《信息、控制与计算》

竝

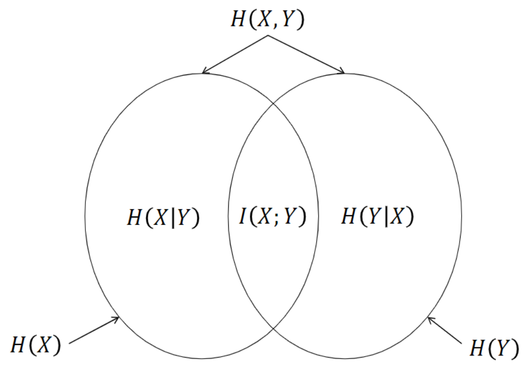
Chapter 1 熵和互信息

- 事件

	表达式	性质
自信息	$I(x)=-\log p(x)$	/
条件自信息	$I(x\|y)=-\log p(x\|y)=I(x,y)-I(y)$	差集
联合自信息	$I(x,y)=-\log p(x,y)$	并集
互信息	$I(x;y)=\log\frac{p(x\|y)}{p(x)}=\log\frac{p(x,y)}{p(x)p(y)}=I(x)-I(x\|y)=I(x)+I(y)-I(x,y)$	交集，有对称性
条件互信息	$I(x;y\|z)=\log\frac{p(x,y\|z)}{p(x\|z)p(y\|z)}=I(x\|z)+I(y\|z)-I(x,y\|z)$	差集的交集
联合互信息	$I(x;y,z)=\log\frac{p(x\|y,z)}{p(x)}=I(x;y)+I(x;z\|y)=I(y;x)+I(z;x\|y)=I(y,z;x)$	并集的交集

- 离散随机变量

> 熵是随机变量的平均自信息，代表了平均不确定值



	表达式	性质
平均自信息（熵）	$H(X)=E[I(x)]=\sum_k p(x_k)I(x_k)=-\sum_k p(x_k)\log p(x_k)$	/
条件熵	$H(X\|Y)=E[I(X\|y)]=\underline{\sum_y p(Y=y)H(X\|Y=y)}=-\sum_x\sum_y p(x,y)\log p(x\|y)$	差集
联合熵	$H(X,Y)=-\sum_x\sum_y p(x,y)\log p(x,y)=H(X\|Y)+H(Y)=H(Y\|X)+H(X)$ $=H(X)+H(Y)-I(X;Y)$	并集
平均互信息	$I(X;Y)=E[I(x;y)]=\sum_x\sum_y p(x,y)\log\frac{p(x\|y)}{p(x)}=\sum_x\sum_y p(x,y)\log\frac{p(x,y)}{p(x)p(y)}$ $=H(X)-H(X\|Y)=H(Y)-H(Y\|X)$	交集，有对称性
条件互信息	$I(X;Y\|Z)=E[I(x;y\|z)]=\sum_x\sum_y\sum_z p(x,y,z)\log\frac{p(x,y\|z)}{p(x\|z)p(y\|z)}$	差集的交集
联合互信息	$I(X;Y,Z)=E[I(x;y,z)]=\sum_x\sum_y\sum_z p(x,y,z)\log\frac{p(x\|y,z)}{p(x)}=I(X;Z)+I(X;Y\|Z)$	并集的交集

- 概率分布的散度——相对熵

  > 令 $\{p(x)\}$ 与 $\{q(x)\}$ 是同一字符表 $\mathcal{X}$ 上的两个概率分布，相对熵表示的是实际分布 $\{p(x)\}$ 与假定分布 $\{q(x)\}$ 间的平均差距，又称“鉴别熵”。

  $D(p//q)=E_p\{\log\frac{p(x)}{q(x)}\}=\sum_x p(x)\log\frac{p(x)}{q(x)}$

  性质：（1）$D(p//q)\geq 0$（2）$D(p//q)\neq D(q//p)$　（3）$I(X;Y)=\sum_x p(x,y)\log\frac{p(x,y)}{p(x)p(y)}=D(p(x,y)//p(x)p(y))\geq 0$

- 关于疑义度的 **Fano** 不等式

  把 $\hat{X}$ 看成是对 X 的估计，定义错误概率为 $P_E=\sum_{k=0}^{K-1}\sum_{\substack{j=0\\k\neq j}}^{K-1}P\{X=k,\hat{X}=j\}$

  于是关于疑义度 $H(X|\hat{X})$ 有如下不等式：$H(X|\hat{X})\leq H(P_E)+P_E\log(K-1)$

  物理意义：已知 $\hat{X}$ 条件下，X 的不确定性。若 $\hat{X}$ 正确，其不确定性为 $H(P_E)$；若 $\hat{X}$ 不正确，此时 X 的取值范围为剩下 $K-1$ 个值，这部分不确定性不会大于 $P_E\log(K-1)$

- 连续随机变量

随机变量 X,Y 的概率密度为 $f_X(x),f_Y(y)$，联合概率密度为 $f_{XY}(x,y)$ 有

$$\begin{cases} f_X(x)=\int_{-\infty}^{\infty}f_{XY}(x,y)dy \\ f_Y(y)=\int_{-\infty}^{\infty}f_{XY}(x,y)dx \\ f_{XY}(x,y)=f_X(x)f_{Y|X}(y|x)=f_Y(y)f_{X|Y}(x|y) \end{cases}$$

	表达式	性质
互信息	$I(X;Y)=\iint f_{XY}(x,y)\log\frac{f_{X\|Y}(x\|y)}{f_X(x)}dxdy=\iint f_{XY}(x,y)\log\frac{f_{XY}(x,y)}{f_X(x)f_Y(y)}dxdy$ $=H_c(X)-H_c(Y\|X)=H_c(Y)-H_c(X\|Y)=H_c(X)+H_c(Y)-H_c(X,Y)$	/
条件互信息	$I(X;Y\|Z)=\iiint f_{XYZ}(x,y,z)\log\frac{f_{XY\|Z}(x,y\|z)}{f_{X\|Z}(x\|z)f_{Y\|Z}(y\|z)}dxdydz$	差集
联合互信息	$I(X;Y,Z)=\iiint f_{XYZ}(x,y,z)\log\frac{f_{XYZ}(x,y,z)}{f_{X,Y}(x,y)f_Z(z)}dxdydz$ $=I(X;Y)+I(X;Z\|Y)=I(X;Z)+I(X;Y\|Z)$	并集
微分熵	$H_c(X)=-\int f_X(x)\log f_X(x)dx$	连续随机变量离散化，不具有线性变换不变性，表征相对不确定性，可正可负。
条件微分熵	$H_c(X\|Y)=-\iint f_{XY}(x,y)\log f_{X\|Y}(x\|y)dxdy$	
联合微分熵	$H_c(X,Y)=-\iint f_{XY}(x,y)\log f_{XY}(x,y)dxdy=H_c(X)+H_c(Y\|X)$	

- 微分熵极大化

  峰值受限：$if\ X\in[-M,M]\quad then\ H_c(X)\le ln(2M)$

  平均功率受限：$if\ \sigma^2=const, X\sim N(\mu,\sigma^2)\quad then\ H_c(X)\le ln(\sqrt{2\pi e}\sigma)=\frac{1}{2}ln(2\pi e\sigma^2)$

  熵功率：$\bar{\sigma_x^2}=\frac{1}{2\pi e}e^{2H_c(x)}\le\sigma_x^2$

  $e^{2H_c(X+Y)}\le e^{2H_c(X)}+e^{2H_c(Y)}$

- 平稳信源的熵及其性质

  设信源发出序列为 N 维随机矢量 $\mathbf{X}=(X_1,X_2,...,X_N)$，则信源熵表示为：

  $H(\mathbf{X})=-\sum p(x_{i_1},x_{i_2}\dots,x_{i_N})\log p(x_{i_1},x_{i_2}\dots,x_{i_N})$

平均每符号熵：$H_N(\mathbf{X}) \triangleq \frac{1}{N}H(\mathbf{X}) = \frac{1}{N}H(X_1, X_2 \ldots, X_N)$

熵速率：$H_\infty(\mathbf{X}) = \lim_{N\to\infty} H_N(\mathbf{X})$ 即最小平均符号熵。

平均条件熵：$H(X_N|X_{N-1}, X_{N-2}\ldots, X_1)$

熵的相对率：$\eta = \frac{H_\infty}{H_0} = \frac{H_\infty}{\log K}$

信源冗余度：$R = 1-\eta$

性质：

（1）平均每符号熵、平均条件熵随 N 增大单调不增。

（2）$H_N(\mathbf{X}) \geq H(X_N|X_{N-1}, X_{N-2}\ldots, X_1)$

（3）$\lim_{N\to\infty} H_N(\mathbf{X}) = \lim_{N\to\infty} H(X_N|X_{N-1}, X_{N-2}\ldots, X_1)$

（4）对于无记忆离散源，当 X 在取值范围内等概率分布时熵最大，且有
$H_\infty \leq \cdots \leq H_1 \leq H_0 = \log K$

Chapter 2 离散无记忆信源的无损编码

- 离散无记忆信道的等长编码
  - **等长编码**

    设信源输出长度为 L 的消息序列，字符表含 K 个字符，则共有 K^L 种不同序列。用长度为 N，包含 D 个字符的编码字符序列来进行无差错等长编码，则要求 $D^N \geq K^L$，即 $N \geq \frac{L\log K}{\log D}$。（所有可能的编码序列数比消息序列数大）

    （1）平均码长：$\bar{N} = \frac{N}{L} \geq \log K \overset{\text{十进制}}{\underset{K=10}{=}} 3.322\ bit$

    （2）当等概率输出序列字符时，信源熵最大，为 $\log K$

    （3）编码速率：$R = \frac{N}{L}\log D \geq \log K$
  - **等长编码定理**

    令离散无记忆信源熵为 $H(U)$，若 $N > L[H(U)+\varepsilon_L]/\log D$，则可实现无损编码。
  - **渐近等分性质 AEP**

    对于 DMS（离散无记忆信源）而言，$\lim_{L\to\infty} I_L = \lim_{L\to\infty} \frac{I(u^L)}{L} \overset{P=1}{\longrightarrow} H(U)$ 表示平均每符号自信息以概率 1 趋于熵

    其中，u^L, L 分别为输出序列及其长度，$I(u^L) = \sum I(u_l)$ 表示序列自信息等于每个字符的自信息之和（由于无记忆性）
  - **典型列**

    $A_\varepsilon^{(L)}(U) = \{u^L; H(U)-\varepsilon \leq I_L \leq H(U)+\varepsilon\}$，$\varepsilon > 0$

    性质：

    （1）L 足够大时，$P(u^L \in A_\varepsilon^{(L)}(U)) > 1-\varepsilon$

    （2）若 $u^L \in A_\varepsilon^{(L)}(U)$，则有 $2^{-L[H(U)+\varepsilon]} \leq p(u^L) \leq 2^{-L[H(U)-\varepsilon]}$，即每个序列出现概率 $p(u^L) \approx 2^{-LH(U)}$

    （3）典型列数目 $|A_\varepsilon^{(L)}(U)|$ 满足 $(1-\varepsilon)2^{L[H(U)-\varepsilon]} \leq |A_\varepsilon^{(L)}(U)| \leq 2^{L[H(U)-\varepsilon]}$，即 $|A_\varepsilon^{(L)}(U)| \approx 2^{LH(U)}$
- 离散无记忆信道的不等长编码

  码唯一可译：后缀分解集中不含码字

  码唯一可译且即时：第一个后缀分解集为空集，即 $S_1 = \emptyset$
  - **Kraft 不等式**

存在长度为 $n_1, n_2, \ldots, n_k$ 的 D 元异字头码的充要条件为：$\sum\limits_{k=1}^{K} D^{-n_k} \le 1$

一个码唯一可译 $\Longleftrightarrow$ Kraft 不等式成立 $\Longleftrightarrow$ 存在一个具有同样长度的异字头码

- 不等长编码定理

  平均码长：$\bar{n} = \sum p_k \cdot n_k$

  任何一个唯一可译码的平均码字长度必须满足：$\bar{n} \ge \frac{H(U)}{\log D}$

  同时一定存在一个D元唯一可译码，其平均码长满足：$\bar{n} \le \frac{H(U)}{\log D} + 1$

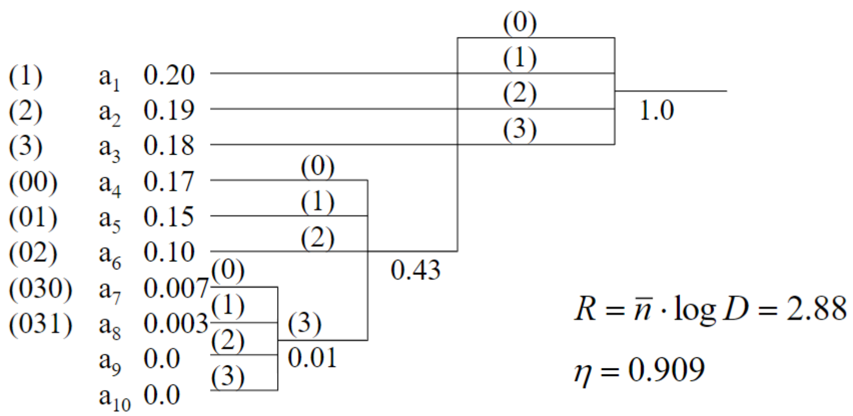
  编码速率：$R = \bar{n} \log D$

  编码效率：$\eta = \frac{H(U)}{R} = \frac{-\sum p_k \log p_k}{R}$

- **最佳不等长编码——Huffman编码**

  对 K 个信源，进行 D 元 Huffman 编码时，每次合并 D 个概率最小的消息，满足 $K = (D-1) \cdot i + 1$ 。

  若 $K = (D-1) \cdot i + M \quad M = 2, 3, \ldots, D-1$，则需添补 $D - M$ 个概率为零的虚拟消息，再合并。



- **Shannon编码**

  对于每个信息，其概率为 p_k ，码长为 l_k 。令 $P_k = \sum\limits_{i=1}^{k-1} p_i$

  用 $l_k = \lceil \log \frac{1}{p_k} \rceil$ 个比特来表示 P_k ，将 P_k 以二进制的形式表示，取小数前 l_k 位，即为对应编码。是前缀码。

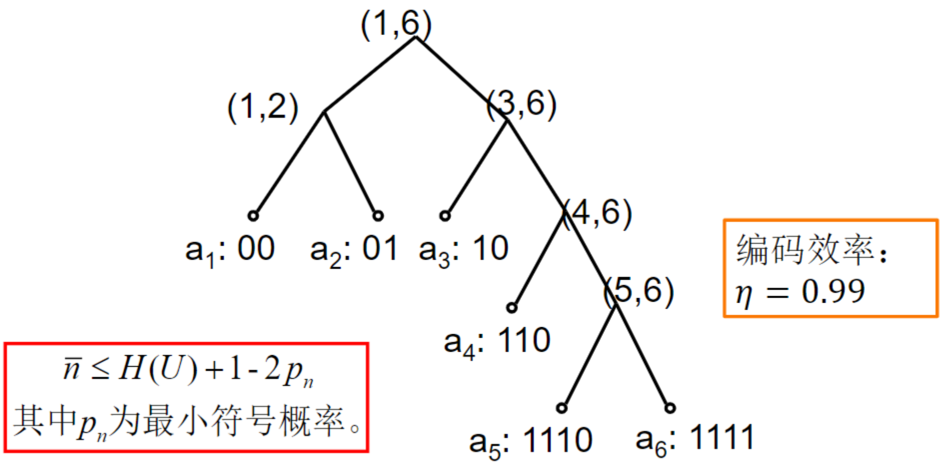
p_k	0.4	0.25	0.2	0.15
l_k	2	2	3	3
P_k	0	0.4	0.65	0.85
二进制	0.00000	0.01100	0.10100	0.11011
码字	00	01	101	110

- **Fano编码**

  将消息按概率降序排列并分成两大组，使这两组的概率差尽可能小，左边赋 0 右边赋 1；

  重复上述步骤直至每组只有一个信息。

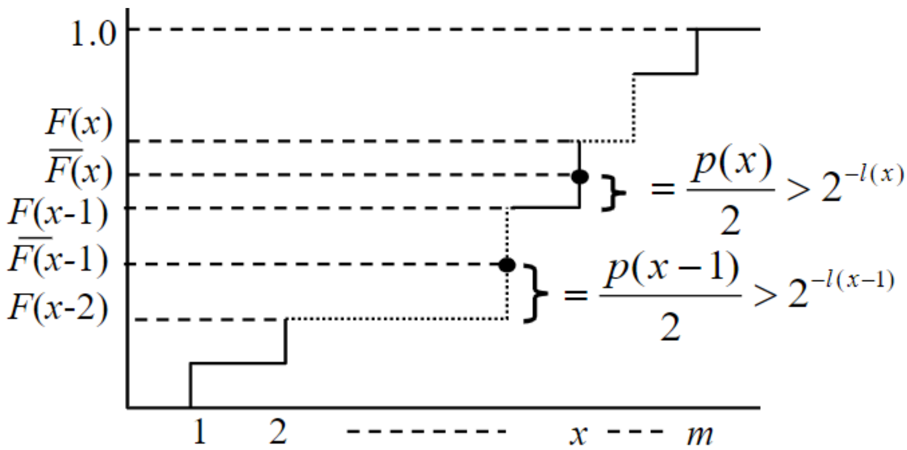
$$U=\begin{pmatrix} a_1 & a_2 & a_3 & a_4 & a_5 & a_6 \\ 0.3 & 0.25 & 0.2 & 0.15 & 0.05 & 0.05 \end{pmatrix}$$



- **S.F.E编码**

令 $U=\begin{bmatrix} a_1 & a_2 & \dots & a_m \\ p(1) & p(2) & \dots & p(m) \end{bmatrix}$，记 $\bar{F}(x) \overset{def}{=} \sum_{i<x} p(i)+\frac{1}{2}p(x)$　$F(x) \overset{def}{=} \sum_{i\le x} p(i)$

（事实上 $\bar{F}(x)=F(x)-\frac{1}{2}p(x)=F(x-1)+\frac{1}{2}p(x)$，即 $\bar{F}(x)=\frac{1}{2}[F(x)+F(x-1)]$ ）



将 $\bar{F}(x)$ 按二进制形式表示，取小数前 $l(x)$ 位，即对应编码。其中 $l(x)=\lceil \log\frac{1}{p(x)} \rceil+1$ 。

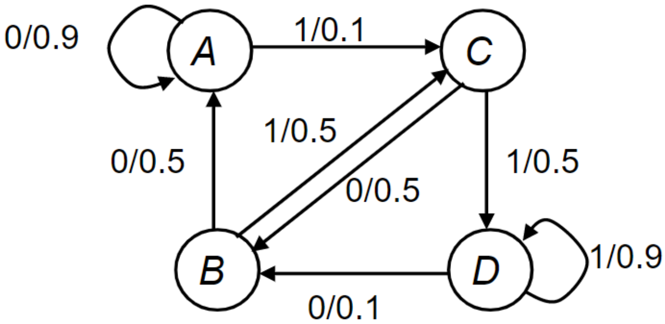
平均码长 $\bar{n}=\sum p(x)l(x)$

x	$p(x)$	$F(x)$	$\overline{F}(x)$	$\overline{F}(x)$的二进表示	$l(x)$	码字	Huffman码
1	0.25	0.25	0.125	0.001	3	001	01
2	0.5	0.75	0.5	0.10	2	10	1
3	0.125	0.875	0.8125	0.1101	4	1101	001
4	0.125	1.0	0.9375	0.1111	4	1111	000

- **马尔可夫信源编码**

$\frac{H_\infty(U)}{\log D} \le \bar{n} < \frac{H_\infty(U)}{\log D}+\frac{1}{L}$　　$H_\infty(U)=\sum q(S=s)H(U|S=s)=H(U|S)$

$q(S=s)$ 表示平稳分布时的概率，$H(U|S=s)=-\sum p(s)\log p(s)$，$p(s)$ 为进入下一步的概率。



$q(A)=q(D)=5/12,\ q(B)=q(C)=1/12$

$H(X\mid S=A)=H(X\mid S=D)=0.469$

$H(X\mid S=C)=H(X\mid S=B)=1$

$\therefore H(X\mid S)=\sum_{s\in\{A,B,C,D\}} q(s)H(X\mid s)=0.558$

$\therefore$ 二元编码时

$\bar{n}_{opt}=H_\infty/\log 2=H(X\mid S)=0.558$

Chapter 3 信道、信道容量以及信道编码定理

- 离散无记忆信道

  若离散信道对于任何 n，都有 $p(y^n|x^n)=\prod_{i=1}^{n}p(y_i|x_i)$，则称该信道为离散无记忆信道（DMC）。

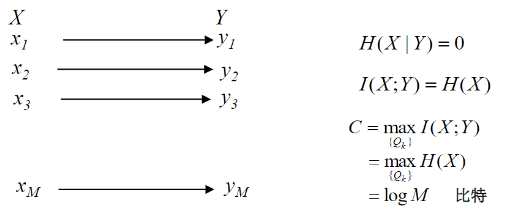
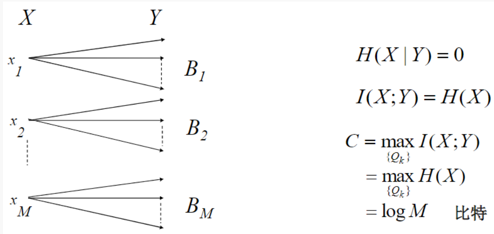
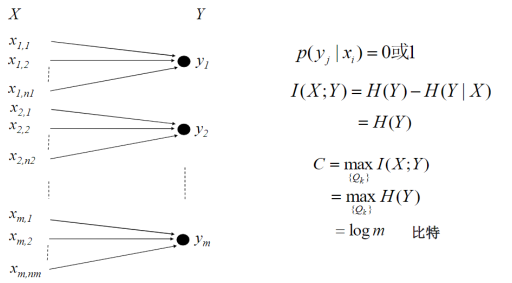
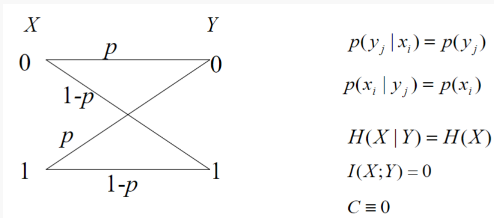
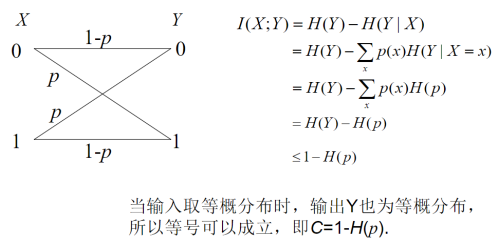
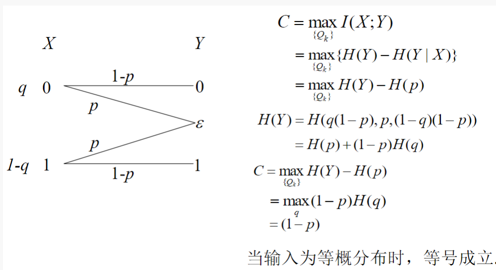
  若对任何 i 和 l 有 $p(y_i=j|x_i=k)=p(y_l=j|x_l=k)$，则称该信道是平稳/恒参的。

- 信道容量

  信道容量指，对于给定的信道而言能传输的最多的信息量，即互信息能够达到的最大值。

  $C=\lim_{n\to\infty}\frac{1}{n}\max I(X_1X_2\dots X_n;Y_1Y_2\dots Y_n)=\max I(X;Y)=\max[H(Y)-H(Y|X)\ or\ H(X)-H(X|Y)]$

  $I(X_1X_2\dots X_n;Y_1Y_2\dots Y_n)\le\sum_{i=1}^{n}I(X_i;Y_i)=nI(X;Y)$ 当输入为独立随机序列时取等号。

信道	信道容量	信道示意图
无噪信道	$C = \log M(bit)$	
无损信道	$C = \log M(bit)$	
确定信道	$C = \log m(bit)$	
无用信道	$C = 0$	
二进制对称信道	$C = 1 - H(p)$（输入等概时成立）	
二进制删除信道	$C = 1 - p$	

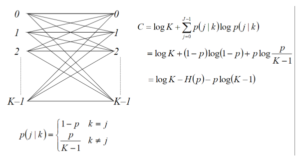
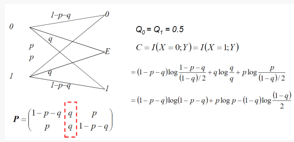
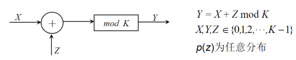
- **对称离散无记忆信道容量**

  若信道转移概率矩阵 $\mathbf{P} \in \mathbb{R}^{K\times J}$ 的每行都是第一行的一个置换，则输入对称；每列是第一列的一个置换，则输出对称。

  若信道输入输出均对称，则称其为**对称信道**，$C = \log J + \sum_{j=0}^{J-1} p(j|k) \log p(j|k)$；若仅仅输入对称，则上式 $=$ 改为 $\leq$

  若把信道输出字符集合划分成若干子集，这些子集对应 $\mathbf{P}$ 中列组成的子阵满足每行是第一行的置换，每列是第一列的置换，则这个信道是**准对称信道**。

  准对称信道容量定理：达到准对称信道容量的输入概率分布为等概分布。

信道	信道容量	信道示意图
K元对称信道	$C = \log K - H(p) - p\log(K-1)$	
二进制删除信道	$C = (1-p-q)\log(1-p-q) + p\log p - (1-q)\log\frac{1-q}{2}$	
模K加法信道	$C = \log K - H(z)$	

- 信道的组合
    - 积信道（平行组合信道）

      当输入独立且各自达到分量信道容量时，积信道就可达到其信道容量 $C = \sum C_i$
    - 和信道（开关信道）

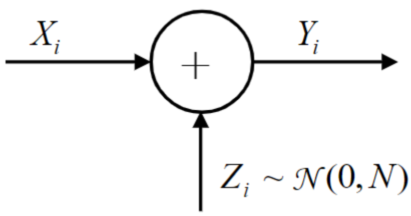
      > 信道被选中的概率之和为 1

      $2^C = 2^{C_1} + 2^{C_2} + \ldots$，即 $C = \log(\sum 2^{C_i})$，这时每个信道的利用概率为 $P_n = 2^{C_i}/2^C = 2^{C_i - C} \quad i = 1, 2\ldots, N$
    - 级联信道

      > 前一个信道的输出作为后一个信道的输入

      $C \le \min\{C_1, C_2, \ldots, C_n\}$，且级联信道容量趋于零。
- 加性高斯噪声信道

> 时间离散，输入输出连续。输入 X 为高斯分布时等号成立。



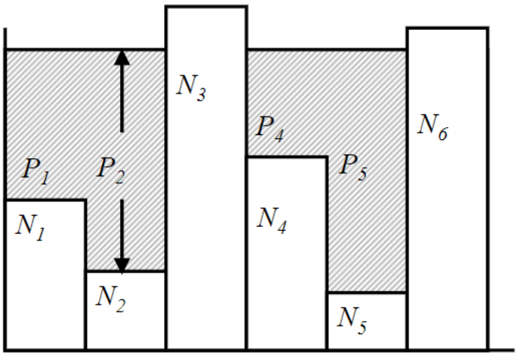
容量：
$\because\ I(X;Y) = h(Y) - h(Y|X) = h(Y) - h(X+Z|X) = h(Y) - h(Z) \le \frac{1}{2}\log(1+\frac{P}{N}) \qquad \therefore\ C = \frac{1}{2}\log(1+\frac{P}{N})$

- 平行高斯信道

$C = \frac{1}{2}\sum_{i=1}^{k}\log(1+\frac{P_i}{N_i})$

注水法则：先把功率分配给噪声小的信道，就像灌水一样进行，直到总的灌水量为总功率 P。（必考！）

附：级联时按噪声比例分配



- **信源信道分离编码**

  若信源、信道编码速率满足 $H < R_s < R_c < C$，则可以通过信源和信道分别编码使总误码率 P_e 趋于零。

- **信源信道联合编码**

  若随机序列熵速率 $H_\infty(U) < C$，则存在一个信源—信道联合编码，使 $P_e^{(n)} \to 0$

Chapter 4 率失真信源编码

最小平均失真度：$D_{\max} = \min\limits_{\hat{x}\in\hat{X}} \sum\limits_x p(x)d(x,\hat{x})$ （分布概率向量与失真矩阵相乘得到的新向量取最小值）

最大失真度：$D_{\min} = \sum\limits_x p(x) \min\limits_{\hat{x}\in\hat{X}} d(x,\hat{x})$ （找失真矩阵第 i 行的最小值与 p_i 相乘并求和）

平均失真度：$D = \sum p(x)q(\hat{x}|x)d(x,\hat{x})$

率失真函数：$R(D) = \min\limits_{Ed(X,\hat{X})\le D} I(X;\hat{X}) = H(\hat{X}) - H(\hat{X}|X)$

例：设二元等概信源 $\begin{pmatrix} X \\ p(x) \end{pmatrix} = \begin{pmatrix} x_0 & x_1 \\ 0.5 & 0.5 \end{pmatrix}$，再生字符表为 $\hat{X} = \{\hat{x}_0, \hat{x}_1, \hat{x}_2\}$，失真度矩阵为 $(\mathbf{d})_{2\times3} = \begin{pmatrix} 0 & \infty & 1 \\ \infty & 0 & 1 \end{pmatrix}$，求 $R(D)$

解：

（1）$D_{\max} = 1$、$D_{\min} = 0$

（2）存在与失真度量矩阵具有同样对称性的概率转移分布矩阵 $Q = \begin{pmatrix} \alpha & \beta & \gamma \\ \beta & \alpha & \gamma \end{pmatrix}, \alpha + \beta + \gamma = 1$

（3）由于 $d(x_0, \hat{x}_1) = d(x_1, \hat{x}_0) = \infty$，因此 $\beta = 0$，即 $Q = \begin{pmatrix} \alpha & 0 & 1-\alpha \\ 0 & \alpha & 1-\alpha \end{pmatrix}$

（4）求得 $D = \sum p(x)q(\hat{x}|x)d(x,\hat{x}) = \frac{1}{2}[\alpha * 0 + (1-\alpha) * 1] + \frac{1}{2}[\alpha * 0 + (1-\alpha) * 1] = 1 - \alpha$，所以 $\alpha = 1 - D$

（5）求得 $R(D) = \min I(X;Y) = H(\frac{\alpha}{2}, \frac{\alpha}{2}, 1-\alpha) - H(\alpha, 1-\alpha) = \alpha = 1 - D$

Chapter 5 计算理论

- **Kolmogorov复杂度**

关于一个通用计算机 u，二元字符串 x 的 Kolmogorov 复杂度定义为即能够输出并且停止的所有程序的最小长度。

$$K_u(x) = \min_{u(p)=x} l(p) \leq K_u(x|l(x)) + 2\log l(x) + c$$

- 机器学习方法分类

  监督学习：从标记的训练数据来推断一个功能的机器学习任务。如分类、回归等。

  非监督学习：在未标记的数据中，试图找到隐藏的结构。如聚类、密度估计。

  强化学习：强调如何基于环境而行动，以取得最大化的预期利益。

- 贝叶斯学习

  极大后验（MAP）策略规则：给定数据 D，在候选假设集合 H 中寻找可能性最大的假设 h。

  $$h_{MAP} = \arg\max_{h\in H} P(h|D) = \arg\max_{h\in H} P(D|h)P(h)$$

  极大似然（ML）策略规则：假设集合 H 的每个假设有相同的先验概率 $P(h_i) = P(h_j)$

  $$h_{ML} = \arg\max_{h\in H} P(D|h)$$

  - 朴素贝叶斯分类

    给定实例的属性 $< a_1, a_2, \ldots, a_n >$，取最可能的目标值 v_{MAP}

    $$v_{MAP} = \arg\max_{v_j\in V} P(v_j|a_1, a_2, \ldots, a_n) = \arg\max_{v_j\in V} P(a_1, a_2, \ldots, a_n|v_j)P(v_j) = \arg\max_{v_j\in V} \prod_i P(a_i|v_j)$$

    $$v_{NB} = \arg\max_{v_j\in V} P(v_j)\prod_i P(a_i|v_j)$$

    $$\begin{aligned} v_{NB} = \arg\max_{v_j\in V} & P(v_j) \\ & * P(\text{Outlook} = \text{Sunny}|v_j)P(\text{Temperature} = \text{Cool}|v_j) \\ & * P(\text{Humidity} = \text{High}|v_j)P(\text{Wind} = \text{Strong}|v_j) \end{aligned}$$

    计算结果：

    $P(\text{PlayTennis} = \text{Yes}) = 9/14 = 0.64;$

    $P(\text{PlayTennis} = \text{No}) = 5/14 = 0.36$

    $P(\text{Wind} = \text{Strong}|\text{PlayTennis} = \text{Yes}) = 3/9 = 0.33$

    $P(\text{Wind} = \text{Strong}|\text{PlayTennis} = \text{No}) = 3/5 = 0.60$

    $P(\text{Yes})P(\text{Sunny}|\text{Yes})P(\text{Cool}|\text{Yes})P(\text{High}|\text{Yes})P(\text{Strong}|\text{Yes}) = 0.0053$

    $P(\text{No})P(\text{Sunny}|\text{No})P(\text{Cool}|\text{No})P(\text{High}|\text{No})P(\text{Strong}|\text{No}) = 0.02$

- 决策树学习

  - 信息增益

    用熵来定义样例集合 S 的纯度：即 $Entropy(S) = -p_+ \log p_+ - p_- \log p_-$。其中 p_+ 为 S 中正例的比例，p_- 同理。

    用信息增益度量期望熵降低：使用属性 A 相对于样例集合 S 的信息增益 $Gain(S, A)$ 定义为：

    $$Gain(S, A) = Entropy(S) - \sum_{v\in Values(A)} \frac{|S_v|}{|S|} Entropy(S_v)$$

    其中 $Values(A)$ 表示属性 A 中所有可能值的集合，S_v 是 S 中属性 A 的值为 v 的子集。

- Values(Wind) = Weak， Strong
- $S = [9+,\ 5-]$
- $S_{Weak} = [6+,\ 2-]$
- $S_{Strong} = [3+,\ 3-]$

$$\mathrm{Gain}(S, \mathrm{Wind}) = \mathrm{Entropy}(S) - \sum_{v\in \mathrm{Values}(A)} \frac{|S_v|}{|S|} \mathrm{Entropy}(S_v)$$

$$= \mathrm{Entropy}(S) - \left(\frac{8}{14}\right) \mathrm{Entropy}(S_{Weak}) - \left(\frac{6}{14}\right) \mathrm{Entropy}(S_{Strong})$$

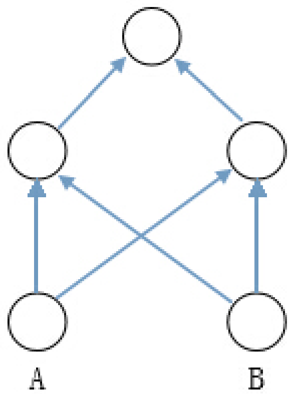
$$= 0.94 - \frac{8}{14} * 0.811 - \frac{6}{14} * 1 = 0.048$$

- 感知器

  例：两层的感知器网络实现异或布尔函数

  第一层：$A \wedge B'$ 与 $B \wedge A'$

  第二层：$A \oplus B = (A \wedge B') \vee (B \wedge A')$



Chapter 6 控制理论

- 传递函数

  状态向量：$\vec{x} = [x_1, x_2, \ldots, x_n]^T$

  输入信号向量：$\vec{u} = [u_1, u_2, \ldots, u_n]^T$

  状态微分方程：$\dot{\vec{x}} = \mathbf{A}\vec{x} + \mathbf{B}\vec{u}$

  系统输出方程：$\vec{y} = \mathbf{C}\vec{x} + \mathbf{D}\vec{u}$

  （1）拉普拉斯变换：$\boxed{\begin{aligned} \dot{x} &= Ax + Bu \\ y &= Cx + Du \end{aligned}} \xrightarrow{\text{拉普拉斯变换}} \boxed{\begin{aligned} sX(s) &= AX(s) + BU(s) \\ Y(s) &= CX(s) + DU(s) \end{aligned}}$

  （2）引入 $\phi(s) = [s\mathbf{I} - \mathbf{A}]^{-1}$，则 $X(s) = \phi(s)\mathbf{B}U(s)$

  （3）从而 $Y(s) = (\mathbf{C}\phi(s)\mathbf{B} + \mathbf{D})U(s)$ $G(s) = \frac{Y(s)}{U(s)} = \mathbf{C}\phi(s)\mathbf{B} + \mathbf{D}$

- 可控性

  充要条件：可控性矩阵 $U_c = [\mathbf{B}, \mathbf{AB}, \cdots, \mathbf{A}^{n-1}\mathbf{B}]$ 的秩为 n

- 可观性

充要条件：可观性矩阵 $U_0 = \begin{bmatrix} \mathbf{C} \\ \mathbf{CA} \\ \vdots \\ \mathbf{CA}^{n-1} \end{bmatrix}$ 的秩为 n

- **稳定性**

  控制系统本身处于平衡状态。受到扰动，产生偏差。扰动消失后，偏差逐渐变小，能恢复到原来的平衡状态，则稳定.偏差逐渐变大，不能恢复到原来的平衡状态，则不稳定。即系统在初始偏差作用下，过渡过程的收敛性。

  充要条件：传递函数所有极点位于左半平面。

  - **劳斯判据**

    > 不求解特征方程的根，直接根据特征方程的系数，判断系统的稳定性，回避了求解高次方程根的困难。

    必要条件：特征方程所有系数大于零。只要有一项小于等于零，则不稳定。

    充要条件：劳斯表中第一列元素全部大于0。若出现小于0的元素，则系统不稳定。且第一列元素符号改变的

    次数等于系统正实部根的个数。

    劳斯表：

s^n	a_n	a_{n-2}	a_{n-4}	a_{n-6}	$\cdots$
s^{n-1}	a_{n-1}	a_{n-3}	a_{n-5}	a_{n-7}	$\cdots$
s^{n-2}	A_1	A_2	A_3	A_4	$\cdots$
s^{n-3}	B_1	B_2	B_3	B_4	$\cdots$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	
s^2	D_1	D_2			
s^1	E_1				
s^0	F_1				

    $A_1 = \frac{a_{n-1}a_{n-2} - a_n a_{n-3}}{a_{n-1}}$，$A_2 = \frac{a_{n-1}a_{n-4} - a_n a_{n-5}}{a_{n-1}}$，$A_3 = \frac{a_{n-1}a_{n-6} - a_n a_{n-7}}{a_{n-1}}$…全零

    $B_1 = \frac{A_1 a_{n-3} - a_{n-1}A_2}{A_1}$，$B_2 = \frac{A_1 a_{n-5} - a_{n-1}A_3}{A_1}$，$B_3 = \frac{A_1 a_{n-7} - a_{n-1}A_4}{A_1}$…全零

    例：已知系统函数 $\phi(s) = \frac{1}{s^3+3s^2+2s+K}$，求使系统稳定时的 K 的取值范围。

    解：

    （1）特征方程 $D(s) = s^3 + 3s^2 + 2s + K$。$a_3 = 1, a_2 = 3, a_1 = 2, a_0 = K$

    （2）画出劳斯表如下：

    $$\begin{array}{c|cc} s^3 & 1 & 2 \\ s^2 & 3 & K \\ s^1 & \frac{3\times2-1\times K}{3} & 0 \\ s^0 & \frac{6K-K^2}{6-K} & 0 \end{array}$$

    （3）有 $\begin{cases} \frac{6-K}{3} > 0 \\ K > 0 \end{cases}$ 得到 $0 < K < 6$

  - **李雅普诺夫方法**

    - **间接法**

      $\mathbf{A}$ 的所有特征值 $\lambda \leq 0$，且为零的特征值无重根。传递函数极点即为特征值。

    - **直接法**

      系统运动需要能量。在非零初始状态作用下的运动过程中，若能量随时间衰减以致最终消失，则系统迟早会达到平衡状态，即系统渐近稳定。反之，系统则不稳定。若能量在运动过程中不增不减，则称为李雅普诺夫意义下的稳定。