7.1)的下界。

对称地,上述条件概率分布对应的反向试验信道的输出 X 将遵循均匀分布。容易得知,当 $D<1-1/m$ 时平均互信息量 $I(X;\hat{X})$达到上述下界 $R(D)$:

$$R(D)=\begin{cases}\log_2 m-H(D)-D\log_2(m-1) & 0\leq D\leq 1-1/m\\ 0 & D>1-1/m\end{cases}$$

对于给定的失真,$R(D)$是 m 的单调增函数。

7.10　一个离散无记忆信源输出符号等概率分布,失真函数定义如下,试求:

$$(d_{ij})=\begin{bmatrix}1 & 2\\ 1 & 1\\ 2 & 1\end{bmatrix}$$

(1) 信息率失真函数 $R(D)$;

(2) 试验信道转移概率。

解:

(1) **由对称性可设信道转移概率矩阵为**

$$P = \begin{bmatrix} p & p & 1-2p \\ 1-2p & p & p \end{bmatrix}$$

则由信源概率分布可得信道输出概率分布为

$$\begin{bmatrix} Y \\ P \end{bmatrix} = \begin{bmatrix} y_1 & y_2 & y_3 \\ \frac{1-p}{2} & p & \frac{1-p}{2} \end{bmatrix}$$

平均失真为

$$D = \sum_{ij} p(x_i)p(y_j \mid x_i)d_{ij} = 2-2p \quad p = \frac{2-D}{2}$$

因此有

$$\begin{aligned} R(D) &= H(Y) - H(P)\big|_{D参量} \\ &= -\left[2\times\frac{D}{4}\log_2\frac{D}{4} + \frac{2-D}{2}\log_2\frac{2-D}{2}\right] + \left[2\times\frac{2-D}{2}\log_2\frac{2-D}{2} + (D-1)\log_2(D-1)\right] \\ &= (D-1)\log_2(D-1) + \frac{2-D}{2}\log_2\frac{2-D}{2} - \frac{D}{2}\log_2\frac{D}{4} \end{aligned}$$

(2) **将** $p = \frac{2-D}{2}$**代入** P **可得试验信道转移概率为**

$$P\big|_{D参量} = \begin{bmatrix} \frac{2-D}{2} & \frac{2-D}{2} & D-1 \\ D-1 & \frac{2-D}{2} & \frac{2-D}{2} \end{bmatrix}$$

7.11 **一个离散** n **进制等概率无记忆信源,且具有对称失真函数**:

$$d_{ij} = \begin{cases} 0 & i = j \\ \alpha & i \neq j \end{cases}$$

试证明:

$$R(D) = \frac{D}{\alpha}\log_2\frac{\frac{D}{\alpha}}{1-\frac{1}{n}} + \left(1-\frac{D}{\alpha}\right)\log_2\frac{1-\frac{D}{\alpha}}{\frac{1}{n}}$$

证明:

由 d_{ij}**的对称性和** P **的归一性可以假设**:

$$P=\begin{bmatrix} p & \frac{1-p}{n-1} & \cdots & \frac{1-p}{n-1} \\ \frac{1-p}{n-1} & p & \cdots & \frac{1-p}{n-1} \\ \cdots & \cdots & & \cdots \\ \frac{1-p}{n-1} & \frac{1-p}{n-1} & \cdots & p \end{bmatrix}$$

平均失真为

$$D=\sum_{ij} p(x_i)p(y_j|x_i)d_{ij}=(1-p)\alpha \quad p=1-\frac{D}{\alpha}$$

因此有

$$p(y_j)=\sum_i p(x_i)p(y_j|x_i)=\frac{1}{n} \qquad j=1,2,\cdots,n$$

所以可得:

$$\begin{aligned} R(D) &= H(Y)-H(P)|_{D参量} \\ &= \log_2 n+p\log_2 p+(1-p)\log_2(1-p)+(p-1)\log_2 n\left(1-\frac{1}{n}\right) \\ &= p\log_2 np+(1-p)\log_2\frac{1-p}{1-\frac{1}{n}} \\ &= \frac{D}{\alpha}\log_2\frac{\frac{D}{\alpha}}{1-\frac{1}{n}}+\left(1-\frac{D}{\alpha}\right)\log_2\frac{1-\frac{D}{\alpha}}{\frac{1}{n}} \end{aligned}$$

第 8 章

上机实验

8.1　常用 C 语言函数简介

本小节主要介绍一下后面将会用到的几个 C 语言函数。

8.1.1　动态内存分配函数

1. malloc 函数用于动态分配内存

void * malloc(size_t size);

其中参数 size 表示所需内存的大小,以字节为单位。

2. calloc 函数用于动态分配内存

void * calloc(size_t nmemb, size_t size);

其中参数 nmemb 表示元素的个数,size 表示每个元素的大小(以字节为单位)。

3. realloc 函数用于动态调整所分配的内存的大小

void * realloc(void * ptr, size_t size);

realloc 函数将指针 ptr 所指向的内存区域的大小修改为 size 个字节。

4 . free 函数用于释放所分配的内存

void free(void *ptr);

free 函数释放指针 ptr 所指向的内存。ptr 必须是由 malloc, calloc 或 realloc 返回的指针。

上述 3 个函数的原型定义在头文件 stdlib .h 中。

malloc 和 calloc 函数将返回一个指向分配出来的内存区域的指针,如果失败将返回一个空指针 NULL。

realloc 函数将返回一个指向调整大小后的内存区域的指针,如果失败将返回一个空指针 NULL。如果参数 size 为 0,则将返回一个空指针或一个适合于作为 free 函数参数的指针。当函数 realloc 失败时,原来的内存区域不会受到影响。

8 .1 2　输入输出函数

1 . fscanf 函数用于输入数据

int fscanf(FILE *fp, const char *fmt, …);

fscanf 函数用于从指定的流 fp 中依照格式 fmt 读入数据,从第三个参数开始是待读入数据对应的地址变量的列表。其中参数“流”通常是 stdin(标准输入,即键盘)或之前以 ASCII 读模式打开的文件。此函数成功时返回读入的变量个数,失败时返回 EOF。

2 . fprintf 函数用于输出数据

int fprintf(FILE *fp, const char *fmt, …);

fprintf 函数将变量的值依照格式 fmt 输出到流 fp 中。其中参数“流”通常是 stdout(标准输出,即显示器)、stderr(标准错误,即显示器)或之前以 ASCII 写模式打开的文件。此函数成功时返回输出的字符个数,失败时返回一个负数。

上述两个函数的原型定义在头文件 stdio .h 中。

参数 fmt 称为格式描述字符串,由普通字符串和格式描述符组成,所有的格式描述符都以百分号(%)开头。常用的格式描述符如表 8 .1 所示。

表 8.1 标准 I/O 格式描述符

描述符	含义
%	输出百分号,即格式字符串%%的输出为%
c	输出字符型数据
s	输出 C 字符串(以'\0'结尾的字符串)
d,i	输出整型数据,缺省为十进制
o,u,x,X	分别表示以八进制、十进制、小写十六进制和大写十六进制输出整型数据
f	输出浮点型数据
e	以科学记数法输出浮点型数据

8.1.3 字符串处理函数

1. strcpy 函数用于字符串复制

```
char * strcpy(char * dst, const char * src);
```

它将由 src 指向的字符串拷贝到 dst 指向的字符串中,包括字符串终止符'\0'。目的缓冲区 dst 的大小必须足以接受复制。

此函数的返回值为指向目的字符串 dst 的指针。

2. strncpy 函数用于指定长度的字符串复制

```
char * strncpy(char * dst, const char * src, size_t n);
```

此函数与 strcpy 的唯一区别就是指定了复制的字符个数 *n*。如果在 src 的前 *n* 个字符中没有字符串终止符,则复制的结果也将没有字符串终止符。当 src 的长度小于 dst 时,dst 中剩余的位置将被空字符填充。

此函数的返回值为指向目的字符串 dst 的指针。

3. strcmp 函数用于字符串比较

```
int strcmp(const char * s1, const char * s2);
```

根据字符串 s1 按照字典顺序大于、等于或小于字符串 s2,此函数分别返回一个大于、等于或小于零的整数。

4. strncmp 函数用于指定长度的字符串比较

```
int strcmp(const char * s1, const char * s2, size_t n);
```

此函数与 strcmp 的唯一区别是只比较两个字符串的前 *n* 个字符。

5. strcat 函数用于字符串串接

```
char * strcat(char * dst, const char * src);
```

此函数首先去掉字符串 dst 末尾的字符串终止符'\0',然后将 src 追加到其后的位

置,串接为一个新的字符串返回。目的字符串的大小必须足够大以容纳串接的结果。

此函数的返回值为指向目的字符串 dst 的指针。

6 . strlen 函数返回字符串的长度

```
size_t strlen(const char * str);
```

此函数计算字符串 str 的长度,不包括字符串终止符 \ 0 ’并返回。

8 2　信道容量的迭代算法

1 . 实验目的

(1) 进一步熟悉信道容量的迭代算法;

(2) 学习如何将复杂的公式转化为程序;

(3) 掌握 C 语言数值计算程序的设计和调试技术。

2 . 实验要求

(1) 已知:信源符号个数 r、信宿符号个数 s、信道转移概率矩阵 P。

(2) 输入:任意的一个信道转移概率矩阵。信源符号个数、信宿符号个数和每个具体的转移概率在运行时从键盘输入。

(3) 输出:最佳信源分布 P^*,信道容量 C。

3 . 算法

Algorithm 1:信道容量的迭代算法

procedure CHANNEL CAPACITY(r, s, (p_{ji}))

initialize:信源分布 $p_i = \frac{1}{r}$,相对误差门限 δ, $C = -\infty$

repeat

$$ {}_{ij} \leftarrow \frac{p_i p_{ji}}{\sum_{i=1}^{r} p_i p_{ji}} $$

$$ p_i \leftarrow \frac{\exp\left(\sum_{j=1}^{s} p_{ji} \log_2 {}_{ij}\right)}{\sum_{i=1}^{r} \exp\left(\sum_{j=1}^{s} p_{ji} \log_2 {}_{ij}\right)} $$

$$ C \leftarrow \log_2\left[\sum_{i=1}^{r} \exp\left(\sum_{j=1}^{s} p_{ji} \log_2 {}_{ij}\right)\right] $$

until $\frac{|\Delta C|}{C} \le \delta$

output $P^* = (p_i)_r, C$

end procedure

4 . 代码

```
/ ***************************************************************
 * Author       : Hop Lee
 * Date         : 2003.06.25
 * Copyright    : GPL
 * Purpose      : Caculate the capacity of a given channel
 *  ***************************************************************/
#include <stdio.h>
#include <math.h>
#include <stdlib.h>
#include <unistd.h>
#include <values.h>

#define DELTA 1e-6/* delta,the threshold */

int main(void)
{
  register int i,j;
  register int k;
  int r,s;

  float *p_i = NULL;
  float **p_ji = NULL;
  float **phi_ij = NULL;
  float C,C_pre,validate;

  float *sum = NULL;
  float p_j;

  /* Read the number of input symbol: r,
```

```
  * and the number of output symbol : s */
fscanf (stdin,"%d", &r);
fscanf (stdin,"%d", &s);

/* Allocation memory for p_i, p_ji and phi_ij */
p_i = (float *) calloc (r, sizeof (float));

p_ji = (float **) calloc (r, sizeof (float));
for (i = 0; i < r; i++)
   p_ji[i] = (float *) calloc (s, sizeof (float));

phi_ij = (float **) calloc (r, sizeof (float *));
for (i = 0; i < r; i++)
   phi_ij[i] = (float *) calloc (s, sizeof (float))

/* Read the channel transition probability matrix p_ji */
for (i = 0; i < r; i++)
   for (j = 0; j < s; j++)
      fscanf (stdin,"%f", &p_ji[i][j]);

/* Validate the input data */
for (i = 0; i < r; i++)
{
   validate = 0.0;
   for (j = 0; j < s; j++)
   {
      validate += p_ji[i][j];
   }
   if (fabs (validate - 1.0) > DELTA)
   {
      fprintf (stdout,"Invalid input data .\n");
      exit (-1);
   }
}
```

```
fprintf (stdout,"Starting ...\n");

/* Initialize the p_i and phi_ij */
for (i = 0; i < r; i++ )
{
   p_i[i] = 1.0/ ( float ) r;
}
/* Initialize C and iteration counter : k,and temprory variable */
C = - MAXFLOAT ; /* MAXFLOAT was defined in <values.h> */
k = 0;
sum = ( float * ) calloc (r,sizeof ( float ));

/* Start iterate */
do
{
   k++;

   /* Calculate phi_ij (k) first */
   for (j = 0; j < s; j++ )
   {
      p_j = 0.0;
      for (i = 0; i < r; i++ )
         p_j += p_i[i] * p_ji [i][j];

      if ( fabs ( p_j ) >= DELTA )
         for (i = 0; i < r; i++ )
            phi_ij [i][j] = p_i[i] * p_ji [i][j]/ p_j;
      else
         for (i = 0; i < r; i++ )
            phi_ij [i][j] = 0.0;
   }

   /* Calculate p_i(k +1) then */
   p_j = 0.0;
   for (i = 0; i < r; i++ )
```

```
    {
        sum[i] = 0.0;
        for (j = 0; j < s; j++)
        {
            /* Prevent divided by zero */
            if (fabs(phi_ij[i][j]) >= DELTA)
                sum[i] += p_ji[i][j] * log2(phi_ij[i][j]) / log2(2.0);
            }
            sum[i] = pow(2.0, sum[i]);
            p_j += sum[i];
    }

    for (i = 0; i < r; i++)
    {
        p_i[i] = sum[i] / p_j;
    }
   /* And C(k + 1) */
    C_pre = C;
    C = log2(p_j) / log2(2.0);
  }
  while (fabs(C - C_pre) / C > DELTA);


  free(sum);
  sum = NULL;

 /* Output the result */
  fprintf(stdout, "The iteration number is %d.\n\n", k);
  fprintf(stdout, "The capacity of the channel is %.6f bit/symbol.\n\n", C);
  fprintf(stdout, "The best input probability distribution is :\n");
  for (i = 0; i < r; i++)
     fprintf(stdout, "%.6f ", p_i[i]);
  fprintf(stdout, "\n");

 /* Free the memory we allocation before with stack sequence */
```

```
for (i = s - 1; i >= 0; i--)
{
    free(phi_ij[i]);
    phi_ij[i] = NULL;
}
free(phi_ij);
phi_ij = NULL;

for (i = r - 1; i >= 0; i--)
{
    free(p_ji[i]);
    p_ji[i] = NULL;
}
free(p_ji);
p_ji = NULL;
free(p_i);
p_i = NULL;

exit(0);
}
```

8.3 唯一可译码判决准则

1. 实验目的

(1) **进一步熟悉唯一可译码判决准则**;

(2) **掌握 C 语言字符串处理程序的设计和调试技术。**

2. 实验要求

(1) **已知:信源符号个数** q、**码字集合** C。

(2) **输入:任意的一个码。码字个数和每个具体的码字在运行时从键盘输入。**

(3) **输出:判决(是唯一可译码/不是唯一可译码)。**

3．算法

Algorithm 2: 唯一可译码判决准则

procedure UNIQUEDECODABLE(C)

for all W_i, $W_j \in C$ do

if W_i **是** W_j **的前缀** then

将相应的后缀作为一个尾随后缀码放入集合 F_0 **中**

end if

end for

loop

for all $W_i \in C$ do

for all $W_j \in F_n$ do

if W_i **是** W_j **的前缀** then

将相应的后缀作为一个尾随后缀码放入集合 F_{n+1} **中**

else if W_j **是** W_i **的前缀** then

将相应的后缀作为一个尾随后缀码放入集合 F_{n+1} **中**

end if

end for

end for

$F \leftarrow \bigcup_i F_i$

ifv $W_i \in F$, $W_i \in C$ then

return False

elss if F **中没有出现新的元素** then

return True

end if

end loop

end procedure

4．代码

```
/***************************************************************
 * Author : Chen Min Ru
 * Date : 2004.03.16
 * Copyright : GPL
 * Purpose : Find out whether a code is unique decodable or not
 ***************************************************************/
```

```
# include < stdio .h >
# include <math .h >
# include < stdlib .h >
# include <unistd .h >
# include < values .h >
# include < string .h >
# include "ourhdr .h"

int scomp ( char * * C , int n , int * l);
void compline ( char * C , int cline , char * D,
                    int dline , char * * E , int * elinenum );
int  judge ( char * * C , int crow , char * * E);
int  erow ;

int  main (void)
{
  register  int i;
  int         n, l;
  int         * l_i;
  char        * * C;

  / * Read the number of input symbol : n */
  fscanf (stdin , "% d", & n);

  / * Allocation memory for l_i and * C[i] */
  l_i = ( int * ) calloc (n , sizeof ( int ));
  C = ( char * * ) calloc (n , sizeof ( char * ));

  / * Read the code and its length in */
  for (i = 0; i < n; i++ )
  {
     fscanf (stdin, "% d", & l_i[i]);
     l = l_i[i];
     C[i] = ( char * ) calloc (l , sizeof ( char ));
      fscanf (stdin , "% s", C[i]);
```

```
  }

  fprintf (stdout , "Starting ...\n\n");

  i = scomp (C , n , l_i );

 /* Output the result */
  if(i == 0)
      fprintf ( stdout , "The code is unique decodable .\n");
  else if(i == 1)
      fprintf ( stdout , "The code is NOT unique decodable .\n");
   else
      fprintf ( stderr , "Error ! \n");

 /* Free the memory we allocation before with stack sequeue */
  for (i = n - 1; i > = 0; i - - )
  {
  free (C[i]);
  }
  free ( *C);
  free (l_i );

  exit (0);
}

/* Compare C , D and E*/
int scomp ( char **C , int n , int *l)
{
  int i;
   int crow,drow ;
   char **D, ** E;
   int * dlinenum , * elinenum ;
   int drownum ;
 /* Allocate space to dlinenum [i] and elinenum [i] */
   dlinenum = ( int * ) calloc (n, sizeof (int));
```

```
  /* For the first time, D has the same length with C */
  D = (char **) calloc (n, sizeof (char *));
  for (i = 0; i < n; i++)
  {
     dlinenum[i] = l[i];
     D[i] = (char *) calloc (l[i], sizeof (char));
  }
  drownum = n;

  /* For the first time, let D = C */
  for (i = 0; i < n; i++)
      strcpy (D[i], C[i]);

/* Compare C and D */
compare:
  erow = 1;
  E = (char **) calloc (n, sizeof (char *));
  for (i = 0; i < n; i++)
  {
     E[i] = (char *) calloc (n, sizeof (char));
  }
  elinenum = (int *) calloc (n, sizeof (int));

  for (crow = 0; crow < n; crow++)
  {
     for (drow = 0; drow < drownum; drow++)
     {
         compline (C[crow], 1[crow], D[drow],
                   dlinenum[drow], E, elinenum);
     }
  }
/* Compare D and C */
for (drow = 0; drow < drownum; drow++)
  {
     for (crow = 0; crow < n; crow++)
```

```
        {
            compline (D[drow], dlinenum [drow], C[crow],
                        l[crow], E , elinenum );
        }
    }

    i = judge (C, crow , E);

    if (i != 1 && i != 0) /* Let D = E and go on comparing */
    {
        drownum = erow - 1;
        realloc (D , drownum );
        realloc (dlinenum , drownum );
        for (i = drownum - 1; i >= 0; i -- )
       {
            dlinenum [i] = elinenum [i];
            realloc (D[i], dlinenum [i]);
            strcpy (D[i], E[i]);
            free (E[i]);
        }
        free ( * E);
        free ( elinenum );

        goto compare ;
    }
    else
    {
        for (i = erow - 2; i >= 0; i -- )
        {
            free (E[i]);
        }
        for (i = drownum - 1; i >= 0; i -- )
        {
            free (D[i]);
        }
```

```
        free ( * D);
        free ( * E);
        free (dlinenum);
        free (elinenum);

        if(i == 1) / *  It 's NOT unique decodable */
            return 1;
        else if(i == 0) / *  It 's unique decodable */
            return 0;
    }
}

int judge ( char ** C , int crow , char ** E)
{
    int i , j;
    if( erow == 1)
        return 0; / *  Stop the processing */

    for (i = 0; i < crow ; i++)
        for (j = 0; j < erow - 1; j ++)
            if( strcmp (C[i], E[j]) == 0)
                return 1; / *  Stop the processing */

    return 2;
}

    void compline (char * C,
                        int cline ,
                        char * D,
                        int dline ,
                        char ** E,
                        int * elinenum )
{
    if( cline < dline )
    {
```

```
        if( strncmp (C , D , cline ) == 0)
      {
        / * Reallocate E */
        realloc (E , erow );
        realloc ( elinenum , erow );

        elinenum [erow - 1] = dline - cline ;
        realloc (E[erow - 1] , dline - cline + 1);
        / * Copy the last dline - cline code into E */
        strncpy (E[erow - 1] , & D[cline], dline - cline + 1);
        erow ++;
      }
   }
}
```

8.4　Huffman 编码

1 . 实验目的

(1) **进一步熟悉** Huffman **编码过程**;

(2) **掌握** C **语言递归程序的设计和调试技术。**

2 . 实验要求

(1) **输入**:**信源符号个数** r、**信源的概率分布** P;

(2) **输出**:**每个信源符号对应的** Huffman **编码的码字。**

3 . 算法描述

Algorithm 3: Huffman **编码算法**

procedure HUFFMAN($\{s_i\}$, $\{p_i\}$)

　　if $q == 2$ then

　　　　return $s_0 \mapsto 0$, $s_1 \mapsto 1$

　　else

　　　　降序排序$\{p_i\}$

　　　　缩减信源:**创建一个符号** s'**以取代** s_{q-2}, s_{q-1}, **其概率为** $p' = p_{q-2} + p_{q-1}$

　　　　递归调用 Huffman **算法以得到** $s_0, \cdots, s_{q-3}, s'$ **的编码**: $w_0, \cdots, w_{q-3}, w'$, **相应的**

概率分布为 $p_0, \cdots, p_{q-3}, p'$

return $s_0 \mapsto w0, \cdots, s_{q-3} \mapsto w_{q-3}, s_{q-2} \mapsto w'0, s_{q-1} \mapsto w'1$

end if

end procedure

4．代码

```
/************************************************************
 * Author : Chen Min Ru
 * Date : 2004.03.15
 * Copyright : GPL
 * Purpose : Huffman recursive coding algorithm
 ************************************************************/
#include <stdio.h>
#include <stdlib.h>
#include <string.h>
#include <math.h>

#define DELTA 1.0e-6

void sort(double *, char **, int *, int);
void code(double *, char **, int *, int);

int
main(void)
{
  float *p_i, *p;
  float sum;
  float temp;
  char **c;
  int *idx;
  int q;
  int i;

  /* Read the number of source symbol in */
```

```
    fscanf(stdin, "%d", &q);

  /* Allocation memory */
    idx = (int *) calloc(q, sizeof(int));
    p_i = (float *) calloc(q, sizeof(float));
    p = (float *) calloc(q, sizeof(float));
    c = (char **) calloc(q, sizeof(char *));

   for (i = 0; i < q; i++)
   {
       c[i] = (char *) calloc(1, sizeof(char));
       c[i][0] = '\0';
   }

    /* Read the probability of each symbol in and validate them */
    sum = 0.0;
    for (i = 0; i < q; i++)
    {
      fscanf(stdin, "%f", &p[i]);
       p_i[i] = p[i];
       idx[i] = i;
      sum += p_i[i];
    }

    if (fabs(sum - 1.0) > DELTA)
    {
       fprintf(stderr, "p_i error\n");
       exit(-1);
    }

    /* Coding */
    code(p_i, c, idx, q);

    /* Output result */
    for (i = 0; i < q; i++)
```

```
    {
        fprintf ( stdout , "%d %.6f %s\n", idx[i], p[idx[i]] , c[idx[i]]);
    }

   /* Free the memory */
    for (i = q; i > 0; --i)
    {
        free (c[i]);
    }
    free (c);
    free (p);
    free (p_i );
    free (idx );

    exit (0);
}

/*
 * Sorting algorithm
 */
void
sort ( float *p , char **c , int *idx , int q)
{
      int finish = 0;
      int i, j;
      int l1, l2;
      char *s;
      float t;

      while (i < q && !finish )
      {
         finish = 1;
          for (j = 0; j < q - i; j ++ )
         {
            if(p[j] < p[j + 1])
```

```
            {
                t = p[j];
                p[j] = p[j + 1];
                p[j + 1] = t;

                l1 = idx[j];
                idx[j] = idx[j + 1];
                idx[j + 1] = l1;

                l1 = strlen(c[j]);
                l2 = strlen(c[j + 1]);
                s = (char *)calloc(l1 + 1, sizeof(char));

                strcpy(s, c[j]);

                realloc(c[j], l2 + 1);
                strcpy(c[j], c[j + 1]);

                realloc(c[j + 1], l1 + 1);
                strcpy(c[j + 1], s);

                free(s);
            }
        }
        i++;
    }
}

/*
 * Huffman recursive coding algorithm
 */
void code(float *p, char **c, int *idx, int q)
{
    int l1, l2;
    char *s;
```

```

    /* If q=2, return the code words */
    if(q == 2)
    {
       realloc (c [0] , 3);
       realloc (c [1] , 3);

       strcat (c[0] , "0");
       strcat (c[1] , "1");
    }
    else
    {
       /* If q>2, reduce the source . */
       p[q - 2] = p[q - 1] + p[q - 2];
       /* Call the coding algorithm recursively */
       sort (p , c , idx , q - 1);
       code (p , c , idx , q - 1);

       /* Resume the source and return the code words */
       l1 = strlen (c[q - 2]);
       l2 = strlen (c[q - 1]);

       s = ( char * ) calloc (l1 + 2 , sizeof ( char * ));
       strcpy (s , c[q - 2]);

       realloc (c[q - 2] , l1 + 2);
       strcat (c[q - 2] , "0");

       realloc (c[q - 1] , l2 + 2);
       strcat (s , "1");
       strcpy (c[q - 1] , s);

       free (s);
    }
}
```

8.5　LZW 编码

1 . 实验目的

(1) **进一步熟悉通用编码算法;**

(2) **掌握 C 语言程序设计和调试过程中数值的进制转换、数值与字符串之间的转换等技术。**

2 . 实验要求

(1) **输入:本程序将从标准输入中读入待压缩的数据流;**

(2) **输出:将压缩结果输出到标准输出上去。**

3 . 算法

Algorithm 4: LZW **编码算法**

1:procedure LZW

2:**字典初始化:将被压缩文件中所有使用到的单字节字符放如字典中,为了压缩任何类型的文件,可以将字典的前** 256 **个位置**(0x000 **到** 0x0FF)**依次分配给** 0x00 **到** 0xFF **的** 256 **个单字节字符。**

3:**动态数据初始化:初始化新单词存放位置指针** P。**将它指向字典的第一个空位置。例如** P = 256(**即** 0x100),**读入被压缩文件的第一个字符** cha,**作为待处理单词** W。**单词的前缀** Q **为空,即** Q = 4095,**尾字符就是** cha,**序号(码字)就是** cha **的序号。**

4:**如果文件再没有字符了,输出当前单词** W **的序号。编码结束。如果文件中还有字符,把当前单词** W **作为前缀,再从被压缩文件中读入一个字符** CH,**把** CH **作为尾字符,得到一个单词** W_1。

5:**如果字典中已有** W_1,**则将** W_1 **看作当前单词** W,**返回第三步。如果字典中没有** W_1 (**发现一个新单词**),**先将原单词** W **的序号输出,再加新单词** W_1,**增加到字典中,然后把刚刚读入的字符** CH **作为当前单词** W,**返回第三步。**

6:end procedure

---

4 . 代码

```
# include < stdio .h >
# include < stdlib .h >
# include < conio .h >

```

```
 struct word {
unsigned int n;
unsigned char c;
} w,wd[4096]; //  Dictionary
unsigned int p , n;
unsigned char h , m , l , f;

/ *  Out put the code  */
void out( int n)
{
        if(f == 0){
        h = n/ 16;
        m = (n<<4) & 0xf0;
        f = 1;
    }
    else {
        m += n/ 256;
        l = n & 0xff;
        fputc (h , stdout );
        fputc (m , stdout );
        fputc (l , stdout );
        h = m = l = f = 0;
    }
 }

/ *  Main compress program  */
void lzw ()
{
int c , i;
   unsigned char ch;
   fprintf (stderr , "\n\n begin compress , please wait !  \n");
   for (i = 0; i < 256; i ++){//  Initialize first 256 word
        wd[i].n = 4095;        //  in dictionary
        wd[i].c = i;
    }
   p = 256;
```

```
    w.n = 4095;
    w.c = n = fgetc( stdin );
    h = m = l = f = 0;
    for(;;){
      c = fgetc( stdin );
       if(c == -1){
         out(n);
         if(f)
            out(4095);
         fprintf(stderr,"\n\n compression is over ! \n");
         return;
       }
       ch = c;
       for(i = n + 1; i < p; i ++){
          if(wd[i].n != n)
             continue;
         if(wd[i].c == ch)
             break;
       }
     if(i != p){
          w.n = n;
          w.c = ch;
          n = i;
       }
       else {
          out(n);
          if(p < 4095){
             wd[p].n = n;
             wd[p].c;
             p++;
          }
          w.n = 4095;
          n = w.c = ch;
       }
    }
 }
```

```
void main ( void )
{
lzw ( ) ;
}
```

8.6 Shannon 编码

1 . 实验目的

(1) **进一步熟悉** Shannon **编码算法**;

(2) **掌握** C **语言程序设计和调试过程中数值的进制转换、数值与字符串之间的转换等技术**。

2 . 实验要求

(1) **输入**: **信源符号个数** q、**信源的概率分布** P;

(2) **输出**: **每个信源符号对应的** Shannon **编码的码字**。

3 . 算法

Algorithm 5: Shannon **编码算法**

procedure SHANNON(q, $\{ p_i \}$)
　　降序排序 $\{ p_i \}$
　　for $i = 1 \mapsto q$ do
　　　　$F(s_i) \leftarrow \sum_{k=1}^{i-1} p(s_k)$
　　　　$l_i \leftarrow \left\lceil \log_2 \frac{1}{p(s_i)} \right\rceil$
　　　　将累加概率 $F(s_i)$ (**十进制小数**) **变换成二进制小数**。
　　　　取小数点后 l_i **个二进制数字作为第** i **个消息的码字**。
　　end for
end procedure

4 . 代码

```
/ **************************************************
 *  Author : Chen Min Ru
 *  Date : 2004 .03 .15
```

```
* Copyright : GPL
* Purpose : Use shannon algorithm to code the source symbols
*****************************************************
#include <stdio.h>
#include <math.h>
#include <stdlib.h>
#include <unistd.h>
#include <values.h>
#include <string.h>

#define DELTA 1e-6

void sort(float *, int);

int main(void)
{
    register int i, j;
    int n; /* Number of the total words */

    int temp;

    float *p_i; /* Probability of the word */
    float *P_i; /* Cumulate probability */
    int *l_i; /* Code length */
    char **C; /* Code set */
   /* Use sum to test the data and p to store the temp data */
    float sum, p;

   /* Read the numer of input symbol : n */
    fscanf(stdin, "%d", &n);

   /* Allocation memory for p_i and *C[i] */
    p_i = (float *) calloc(n, sizeof(float));
    P_i = (float *) calloc(n, sizeof(float));
    l_i = (int *) calloc(n, sizeof(int));
```

```
    /* Read the channel transition probability matrix p_i in */
    for (i = 0; i < n; i++)
        fscanf(stdin, "%f", &p_i[i]);

    /* Validate the input data */
    sum = 0.0;
    for (i = 0; i < n; i++)
        sum += p_i[i];
    if( fabs( sum - 1.0) > DELTA )
        fprintf( stderr, "Invalid input data \n");

    fprintf(stdout, "Starting ...\n\n");

    /* Sorting the p_i descend */
    sort(p_i, n);

    /* Calculate the binary number's length */
    for (i = 0; i < n; i++)
    {
        p = ( - ( log2( p_i[i])))/ log2(2.0);
        l_i[i] = (int) ceil(p);
    }

    /* Allocate C[i] */
    C = (char **) calloc(n, sizeof(char *));
    for (i = 0; i < n; i++)
    {
        C[i] = (char *) calloc(l_i[i] + 1, sizeof(char));
        C[i][0] = '\0';
    }

   /* Calculate the P_i[i] */
    P_i[0] = 0.0;
    for (i = 1; i < n; i++)
```

```
      P_i[i] = P_i[i - 1] + p_i[i - 1];

   /* Transform P_i to binary mode */
   for (i = 0; i < n; i++)
   {
      for (j = 0; j < l_i[i]; j++)
      {
         P_i[i] = P_i[i] * 2;
         temp = (int)(P_i[i]);
         P_i[i] = P_i[i] - temp;

         if( temp == 0)
            C[i] = strcat (C[i], "0");
         else
            C[i] = strcat (C[i], "1");
      }
   }

   /* Output the result */
   fprintf (stdout , "The output coding is : \n");
   for (i = 0; i < n; i++)
      fprintf ( stdout , "%s", C[i]);
   fprintf (stdout , "\n\n");

   /* Free the memory we allocation */
   for (i = n - 1; i >= 0; i--)
      free (C[i]);
   free (C);
   free (p_i );
   free (P_i );
   free (l_i );

   exit (0);
}
```

```
/ * Bubble sorting */
void sort ( float *k , int m)
{
    int i = 1;
    int j = 1;
    int finish = 0;
    float temp ;

    while (i < m && !finish )
    {
        finish = 1;
        for (j = 0; j < m - i; j ++ )
        {
            if(k[j] < k[j + 1])
            {
                temp = k[j];
                k[j] = k[j + 1];
                k[j + 1] = k[j];
                finish = 0;
            }
            i ++ ;
        }
    }
}
```

参 考 文 献

1 傅祖芸 . 信息论——基础理论与应用 . 北京: 电子工业出版社, 2001

2 周荫清 . 信息理论基础 . 北京: 北京航空航天大学出版社, 2002

3 T M Cover & J A Thomas . Elements of Information Theory .New York: John Wiley & Sons . Inc ., 1991

4 [美] R W 汉明著, 朱雪龙译 . 编码和信息理论 . 北京: 科学出版社, 1984

5 朱雪龙 . 应用信息论基础 . 北京: 清华大学出版社, 2001

6 姜丹 . 信息论与编码 . 合肥: 中国科技大学出版社, 2001

7 陈运 . 信息论与编码 . 北京: 电子工业出版社, 2002

8 曹雪虹 . 信息论与编码 . 北京: 北京邮电大学出版社, 2001

9 吴伟陵 . 信息处理与编码 . 北京: 人民邮电出版社, 2003