第4章　信息率失真函数

■问题

信源熵$H(X)$的物理含义是什么？

为什么要研究信源熵？

信源无失真传输所需的最小信息率为$R \geq H(X)$；

允许信源有失真时，输出的最小速率可降低为$R < H(X)$；

失真D越大，R可以越小，因此R是D的函数，且为单调递减函数。

$R(D)$就叫做信息率失真函数。

4.1 信息率失真函数的概念和性质

4.1 信息率失真函数的概念和性质

在实际问题中，信号有一定的失真是可以容忍的。但是当失真大于某一限度后，信息质量将被严重损伤，甚至丧失其实用价值。要规定失真限度，必须先有一个定量的失真测度。为此可引入失真函数。

4.1.1 失真函数和平均失真

$X=\{x_i\}$，$x_i \in \{a_1,\ldots a_n\}$ $Y=\{y_j\}$，$y_j \in \{b_1,\ldots b_m\}$

失真函数$d(x_i, y_j)$

$$d(x_i, y_j) = \begin{cases} 0 & x_i = y_j \\ \alpha \quad \alpha > 0 & x_i \neq y_j \end{cases}$$

失真矩阵

单个符号的失真度的全体构成的矩阵 $\left[d(x_i, y_j)\right]$，称为失真矩阵

$$\mathbf{d} = \begin{bmatrix} d(a_1,b_1) & d(a_1,b_2) & \cdots & d(a_1,b_m) \\ d(a_2,b_1) & d(a_2,b_2) & \cdots & d(a_2,b_m) \\ \vdots & \vdots & & \vdots \\ d(a_n,b_1) & d(a_n,b_2) & \cdots & d(a_n,b_m) \end{bmatrix}$$

最常用的失真函数

均方失真：$$d(x_i, y_j) = (x_i - y_j)^2$$

绝对失真：$$d(x_i, y_j) = |x_i - y_j|$$

相对失真：$$d(x_i, y_j) = |x_i - y_j| / |x_i|$$

误码失真：$$d(x_i, y_j) = \delta(x_i, y_j) = \begin{cases} 0, & x_i = y_j \\ 1, & \text{其它} \end{cases}$$

前三种失真函数适用于连续信源，后一种适用于离散信源。

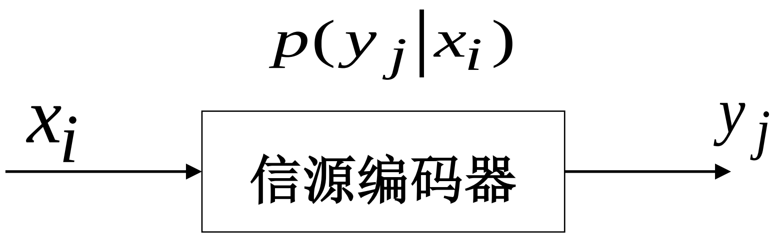
序列编码情况失真函数定义为：

$$d_L(\mathbf{x}_i,\mathbf{y}_j)=\frac{1}{L}\sum_{l=1}^{L}d(x_{il},y_{jl})$$

其中$d(x_{il}, y_{jl})$是信源输出L长符号样值x_i中的第l个符号x_{il}时，编码输出L长符号样值y_j中的第l个符号y_{jl}的失真函数。

失真函数的数学期望称为平均失真，记为

$$\overline{D}=\sum_{i,j}p(x_i,y_j)d(x_i,y_j)=\sum_{i,j}p(x_i)p\left(y_j/x_i\right)d(x_i,y_j)$$



已知$p(x_i)$和$d(x_i，y_j)$，平均失真只是符号转移概率$p(y_j/x_i)$的函数。$p(y_j/x_i)$在此实质上代表编码方式。

如：

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

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

$$x_1 \to y_1$$

$$x_2 \to y_2$$

$$x_1 \to y_1$$

$$x_2 \to y_1$$

对于连续随机变量同样可以定义平均失真

$$\overline{D} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} p_{xy}(x,y)d(x,y)dxdy$$

对于L长序列编码情况，平均失真为

$$\overline{D}_L = \frac{1}{L}\sum_{l=1}^{L} E[d(x_{il}, y_{jl})]$$

$$= \frac{1}{L}\sum_{l=1}^{L} \overline{D}_l$$

4.1.2 信息率失真函数R(D)

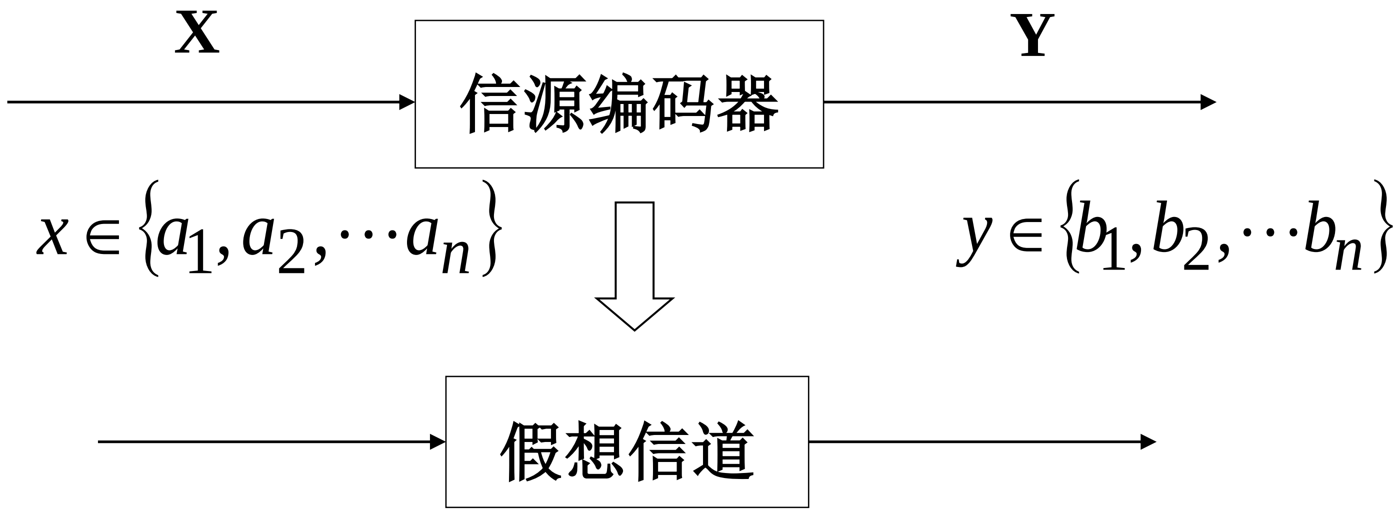
信源编码器的目的是使编码后所需的信息传输率**R**尽量小，

$$R\downarrow \longrightarrow \overline{D}\uparrow$$

给定失真的限制值**D**，使 $\overline{D}\le$ **D**，找最小**R**，

$\longrightarrow$ **R(D)**，定义为信息率失真函数。

4.1.3 信息率失真函数R(D)



$p(y_j/x_i)$ 信源符号编码概率 $\Longrightarrow$ 信道转移概率

将信源编码器看作信道，信源编码器输出的信息率R对应到信道，即为接收端Y需要获得的有关X的信息量，也就是互信息$I(X;Y)$。

D允许试验信道

$$\overline{D}=\sum_{i,j}p(x_i)p\left(y_j/x_i\right)d(x_i,y_j)$$

若$p(x_i)$和$d(x_i, y_j)$已定，则可给出满足$\overline{D}\le D$条件的所有转移概率分布p_{ij}，它们构成了一个信道集合P_D

$$P_D=\left\{p(y_j/x_i):\overline{D}\le D\right\}$$

称为D允许试验信道。

信息率失真函数R(D)

$I[p(x_i),\ \ p(y_j/x_i)]$

当$p(x_i)$一定时，互信息I是关于$p(y_j/x_i)$ 的U型凸函数，存在极小值(2.2节）。

在上述允许信道P_D中，可以寻找一种信道p_{ij}，使给定的信源$p(x_i)$经过此信道传输后，互信息$I(X;Y)$达到最小。

$$R(D)=\min_{P_D} I(X;Y)$$

D=? $p(y_j/x_i)=p_{ij}$? R(D)=?

对于离散无记忆信源，R(D)函数可写成

$$R(D)=\min_{P_{ij}\in P_D}\sum_{i=1}^{n}\sum_{j=1}^{m}p(a_i)p(b_j/a_i)\log\frac{p(b_j/a_i)}{p(b_j)}$$

$p(a_i)$，$i=1，2，…，n$ 是信源符号概率分布；

$p(b_j/a_i)$，$i=1，2，…，n$，$j=1，2，…，m$ 是转移概率分布；

$p(b_j)$，$j=1，2，…，m$ 是接收端收到符号概率分布。

R(D)的物理意义

无失真时：R=H(X)

有失真时：R=R(D)=H(X)－H(X/Y)≤H(X)

H(X/Y)：由于压缩编码损失的信息

对于给定信源，在平均失真不超过失真限度D的条件下，信息率容许压缩的最小值R(D)

例 4-2

设信源的符号表为A＝{a_1，a_2，…，a_{2n}}，概率分布为$p(a_i)$＝1/2n，i＝1，2，…，2n，失真函数规定为

$$d(a_i,a_j)=\begin{cases}1 & i\neq j\\ 0 & i=j\end{cases}$$

即符号不发生差错时失真为0，一旦出错，失真为1，试研究在一定编码条件下信息压缩的程度。

$$R(1/2)=H(Y)=H(\underbrace{\frac{1}{2n},\cdots,\frac{1}{2n}}_{n-1},\frac{1+n}{2n})=\underbrace{\log 2n}_{H(X)}-\underbrace{\frac{n+1}{2n}\log(n+1)}_{H(X/Y)}$$

H(X)

H(X/Y)
可压缩的
信息量

4.1.3 信息率失真函数的性质

1. **R(D)函数的定义域**

（1）**Dmin和R(Dmin)**

Dmin＝0

$$R(D_{\min}) = R(0) = H(X)$$

对于连续信源

$$R(D_{\min}) = R(0) = H_c(x) = \infty$$

讨论

何时$D_{min}=0$?

– 只有当失真矩阵中每行至少有一个零元素。

何时$R(0)=H(X)$?

– 只有当失真矩阵中每行至少有一个零，并每一列最多只有一个零。
– 否则R(0)可以小于H(X)，表示这时信源符号集中有些符号可以压缩、合并而不带来任何失真。

(2) D_{max}和$R(D_{max})$

$R(D_{max})=0$

选择所有满足R(D)＝0中D的最小值，定义为R(D)定义域的上限D_{max}，即

$$D_{max} = \min_{R(D)=0} D$$

因此可以得到R(D)的定义域为

$$D \in \left[0, D_{max}\right]$$

D_{max}=?

R(D)＝0就是*I*(*X*;*Y*)＝0，这时试验信道输入与输出是互相独立的，所以条件概率*p*(y_j/x_i)与x_i无关。即 $p_{ij} = p(y_j / x_i) = p(y_j) = p_j$

$$D_{\max} = \min_{p_{ij}} D = \min_{p_{ij}} \sum_{i=1}^{n} \sum_{j=1}^{m} p_i p_{ij} d_{ij}$$

$$= \min_{p_j} \sum_{i=1}^{n} \sum_{j=1}^{m} p_i p_j d_{ij} = \min_{p_j} \sum_{j=1}^{m} p_j \sum_{i=1}^{n} p_i d_{ij}$$

需满足条件 $\sum_{j=1}^{m} p_j = 1$

$$D_{\max} = \min_{p_j} \sum_{j=1}^{m} p_j \sum_{i=1}^{n} p_i d_{ij}$$

从上式观察可得：在**j=1，…，m**中，可找到 $\sum_{i=1}^{n} p_i d_{ij}$ 值最小的**j**，当该**j**对应的p_j=**1**，而其余p_j为零时，上式右边达到最小，这时上式可简化成

$$D_{\max} = \min_{j=1,2,\cdots,m} \sum_{i=1}^{n} p_i d_{ij}$$

例4-3

设输入输出符号表为$X=Y\in\{0，1\}$，输入概率分布$p(x)=\{1/3，2/3\}$，失真矩阵为

$$\mathbf{d}=\begin{bmatrix} d(a_1,b_1) & d(a_1,b_2) \\ d(a_2,b_1) & d(a_2,b_2) \end{bmatrix}=\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$$

解：

当D_{min}＝0时，$R(D_{min})=H(X)=H(1/3，2/3)$＝0.91比特/符号，这时信源编码器无失真，所以该编码器的转移概率为

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

当R(D_{max})＝0时

$$D_{\max} = \min_{j=1,2} \sum_{i=1}^{2} p_i d_{ij}$$
$$= \min_{j=1,2} \{p_1 d_{11} + p_2 d_{21}, p_1 d_{12} + p_2 d_{22}\}$$
$$= \min_{j=1,2} \left\{\frac{1}{3} \times 0 + \frac{2}{3} \times 1, \frac{1}{3} \times 1 + \frac{2}{3} \times 0\right\}$$
$$= \min_{j=1,2} \left\{\frac{2}{3}, \frac{1}{3}\right\} = \frac{1}{3}$$

此时输出符号概率$p(b_1)$＝0，$p(b_2)$＝1，

$$a_1 \to b_2, a_2 \to b_2$$

所以这时的编码器的转移概率为 $\mathbf{P} = \begin{bmatrix} 0 & 1 \\ 0 & 1 \end{bmatrix}$

2. R(D)的下凸性

证明思路:

设 $D=\theta D_1+(1-\theta)D_2$, $(0\le\theta\le 1)$

再证: $R(\theta D_1+(1-\theta)D_2)\le\theta R(D_1)+(1-\theta)R(D_2)$

3. R(D)的单调递减性和连续性

若$D>D'$, $\rightarrow P_D\supset P_{D'}$

(选择p_{ji}的范围大)

$$R(D)=\min_{p_{ji}\in P_D} I(X;Y)\le\min_{p_{ji}\in P_{D'}} I(X;Y)=R(D')$$

(连续性证明从略)

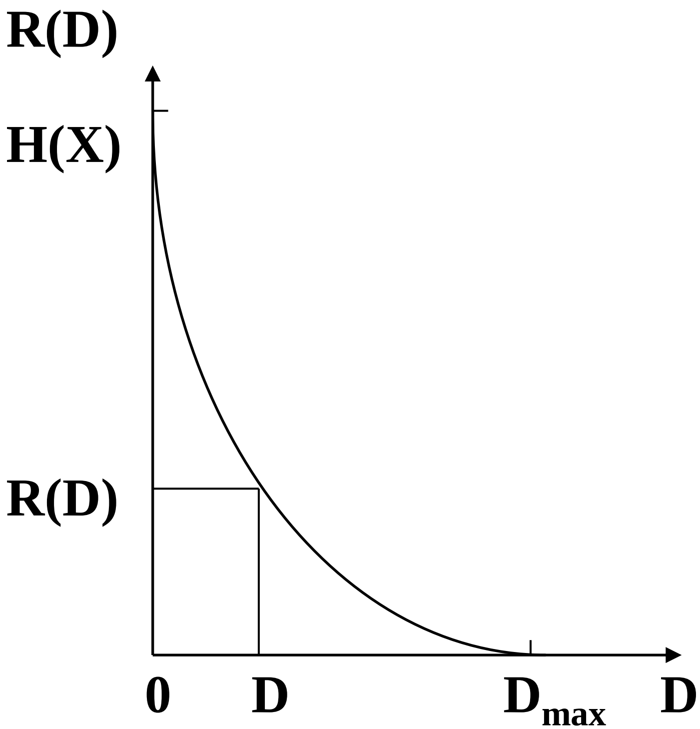
综上所述，可以得出如下结论：

R(D)是非负的实数，即R(D) $\geq$ 0。其定义域为0～Dmax，其值为0～H(X)。当D>Dmax时，R(D) $\equiv$ 0。

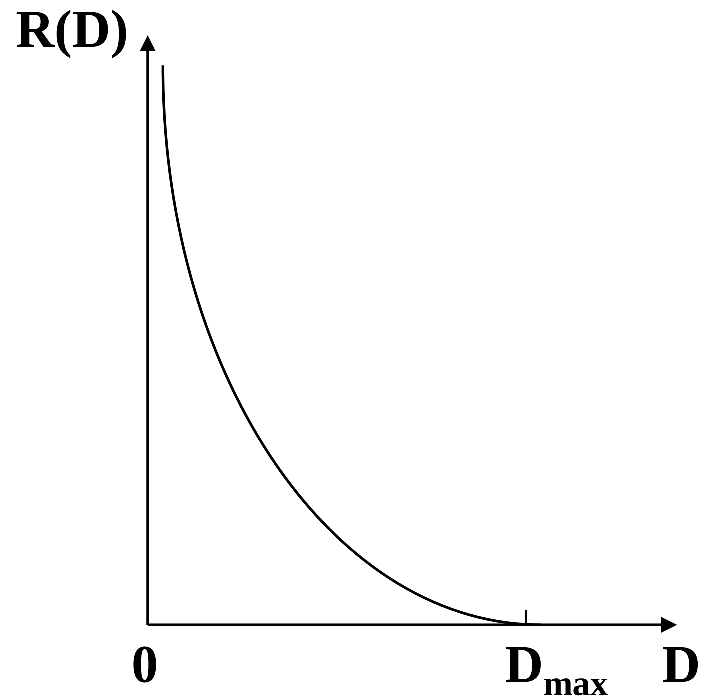
R(D)是关于D的下凸函数，因而也是关于D的连续函数。

R(D)是关于D的严格递减函数。容许的D越大，所要求的R越小。反之亦然。

由以上三点结论，对一般信息率失真R(D)曲线的形态可以画出来：



离散系统



连续系统

4.1.4 R(D)与C

	信道容量C	率失真函数R(D)
研究对象	信道	信源
给定条件	信道转移概率$p(y_j/x_i)$	信源分布$p(x_i)$
选择参数	信源分布$p(x_i)$	信源编码器编码方法$p(y_j/x_i)$
限制条件	$\sum_i p(x_i)=1$	$P_D=\{p(y_j/x_i):\bar{D}\le D\}$
结论	$C=\max\limits_{p(x_i)} I(X;Y)$	$R(D)=\min\limits_{P_D} I(X;Y)$
H(X/Y) =H(X)-I(X;Y)	噪声干扰丢失的信息量	编码压缩损失的信息量

4.2 离散信源和连续信源的R(D)计算

某些特殊情况下R（D）的表示式为：

（1）当$d(x,y)=(x-y)^2$， $p(x)=\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{x^2}{2\sigma^2}}$ 时，

$$R(D)=\log\frac{\sigma}{\sqrt{D}}$$

(2)当$d(x,y)=|x-y|$，$p(x)=\frac{\lambda}{2}e^{-\lambda|x|}$ 时，

$$R(D)=\log\frac{1}{\lambda D}$$

(3)当$d(x,y)=\delta(x,y)$，$p(x=0)=p$，$p(x=1)=1-p$时，

R(D)=H(*p*)－H(D)

这些R(D)可画成三条曲线

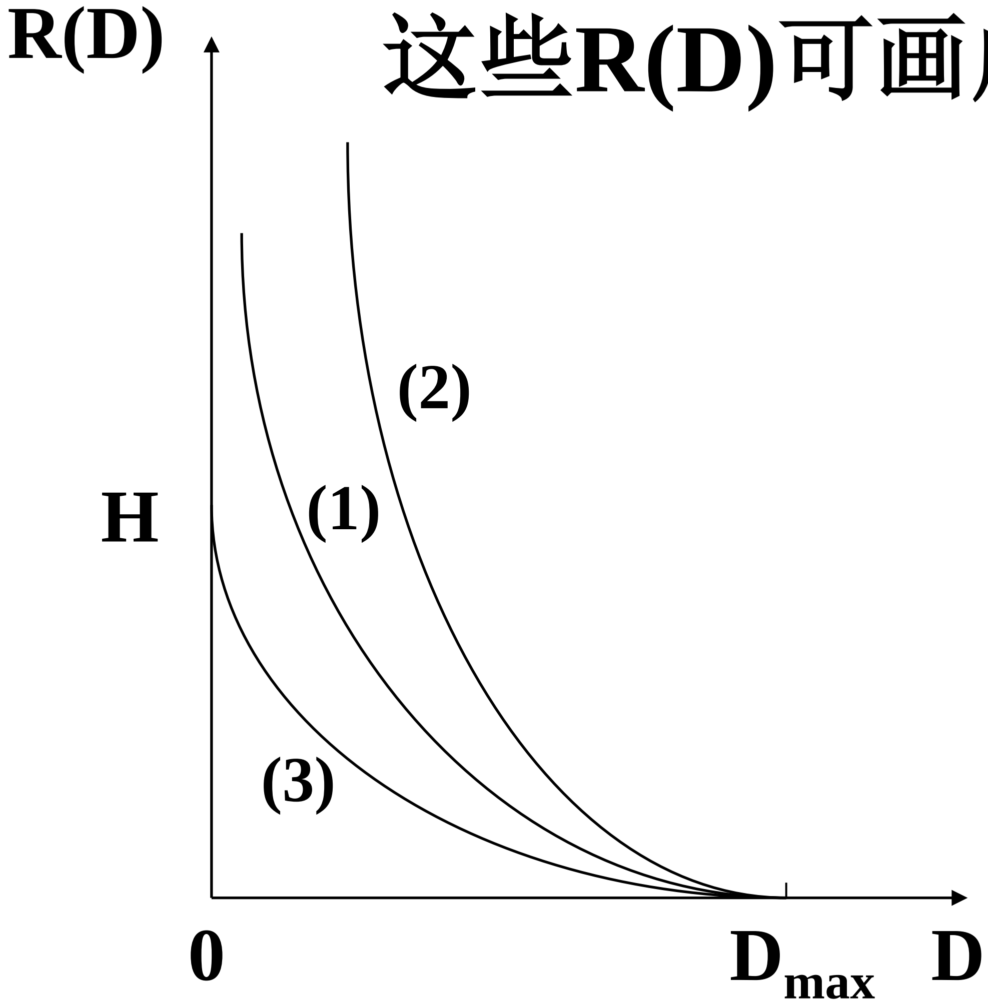


图4-5 信息率失真函数R(D)

R(D)的参量求法

已知信源概率分布p_i和给定失真矩阵$[d_{ij}]$，求R(D)。

- $I(p_{ij})$是p_{ij}的下凸函数，但通过求条件极值难以得到R(D)显式，通常先求参量表示式。

$$I(X;Y)=\sum_{ij}p_i p_{ij}\log\frac{p_{ij}}{q_j} \qquad (1) \qquad \left(q_j=\sum_i p_i p_{ij}\right)$$

$$\begin{cases} D=\sum_{ij}p_i p_{ij}d_{ij} & (2) \\ \sum_{j=1}^{m}p_{ij}=1,(i=1,2,\cdots,n) & \end{cases}$$

用拉格朗日乘子法有

$$\frac{\partial}{\partial \mathbf{p}_{ij}}[\mathbf{I}-\mathbf{sD}-\boldsymbol{\mu}_i\sum_j \mathbf{p}_{ij}]=0,(\mathbf{i}=1,2,\cdots,\mathbf{n}) \qquad (3)$$

令 $\log\lambda_i=\frac{\mu_i}{p_i}$

可解得： $p_{ij}=q_j\lambda_i\exp(sd_{ij}),(i=1,\cdots,n;j=1,\cdots,m) \quad (4)$

(4)式两边对j求和得 $\lambda_i=[\sum_l q_l\exp(sd_{il})]^{-1} \qquad (5)$

(4)式两边乘以p_i再对i求和得 $\sum_i \lambda_i p_i\exp(sd_{ij})=1 \qquad (6)$

解(6)式得λ_i，解(5)式得q_j。

将(4)式代入(2)式得到以s为参量的D(s)

$$D(s) = \sum_{ij} d_{ij} p_i q_j \lambda_i \exp[s d_{ij}] \qquad (7)$$

将(4)式代入(1)式得到以s为参量的R(s)

$$R(s) = sD(s) + \sum_{i} p_i \log \lambda_i \qquad (8)$$

将λ_i和q_j代入(7)式得到S(D)，再代入(8)式消去参量S，就可得到R(D)。

注意: q_j应确保为非负，否则计算失效。

例4-5 二元信源的R(D)函数

已知: $P(a_1)=p$, $P(a_2)=1-p$, $p\le 1/2$, $\left[d_{ij}\right]=\begin{bmatrix}0 & 1\\ 1 & 0\end{bmatrix}$

求：R(D)

解：(1) $D_{min}=0$，$R(0)=H(X)=H(p)$，$P=\begin{bmatrix}1 & 0\\ 0 & 1\end{bmatrix}$

$$(2) D_{\max} = \min_j \sum_i p_i d_{ij} = \min_j \left[(1-p), p\right] = p$$

$R(D_{max})=0$，$P=\begin{bmatrix}0 & 1\\ 0 & 1\end{bmatrix}$

(3) 当0<D<D_{max} 时，用参量法：

解(6)式方程组可求得：

$$\lambda_1=\left[(1+e^s)p\right]^{-1},\lambda_2=\left[(1+e^s)(1-p)\right]^{-1}$$

解(5)式方程组可求得：

$$q_1=\frac{p-(1-p)e^s}{1-e^s},q_2=\frac{(1-p)pe^s}{1-e^s}$$

代入(7)、(8)式得：

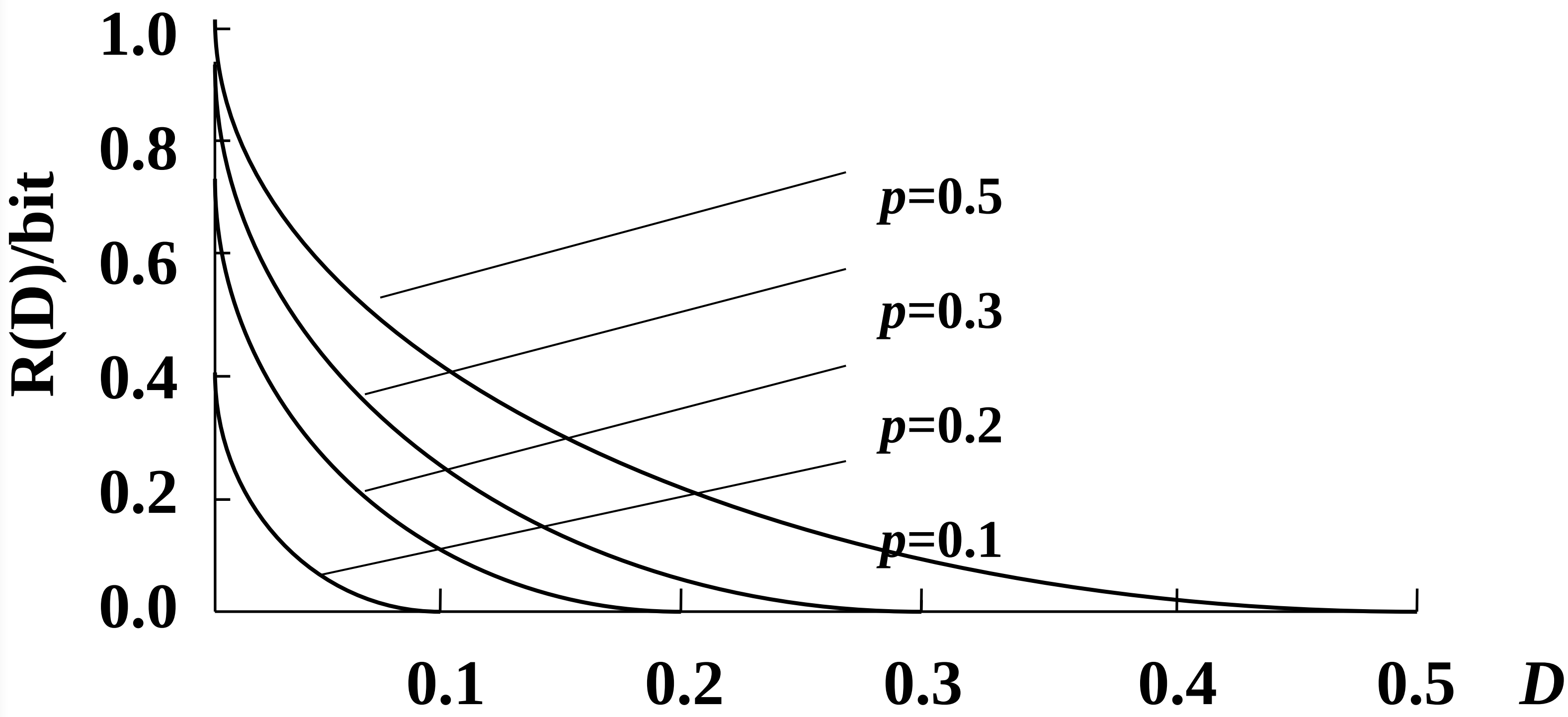
$$D(s)=\frac{e^s}{1+e^s},R(s)=\frac{e^s}{1+e^s}-\log(1+e^s)+H(p)$$

消去s得：

$$R(D)=H(p)-H(D)$$

从(4)式得：

$$\boldsymbol{P}=\begin{bmatrix}1-D & D\\ D & 1-D\end{bmatrix}$$



- **R(D)=H(*p*)－H(D)，*p*为参数**
- 信源分布越均匀，**R(D)**越大，信源压缩的可能性越小；
- 信源分布不均匀，冗余度大，**R(D)**小，信源压缩的可能性越大；

总结　第4章

无失真时：$R=H(X)$

有失真时：$R=R(D)=H(X)-H(X/Y)\le H(X)$

$H(X/Y)$：由于压缩编码损失的信息

$$R(D)=\min_{p_{ij}\in P_D} I(X;Y)$$

总结

失真函数d(x_i,y_j)

平均失真$\overline{D}$

信息率失真函数R(D)

$$R(D)=\min_{P_D} I(X;Y) \qquad P_D=\left\{p(y_j/x_i):\overline{D}\le D\right\}$$

$$D_{min}=0, \qquad R(D_{\min})=R(0)=H(X)$$

$$D_{\max}=\min_{R(D)=0} D \qquad R(D_{\max})=0$$

总结

R(D)的性质

1. R(D)是非负的实数，即$R(D) \geq 0$。
2. 其定义域为0～D_{max}，其值域为0～$H(X)$。当$D>D_{max}$时，$R(D) \equiv 0$。
3. R(D)是下凸的、连续的、严格递减的函数。