应用信息论基础——第三、四章作业

2022年12月14日

第三章重点总结

- AEP
  - “几乎一切事件都令人等同的意外。”具体讲，若$X_1, X_2, \dots, X_n$为服从$p(x)$的i.i.d序列，则

$$-\frac{1}{n}\log p(X_1, X_2, \dots, X_n) \to H(X) \qquad \text{依概率}$$

- **典型集：**
  - 典型集$A_\varepsilon^{(n)}$为满足如下条件的序列$x_1, x_2, \dots, x_n$的集合

$$2^{-n(H(X)+\varepsilon)} \le p(x_1, x_2, \dots, x_n) \le 2^{-n(H(X)-\varepsilon)}$$

- **典型集的性质**

定理 3.1.2

这些序列几乎是等概的

1. 如果$(x_1, x_2, \cdots, x_n) \in A_\epsilon^{(n)}$，则 $H(X) - \epsilon \leqslant -\frac{1}{n}\log p(x_1, x_2, \cdots, x_n) \leqslant H(X) + \epsilon$
2. 当 n 充分大时，$\Pr\{A_\epsilon^{(n)}\} > 1 - \epsilon$。 序列集概率和的值接近1

扩张稳定性：
单符号的熵是一致的

3. $|A_\epsilon^{(n)}| \leqslant 2^{n(H(X)+\epsilon)}$，其中$|A|$表示集合 A 中的元素个数。

序列的数目

4. 当 n 充分大时，$|A_\epsilon^{(n)}| \geqslant (1-\epsilon)2^{n(H(X)-\epsilon)}$。

- 典型集选择的一致性理论

第三章重点总结

Tsinghua University

- **Kraft不等式：**唯一可译码的 充要条件

$$\sum_i D^{-l_i} \leq 1$$

- **数据压缩的熵界**

$$L = \sum_i p_i l_i \geq H_D(X)$$

- **香农码**

$$l_i = \left\lceil log_D \frac{1}{p_i} \right\rceil$$

$$H_D(X) \leq L^* < H_D(X) + 1$$

- **哈夫曼码**

$$L^* = \min_{\sum_i D^{-l_i} \leq 1} p_i l_i$$

$$H_D(X) \leq L^* < H_D(X) + 1$$

- **随机过程**

$$\frac{H(X_1, X_2, \ldots, X_n)}{n} \leq L_n < \frac{H(X_1, X_2, \ldots, X_n)}{n} + \frac{1}{n}$$

  平稳过程

$$L_n \to H_\infty(X)$$

- **竞争最优性**

1：典型集的性质

1.设有二元离散无记忆信源，熵率为$H_\infty(U)$，试求，在序列长$N \to \infty$时

（1）不同典型序列的总数在全部可能序列中所占的比例

（2）信源序列中出现典型序列的概率

解：

$H_\infty(U) = 1$，等概，所有序列都是典型序列， 占比为1

(1) 设典型序列集合为G，则对任意小的正整数$\varepsilon > 0$，$\delta > 0$，我们有

$$(1-\delta)2^{N[H_\infty(U)-\varepsilon]} \le |G| \le 2^{N[H_\infty(U)+\varepsilon]}$$

注意所有可能的序列个数为2^N，且$H_\infty(U) < 1$，因此我们有

$$\lim_{N\to\infty} \frac{|G|}{2^N} = 0$$

(2) 信源中出现典型序列的概率为$\Pr(G) = 1-\delta$，其中$\delta > 0$为任意小的正整数。当$N \to \infty$时，我们有$\Pr(G) = 1$。

2. 设(X_i, Y_i)为 i.i.d~$p(x, y)$。假设X和Y独立与假设X和Y相关的对数似然比为

$$\frac{1}{n} log \frac{p(X^n)p(Y^n)}{p(X^n, Y^n)}$$

求$n \to \infty$时的极限。

解：

$$\begin{aligned}\frac{1}{n}\log\frac{p(X^n)p(Y^n)}{p(X^n,Y^n)} &= -\frac{1}{n}\sum_{i=1}^{n}\log\frac{p(X_i,Y_i)}{p(X_i)p(Y_i)}\\ &\to -E\left[\log\frac{p(X_i,Y_i)}{p(X_i)p(Y_i)}\right] \qquad \text{依概率}\\ &= -I(X;Y)\end{aligned}$$

3.设随机变量 X 取 4 个值，其概率分布为$\left(\frac{1}{3},\frac{1}{3},\frac{1}{4},\frac{1}{12}\right)$。

（1）请构造此随机变量的 Huffman 码

（2）证明存在两个不同的码字最优长度集，即证明码字长度分配(1,2,3,3)和(2,2,2,2)均是最优的

（3）由此可见，某些最优码的一些字符的相应码长有可能超过香农码的相应码长$\left\lceil log\frac{1}{p(x)}\right\rceil$。

解：

（1）一种可能的编码，

概率	1/3	1/3	1/4	1/12
码字	1	00	010	011

（2）（1,2,3,3）平均码长$L_1=\frac{1}{3}+\frac{1}{3}*2+\frac{1}{4}*3+\frac{1}{12}*3=2$

（2,2,2,2）对应码字为（00,01,11,10），平均码长$L_2=2=L_1$

（3）对概率为1/4的码字而言，按（1）中的最优编码得到码字长3，大于其香农码码长$\left\lceil -\log\frac{1}{4}\right\rceil=2$

一些基本概念

- **信息率：** 一个符号所能传输（或携带）的信息量（需要多少信息量描述），衡量信息传输的有效性

（1）信源：编码后

信息输出率： 编码后传输一个信源符号需要的信息量（一个符号需要多少信息量描述）。用 R 表示（bit/信源符号）。

编码后表示一个信源符号需要的信息量

定长：$\frac{M}{N}\log J$　变长 对字母：$\bar{l}\log J$　N次扩展：$\frac{\bar{l}_N \log J}{N}$

信息输出速率： 单位时间的信息输出率

$$\frac{R}{t}(bit/秒)(输出一个信源符号需要t秒)$$

25

4. 设有一个无记忆信源发出符号 A 和 B，已知$p(A)=\frac{1}{4},p(B)=\frac{3}{4}$。

（1）针对该信源的二次扩展信源，给出香农-费诺编码方案并求编码后信息输出速率；

（2）针对该信源的三次扩展信源，给出霍夫曼编码方案并求编码后信息输出速率；

(1) 二次扩展信源的符号，概率以及 Shannon-Fano 码编码为

符号	AA	AB	BA	BB
概率	1/16	3/16	3/16	9/16
码率	101	100	11	0

（注意，由于 0,1 分布方式不同，可能会有多种编码）

信息输出速率为

$$R=\frac{\bar{l}}{2}=\frac{27}{32}=0.84375\,\text{bits/symbol}$$

4-各种编码方法

Tsinghua University

(2) 类似于上题，可以得到三次扩展信源的符号和概率分布，在这种情况下，可以得到如下的 Huffman 编码树（如图 2 所示）：

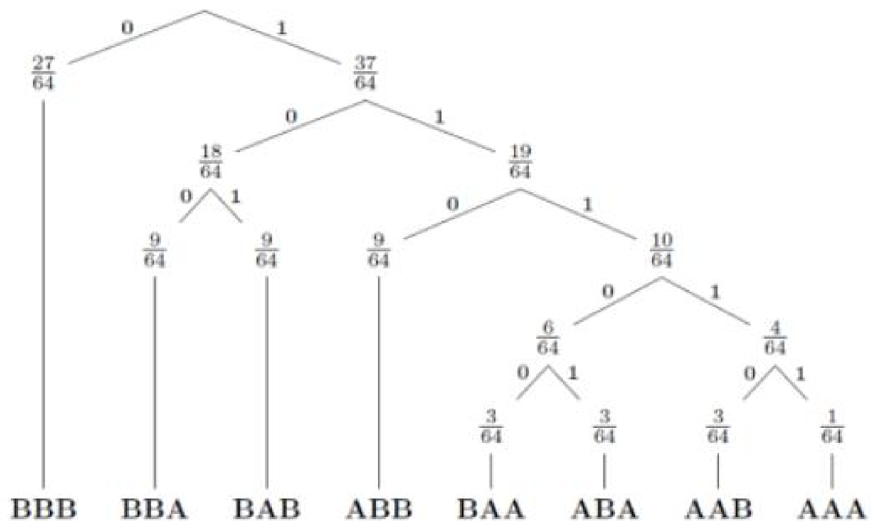


图 2. Huffman 编码树

平均码长

$$\bar{l} = \sum_{i=1}^{8} p(s_i) l_i = \frac{79}{32}$$

信息输出速率

$$R = \frac{\bar{l}}{3} = \frac{79}{96} = 0.8229 \text{ bits/symbol}$$

对(1)(2)题，信息输出速率为R/t，其中t是输出一个信源符号的时间

5. 设信源由一个离散随机变量 X表示，其取值范围是$A=\{a_1,a_2,\ldots,a_K\}$,熵为$H(X)$，若对该信源进行三元变长编码，其前缀码的平均码长为$\bar{l}=\frac{H(X)}{log3}$，试证明：

（1）对于每一个$a_k\in A, p(X=a_k)=3^{-l_k}, k=1,2,\ldots,K$

（2）证明K为奇数

解：（1）题可直接使用书定理5.3.1（ppt3-2 P20）

(1) 令$P(a_1),\cdots,P(a_K)$是各信源符号的概率，$l_1,\cdots,l_K$是码字长度，则有

$$H(\mathbf{X})-\bar{l}\log_2 3=\sum_{k=1}^{K}P(a_k)\log_2\frac{1}{P(a_k)}-\sum_{k=1}^{K}P(a_k)l_k\log_2 3=\sum_{k=1}^{K}P(a_k)\frac{3^{-l_k}}{P(a_k)}$$

根据不等式$\log_2 Z\le(Z-1)\log_2 e$，$Z>0$，有：

$$H(\mathbf{X})-\bar{l}\log_2 3\le\sum_{k=1}^{K}3^{-l_k}-\sum_{k=1}^{K}P(a_k)\le 0$$

Kraft不等式

两边等号同时成立的条件是

$$P(a_k)=3^{-l_k}$$

即当平均码长为$\bar{l}=\frac{H(\mathbf{X})}{\log_2 3}$时，对每一个$a_k\in A$，有$P(a_k)=3^{-l_k}$。

(2) 由上题知，满足 $P(a_k)=3^{-l_k}$ 恰好是 Kraft 不等式取等号的条件。假 K 是偶数，那么进行三元变长编码得到的码树是不完全树，即可以在不违背 Kraft 不等式的情况下，增加一个码字。而这样就与取等号的情形相矛盾（取等号时无法在码树上添加码字），因此 K 为奇数。

另解：由（1）可知

$$\sum_{i=1}^{K} 3^{-l_i} = 1$$

令

$$l_{max} = \max_i l_i$$

则

$$\sum_{i=1}^{K} 3^{l_{max}-l_i} = 3^{l_{max}}$$

显然上式右侧为奇数，而左侧累加中的每一项也为奇数，所以K也必为奇数

6. 对于等概信源

（1）证明：对于一个具有 n 个等概可能输出的信源，任意一个最优的前缀码的码字长度最多相差 1

（2）可变长度码的冗余被定义为$L-H$。一个具有 n 个等可能输出的随机变量，其中$2^m \le n \le 2^{m+1}$，对于最优二元可变长码，求出使得冗余$L-log_2 n$最大的 n 值。当$n\to\infty$时，在这种最坏的情形下，冗余的极限值（提示：0.0861）

6-等概信源的最优编码

(1) 对于一个具 n 个等概可能输出的信源，如果其最大码字长度相差大于 1，那么这个前缀码不是最优的。理由如下:

如图 1，假定在某个前缀码中，码字 c 最长，它和码字 a 的长度之差为 2。根据 Huffman 编码原理，最长码字 c 至少有一个邻居。如果我们可以把码字 c 删去一位，通过如上图的变化，可以减小码字长度。如图所示，在新的编码方案中，至少有两个码字的长度减少了 1，而只有一个码字的长度增加 1。因此，平均码字长度至少减少 $1/n$，这与最优的 Huffman 编码相矛盾，因此码字长度最多相差 1。

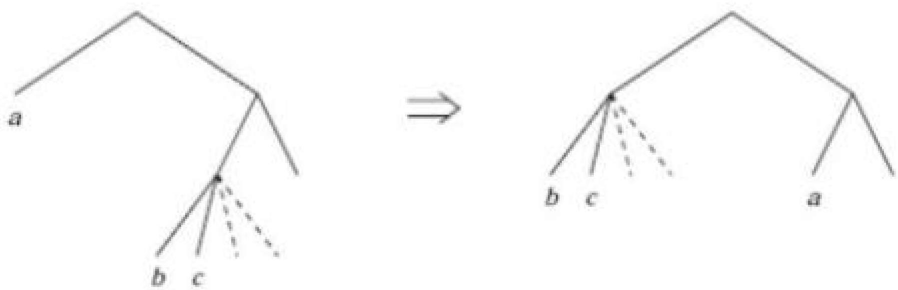


图 1. 码字长度变化

6-等概信源的最优编码

(2) 由第一问知，对于一个具有 n 个等概输出的信源，任意最优二元前缀码的最大码长和最小码长分别是 $\lceil \log_2 n \rceil$ 和 $\lfloor \log_2 n \rfloor$。令

$$d = n - 2^{\lfloor \log_2 n \rfloor}$$

最优码有 $2d$ 个码长为 $\lceil \log_2 n \rceil$，$n-2d$ 个码长为 $\lfloor \log_2 n \rfloor$，这样得到平均码长：

$$L = \frac{1}{n}\left(2d\lceil \log_2 n \rceil + (n-2d)\lfloor \log_2 n \rfloor\right) = \frac{1}{n}\left(n\lfloor \log_2 n \rfloor + 2d\right) = \lfloor \log_2 n \rfloor + \frac{2d}{n}$$

对 $n = 2^m + d$，编码冗余 $r = L - H$ 可写为

$$\begin{aligned} r &= L - H \\ &= \lfloor \log_2 n \rfloor + \frac{2d}{n} - \log_2 n \\ &= m + \frac{2d}{2^m + d} - \log_2\left(2^m + d\right) \\ &= m + \frac{2d}{2^m + d} - \frac{\ln\left(2^m + d\right)}{\ln 2} \end{aligned}$$

因此，令

$$\frac{\partial r}{\partial d} = \frac{2\left(2^m + d\right) - 2d}{\left(2^m + d\right)^2} - \frac{1}{\ln 2}\frac{1}{2^m + d} = 0$$

解得

$$d^* = 2^m\left(2\ln 2 - 1\right)$$

d 取离 d^* 最近的两个整数中使得 r 最大的一个。此时最大的冗余为

$$\begin{aligned} r &\approx m + \frac{2d^*}{2^m + d^*} - \frac{\ln\left(2^m + d^*\right)}{\ln 2} \\ &= m + \frac{2(0.3862)2^m}{2^m + (0.3862)2^m} - \frac{\ln\left(2^m + (0.3862)2^m\right)}{\ln 2} \\ &= 0.0861 \end{aligned}$$

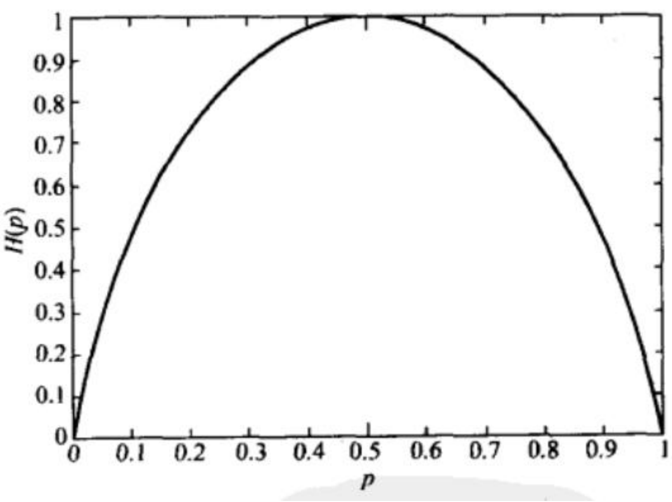
7.信源编码定理表明，随机变量 X 的最优码的期望长度小于$H(X)+1$。请列举出一个随机变量，要求其最优码的期望长度近似等于$H(X)+1$，即对任意$\varepsilon>0$，试构造一个分布，使得其最优码的期望长度满足$L>H(X)+1-\varepsilon$

解：设随机变量X满足，$P(X=1)=p, P(X=0)=1-p$。显然，其最优码的期望长度$L=1$。而当$p\in(0,1)$时，$H(X)\in(0,1]$。
对任意$\varepsilon>0$，存在$p\in(0,1)$，使得

$$H(X)<\varepsilon$$

所以

$$L>H(X)+1-\varepsilon$$



第四章重点总结

- **信道容量**

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

- **例子**
  - 二元对称信道：$C = 1 - H(\mathrm{p})$
  - 二元擦除信道：$C = 1 - \alpha$
  - 对称信道：$C = \log|Y| - \mathrm{H}$ (转移矩阵的行)

- **C的性质**
  - $0 \le C \le \min\{\log|X|, \log|Y|\}$
  - $I(X;Y)$ 是关于$p(x)$的连续凹函数
- **联合典型性**

服从分布$p(x,y)$的联合典型序列$\{(x^n,y^n) \in \chi^n \times \gamma^n\}$的集合$A_\epsilon^{(n)}$为:

$$A_\epsilon^n = \{(x^n,y^n) \in \chi^n \times \gamma^n\}:$$

$$\left|-\frac{1}{n}\log p(x^n) - H(X)\right| < \epsilon\}$$

$$\left|-\frac{1}{n}\log p(y^n) - H(Y)\right| < \epsilon\}$$

$$\left|-\frac{1}{n}\log p(x^n,y^n) - H(X,Y)\right| < \epsilon\}$$

其中$p(x^n,y^n) = \prod_{i=1}^n p(x_i,y_i)$

第四章重点总结

- **联合AEP: 设(X^n, Y^n)为i.i.d.服从分布$p(x^n, y^n) = \prod_{i=1}^{n} p(x_i, y_i)$且长度为$n$的序列，则**
  - $\Pr\left((X^n, Y^n)\right) \in A_\epsilon^{(n)}) \to 1, n \to \infty$
  - $\left|A_\epsilon^{(n)}\right| \leq 2^{n(H(X,Y)+\epsilon)}$
  - 如果$\left(\tilde{X}^n, \tilde{Y}^n\right) \sim p(x^n)p(y^n)$，则$\Pr\left((X^n, Y^n)\right) \in A_\epsilon^{(n)}) \leq 2^{-n(I(X;Y)-3\epsilon)}$
- **信道编码定理**

所有小于信道容量C的码率都是可达的，而所有大于信道容量的码率是不可达的；也就是说，对任意的$R < C$，存在误差概率满足$\lambda^{(n)} \to 0$的一个$(2^{nR}, \mathrm{n})$码序列。反之，如果码率$R > C$，那么λ^n将远离0。

- **反馈容量**

对于离散无记忆信道，反馈并不能增加信道容量，即$C_{FB} = C$

- **最大熵：**

$$\boldsymbol{max}_{\boldsymbol{E\,X^2=\alpha}}\ \mathrm{h(X)} = \frac{1}{2}\log 2\pi e\alpha$$

- AWGN信道：$Y_i = X_i + Z_i$, $Z_i \sim \mathrm{N}(0, N)$，且满足功率限制$\frac{1}{n}\sum_{i=1}^{n} x_i^2 \leq P$

$$C = \frac{1}{2}\log\left(1 + \frac{P}{N}\right) \text{ 比特/传输}$$

- 带宽有限的可加高斯白噪声信道：带宽为W，双边功率谱密度为$\frac{N_0}{2}$，信号功率为P,

$$C = W\log\left(1 + \frac{P}{N_0 W}\right)$$

1：信道容量计算

1.试求以下信道的信道容量

$$Q_1=\begin{bmatrix}1/3 & 1/3 & 1/6 & 1/6\\ 1/6 & 1/3 & 1/6 & 1/3\end{bmatrix},\quad Q_2=\begin{bmatrix}1/2 & 0 & 0 & 1/2\\ 0 & 0 & 0 & 1\\ 1/2 & 1/2 & 0 & 0\\ 0 & 0 & 1 & 0\end{bmatrix}$$

解：

（1）输入等概时达到信道容量$C=\frac{5}{6}-\frac{\log 3}{2}=0.04\ bit$

（2）

假设 $p(x_i)>0(i=1,2,3,4)$，则

由 $Q_3\begin{bmatrix}\beta_1\\ \beta_2\\ \beta_3\\ \beta_4\end{bmatrix}=-\begin{bmatrix}H(Y|a_1)\\ H(Y|a_2)\\ H(Y|a_3)\\ H(Y|a_4)\end{bmatrix}$ 得：$\begin{cases}\beta_1=-2\\ \beta_2=0\\ \beta_3=0\\ \beta_4=0\end{cases}$

$$C=\log_2\left(\sum_{j=1}^{4}2^{\beta_j}\right)=\log_2 13-2bit$$

输出分布为： $p(y_j)=2^{\beta_j-C}$

$$p(y_1)=2^{-\log_2 13}=\frac{1}{13}$$

$$p(y_2)=p(y_3)=p(y_4)=\frac{4}{13}$$

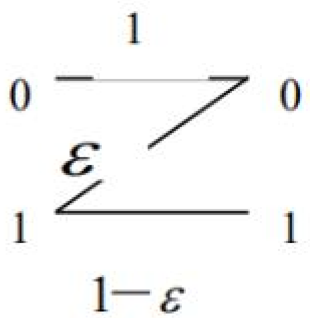
求解发现 $p(x_1)<0$，与 $p(x_i)>0(i=1,2,3,4)$ 矛盾，令 $p(x_1)=0$，显而易见

$$C=\log_2 3\,bit$$

此时的输入分布为： $p(x_1)=0;p(x_2)=p(x_3)=p(x_4)=1/3$

2. Z信道的信道容量

2. Z 信道有二进制的输入和输出，如下所示



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

（1）求 Z 信道的最大的输入的 pdf。提示$\frac{dH(x)}{dx}=ln\frac{1-x}{x}$

（2）当$\varepsilon=0.5$时，求 Z 信道的信道容量。

2. Z信道的信道容量

(1)设$Pr(X=1)=\alpha$，则Z信道的输入输出的互信息是关于α的函数。

$$H(Y|X)=\sum_x p(x)H(Y|X=x)=\alpha H(\epsilon)$$
$$H(Y)=H(\alpha(1-\epsilon))$$

因此

$$I(X;Y)=H(Y)-H(Y|X)=H(\alpha(1-\epsilon))-\alpha H(\epsilon)$$

令

$$\frac{\partial I(X;Y)}{\partial\alpha}=0\Rightarrow(1-\epsilon)\log_2\frac{1-\alpha(1-\epsilon)}{\alpha(1-\epsilon)}=H(\epsilon)$$

解得

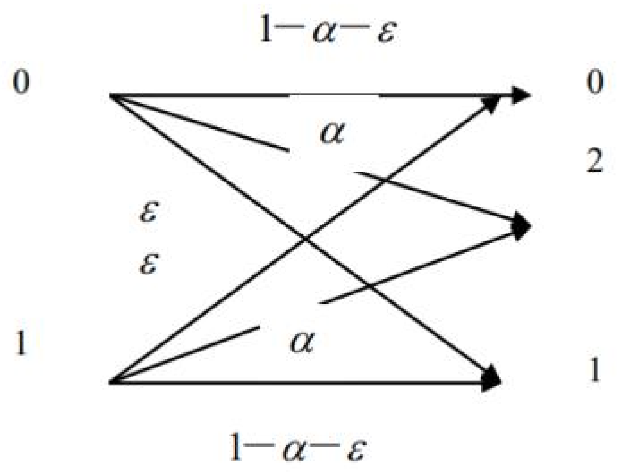
$$\alpha=\frac{\epsilon^{\frac{\epsilon}{1-\epsilon}}}{1+(1-\epsilon)\epsilon^{\frac{\epsilon}{1-\epsilon}}}$$

(2)由第一问知，当$\epsilon=0.5$时，$\alpha=2/5$时达到信道容量，此时信道容量为

$$C=H(\alpha/2)-\alpha=H(1/5)-2/5=0.322\ bit$$

3-对称删除信道的信道容量

3. 求下面二进制对称删除信道的容量，其中错误概率为ε，删除概率为α



$$p(y\mid x)=\begin{bmatrix}1-\alpha-\varepsilon & \varepsilon & \alpha \\ \varepsilon & 1-\alpha-\varepsilon & \alpha\end{bmatrix}$$

$$x\in\{0,1\}$$
$$y\in\{0,1,2\}$$

3-对称删除信道的信道容量

Tsinghua University

假定$p(0)=p$和$p(1)=1-p$是使得$I(X;Y)$最大化的一个分布。由于信道对符号0和1是对称的，那么分布$p(0)=1-p$和$p(1)=p$也是互信息最大化的分布。又因为$I(X;Y)$是关于输入分布$p(x)$的凸函数，因此平均分布$p(0)=p(1)=1/2$也是最大化的分布。因此

$$
\begin{aligned}
C &= \sum_x \sum_y p(x,y)\log\frac{p(x,y)}{p(x)p(y)} \\
&= 2\sum_y p(0,y)\log\frac{2p(0,y)}{p(y)} \\
&= 2[\frac{1}{2}(1-\alpha-\epsilon)\log(\frac{2(1-\alpha-\epsilon)}{1-\alpha}) + \frac{1}{2}\alpha\log 1 + \frac{1}{2}\epsilon\log(\frac{2\epsilon}{1-\alpha})] \\
&= (1-\alpha) - (1-\alpha)\log(1-\alpha) + (1-\alpha-\epsilon)\log(1-\alpha-\epsilon) + \epsilon\log\epsilon
\end{aligned}
$$

4. 二元离散对称无记忆信道的容量为$C = 0.5\text{bit/symbol}$，现按香农随机编码的方法对每一消息$m(m = 0,1,2,\dots,M-1)$给以相应的长为N的码字$C_m = (c_{m1}, c_{m2}, \dots, c_{mN})$。全部$M = 2^{NR}$个码字组成一分组码。现设$N = 10^{10}$，$R = 0.5b/s$，设信道输入和输出分别用随机变量$X$和$Y$表示。试求在输出序列中与某一特定输入码字构成联合典型的序列的数目为。

解：

二元对称信道的信道容量为

$$C = \max_X \left(H(Y) - H(Y|X) \right) = 1 - H(Y|X)$$

由$C = 0.5$知$H(Y|X) = 0.5$。又因为$N \to \infty$时，典型序列的个数为$|G| = 2^{NH(Y|X)}$，

代入参数可知$|G| = 2^{5\times10^9}$。

5-和信道的信道容量

5.和信道：设$(X_1, p(y|x), Y_1)$和$(X_2, p(y|x), Y_2)$均为离散无记忆信道，信道容量分别为C_1和C_2。输入字母表X_1和X_2是不同的，输出字母表Y_1和Y_2也是不同的。一个新的信道，输入$x \in X_1 \cup X_2$，输出$y \in Y_1$或$y \in Y_2$，构造了一个输入字母表为$X_1 \cup X_2$，输出字母表为$Y_1 \cup Y_2$的信道，求其信道容量。

条件互信息

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

解:

利用和信道传输信息时，假设每次只使用其中一条信道发送，设随机变量θ

$$\theta=\begin{cases}1, & \text{以概率}\alpha\text{通过信道1发送}\\ 2, & \text{以概率}1-\alpha\text{通过信道2发送}\end{cases}$$

由于两个信道的输入字母表不同，输出字母表也不同，因此θ可由X或Y唯一确定

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

5-和信道的信道容量

因此，信道容量为

$$C=\max I(X;Y)=\max\left(H(\theta)+\alpha I(X_1;Y_1)+(1-\alpha)I(X_2;Y_2)\right)$$

C关于α严格凸，因此令$\frac{\partial C}{\partial \alpha}=0$，得

$$\alpha=\frac{1}{1+2^{C_2-C_1}}$$

代入得到信道容量为

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

给出一个直观解释。如果$M=2^C$是有效无噪符号数，那么$M_1=2^{C_1}$和$M_2=2^{C_2}$分别是信道 1 和信道 2 的有效无噪符号数。因为每次发送我们只利用一个信道发送，因此对新信道有$M_1+M_2=2^{C_1}+2^{C_2}$个无噪符号数。因此信道容量为$C=\log_2\left(2^{C_1}+2^{C_2}\right)$。

6. 设二进制对称信道$C_1, C_2, \ldots, C_m$，所有信道的错误转移概率都为$\varepsilon(0 < \varepsilon < 1)$，并且构成级联信道，即$C_i$的输出是$C_{i+1}$的输入$(i = 1,2, \ldots, m-1)$。设信道$C_1$输入0的概率为$p_0$，在不引入任何编译码机制的基础上，试求

（1）在C_1的输入端给定任何概率分布下C_m的输出端出现0的概率

（2）该级联信道的容量C以及$\lim_{m \to +\infty} C$

6-级联信道

设一个 BSC 的转移概率矩阵是

$$A=\begin{bmatrix}1-p & p\\ p & 1-p\end{bmatrix}$$

那么 m 个 BSC 级联的转移概率矩阵是

$$A_m=A^m$$

注意到

$$A=T^{-1}\begin{bmatrix}1 & 0\\ 0 & 1-2p\end{bmatrix}T$$

其中

$$T=\begin{bmatrix}1 & 1\\ 1 & -1\end{bmatrix}$$

因此

$$\begin{aligned}A_m&=A^m\\ &=T^{-1}\begin{bmatrix}1 & 0\\ 0 & (1-2p)^m\end{bmatrix}T\\ &=\begin{bmatrix}\frac{1}{2}\left(1+(1-2p)^m\right) & \frac{1}{2}\left(1-(1-2p)^m\right)\\ \frac{1}{2}\left(1-(1-2p)^m\right) & \frac{1}{2}\left(1+(1-2p)^m\right)\end{bmatrix}\end{aligned}$$

显然，m 个 BSC 级联的信道仍是一个 BSC 信道，其差错概率为

$$p_m=\frac{1}{2}\left(1-(1-2p)^m\right)$$

(1) 出现 0 的概率

$$p_0(1-p_m)+(1-p_0)p_m=\frac{1}{2}+\left(p_0-\frac{1}{2}\right)(1-2p)^m$$

(2) 该级联信道的信道容量

$$C_m=1-H\left(\frac{1}{2}\left(1+(1-2p)^m\right)\right)$$

而 $1-2p\le 1$，因此

$$\lim_{m\to\infty}C_m=1-H\left(\frac{1}{2}\right)=0$$

7.考虑一个$x, y \in \{0,1,2,3\}$的信道，转移概率矩阵$p(y|x)$由下列矩阵给出：

$$\begin{bmatrix} 0.5 & 0.5 & 0 & 0 \\ 0 & 0.5 & 0.5 & 0 \\ 0 & 0 & 0.5 & 0.5 \\ 0.5 & 0 & 0 & 0.5 \end{bmatrix}$$

（1）求该信道的信道容量

（2）定义随机变量$z = g(y)$，其中

$$g(y) = \begin{cases} A, & if\ y \in \{0,1\} \\ B, & if\ y \in \{2,3\} \end{cases}$$

对以下的概率分布，计算$I(X;Z)$: $p(x) = \begin{cases} \frac{1}{2}, & if\ x \in \{1,3\} \\ 0, & if\ x \in \{0,2\} \end{cases}$

（3）求x和z之间的信道容量

（4）对于（2）中的分布，证明或证伪下述命题

“$X \to Z \to Y$形成一个马尔科夫链”

清華大學 Tsinghua University

(1) 对称信道的信道容量在输入等概时达到，此时输出也等概，此时信道容量

$$C = \log 4 - \log 2 = 1 \ \ bit$$

(2) $I(X;Z) = 0$，可通过X与Z的联合概率分布求出。其联合分布为

X \ Z	A	B
0	0	0
1	1/4	1/4
2	0	0
3	1/4	1/4

(3) $C = 1$ bit，可通过X和Z的转移概率密度矩阵求得

$$Q = \begin{bmatrix} 1 & 0 \\ 1/2 & 1/2 \\ 0 & 1 \\ 1/2 & 1/2 \end{bmatrix}$$

达到信道容量时输入分布为$p(x=0) = p(x=2) = 1/2$。

(4) $X \to Z \to Y$ 不构成马尔科夫链。因为其不满足$I(X;Y) \le I(X;Z)$

8. 设离散无记忆信道的输入和输出皆为 r 个符号，每一个输入符号对应 r 个输出符号，其传输概率为概率组$p_1, p_2, \ldots, p_r$的一个置换，并且每一个输出符号对应的转移概率的和是相同的。试证明上述信道的信道容量为：

$$logr + \sum_{j=1}^{r} p_j logp_j$$

解：设输入输出为X, Y，则

$$I(X;Y) = H(Y) - H(Y|X) = H(Y) - \sum_{j=1}^{r} p(x_j) H(Y|X = x_j) = H(Y) - H(p_1, p_2, \ldots, p_j)$$

$$\leq logr + \sum_{j=1}^{r} p_j logp_j$$

等号成立当且仅当输出Y均匀分布时，且令输入X也为均匀分布时可以做到，这是由于

$$p(y) = \sum_{j=1}^{r} p(x_j) p(y|x_j) = \frac{1}{r} \sum_{j=1}^{r} p(y|x_j) = \frac{1}{r}$$

清華大學 Tsinghua University

9. 等噪声信道：一个无记忆信道有离散的输入X和连续的输出$Y = X + Z$，其中Z在区间$[-a, +a]$上服从均匀分布。

（1）当$X \in \{-1, +1\}$等概时，求$I(X;Y)$（为a的函数）

（2）当$a = 1/2$时，在区间$[-1, +1]$上选择X的离散值和他们的概率分布，使得$I(X;Y)$最大

（3）当$\frac{1}{2} < a < 1$时，求限幅信道（$|X| \le 1$）的容量

(1) 当 $\alpha \le 1$ 时，由输入等概可得输出 Y 的分布为

$$f_Y(y)=\begin{cases}\frac{1}{4a}, & -1-a\le y\le -1+a\\ \frac{1}{4a}, & 1-a\le y\le 1+a\end{cases}$$

容易得到

$$h(Y)=-\int f_Y(Y)\log_2 f_Y(Y)dy=2+\log_2 a$$

而 $h(Y|X)=h(Z)=1+\log_2 a$，因此 $I(X;Y)=h(Y)-h(Y|X)=1$。

当 $\alpha>1$ 时，由输入等概可得输出 Y 的分布为

$$f_Y(y)=\begin{cases}\frac{1}{4a}, & -1-a\le y\le 1-a\\ \frac{1}{2a}, & 1-a\le y\le a-1\\ \frac{1}{4a}, & a-1\le y\le 1+a\end{cases}$$

此时，容易得到

$$h(Y)=-\int f_Y(Y)\log_2 f_Y(Y)dy=\frac{1}{a}+1+\log_2 a$$

因此 $I(X;Y)=h(Y)-h(Y|X)=\frac{1}{a}$。

换个角度

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

当 $a\le 1$ 时，由 Y 可以唯一确定 X,

$$I(X;Y)=H(X)=1$$

当 $a>1$ 时，有重叠

$$H(X|Y)=P(Y\text{处于重叠区})$$

$$=\frac{2a-2}{2a}=1-\frac{1}{a}$$

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

$$=1-\left(1-\frac{1}{a}\right)=\frac{1}{a}$$

（2）

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

而

$$h(Z)=1+\log_2 a=0$$

所以

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

又

$$1-a\leq Y\leq 1+a$$

即$Y\in\left[-\frac{3}{2},+\frac{3}{2}\right]$

所以

$$h(Y)\leq\log_2 3$$

给定区间无约束，
均匀分布熵最大。

等号取到当且仅当Y在$\left[-\frac{3}{2},+\frac{3}{2}\right]$上均匀分布时

易得，选择X在$\{-1,0,1\}$三点，且概率均为1/3 时，满足上述条件，此时

$$C=\max I(X;Y)=\max h(Y)=\log_2 3$$

9-等噪声信道

(3) 当$0.5<a\le 1$，信道容量在$\log_2 2$与$\log_2 3$之间，也就是$a=1$和$a=1/2$的容量。第一个猜想是X取值为$\{-1,0,+1\}$时，可以达到信道容量。但是，如果X取值为$\{-1,0,+1\}$，那么输入值 0 的可靠性不如其他输入值，因为它的输出区域有更多的重叠部分，因此取值为 0 的概率要小一些。当$0.5<a\le 1$，能使$I(X;Y)$最大的X的取值是$\{-1,1-2a,-1+2a,+1\}$，最优概率分布是$\{1/3,1/6,1/6,1/3\}$。此时，输出Y的分布为

$$f_Y(y)=\begin{cases}\dfrac{1}{6a}, & -1-a\le y\le 1-3a\\ \dfrac{1}{4a}, & 1-3a\le y\le -1+a\\ \dfrac{1}{6a}, & -1+a\le y\le 1-a\\ \dfrac{1}{4a}, & 1-a\le y\le -1+3a\\ \dfrac{1}{6a}, & -1+3a\le y\le 1+a\end{cases}$$

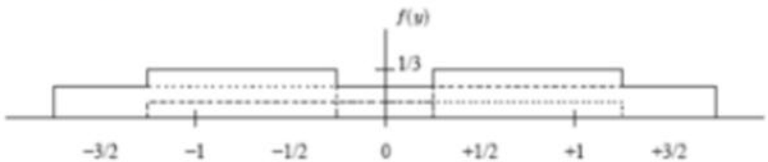
求得输出微分熵为

$$h(Y)=\frac{1-a}{a}\log_2 6a+\frac{2a-1}{a}\log_2 4a$$

得到，对$0.5<a\le 1$

$$I(X;Y)=h(Y)-h(Y\mid X)=\left(\frac{1}{a}-1\right)\log_2 3+\left(2-\frac{1}{a}\right)$$

例如，当$a=0.75$，四个输出值是± 1，± 0.5。$f(y)$的密度如图 1 所示，$f(y)$只有两个取值$2/9$和$1/3$。



$a=0.75$时的微分熵为

$$h(Y)=-\frac{3}{2}\cdot\frac{2}{9}\log_2\frac{2}{9}-2\cdot\frac{1}{3}\log_2\frac{1}{3}=1.7800\text{ bits}$$

$$I(X;Y)=h(Y)-h(Y\mid X)=h(Y)-\log_2\frac{3}{2}=1.0950\text{ bits}$$

10. 设两个n长离散随机矢量x^n和y^n构成随机序列对(x^n, y^n)，其联合分布$p(x^n, y^n)$满足：$p(x^n, y^n) = \prod_{i=1}^{n} p(x_i, y_i)$。当$x^n$和$y^n$构成联合典型序列时，试证明，对于任意正实数$\varepsilon$，存在足够大的正整数$N$，当$n > N$时，满足

$$2^{-n[H(Y|X)+\varepsilon]} < p(y^n|x^n) < 2^{-n[H(Y|X)-\varepsilon]}$$

解

由典型序列和联合典型序列的定义以及大数定律，对于任意的$\delta = \varepsilon/2 > 0$，存在$N_1$、$N_2$，使得当$n > N_1$时，有$|\frac{1}{n}logp(x^n) + H(X)| < \delta$，当$n > N_2$时，有$|\frac{1}{n}logp(x^n y^n) + H(XY)| < \delta$。令$N = max\{N_1，N_2\}$，即对于任意的$\delta = \varepsilon/2 > 0$，存在$N$，使得当$n > N$时，有

$$2^{-n(H(X)+\delta)} < p(x^n) < 2^{-n(H(X)-\delta)}$$
$$2^{-n(H(XY)+\delta)} < p(x^n y^n) < 2^{-n(H(XY)-\delta)}$$

而$p(y^n|x^n) = p(x^n y^n)/p(x^n)$，因此有

$$2^{-n(H(XY)+\delta)}2^{n(H(X)-\delta)} < p(y^n|x^n) < 2^{-n(H(XY)-\delta)}2^{n(H(X)+\delta)}$$
$$2^{-n(H(XY)-H(X)+2\delta)} < p(y^n|x^n) < 2^{-n(H(XY)-H(X)-2\delta)}$$
$$2^{-n(H(Y|X)+\varepsilon)} < p(y^n|x^n) < 2^{-n(H(Y|X)-\varepsilon)}$$

11- Fano不等式的证明

清華大學 Tsinghua University

11.设 X 和 Y 为信道的输入和输出，均取值于集合$a=\{a_1,a_2,\dots,a_K\}$。已知 $p(x=a_k)=p_k$，$p(y=a_j|x=a_k)=p_{kj}$，定义$P_e=\sum_k p_k\sum_{j\neq k}p_{kj}$，求证：

$$H(X|Y)\le P_e\,log(K-1)+H(P_e)$$

解：令误差随机变量

$$E=\begin{cases}1, & X\neq Y\\ 0, & X=Y\end{cases}$$

则由熵的链式法则，可将$H(E,X|Y)$拆成两种形式

$$\begin{aligned}H(E,X|Y)&=H(X|Y)+H(E|X,Y)\\&=H(E|Y)+H(X|E,Y)\end{aligned}$$

一方面，由于E是X,Y的函数，所以$H(E|X,Y)=0$

另一方面

$$H(E|Y)\le H(E)=H(P_e)$$

$$\begin{aligned}H(X|E,Y)&=P(E=0)H(X|Y,E=0)+P(E=1)H(X|Y,E=1)\\&\le(1-P_e)*0+P_e\log(K-1)\end{aligned}$$

综上

$$H(X|Y)\le P_e\log(K-1)+H(P_e)$$

附件：拉格朗日乘子法

Tsinghua University

- 引进一个新函数

$$\Phi = I(X;Y) - \lambda\left[\sum_{i=1}^{n} p(x_i) - 1\right] \qquad (4.2.20)$$

其中 λ 为拉格朗日乘子，解方程组

$$\frac{\partial \Phi}{\partial p(x_i)} = \frac{\partial\left\{I(X;Y) - \lambda\left[\sum_{i=1}^{n} p(x_i) - 1\right]\right\}}{\partial p(x_i)} = 0 \qquad (4.2.21)$$

可得一般信道容量C。

由$p(y_j) = \sum_{i=1}^{n} p(x_i)p(y_j / x_i)$ (4.2.22) 有 $\frac{dp(y_j)}{dp(x_i)} = p(y_j / x_i)$

附件：拉格朗日乘子法

清華大學 Tsinghua University

- 将$I(X;Y)$的表达式代入(4.2.21)得

$$\frac{\partial \Phi}{\partial p(x_i)} = \left\{ -\sum_{j=1}^{m} p(y_j)\log_2 p(y_j) + \sum_{i=1}^{n}\sum_{j=1}^{m} p(x_i)p(y_j/x_i)\log_2 p(y_j/x_i) - \lambda\left[\sum_{i=1}^{n} p(x_i) - 1\right]\right\} = 0$$

求偏导得

$$-\sum_{j=1}^{m}\left[p(y_j/x_i)\log_2 p(y_j) + p(y_j/x_i)\log_2 e\right] + \sum_{j=1}^{m} p(y_j/x_i)\log_2 p(y_j/x_i) - \lambda = 0$$

- 整理得

$$\sum_{j=1}^{m} p(y_j/x_i)\log_2 \frac{p(y_j/x_i)}{p(y_j)} = \log_2 e + \lambda \qquad (4.2.23)$$

其中 $\sum_{j=1}^{m} p(y_j/x_i) = 1$ 　上式两边乘以$p(x_i)$并求和

$$\sum_{i=1}^{n}\sum_{j=1}^{m} p(x_i)p(y_j/x_i)\log_2 \frac{p(y_j/x_i)}{p(y_j)} = \sum_{i=1}^{n} p(x_i)(\log_2 e + \lambda) = \log_2 e + \lambda \qquad (4.2.24)$$

附件：拉格朗日乘子法

Tsinghua University

- 式(4.2.24)左边为平均互信息的极大值，即

$$C=\log_2 e+\lambda \qquad (4.2.25)$$

式(4.2.25)代入(4.2.23)得

$$\sum_{j=1}^{m} p(y_j/x_i)\log_2 p(y_j/x_i)=\sum_{j=1}^{m} p(y_j/x_i)\log_2 p(y_j)+C$$

$$=\sum_{j=1}^{m} p(y_j/x_i)[\log_2 p(y_j)+C]$$

令 $\beta_j=\log_2 p(y_j)+C \qquad (4.2.26)$

则 $\sum_{j=1}^{m} p(y_j/x_i)\log_2 p(y_j/x_i)=\sum_{j=1}^{m} p(y_j/x_i)\beta_j \qquad (4.2.27)$

由(4.2.27)和信道矩阵求出β_j，再由(4.2.26)得

$$p(y_j)=2^{\beta_j-C} \qquad (4.2.28)$$

附件：拉格朗日乘子法

上式两边对j求和得$\sum_{j=1}^{m} p(y_j) = \sum_{j=1}^{m} 2^{\beta_j} \cdot 2^{-C} = 1, \quad 2^C = \sum_{j=1}^{m} 2^{\beta_j}$

求出信道容量 $C = \log_2\left(\sum_{j=1}^{m} 2^{\beta_j}\right)$ (4.2.29)

再根据(4.2.22)求出对应对输入概率分布$p(x_i)$。

$$p(y_j) = \sum_{i=1}^{n} p(x_i) p(y_j / x_i) \qquad (4.2.22)$$

附件：拉格朗日乘子法

Tsinghua University

- 一般离散信道容量对计算步骤总结如下：

①由$\sum_{j=1}^{m} p(y_j/x_i)\beta_j = \sum_{j=1}^{m} p(y_j/x_i)\log_2 p(y_j/x_i)$，求$\beta_j$；

②由$C=\log_2\left(\sum_{j=1}^{m} 2^{\beta_j}\right)$，求$C$；

③由$p(y_j)=2^{\beta_j - C}$，求$p(y_j)$；

④由$p(y_j)=\sum_{i=1}^{n} p(x_i)p(y_j/x_i)$,求$p(x_i)$。

附件：拉格朗日乘子法

- **注意：**
  - 在第②步信道容量C被求出后，计算并没有结束，必须解出相应的$p(x_i)$，并确认所有的$p(x_i)\geqslant 0$时，所求的C才存在。
  - 在对$I(X;Y)$求偏导时，仅限制 $\sum_{i=1}^{n} p(x_i)=1$ ，并没有限制$p(x_i)\geqslant 0$，所以求出的$p(x_i)$有可能为负值，此时C就不存在，必须对$p(x_i)$进行调整，再重新求解C。
  - 近年来人们一般采用计算机，运用迭代算法求解。