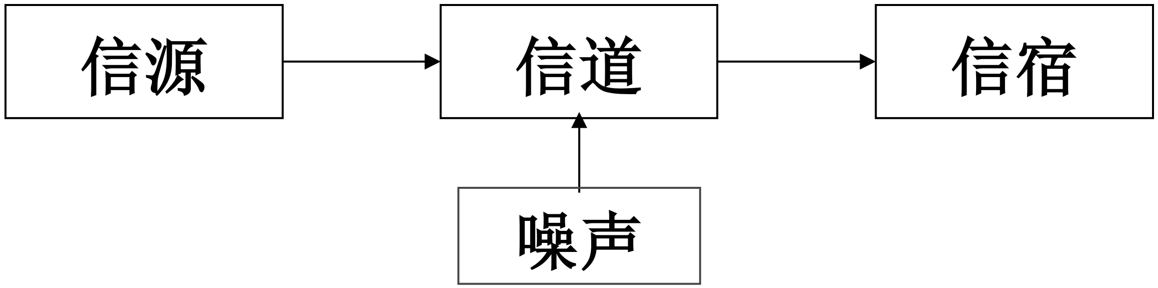
2022-2023学年秋季学期

信息论与编码

第五讲

复习



- 信道表示为 $[X,P(Y|X),Y]$

- **信道转移矩阵(n行m列)完全描述了信道在干扰作用下的统计特性。**
  - **行表示输入x，列表示输出y；**
  - **每个元素均非负，每一行元素之和为1。**
  - **每列若有多个非零，则疑义度H(X|Y)非零。**

- 信道信息传输率R：信道中平均每个符号所能传送的信息量=平均互信息

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

信道容量==平均互信息的最大值：

$$C = \max_{p(x_i)} I(X;Y) \qquad \text{bit/信道符号}$$

- **离散对称信道：** $C = \max_{p(x_i)}[H(Y) - H(Y \mid X)] = \log_2 m - H(\mathrm{Y|X})$

当输入为等概率分布时，达到信道容量C。

- **强对称离散信道** $C = \log n - p\log(n-1) - H(p)$ bit/信道符号
- **准对称离散信道**

$$C = -\sum_{k=1}^{s} m_k \bar{p}(b_k)\log_2 \bar{p}(b_k) - H(q_1, q_2, \cdots, q_m)$$

$$= \log n - \sum_{k=1}^{s} N_k \log M_k - H(q_1, q_2, \cdots, q_m)$$

其中，n为输入符号数，N_k为第k个子阵中的行元素和，M_k为第k个子阵中的列元素和。

离散信道容量的一般计算方法

对于固定的信道，平均互信息$I(X;Y)$是信源概率分布$p(x_i)$的上凸函数, 所以极大值一定存在。

$I(X;Y)$是n个变量$p(a_i)(i=1,2,\ldots,n)$的多元函数，并满足

$$\sum_{i=1}^{n} p(a_i) = 1$$

所以可以用拉格朗日乘子法计算条件极值:引进新函数

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

其中，λ 为拉格朗日乘子

解方程组 $\dfrac{\partial \Phi}{\partial p(a_i)} = \dfrac{\partial \left\{ I(X;Y) - \lambda [\sum_{i=1}^{n} p(a_i) - 1] \right\}}{\partial p(a_i)} = 0$

可得一般信道容量*C*。

由 $p(b_j) = \sum_{i=1}^{n} p(a_i) p(b_j \mid a_i) \longrightarrow \dfrac{dp(b_j)}{dp(a_i)} = p(b_j \mid a_i)$

$$\frac{\partial}{\partial p(a_i)} \left\{ -\underbrace{\sum_{j=1}^{m} p(b_j) \log_2 p(b_j)}_{H(Y)} + \underbrace{\sum_{i=1}^{n} \sum_{j=1}^{m} p(a_i) p(b_j \mid a_i) \log_2 p(b_j \mid a_i)}_{H(Y|X)} \right.$$

$$\left. -\lambda [\sum_{i=1}^{n} p(a_i) - 1] \right\} = 0 \longrightarrow$$

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

$$-\sum_{j=1}^{m} [p(b_j \mid a_i) \log_2 p(b_j) + p(b_j \mid a_i) \log_2 e] + \sum_{j=1}^{m} p(b_j \mid a_i) \log_2 p(b_j \mid a_i) - \lambda = 0$$

$$\longrightarrow \sum_{j=1}^{m} p(b_j \mid a_i) \log_2 \frac{p(b_j \mid a_i)}{p(b_j)} = \log_2 e + \lambda$$

$$\sum_{i=1}^{n}\sum_{j=1}^{m} p(a_i)p(b_j|a_i)\log_2\frac{p(b_j|a_i)}{p(b_j)} = \sum_{j=1}^{m} p(a_i)(\log_2 e+\lambda) = \log_2 e+\lambda$$

$I(X;Y)$最大值 $\quad C=\log_2 e+\lambda$

$$\sum_{j=1}^{m} p(b_j|a_i)\log_2 p(b_j|a_i) = \sum_{j=1}^{m} p(b_j|a_i)\log_2 p(b_j) + C$$

$$= \sum_{j=1}^{m} p(b_j|a_i)[\log_2 p(b_j) + C]$$

令 $\beta_j = \log_2 p(b_j) + C$

则 $\sum_{j=1}^{m} p(b_j|a_i)\beta_j = \sum_{j=1}^{m} p(b_j|a_i)\log_2 p(b_j|a_i)$ **若*m*=*n*,则可求解**

$$\beta_j \longrightarrow p(b_j) = 2^{\beta_j - C} \longrightarrow 2^C = \sum_{j=1}^{m} 2^{\beta_j} \longrightarrow C = \log_2(\sum_{j=1}^{m} 2^{\beta_j})$$

一般离散信道容量的求解步骤：

(1) $$\sum_{j=1}^{m} p(b_j \mid a_i)\beta_j = \sum_{j=1}^{m} p(b_j \mid a_i)\log_2 p(b_j \mid a_i) \longrightarrow \beta_j$$

(2) $$C = \log_2(\sum_{j=1}^{m} 2^{\beta_j})$$

(3) $$p(b_j) = 2^{\beta_j - C}$$

(4) $$p(b_j) = \sum_{i=1}^{n} p(a_i) p(b_j \mid a_i) \longrightarrow p(a_i)$$

需要确认所有的 $p(a_i) \geq 0$ ，所求的***C***才存在。

例：

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

可列方程组：

$$\begin{cases} \frac{1}{2}\beta_1+\frac{1}{4}\beta_2+\frac{1}{4}\beta_4=\frac{1}{2}\log\frac{1}{2}+\frac{1}{4}\log\frac{1}{4}+\frac{1}{4}\log\frac{1}{4} \\ \beta_2=0 \\ \beta_3=0 \\ \frac{1}{4}\beta_1+\frac{1}{4}\beta_3+\frac{1}{2}\beta_4=\frac{1}{4}\log\frac{1}{4}+\frac{1}{4}\log\frac{1}{4}+\frac{1}{2}\log\frac{1}{2} \end{cases}$$

3.2单符号离散信道的信道容量

解之得： $\beta_2=\beta_3=0,\ \beta_1=\beta_4=-2$

$$C=\log(2^{-2}+2^{0}+2^{0}+2^{-2})=\log\frac{5}{2}=\log 5-1$$

$$P(b_1)=P(b_4)=2^{-2-\log 5+1}=\frac{1}{10} \qquad P(b_2)=P(b_3)=2^{0-\log 5+1}=\frac{4}{10}$$

$$P(a_1)=P(a_4)=\frac{4}{30},\ \ P(a_2)=P(a_3)=\frac{11}{30}$$

3.3多符号离散信道的信道容量

■ 多符号离散信道的数学模型

➢ 基本概念

多符号离散信源$X^N=X_1X_2\ldots X_N$在N个不同时刻分别通过单符号离散信道 $\{X,P(Y|X),Y\}$，

在输出端出现$Y^N=Y_1Y_2\ldots Y_N$，形成一个新的信道，此即多符号离散信道。

由于新信道相当于单符号离散信道在N个不同的时刻连续运用了N次，也称为单符号离散信道的N次扩展信道

3.3多符号离散信道的信道容量

- 多符号离散信道的数学模型

$$X^N=X_1X_2\ldots X_N \longrightarrow \boxed{P(Y^N|X^N)} \longrightarrow Y^N=Y_1Y_2\ldots Y_N$$

X_k取值: $\{a_1,a_2,\ldots,a_n\}$,则X^N共有n^N种 α_i， $i=1,\ldots,n^N$

Y_k取值: $\{b_1,b_2,\ldots,b_m\}$,则Y^N共有m^N种β_j， $j=1,\ldots, m^N$

$$\alpha_i=(a_{i_1}a_{i_2}\ldots a_{i_N}),\qquad a_{i_1},a_{i_2},\ldots,a_{i_N}\in\{a_1,a_2,\ldots,a_n\}$$

$$\beta_j=(b_{i_1}b_{i_2}\ldots b_{i_N}),\qquad b_{i_1},b_{i_2},\ldots,b_{i_N}\in\{b_1,b_2,\ldots,b_n\}$$

3.3多符号离散信道的信道容量

- 信道矩阵

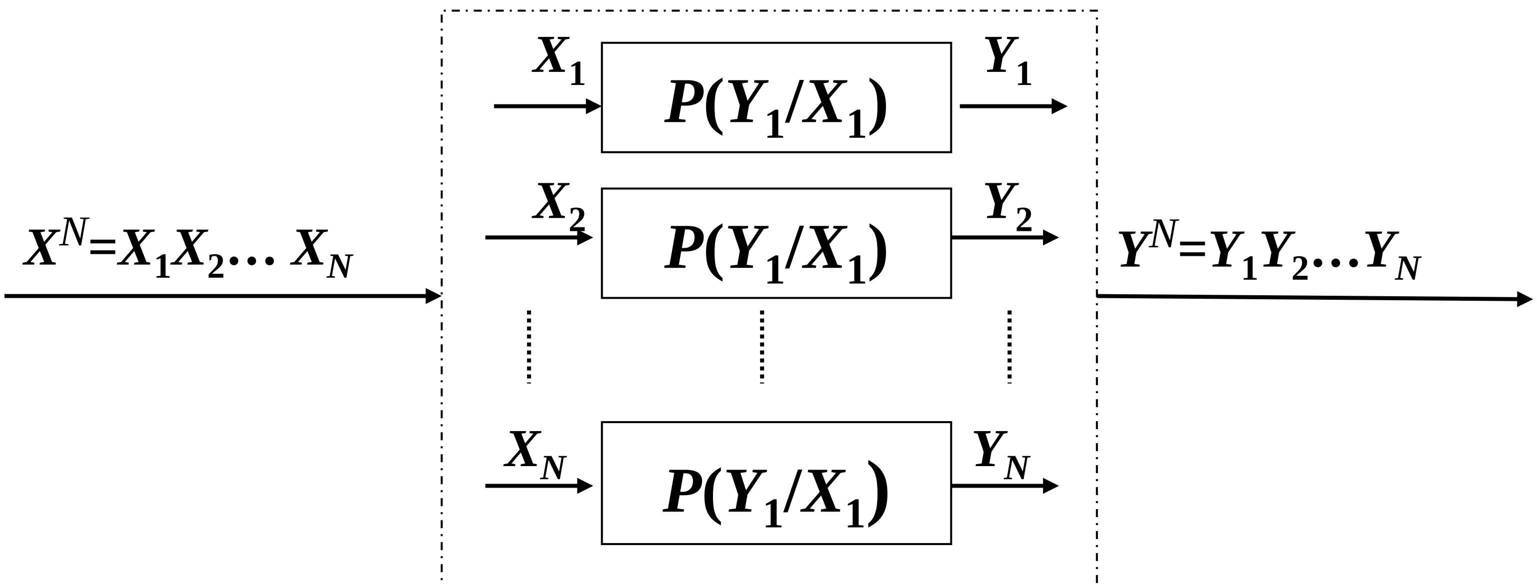
$$
\boldsymbol{P}(\boldsymbol{Y}^N \mid \boldsymbol{X}^N) = \begin{bmatrix} p(\beta_1/\alpha_1) & p(\beta_2/\alpha_1) & \cdots & p(\beta_{m^N}/\alpha_1) \\ p(\beta_1/\alpha_2) & p(\beta_2/\alpha_2) & \cdots & p(\beta_{m^N}/\alpha_2) \\ \vdots & \vdots & \vdots & \vdots \\ p(\beta_1/\alpha_{n^N}) & p(\beta_2/\alpha_{n^N}) & \cdots & p(\beta_{m^N}/\alpha_{n^N}) \end{bmatrix}
$$

每行元素之和等于1

3.3多符号离散信道的信道容量

- 离散无记忆扩展信道的信道容量

把多符号离散信道理解成单符号离散信道在每一个单位时间传递一个随机变量的时候，需要考虑k时刻的输出变量Y_k与时刻k之前的输入变量$X_1X_2\ldots X_{k-1}$和输出变量$Y_1Y_2\ldots Y_{k-1}$之间有无依赖关系。



单符号离散信道的N次扩展信道的数学模型

若多符号离散信道的转移概率满足

$$\boldsymbol{P}(\boldsymbol{Y}\mid\boldsymbol{X})=P(Y_1Y_2\ldots Y_N\mid X_1X_2\ldots X_N)=$$

$$P(Y_1\mid X_1)P(Y_2\mid X_2)\ldots P(Y_N\mid X_N)=\prod_{k=1}^{N}p(Y_k\mid X_k)$$

则称之为离散无记忆信道的N次扩展信道

注：扩展信道的转移概率

=各时刻单符号信道转移概率的连乘

- 无记忆性：k时刻输出Y_k只与k 时刻输入X_k有关,与k时刻之前输入$X_1X_2\ldots X_{k-1}$无关
- 无预感性：k时刻之前输出$Y_1Y_2\ldots Y_{k-1}$只与k时刻之前输入$X_1X_2\ldots X_{k-1}$有关，与X_k无关

⇒ 离散无记忆N次扩展信道——无记忆，无预感

➢ 互信息和信道容量

离散无记忆N次扩展信道两端的平均互信息

$$\boldsymbol{I(X^N;Y^N)=H(Y^N) - H(Y^N|X^N)}$$

由于信道无记忆 $H(\boldsymbol{Y}^N|\boldsymbol{X}^N)=\sum_{k=1}^{N}H(Y_k|X_k)$

$$I(\boldsymbol{X}^N;\boldsymbol{Y}^N)=H(Y_1Y_2\ldots Y_N)-\sum_{k=1}^{N}H(Y_k|X_k)$$

当第$\boldsymbol{k}$个随机变量$\boldsymbol{X_k}$单独通过单符号离散信道时，在输出端得到的关于$\boldsymbol{X_k}$的平均互信息量

$$I(X_k;Y_k)=H(Y_k)-H(Y_k|X_k),\qquad k=1,2,\ldots,N$$

$$\sum_{k=1}^{N}I(X_k;Y_k)=\sum_{k=1}^{N}[H(Y_k)-H(Y_k|X_k)]=\sum_{k=1}^{N}H(Y_k)-\sum_{k=1}^{N}H(Y_k|X_k)$$

相减可得

$$I(\boldsymbol{X}^N;\boldsymbol{Y}^N)-\sum_{k=1}^{N}I(X_k;Y_k)=H(Y_1Y_2\ldots Y_N)-\sum_{k=1}^{N}H(Y_k)\le 0$$

$$I(\boldsymbol{X}^N;\boldsymbol{Y}^N)\le\sum_{k=1}^{N}I(X_k;Y_k)$$

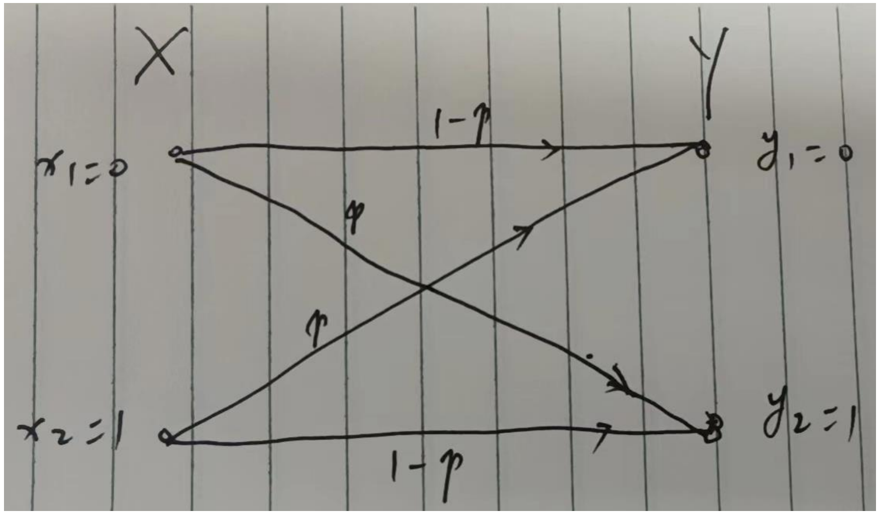
离散无记忆信道的N次扩展信道的平均互信息，不大于N个随机变量$X_1X_2\ldots X_N$单独通过信道{X,P(Y|X),Y}的平均互信息之和。

当且仅当信源$X^N=X_1X_2\ldots X_N$无记忆,或信源X^N是离散无记忆信源X的N次扩展信源时，等号成立。

事实上，$$H(\boldsymbol{Y}^N)=H(Y_1Y_2\ldots Y_N)=\sum_{k=1}^{N}H(Y_k)$$

离散无记忆信道的N次扩展信道，当输入端的N个输入随机变量统计独立时，信道的总平均互信息等于这N个变量单独通过信道的平均互信息之和。

- **例 二次无记忆对称信道的二次扩展信道的信道矩阵和信道容量。**



- **若p=0.1,则C=？**

信道+信源=无记忆+N次扩展

对于离散无记忆信源的N次扩展信源

$$X_k \in \{a_1, a_2, ..., a_n\}, \qquad \{p(a_1), p(a_2), ..., p(a_n)\}$$

通过同一个离散无记忆信道信道$\{\boldsymbol{X}, \boldsymbol{P(Y|X)}, \boldsymbol{Y}\}$
在信道输出端，随机变量序列$\boldsymbol{Y=Y_1Y_2\ldots Y_N}$中的随机变量$\boldsymbol{Y_k}$

$$Y_k \in \{b_1, b_2, ..., b_m\}, 有 I(X_k; Y_k) = I(X;Y)$$

$$\Rightarrow I(\boldsymbol{X};\boldsymbol{Y}) = \sum_{k=1}^{N} I(X_k; Y_k) = NI(X;Y)$$

$$\Rightarrow C^N = NC$$

离散无记忆信道的N次扩展信道，如果信源也是离散无记忆信源的N次扩展信源，则信道总的平均互信息量是单符号离散无记忆信道的平均互信息量的N倍。

■独立并联信道

➢ 定义

信道输入序列的各随机变量取值于不同符号集

信道输出序列的各随机变量亦取值于不同符号集

也称为独立并列信道、独立平行信道或积信道

➢信道容量 $C^N \le C_1 + C_2 + \cdots + C_N = \sum_{k=1}^{N} C_k$

当N个输入随机变量之间统计独立，并且每个输入随机变量X_k的概率分布为达到各自信道容量C_k的最佳分布时，C^N达到最大值

$$C_{\max}^N = \sum_{k=1}^{N} C_k$$

N个独立并联信道的信道容量等于各个信道容量之和

■ 信道编码定理(Shannon第二定理)

对离散平稳无记忆信道，其信道容量为***C***，输入序列长度为***L***。只要实际信息率***R***<***C***，就必可找到一种编码，当***L***足够长时，译码差错概率***Pe*** < ***ε***，***ε***为任意大于零的正数。反之，若实际信息率***R***>***C***，则对任何编码，***Pe***必大于零；且当 $L \to \infty$ 时，$P_e \to 1$ 。

[说明]

①给出了信息传输率的极限

只要***R***<***C***，必可无失真传输

若***R***>***C***，必为有失真传输

②存在性定理

失真不可避免，甚至必要；允许一定程度的失真，可以压缩信息。

信息率的失真函数

唯一可译，称之为保真码

- **Shannon第一定理：**

信息率(码率)$R=\frac{\overline{L}_N}{N}\log r \geq H(X)$时可实现无失真编码

- **现实：**

现实中所用信源编码几乎都是失真编码

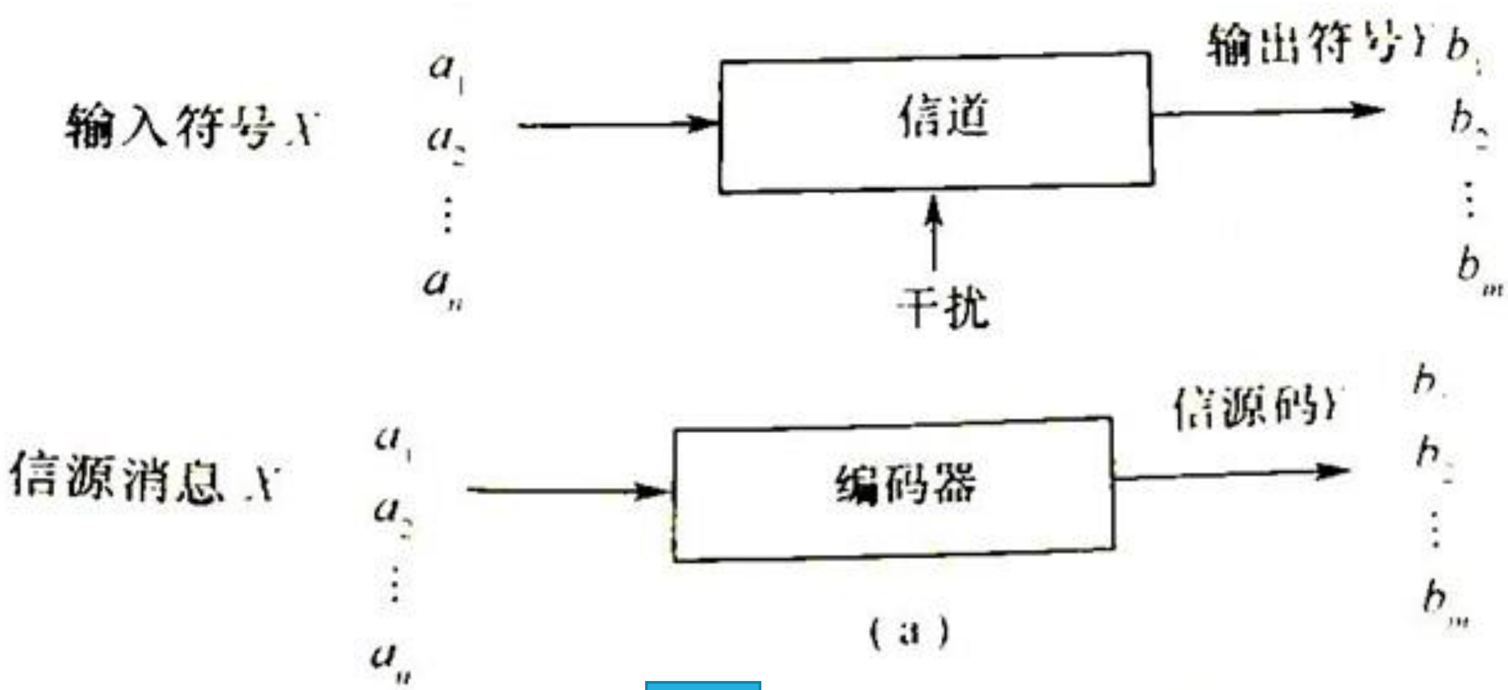
原因如下：

一是不得已而为之。现实中多数场合要想对原始消息编出保真码几乎不可能。编码的第一步是记录消息，到目前为止，我们还没有绝对精确的记录设备。或者即使能够记录下原始消息，如果是连续消息，Shannon 第一定理要求每个符号的平均码长携带信息不能小于原始信息熵，即，$R=\frac{\overline{L}_N}{N}\log r \geq H(X)$ 。但连续信源的信息熵往往无穷大，这对我们不可能。

其次，很多情况对原始消息完全保真也没必要，如，视觉暂留也会弥补连续画面的一定程度的断续。因此，从技术实现的可行性而言，失真编码的意义远大于无失真编码。

如果信源编码可以做到唯一可译时，则要求无失真编码。否则，我们适当引入失真度进行限失真编码，达到压缩数据等目的。

信源失真编码与信道的关系



编码引入的失真编码相当于信道有噪声干扰的结果。即，引入失真编码器可以等效于一个有噪信道。

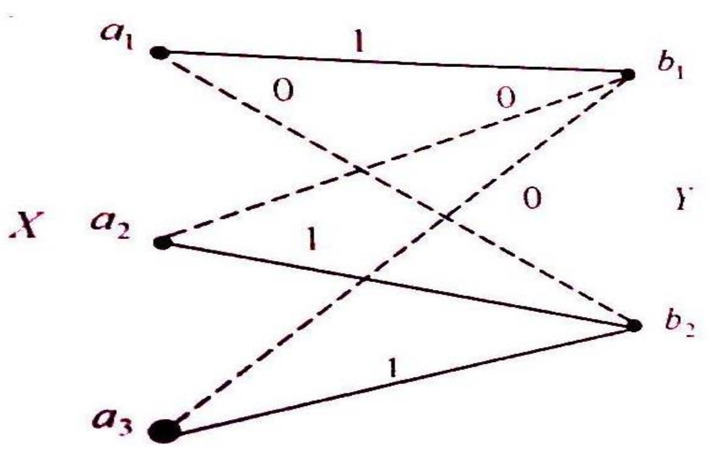
用信道理论研究限失真信源编码

❍编码器的行为与信道运行还是不同

编码器把一个消息变成某个信源码时，无失真时，转移概率$p(y|x)=0/1$;有失真时， $0\leqslant p(y|x)\leqslant 1$。如，失真编码是压缩性的归并编码，转移概率仍为0或1，如图。

对于归并性确定信道，信息率$R=I(X;Y)=H(Y)$，即，互信息等于编码后的新熵。

把信源编码和信道的互信息联系起来。当输入消息固定时，$I(X;Y)$由$p(y|x)$决定（由编码方法决定）。



❍信道与编码的逻辑关系

消息X进入 → (视为信道) → 按转移概率 $p(y|x)$ 处理 → 确定 $I(X;Y)$

|| ||

消息X进入 → (视为编码器) → 按 $p(y|x)$ 方法编码 → 信源编码后熵 $H(Y)$

输出Y的分布 $P(Y)=\sum_{X}P(X)P(Y|X)$

❍4.1失真度

定义 4.1 设单符号离散信源

$$\begin{pmatrix} X \\ P(X) \end{pmatrix} = \begin{pmatrix} x_1 & x_2 & \cdots & x_n \\ p(x_1) & p(x_2) & \cdots & p(x_n) \end{pmatrix},$$

通过信道传送到接收端的离散变量 Y 的概率空间为

$$\begin{pmatrix} Y \\ P(Y) \end{pmatrix} = \begin{pmatrix} y_1 & y_2 & \cdots & y_n \\ p(y_1) & p(y_2) & \cdots & p(y_n) \end{pmatrix}。$$

指定一个非负的实函数$d: X \times Y \to \overline{\mathbb{R}^-}$，满足，

对于每一对(x_i, y_j)，称$d(x_i, y_j)$为单个符号的失真度或失真函数，

$i = 1, 2, \cdots, n$； $j = 1, 2, \cdots, m$

❍说明：

1) $d(x_i, y_j)$表示：信源发出一个符号x_i，而接收端收到y_j所引起的误差或失真的大小。

$d(x_i, y_j)$越小表示引起的失真越小；$d(x_i, y_j) = 0$表示没有失真。

2) |X|=n,|Y|=m ，令

$$\mathcal{D} = \begin{pmatrix} d(x_1, y_1) & d(x_1, y_2) & \cdots & d(x_1, y_m) \\ d(x_2, y_1) & d(x_2, y_2) & \cdots & d(x_2, y_m) \\ \vdots & \vdots & \ddots & \vdots \\ d(x_n, y_1) & d(x_n, y_2) & \cdots & d(x_n, y_m) \end{pmatrix}_{n \times m},$$

称$\mathcal{D}$为失真矩阵。

3) 常用的失真函数——人为规定。

失真函数是根据实际需要、失真引起的损失风险、主观感觉上的差别大小等因素而人为规定。

Hamming失真度

离散对称信道：n=m，$d(x_i,y_j)=\begin{cases}0, & x_i=y_j\\ 1, & x_i\neq y_j\end{cases}$。

这表示：当接收符号与发送的信源符号相同时，不存在失真/错误；当接收符号和发送符号不同时，失真存在且失真相同，均取值为**1**。

失真矩阵

$$\mathcal{D}=\begin{pmatrix}0 & 1 & \cdots & 1\\ 1 & 0 & \cdots & 1\\ \vdots & \vdots & \ddots & \vdots\\ 1 & 1 & \cdots & 0\end{pmatrix}_{n\times m}$$

称为 Hamming 失真矩阵。

平方误差失真度（函数）

$$d(x_i, y_j) = (y_j - x_i)^2,\ \ i = 1, 2, \cdots, n;\ \ j = 1, 2, \cdots, m$$

表示：若信源符号代表输出信号的幅度值，则“较大幅度失真”比“较小幅度失真”引起的错误更严重。

- 绝对失真：$d(x_i, y_j) = |y_j - x_i|$。

序列失真度：N次扩展信源的失真函数。

$X^N = X_1 \cdots X_N$，$X_i \in \{x_1, \cdots, x_n\}$。于是$|X^N| = n^N$。

在信道传递中，传递作用相当于单符号无记忆信道的 N 次扩展信道，接收符号序列 $Y^N = Y_1 \cdots Y_N$，$Y_i \in \{y_1, \cdots, y_m\}$，$|Y^N| = m^N$。

对于$X^N=\alpha_i=x_{i_1}\cdots x_{i_N}$，$Y^N=\beta_j=y_{j_1}\cdots y_{j_N}$，
定义失真函数为

$$d(\alpha_i,\beta_j)=d(x_{i_1}\cdots x_{i_N},y_{j_1}\cdots y_{j_N})=\sum_{k=1}^{N}d(x_{i_k},y_{j_k})\text{。}$$

表明：离散无记忆信道的 N 次扩展信道输入/输出之间的失真度等于序列中对应单个符号失真度之和。失真矩阵为$\mathcal{D}=(d(\alpha_i,\beta_j))_{n^N\times m^N}$。

❍4.1.2 平均失真度

1) 单符号的平均失真度：

定义平均失真度为失真函数的数学期望，即，$d(x_i, y_j)$在随机变量 X 和 Y 的联合概率空间的统计平均值

$$\bar{D} = E(d(x_i, y_j)) = \sum_{i=1}^{n} \sum_{j=1}^{m} p(x_i, y_j) d(x_i, y_j)$$

显然，$\bar{D} = f(p(x), p(y|x), d(x, y))$为信源、信道、失真度三者函数，此值描述了某一信源在试验信道传输下的失真大小。(讨论中需要确定试验信道)

2) 序列信源的平均失真度：

对于X^N，定义平均失真度为

$$\bar{D}(N) = E(d(\alpha_i, \beta_j)) = \sum_{i=1}^{n^N} \sum_{j=1}^{m^N} p(\alpha_i, \beta_j) d(\alpha_i, \beta_j)$$

当信源与信道均无记忆时，有$\bar{D}(N)=\sum_{k=1}^{N}\bar{D}_k$，其中$\bar{D}_k$为同一信源 X 在 N 个不同时刻通过同一信道所造成的平均失真度，因此等于单符号信源 X 通过信道所造成的平均失真度,即,$\bar{D}_k=\bar{D}$。

于是，$\bar{D}(N)=N\bar{D}$。即，离散无记忆 N 次扩展信源通过离散无记忆 N 次扩展信道的平均失真度是单符号的 N 倍。

4.2 信息率失真函数

4.2.1保真度准则

设 D 为所允许的失真度。离散无记忆信源的平均失真度$\overline{D}\le D$，称为保真度准则。

称 D 为允许平均失真度。

N 次扩展信源的保真度准则为$\bar{D}(N)\le ND$。

注：当信源和单个符号失真度固定时，选择不同的试验信道，相当于选择不同的信道编码方法，所得平均失真度不同：

$$\text{试验信道} \Leftrightarrow \text{信道编码} \Rightarrow \text{不同的平均失真度} \Rightarrow \begin{cases} \text{满足准则} \\ \text{不满足} \end{cases}$$

- D 失真许可的试验信道：满足保真度准则的信道。

令 $\mathcal{P}_D$ 表示所有 D 失真许可的试验信道的集合 $\mathcal{P}_D = \{p(y|x) : \bar{D} \leq D\}$。

对于离散无记忆信源的 N 次扩展信源和离散无记忆信道的 N 次扩展信道，$\mathcal{P}_{D(N)} = \{p(\beta|\alpha) : \bar{D}(N) \leq ND\}$。

4.2.2信息率失真函数

已知，当信源给定后，平均互信息量是信道转移概率的下凸函数，即有最小值。

定义：在许可试验信道集合中，总可以找到某一试验信道P(Y|X)，使信道信息传输率 I(X;Y)达到最小值，记作 R(D)，即，

$$R(D)=\min_{p(y|x)\in P_D} I(X;Y)。$$

称 R(D)为信息率失真函数，简称为率失真函数。

单位：比特/信源符号；D 为允许平均失真度。

- 离散无记忆 N 次扩展“信源＋信道”：

信息率失真函数$R_N(D)=\min_{p(\beta_j|\alpha_i)\in P_D} I(X^N;Y^N)$。

由于均无记忆，故$I(X^N;Y^N)=N\cdot I(X;Y)$，

即，$R_N(D)=NR(D)$。

对于单符号信源和单符号信道：信源给定，定义失真度后，在接收端来看，在满足保真准则条件下，寻找再现信源消息所需的最低平均信息量的信道，即，平均互信息量最小值。但我们没有一般方法寻找该信道，只能凭经验。

$$D \to P_D \to R(D) = \min_{p(y|x)\in P_D} I(X;Y)$$

4.2.3信息率失真函数的性质

下面讨论R(D)的一些性质。

1) 定义域:

自变量 D 是允许平均失真度，是人为规定的平均失真度$\bar{D}$的上界。

D 的取值范围：根据信源的概率分布和选定的失真函数来确定，即，不同试验信道下，要求$\bar{D}$的可能取值范围（D 的区间）。

- D_{min}和$R(D_{min})$:

$\bar{D} = \sum_{i=1}^{n}\sum_{j=1}^{m} p(x_i, y_j) d(x_i, y_j) \geq 0$。于是，允许平均失真度的下限$D_{min} = 0$。

$D_{min}=0$**表明：**

➢ 不允许有任何失真的情况。

➢ 对于通信信道而言，信道中没有干扰和噪声存在。或对信源进行无失真编码，输入、输出形成了一一映射关系。此时，信道或者试验信道传输的信息量即为信源的熵，有

$R(D_{min})=R(0)=H(X)$。

注：一般情况，给定信源和失真矩阵，最小平均失真计算公式为

$$D_{min}=min\{\sum_{i=1}^{n}\sum_{j=1}^{m}p(x_i)p(y_j|x_i)d(x_i,y_j)\}。$$

信源给定后，$p(x_i)$是固定的，故

$$D_{min}=\sum_{i=1}^{n}p(x_i)min\{\sum_{j=1}^{m}p(y_j|x_i)d(x_i,y_j)\}。$$

由于$d(x_i,y_j)$是已知的，于是只需选择试验信道的转移概率$p(y_j|x_i)$，对于每个x_i，使得$\sum_{j=1}^{m}p(y_j|x_i)d(x_i,y_j)$最小。

由$0 \le p(y_j|x_i) \le 1$，且$\sum_{j=1}^{m} p(y_j|x_i) = 1$，知，

适当选择每个x_i使得

$$\begin{cases} \sum_j p(y_j|x_i) = 1, & \text{对所有}d(x_i,y_j)\text{为最小值的}y_j \\ \sum_j p(y_j|x_i) = 0, & \text{对所有}d(x_i,y_j)\text{不为最小值}y_j \end{cases}$$

（信道矩阵行中找最小 d，使得$\sum_j p(y_j|x_i) = 1$，其余为 0，从而得到新的信道矩阵，可计算D_{min}）→ 得到对应的试验信道

于是，

$$\begin{aligned} D_{min} &= \sum_{i=1}^{n} p(x_i)\, min \left\{ \sum_{j=1}^{m} p(y_j|x_i) d(x_i,y_j) \right\} \\ &= \sum_{i=1}^{n} p(x_i) (\sum_{j=1}^{m} p(y_j|x_i)) \min_j d(x_i,y_j) \\ &= \sum_{i=1}^{n} p(x_i) \min_j d(x_i,y_j). \end{aligned}$$

从失真矩阵来看，平均失真最小值就是矩阵每行元素的最小值乘以对应符号概率，再求和（求平均------对信源符号求平均）。

- 当失真矩阵的每一行至少有一个 0 元素时，信源的平均失真度为 0。
- 当失真矩阵满足$D_{min}=0$（每行至少一个 0 外），某些列还有不止一个 0 时，信源符号有些符号是可以压缩合并而不带来任何失真，压缩后的信息率必然减小，此时 R(D)<H(x)。
- 给定 x，最小失真 d(x,y) 可能不唯一，故试验信道选择也不唯一。但是，只要满足下列条件最小平均失真即可。

$$\begin{cases}\text{对于所有}d(x_i,y_j)\text{为最小值的}y_j, & \text{取}\sum_j p(y_j|x_i)=1\\ \text{对于所有}d(x_i,y_j)\text{不为最小值的}y_j, & \text{取}\sum_j p(y_j|x_i)=0\end{cases}$$

例： 信源$X=\{0,1\},Y=\{0,1,2\}$，失真矩阵$D=\begin{pmatrix}0 & 1 & \frac{1}{2}\\ 1 & 0 & \frac{1}{2}\end{pmatrix}$。

求：(1) D_{min}。

(2) 给出有最小允许失真度的试验信道。

(3) $R(D_{min})$。

例. 设信源$\begin{pmatrix}X\\ P(X)\end{pmatrix}=\begin{pmatrix}0 & 1 & 2\\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3}\end{pmatrix}$。

信宿$Y=\{0,1\}$，失真矩阵$D=\begin{pmatrix}0 & 1\\ \frac{1}{2} & \frac{1}{2}\\ 1 & 0\end{pmatrix}$，求$D_{min}$。

- D_{max}和$R(D_{max})$

由 R(D)的定义，知在保真度准则下，R(D)是平均互信息量的最小值。

由于$I(X;Y)\geq 0$，即有$R(D)\geq 0$。当 R(D)=0 时，对应的平均失真最大。（这时 R(D)取最小值，最值为 0）这即为 R(D)函数定义域的上限值D_{max}。(对应信道最差！)

Problem： I(X;Y)= 0 的条件？

事实上，满足 R(D)=0 的 D 可有无限多个，只要满足$D\geq D_{max}$，取最小者定义为D_{max}即可。（对于其它$D\geq D_{max}$，有 R(D)=0）。

当 I(X;Y)= 0 时，X 与 Y 互相独立；这等价于通信中断，即，p(y|x)=p(y) 。这样，不同的$p(y_j)$均可使 R(D)=0，但是所造成的$\bar{D}$不同。

选取其中最小者定义为D_{max}，即，

$$D_{max}=\min_{p(y_j)}\bar{D}=\min_{p(y)}\sum_y p(y)\left(\sum_x p(x)d(x,y)\right)。$$

由于信源和失真函数给定，因此求D_{max}相当于寻找一种信道，其输出分布 p(y) 使上式右边最小。即，

$$\begin{cases} 选取使\sum\limits_i p(x_i)d(x_i,y_j)最小的p(y)=1. \\ 令其它的\sum\limits_i p(x_i)d(x_i,y_j)对应的p(y)=0. \end{cases}$$

于是有

$$D_{max}=\min_{p(y)}\sum_{i=1}^{n}\sum_{j=1}^{m}p(x_i)p(y_j)d(x_i,y_j)=\min_{p(y)}\sum_{i=1}^{n}p(x_i)d(x_i,y_j)$$

例： 二元信源$\begin{pmatrix} X \\ P(X) \end{pmatrix}=\begin{pmatrix} x_1 & x_2 \\ 0.4 & 0.6 \end{pmatrix}$。失真矩阵为$D=\begin{pmatrix} \alpha & 0 \\ 0 & \alpha \end{pmatrix}$，

计算D_{max}。

2) 信息率失真函数对允许平均失真度的下凸性

在定义域(D_{min}, D_{max})上，R(D)关于 D 是下凸函数，即，$\forall 0 \le \alpha \le 1$和任意平均失真度$D_1, D_2 \le D_{max}$，有

$$R(\alpha D_1 + (1-\alpha)D_2) \le \alpha R(D_1) + (1-\alpha)R(D_2)。$$

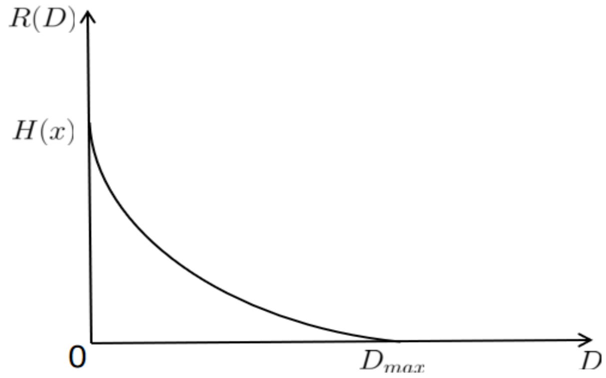
3) R(D)的单调性和连续性

- 连续性：由凸性可得。
- 单调性：严格递减，即，在$D_{min} \le D \le D_{max}$范围内，R(D)不可能为常数。

说明：直观地，R(D)的非递增性，表明：允许失真越大，所要求的信息率越小。

$$\begin{cases} R(0) = H(x)\text{：无失真} \\ R(D_{max}) = 0\text{：（完全失真）}X\text{与}Y\text{无关} \end{cases}$$

于是，R(D)的函数图像如图：



4.3 离散信源的信息率失真函数

对于离散信源，计算

$$\left.\begin{cases}\text{信息率失真函数}\\ \text{信道容量}\end{cases}\right\}$$在约束条件下求平均互信息的极值问题。

取极大值后仅与信道有关。

差别：约束条件不同

1) 信道容量 C：信源（信道固定），求平均互信息的条件极大值。

2) 信息率失真函数 R(D)：信道（信源固定），已知信源的概率分布和失真度函数条件下，求平均互信息的条件极小值。

取极小值后仅与信源有关。

用 Langrange 乘子法：

- 原则上可求。
- 求显式表达式困难。
- 仅能求出信息率失真函数的参量表达式。

❍4.3.1信息率失真函数的计算——拉格朗日乘子法

引入变量 S 作为参数，信息率失真函数为 R(S)，失真函数为 D(S)。

设给定信源 X，$P(X)=\{p(x_i):i=1,2,\cdots,n\}$；规定失真函数为$d(x_i,y_j)$,$i=1,2,\cdots,n;j=1,2,\cdots,m$。选定允许平均失真度为$\bar{D}$，则信源 X 的信息率失真函数 R(D)的约束条件：

$$\begin{cases} p(y_j|x_i)\geq 0, \quad i=1,2,\cdots,n;j=1,2,\cdots,m. \\ \sum\limits_{j=1}^{m} p(y_j|x_i)=1, \quad i=1,2,\cdots,n. \\ \sum\limits_{i=1}^{n}\sum\limits_{j=1}^{m} p(x_i)p(y_j|x_i)d(x_i,y_j)=\bar{D} \\ p(x)\text{固定} \end{cases}$$

$R(D)=min\,I(X;Y)$，其中

$I(X;Y)=\sum\limits_{i=1}^{n}\sum\limits_{j=1}^{m} p(x_i)p(y_j|x_i)\log\dfrac{p(y_j|x_i)}{\sum\limits_{k=1}^{n} p(x_k)p(y_j|x_k)}$为 p(y|x)的下凸函数。

取拉格朗日常数$\mu_i, i = 1, 2, \cdots, n$，对应于

$\sum_{j=1}^{m} p(y_j|x_i) = 1$，$i = 1, 2, \cdots, n$。

构成辅助函数

$$F = I(X;Y) - \sum_i \mu_i \left(\sum_{j=1}^{m} p(y_j|x_i) - 1 \right) - S \left(\sum_{i=1}^{n} \sum_{j=1}^{m} p(x_i) p(y_j|x_i) d(x_i, y_j) - \bar{D} \right)$$

计算$\frac{\partial F}{\partial p(y|x)} = 0$

{eq1} $$p(x_i) \ln \frac{p(y_j|x_i)}{p(y_j)} - \mu_i - S \cdot p(x_i) d(x_i, y_j) = 0$$

{eq*} $$p(y_j|x_i) = p(y_j) \lambda_i \cdot e^{S \cdot d(x_i, y_j)}$$

{eq2} $$\lambda_i = \frac{1}{\sum_{j=1}^{m} p(y_j) e^{S \cdot d(x_i, y_j)}}$$

{eq3} $$\sum_{i=1}^{n} \lambda_i p(x_i) e^{S \cdot d(x_i, y_j)} = 1$$

{eq4} $$p(y_j|x_i)=p(y_j)\lambda_i e^{S\cdot d(x_i,y_j)}=\frac{p(y_j)\cdot e^{S\cdot d(x_i,y_j)}}{\sum\limits_{j=1}^{m}p(y_j)e^{S\cdot d(x_i,y_j)}}$$

{eq5} $$D(S)=\sum_{i=1}^{n}\sum_{j=1}^{m}p(x_i)p(y_j)d(x_i,y_j)\cdot\lambda_i\cdot e^{S\cdot d(x_i,y_j)}$$

{eq6} $$R(D)=\sum_{i=1}^{n}\sum_{j=1}^{m}p(x_i)p(y_j)d(x_i,y_j)\cdot\lambda_i\cdot e^{S\cdot d(x_i,y_j)}\ln(\lambda_i\cdot e^{S\cdot d(x_i,y_j)})$$
$$=S\cdot D(S)+\sum_{i=1}^{n}p(x_i)\ln\lambda_i.$$

一般情况下，S无法消失，因此得不到R(D)的显式解。但对某些特定的简单问题，可以消去S。

利用{eq2}-{eq6})式求R(D)。

由{eq3}，解得λ_i;

由{eq2}，解得$p(y_i)$;

再由{eq4},解得$p(y_j|x_i)$;

将上述代入{eq5}和{eq6}，得$D(S)$，由此知，S，$R(D)$;

将 S 代入各式得$\lambda_i, p(y_i), p(y_j|x_i)$。

S 性质：

- S 取值有限：由$p(y_j) \geq 0$可知。
- 将 S 视为 D 的函数，对 R(D)求导$\frac{dR(D)}{dD}$，整理可得，$\frac{dR(D)}{dD} = S$。由此

可知，S 为 R(D)的斜率。

$$D \in (0, D_{max}), \begin{cases} D \to 0\text{时}, S \to -\infty \\ D > D_{max}, R(D) = 0, S = 0. \end{cases}，即，S \in (-\infty, 0).。$$

4.3.2二元离散信源信息率失真函数的计算

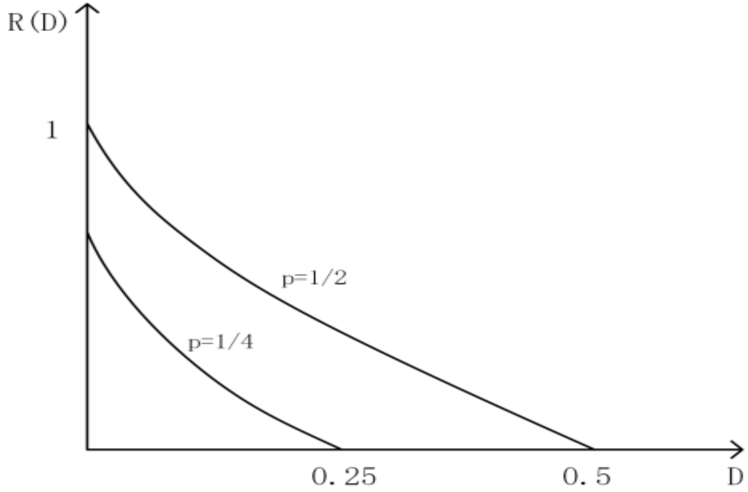
考虑二元离散无记忆信源

$$\begin{pmatrix} X \\ P(X) \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ p & 1-p \end{pmatrix}, 0 \leq p \leq \tfrac{1}{2}，Y = \{0, 1\}。$$

失真矩阵为 Hamming 失真矩阵$D = \begin{pmatrix} 0 & \alpha \\ \alpha & 0 \end{pmatrix}$。求 R(D)。

$$R(D) = \begin{cases} H(p) - H(D/\alpha), & 0 \leq D \leq D_{max} = \alpha p \\ 0, & D > \alpha p \end{cases}$$

如图，当$p=\frac{1}{2}$和$p=\frac{1}{4}$时 R(D)的曲线。



说明：

1) 信源的 R(D)计算困难。
2) 信源的概率空间仅是一种模型，很难准确计算概率分布。
3) 精确计算 R(D)函数是没有必要的。一般，从失真要求出发，估算几个特殊点，如$D=0, D_{max}$以及几个 D 值对应的 R(D)，可以画出 R(D)函数曲线。
4) 信息率失真理论给出，给定失真度 D 条件下，信源输出的信息率所能压缩的极限 R(D)，但没有给出压缩方法，但可作为一尺度，衡量压缩编码方法的压缩效果。

4.4 保真度准则下的信源编码定理

Shannon 第三定理，即为保真准则下的信源偏码定理：

设一离散平稳无记忆信源，输出随机变量序列为$\vec{X}=(X_1X_2\cdots X_L)$。若

该信源的信息率失真函数是 R(D)，并选定有限的失真函数，对于任意允许平均失真度$D\geq 0$和$\forall\epsilon>0$，当信息率 R>R(D)时，只要信源序列长度 L 足够大，一定存在一种编码方式 C,使得译码后的平均失真度$\bar{D}(C)\leq D+\epsilon$。反之，若 R<R(D),则对$\forall C$有$\bar{D}(C)>D$,即，译码平均失真度必大于允许平均失真度。

说明：

1) 这是有失真信源压缩的理论基础：允许失真度 D 确定后，总存在一种编码方法，使编码的信息率 R 可任意接近于 R(D)函数，而且平均失真度$\bar{D}(C) \leq D + \epsilon$。反之，若 R<R(D),则编码的平均失真度将大于 D。总之，信息率失真函数是一个界限。

$$\begin{cases} R(D) \leq R, & \text{则}\bar{D}(C) \leq D + \epsilon\text{，失真可控，即，即使有失真，仍满足要求。} \\ R(D) > R, & \text{不可控，不能满足要求} \end{cases}$$

Shannon Th1. 无失真信源编码 H(X);

Shannon Th2. 信道编码;

Shannon Th3. 限失真信源编码 R(D)<H(X)。

均为存在性定理。

信道容量与信息率失真函数的对偶性：

二者都是求平均互信息极值问题：

$$C=\max_{p(x_i)} I(X;Y),\ R(D)=\min_{p(y_j|x_i)\in P_D} I(X;Y)。$$

1) 二者都是求平均互信息量，单位：比特/符号。

- 信道容量：信道固定，计算 I(X;Y)的极大值；其理论依据为 I(X;Y)是输入信源概率分布$\{p(x_i)|i\}$的上凸函数。
- 信息率失真函数：在试验信道（满足保真准则的信道）中求 I(X;Y)的极小值。其理论依据：I(X;Y)为信道转移概率分布$p(y|x)$的下凸函数。

2) 信道容量一旦求出后，就仅与 p(y|x)有关了，与信源无关，反映信道特性。信息率失真函数一旦求出后，只与输入信源的概率分布$\{p(x_i)|i\}$有关了，与信道无关，反映信源特性。

3) 信道容量，解决通信的可靠性问题，是信息传输的理路基础，通过信道编码增加信息冗余度来实现。信息率失真函数是为解决通信的有效性问题，是压缩理论的基础，通过信源编码，减少信息的冗余度来实现。