基础信息论

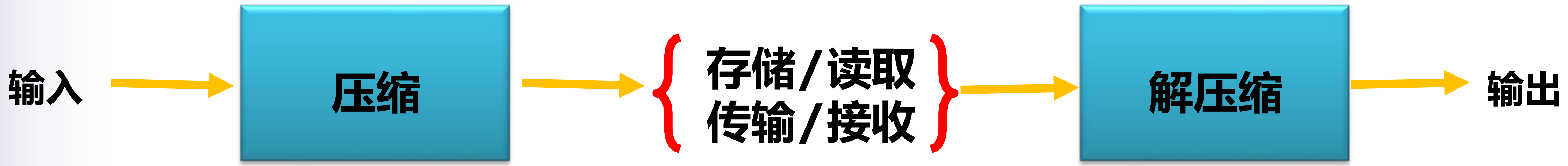
信息率失真问题建模

华中科技大学电信学院

简要回顾

- 绪论
  - 通信系统模型、香农信息论的研究方法及研究内容
- 离散信源熵
  - 单符号离散信源自信息量、互信息量、熵和平均互信息的定义及性质；多符号离散信源
- 无失真离散信源编码
  - 信源编码定理、香农编码、赫夫曼编码、费诺编码
- 离散信道容量
  - 单符号离散信道容量的定义、几种特殊信道的信道容量的计算、一般信道的信道容量的计算；多符号离散信道容量的计算
- 纠错编码
  - 平均错误概率、译码规则、香农第二定理
- 连续信源熵和信道容量
  - 连续信源、熵功率、连续信道信道容量的定义、高斯信道的信道容量的计算（香农公式及香农限）
- 率失真函数
  - 率失真函数的定义、定义域和值域、率失真函数的参量表达式

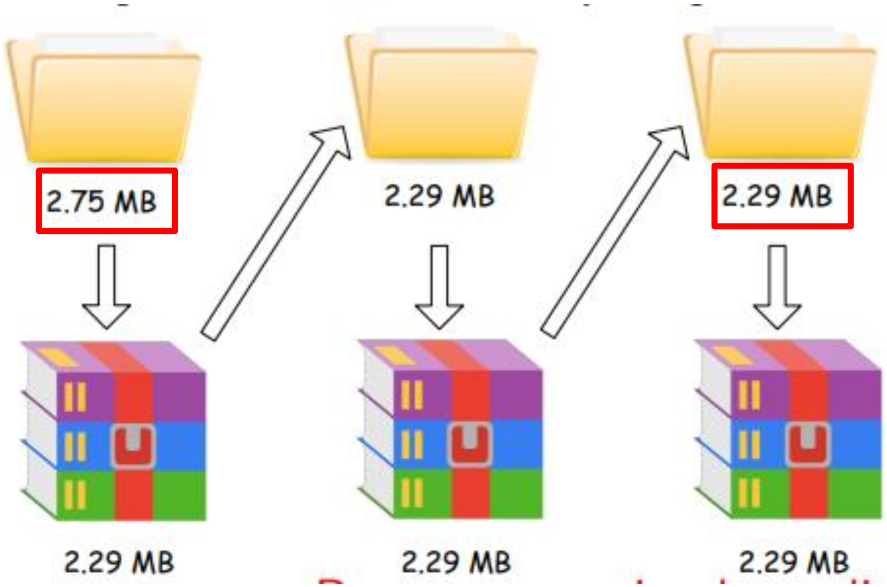
无损信源编码 vs 有损信源编码



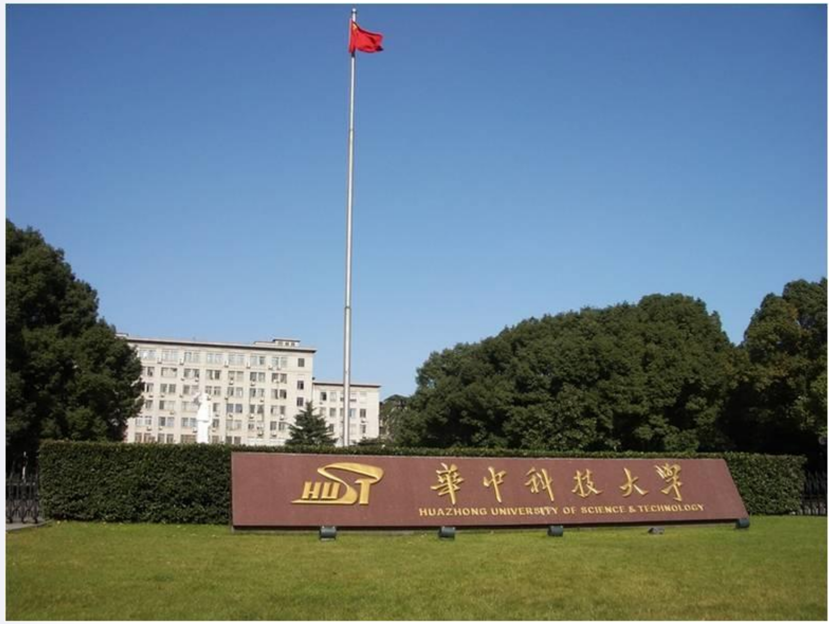
- 无损压缩
  - 输出信息流与输入信息论完全相同.
  - 一般用于文档、消息、数据集等
- 有损压缩
  - 输出信息流与输入信息流相近或者相似.
  - 一般适用于多媒体数据流，因为人的耳朵和眼睛能够容忍一定错误

无损压缩

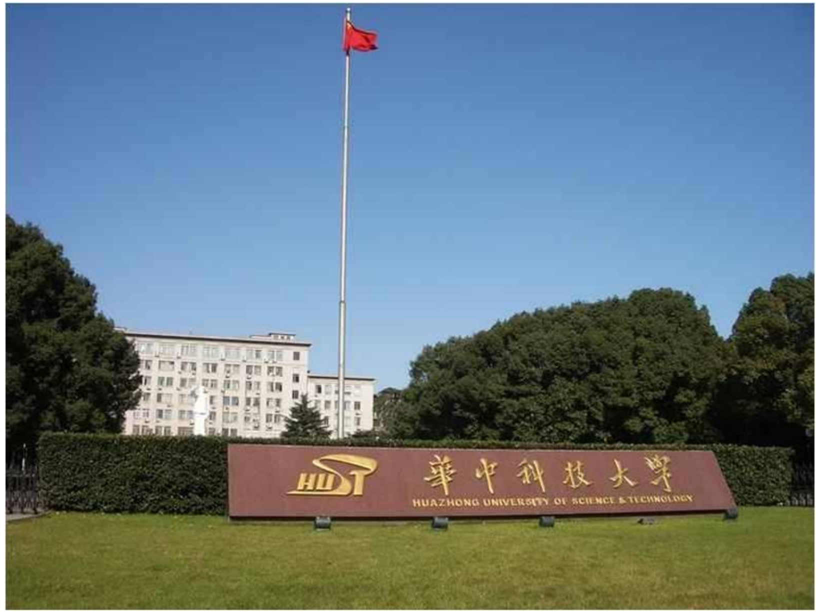
- 演示：使用WinRAR或者7Zip工具压缩文件
- 视频播放



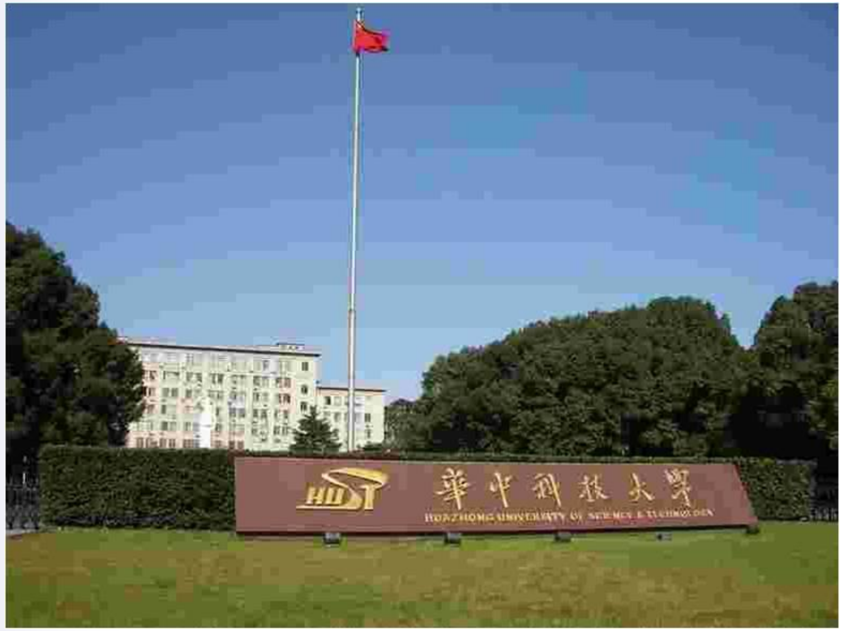
有损压缩



340 KB



33.8 KB



14.5 KB

8.87 KB

有损压缩编码应用

- 实际生活中，并不要求获得完全无失真的消息。
- 通常只要求近似地再现原始消息，即允许一定的失真存在

例1：放电影

- 视觉有暂留性，故传送每秒25帧的图像就能满足人类通过视觉感知信息的要求，而不必占用更大的信息传输率；

例2：听音乐

- 大多数人只能听到几千赫兹到十几千赫兹，对于经过专业训练的音乐家，一般也不过听到20kHz的声音。

例3：打电话

- 由于人耳的听觉特性（接收信号带宽和分辨率均有限），话音有失真，人也可以听懂

信道传输允许一定失真

前面

- 基本出发点都是如何保证信息的无失真传输。

发现

- 许多实际应用，人们并不要求完全无失真地恢复消息，而是只要满足一定的条件，近似地恢复信源发出的消息就可以了。

结论

- 实际应用要求在保证一定质量前提下在信宿近似地再现信源输出的信息，或者说在保真度准则下允许信源输出存在一定的失真。

需要解决的问题

什么是允许的失真?

如何对失真进行描述?

信源输出信息率被压缩的最大程度是多少?

信息率失真理论回答了这些问题，其中香农的限失真编码定理定量地描述了失真，研究了信息率与失真的关系，论述了在限失真范围内的信源编码问题，已成为量化、数据转换、频带压缩和数据压缩等现代通信技术的理论基础。

本章讨论的问题：

1. 对于给定的信源（即给定信源熵H(X)），在允许的失真条件下，信源熵所能压缩的极限（即信息率失真函数 R(D)）**理论**值是多少?

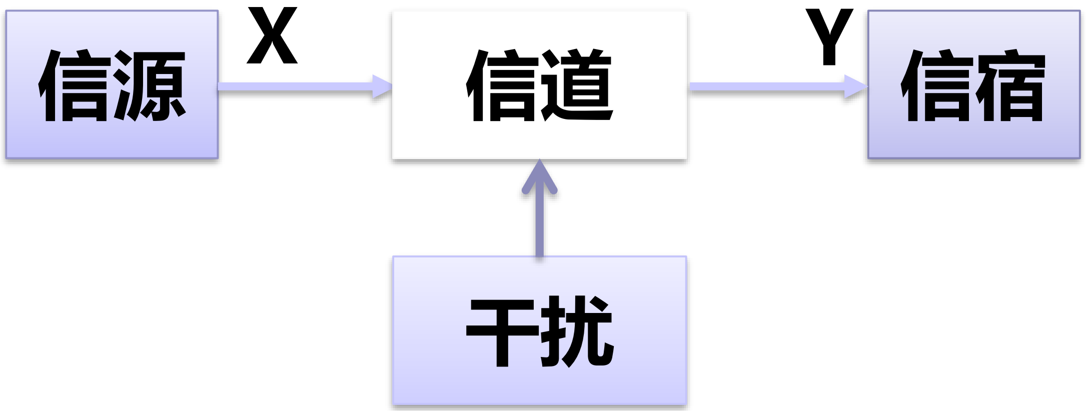
2. 信息率失真理论是研究信源熵的**压缩问题**，采用研究信道的方法，即在数学上将信源熵压缩看成通过一个信道，寻找在保真度准则下的最小的平均互信息。

3. 信息率失真理论是信号量化、模数转换、频带压缩和数据压缩的理论基础，在图像处理、数字通信等领域得到广泛**应用**。

失真函数与平均失真度

失真的测度

简化的通信系统模型



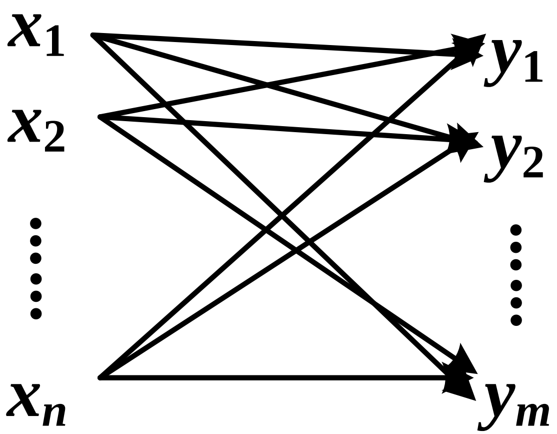
- 信道中固有的噪声和不可避免的干扰，信源发出的消息在信道传输过程中可能会出现误差和失真。
- 失真如何定量表述?

失真的测度-失真函数

- 设信源 $$\begin{bmatrix} X \\ P(X) \end{bmatrix} = \begin{bmatrix} x_1 & x_2 & \cdots & x_n \\ p(x_1) & p(x_2) & \cdots & p(x_n) \end{bmatrix}$$
- 信源符号通过信道传送信宿
$$\begin{bmatrix} Y \\ P(Y) \end{bmatrix} = \begin{bmatrix} y_1 & y_2 & \cdots & y_m \\ p(y_1) & p(y_2) & \cdots & p(y_m) \end{bmatrix}$$
- 定义失真函数：
  - 对每一对(x_i, y_j)，指定一个非负函数 $\boxed{d(x_i, y_j) \geq 0}$
  - 表示信源发出符号x_i，接收端收到符号为y_j的失真
  - 单个符号的失真度/失真函数

失真测度：失真矩阵

- 失真矩阵 D：所有的失真函数排列起来



$$[D]=\begin{bmatrix} d(x_1,y_1) & d(x_1,y_2) & \cdots & d(x_1,y_m) \\ d(x_2,y_1) & d(x_2,y_2) & \cdots & d(x_2,y_m) \\ \vdots & \vdots & & \vdots \\ d(x_n,y_1) & d(x_n,y_2) & \cdots & d(x_n,y_m) \end{bmatrix}$$

失真函数的形式

- 失真函数的形式可根据需要任意选取，通常由失真引起的损失、风险、主观感觉上的差别大小等因素人为规定的。

例1 设信源符号序列为$X=[0,1]$，接收端收到符号序列为$Y=[0,1,2]$规定失真函数为$d(0,0)=d(1,1)=0$，$d(0,1)=d(1,0)=1$，$d(0,2)=d(1,2)=0.5$，求失真矩阵。

解：$$[D]=\begin{bmatrix} d(x_1,y_1) & d(x_1,y_2) & d(x_1,y_3) \\ d(x_2,y_1) & d(x_2,y_2) & d(x_2,y_3) \end{bmatrix}$$

$$=\begin{bmatrix} d(0,0) & d(0,1) & d(0,2) \\ d(1,0) & d(1,1) & d(1,2) \end{bmatrix}=\begin{bmatrix} 0 & 1 & 0.5 \\ 1 & 0 & 0.5 \end{bmatrix}$$

常见失真函数

平方代价函数、绝对代价函数、均匀代价函数等。

第一种常见失真函数：

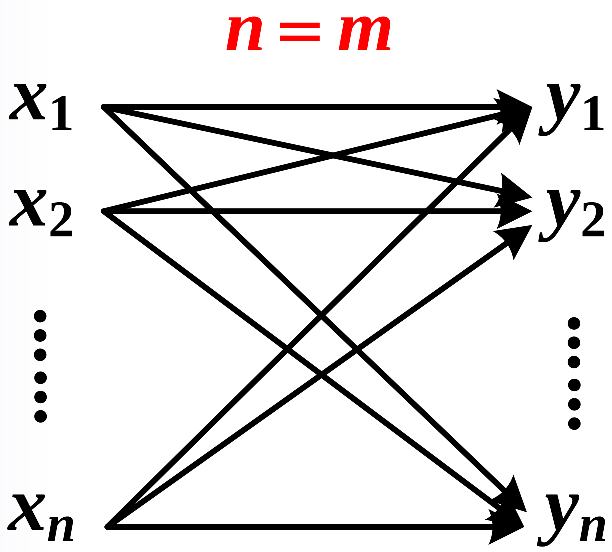
$$d(x_i, y_j) = \begin{cases} 0, & i = j \\ a, & a \ge 0, i \ne j \end{cases}$$

$$[D] = \begin{bmatrix} 0 & a & a & \dots & a \\ a & 0 & a & \dots & a \\ & \dots & \dots & & \\ a & a & a & \dots & 0 \end{bmatrix}$$

- **特点：**
  - 当$i=j$,X与Y取值一样，表示用Y来代表 X 就没有误差，定义失真度为0
  - $i \ne j$时，用Y代表X就有误差
  - 这种定义认为对所有不同的i和j引起的误差都一样，所以定义失真度为常数a

常见失真函数

- 当$a = 1$时，失真函数称为汉明失真函数。



汉明失真矩阵
$$\begin{bmatrix} 0 & 1 & \dots & 1 \\ 1 & 0 & \dots & 1 \\ \dots & \dots & \dots & \dots \\ 1 & 1 & \dots & 0 \end{bmatrix}$$

常见失真函数

- 第二种常见失真函数：

$$d(x_i,y_j)=(y_j-x_i)^2$$

- 这种函数称为平方误差失真函数，失真矩阵称为平方误差失真矩阵。

特点：

- 一般用于表示由幅度变化引起的失真，若信源符号代表输出信号的幅度值，则较大的幅度失真比较小的幅度失真引起的错误更为严重，严重程度用平方表示。
- 多用于连续信源。

失真函数的定义推广到矢量传输

- **假定**： 离散矢量信源的N长符号序列为 $X=[X_1,X_2,X_3,\ldots,X_N]$ 其中，第i个符号 X_i 的取值为 $\{x_1,x_2,\ldots,x_n\}$ ，经离散无记忆信道 $P(Y/X)$ 传输后，接收端收到的N长符号序列为 $Y=[Y_1,Y_2,Y_3,\ldots,Y_N]$ ，其中第i个符号 Y_i 的取值为 $\{y_1,y_2,\ldots,y_m\}$ ，
- 其失真函数的定义为：

$$d_N(X,Y)=\sum_{i=1}^{N}d(X_i,Y_i)$$

或

$$\boxed{d(a_i,b_j)=d(x_{i_1}x_{i_2}\ldots x_{i_N},y_{j_1}y_{j_2}\ldots y_{j_N})=\sum_{k=1}^{N}d(x_{i_k},y_{j_k})}$$

- 对应的失真矩阵共有 $n^N\times m^N$ 个元素

失真函数举例

- 假定 离散矢量信源$N=3$，输出矢量序列为$\mathbf{X}=X_1X_2X_3$，其中X_i，$i=1,2,3$的取值为$\{0,1\}$，经信道传输后的输出为$\mathbf{Y}=Y_1Y_2Y_3$，其中Y_i，$i=1,2,3$的取为$\{0,1\}$，定义失真函数$d(0,0)=d(1,1)=0$，$d(0,1)=d(1,0)=1$，求矢量失真矩$[D_N]$。
- 解：由失真函数的定义得：

$$d_N\ (\mathbf{X},\mathbf{Y})\ =\sum_{i=1}^{N}d(X_i,Y_i)=[d(X_1,Y_1)+d(X_2,Y_2)+d(X_3,Y_3)]$$

$$d_N(000,000)=[d(0,0)+d(0,0)+d(0,0)]=[0+0+0]=0$$

$$d_N(000,001)=[d(0,0)+d(0,0)+d(0,1)]=[0+0+1]=1$$

失真函数例2-续

- 类似可以得到其它元素数值，矢量失真矩阵为

$$[d_N]=\begin{bmatrix} 0 & 1 & 1 & 2 & 1 & 2 & 2 & 3 \\ 1 & 0 & 2 & 1 & 2 & 1 & 3 & 2 \\ 1 & 2 & 0 & 1 & 2 & 3 & 1 & 2 \\ 2 & 1 & 1 & 0 & 3 & 2 & 2 & 1 \\ 1 & 2 & 2 & 3 & 0 & 1 & 1 & 2 \\ 2 & 1 & 3 & 2 & 1 & 0 & 2 & 1 \\ 2 & 3 & 1 & 2 & 1 & 2 & 0 & 1 \\ 3 & 2 & 2 & 1 & 2 & 1 & 1 & 0 \end{bmatrix}$$

平均失真$\bar{D}$

由于a_i和b_j都是随机变量，所以失真函数$d(a_i, b_j)$也是随机变量，且只能表示两个特定的具体符号

- 限失真时的失真值，只能用它的数学期望或统计平均值。将失真函数的数学期望称为**平均失真度**，记为

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

$$= \sum_{i=1}^{n}\sum_{j=1}^{m} p(x_i) \cdot p(y_j/x_i) \cdot d(x_i, y_j)$$

其中，$p(x_i y_j)$是联合概率分布，$i = 1,2,\cdots,n$；$j = 1,2,\cdots,m$；
$p(x_i)$是信源符号概率分布，$i = 1,2,\cdots,n$；
$p(y_j|x_i)$是转移概率分布 $i = 1,2,\cdots,n$，$j = 1,2,\cdots,m$；

平均失真度的意义

- 是对给定信源分布$\{p(x_i)\}$在给定转移概率分布$\{p(y_j|x_i)\}$的信道中传输时的失真的**总体度量**。

- 在平均意义上衡量信道每传递一个符号所引起的失真的大小。

- 它是信源统计特性$p(x_i)$、信道统计特性$p(y_j|x_i)$和失真度$d(x_i,y_j)$的函数。当$p(x_i)$，$p(y_j|x_i)$和$d(x_i,y_j)$给定后，平均失真度就不是一个随机变量了，而是一个**确定的量**。

- 如果信源和失真度一定，就只是信道统计特性的函数。信道传递概率不同，平均失真度随之改变。

矢量传输的平均失真定义

- 若信道输入和输出均为N长的符号序列

$$\mathbf{X}=[X_1,X_2,X_3,\cdots,X_N]\ ,\quad \mathbf{Y}=[Y_1,Y_2,Y_3,\cdots,Y_N]$$

- 其中，第i个位置上的符号取值为

$$X_i\in\{x_1,x_2,\cdots,x_n\}\quad Y_i\in\{y_1,y_2,\cdots,y_m\}$$

- 则平均失真度为

$$\overline{\mathrm{D_N}}=E[D_N]=\sum_{i=1}^{N}E[d(X_i,Y_i)]=\sum_{i=1}^{N}\overline{D_i}$$

其中$\overline{\mathrm{D_i}}$是第i个位置上符号的平均失真。

离散平稳无记忆信源的N次扩展

离散平稳无记忆信源：　　　　离散单符号信源

每次输出一个符号

时刻1　X_1　　$\xrightarrow{\cdots\ X_2\ X_1}$

时刻2　X_2

$\vdots$

$$\begin{bmatrix} x_1 & x_2 & \dots & x_n \\ p(x_1) & p(x_2) & \dots & p(x_n) \end{bmatrix}$$

以此为基础进行N次扩展，得：

每次输出一组，组内独立

$\cdots\ \boxed{X_{2N}\dots X_{N+2}X_{N+1}}\ \ \boxed{X_N\dots X_2X_1}$

第一次　$X_1X_2\dots X_N$　$\longrightarrow$

第二次　$X_{N+1}X_{N+2}\dots X_{2N}$

$\vdots$

$$\begin{bmatrix} x_1 & x_2 & \dots & x_n \\ p(x_1) & p(x_2) & \dots & p(x_n) \end{bmatrix}$$

- 如果信源是离散无记忆N次扩展信源，且信道是离散无记忆N次扩展信道，**求N次扩展信道的平均失真度：**

信源

$$X^N = X_1 X_2 \cdots X_N$$

$$a_i = x_{i_1} \cdots x_{i_N}$$

$$x_{i_1}, \cdots, x_{i_N} \in \{a_1 \cdots a_n\}$$

$$i_1, \cdots, i_N = 1, \cdots, n$$

$$i = 1,2, \cdots, n^N$$

信宿

$$Y^N = Y_1 Y_2 \cdots Y_N$$

$$b_j = y_{j_1} \ldots y_{j_N}$$

$$y_{j_1}, \cdots, y_{j_N} \in \{b_1 \cdots b_m\}$$

$$j_1, j_2, \cdots, j_N = 1 \cdots m$$

$$j = 1,2, \cdots, m^N$$

失真度为： $d(a_i,b_j)=d(x_{i_1}\cdots x_{i_N},y_{j_1}\cdots y_{j_N})$

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

由信源和信道的无记忆性：

$$p(a_i)=\prod_{k=1}^{N}p(x_{j_k});\quad p(b_j/a_i)=\prod_{k=1}^{N}p(y_{j_k}/x_{j_k})$$

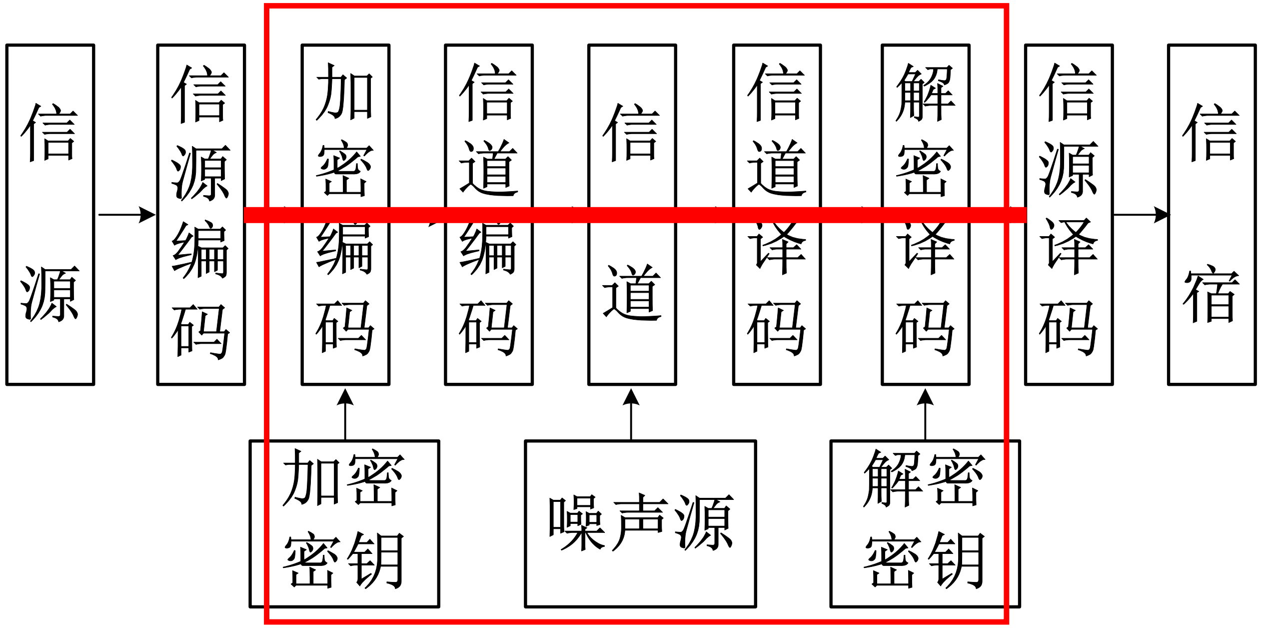
计算平均失真度：

$$\bar{D}(N)=\sum_{i=1}^{n^N}\sum_{j=1}^{m^N}p(a_i)p(b_j/a_i)d(a_i,b_j)=\bar{D}_1+\cdots+\bar{D}_N$$

- **如果信源是离散无记忆N次扩展信源，且信道是离散无记忆N次扩展信道，则每个位置上符号的平均失真$\overline{D_i}$相等，且等于矢量平均失真**
$\overline{D_N}=N\overline{D_i}$， $i=1,2,\ldots,N$。

信息率失真函数

以通信系统角度研究信源编码



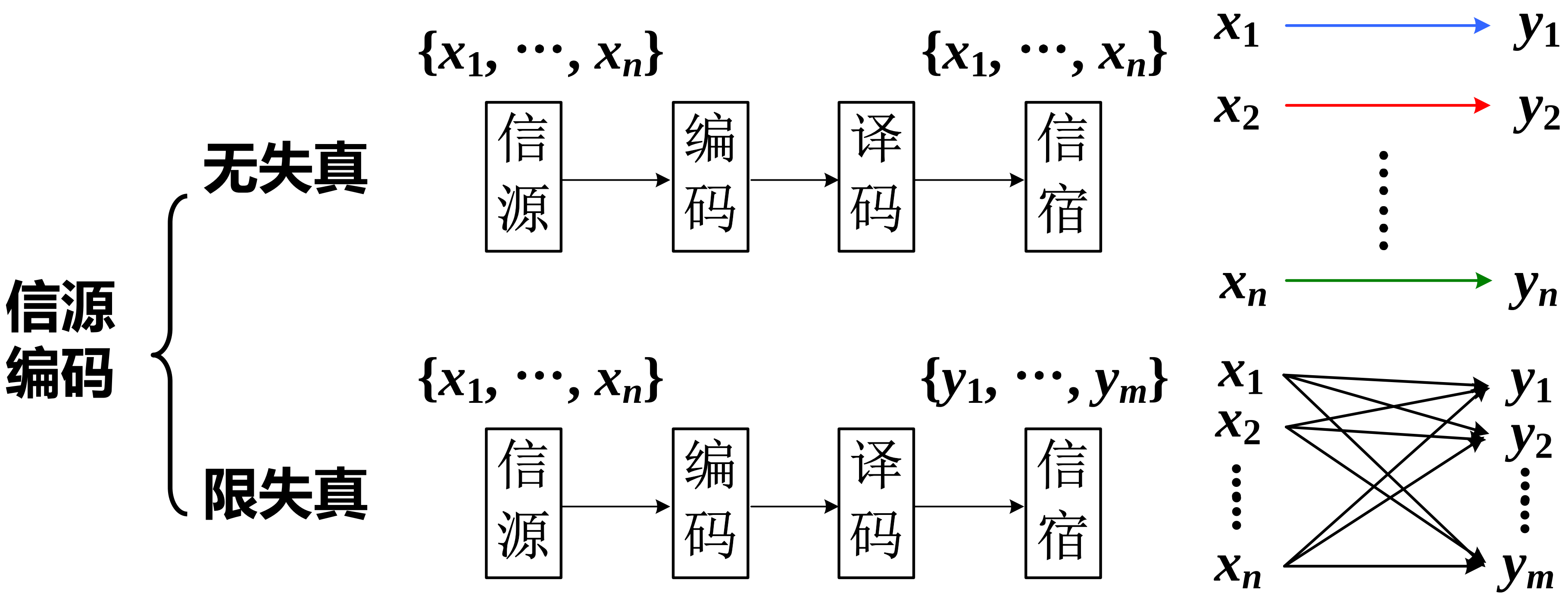
信息论中的通信系统模型

研究信源编码时所做的简化：

(1) 无需考虑信道编码和保密编码。

(2) 传输信道是理想的。

研究信源编码时的通信系统模型



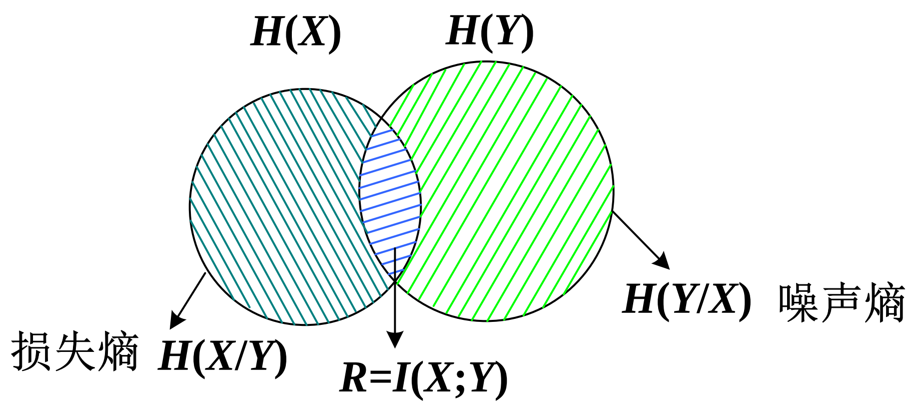
只考虑编码

X → **信源编码器** → Y　$y \in \{b_1, b_2, \cdots b_n\}$

$x \in \{a_1, a_2, \cdots a_n\}$

⇩ **将信源编码器看作信道**

→ **假想信道** →

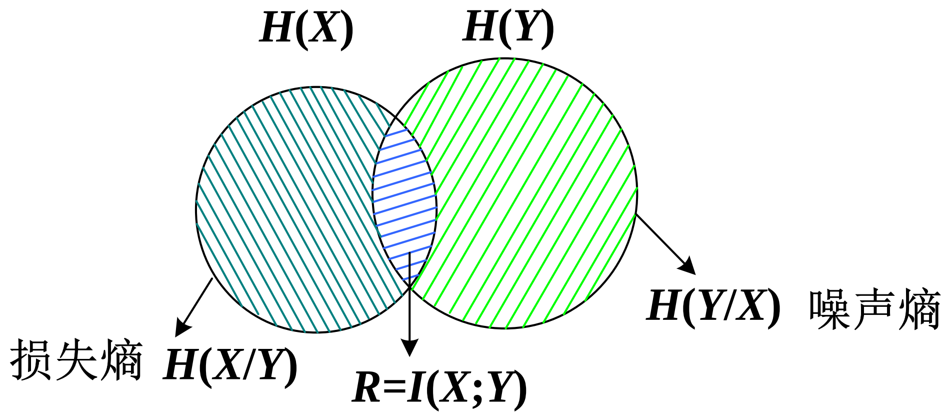


信源编码器的目的：通过尽可能的压缩，提高效率。

——使编码后所需的信息传输率*R*尽量小，对信道容量的需求低。

问题：

信源编码器的目的：使编码后所需的信息传输率R尽量小，即对信道容量的需求低。



矛盾：使编码后所需的信息传输率R尽量小，然而R越小，熵损失越多，从失真的直观意义上理解，引起的平均失真就越大。

研究目标：给出一个失真的限制值D，在满足平均失真 $\overline{D}\le D$ 的条件下，选择一种编码方法使信息率R尽可能小。

$\overline{D}\le D$的条件下，选择编码方法使信息率R尽可能小

- **分析：**

信息率R 就是所需输出的有关信源X 的信息量。

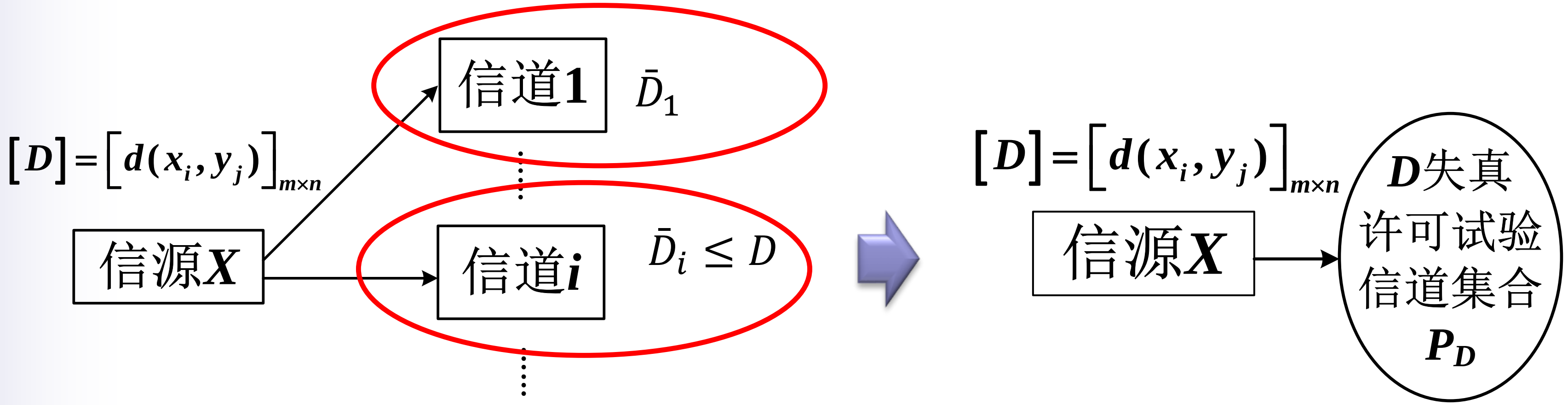
↓

将此问题对应到信道，即为接收端Y 需要获得的有关X的信息量，也就是互信息$I(X;Y)$。

↓

选择信源编码方法的问题就变成了选择假想信道的问题，符号转移概率 $p(y_j\,|\,x_i)$ 就对应信道转移概率。

由平均失真度可引出信息率失真函数的概念



① 固定信源，压缩方法可变，即信道可调。

②$p(x_i)$已知，设$d(x_i, y_j)$ 给定，当选定某信道后，则该信源/信道的平均失真度可计算。

③ 每设定一个允许的失真度为D，在所有信道中选择满足$\bar{D} \le D$的信道，构成D失真许可实验信道集合P_D。

④ 在P_D中，求最小的平均互信息，*即保真度准则下的最有效的压缩方法*。

$$R \triangleq \min_{p(y_j/x_i)\in P_D} I(X;Y)$$

几个重要概念

- 保真度准则

$\overline{D} \le D$ → 预先规定的限定失真度，是允许失真的上界

信源压缩后的平均失真度，
若信源和失真度一定，就只是信道统计特性的函数。
传递概率不同，平均失真度随之改变

- D失真许可信道(D允许的试验信道)
  - 满足保真度准则的所有信道

$$P_D = \{p(y|x) : \overline{D} \le D\}$$

  - 对于离散无记忆信道，相应有：

$$P_D = \{p(y_j|x_i) : \overline{D} \le D \quad i = 1,2,\cdots,n ; j = 1,2,\cdots,m\}$$

给定D
↓
P_D 可确定
↓
$R \triangleq \min\limits_{p(y_j/x_i)\in P_D} I(X;Y)$ 可确定
↓
R是D的函数
↓
$R(D)$：信息率失真函数

信息率失真函数的定义

- **在D允许信道P_D中，寻找一个信道$p(Y|X)$，使给定的信源经过此信道传输时，其信道传输率$I(X,Y)$最小**

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

- **对于离散无记忆信道，率失真函数可以写成：**

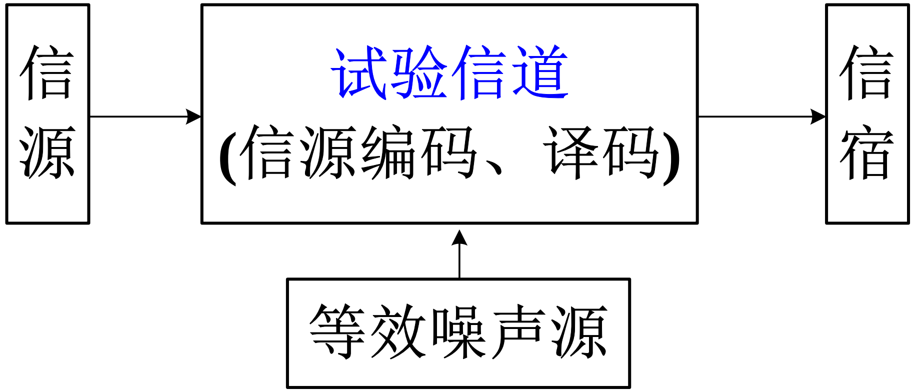
$$R(D)=\min_{p(y_j|x_i)\in P_D}\sum_{i=1}^{n}\sum_{j=1}^{m}p(x_i)\,p(y_j|x_i)\cdot\log\frac{p(y_j|x_i)}{p(y_j)}$$

其中，$p(x_i)$是信源符号概率分布，$i=1,2,\cdots,n$；

$p(y_j|x_i)$是转移概率分布，$i=1,2,\cdots,n$，$j=1,2,\cdots,m$；

$p(y_j)$是接收端收到符号概率分布，$j=1,2,\cdots,m$。

信息率失真函数$R(D)$的意义



研究限失真信源编码时的通信系统模型

将实际编码过程视为等效信道(试验信道)

信息率失真函数 $R(D)$的物理意义

对于某给定信源而言，任何限失真编、译码方法，必须保证系统的平均互信息量$I(X;Y) \geq R(D)$, 才有可能满足失真条件$\bar{D} \leq D$; 否则，一定有$\bar{D} > D$。

说明：

1. 信息率失真函数是在保真度准则下，信宿必须获得的平均信息量的最小值，是信源必须输出的最小信息率。
2. 信息传输速率本质上是描述信源特性的，因此$R(D)$也应该是仅仅用于描述信源。
3. 若信源消息经无失真编码后的信息传输速率为R，则在保真度准则下信源编码输出的信息率$R(D) < R$

- 说明在保真度准则条件下的信源编码比无失真情况得到了压缩，同时$R(D)$是保真度条件下对信源进行压缩的极限值，亦即信源信息率可压缩的最低限度，它仅取决于信源特性和保真度要求，与信道特性无关。

求信息率失真函数的方法

- 与信道容量求解比较：

- 平均互信息$I(X;Y)$既是信源概率分布$p(x_i)$的上凸函数，又是信道传递概率$p(y_j/x_i)$的下凸函数。

- 信道容量C是在信道特性$p(y_j/x_i)$已知的条件下求平均互信息的极大值（最大）问题；而率失真函数$R(D)$是在允许失真D和信源概率分布$p(x_i)$已给的条件下，求平均互信息的极小值（最小）问题。

- 这两个问题是对偶问题。 “棋逢对手，将遇良才”

信道容量与率失真函数的比较（对偶问题）

$C = \max\limits_{p(x_i)}\{I(X;Y)\}$	$R(D) \triangleq \min\limits_{p(y_j \mid x_i)\in P_D}\{I(X;Y)\}$
$I(X;Y)$的上凸函数	$I(X;Y)$的下凸函数
$I(X;Y)$的极大值	$I(X;Y)$的条件极小值
$\{p(b_j \mid a_i)\}$的函数	$\{p(a_i)\}$的函数
仅与信道特性有关	仅与信源特性有关
解决可靠性问题	解决有效性问题
信息传输的基础	信源压缩的基础

求解方法对比

- 求信道容量的方法
- 信道容量是假定信道固定的前提下，选择一种试验信源，使信息率最大。一旦找到了这个信道容量，它就与信源不再有关，而是信道特性的参量，随信道特性的变化而变化。

- 求信息率失真函数的方法
- 信息率失真函数$R(D)$是假定信源给定的情况下，在用户可以容忍的失真度D内，再现信源消息所必须获得的最小平均信息量。它反映的是信源可压缩程度。率失真函数一旦找到，就与求极值过程中选择的试验信道不再有关，而只是信源特性的参量。不同的信源，其$R(D)$是不同的。

研究信道编码和率失真函数的意义

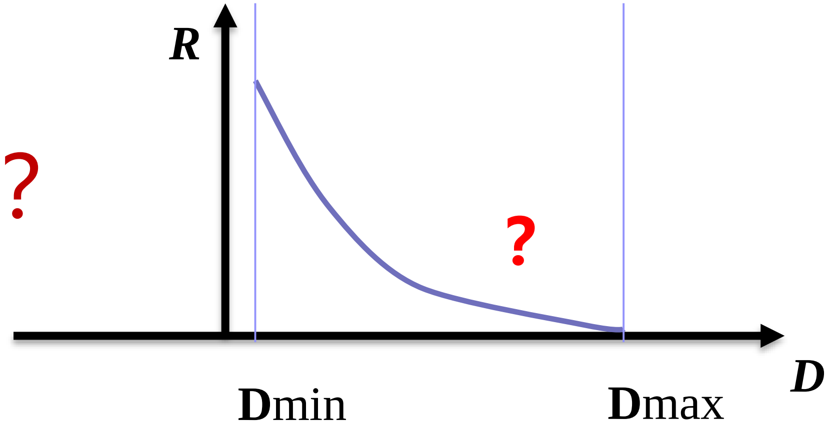
- 研究信道容量的意义：
- 在实际应用中，研究信道容量是为了解决在已知信道中传送最大信息率问题。目的是充分利用已给信道，使传输的信息量最大而发生错误的概率任意小，以提高通信的可靠性。信道编码问题。

- 研究信息率失真函数的意义：
- 研究信息率失真函数是为了解决在已知信源和允许失真度*D*的条件下，使信源必须传送给信宿的信息率最小。即用尽可能少的码符号尽快地传送尽可能多的信源消息，以提高通信的有效性。信源编码问题。

信息率失真函数的性质

信息率失真函数的性质

- 信息率失真函数$R(D)$是D的函数
- 问题：R随D的变化规律是怎样的？



1. D的有效取值范围是多少？

2. R(D)是关于D的形态是上凸还是下凸，还是其他？

3. R(D)在区间是递增还是递减，或是波动？

回答：率失真函数性质

- R(D)的定义域(0,Dmax)
- R(D)是关于D的下凸函数
- R(D)在区间(0,Dmax)上是严格递减函数

什么是率失真函数的定义域?

- 率失真函数中的自变量：允许平均失真度D，也就是平均失真度$\overline{D}$的上限值。
- 率失真函数的定义域问题：
- 就是在信源和失真函数已知的情况下，讨论允许平均失真度D的最小和最大值问题。
- D的选取必须根据固定信源X的统计特性$P(X)$和选定的失真函数$d(x_i, y_j)$，在平均失真度的可能取值范围内。

信源最小平均失真度D_{min}

- 分析
- 平均失真度$\overline{D}$是失真函数$d(x,y)$的数学期望

  而失真函数非负$d(x,y) \geq 0$，故有$\overline{D} \geq 0$
- 因此允许平均失真度D的下限为：

$$\boxed{D_{min} = 0}$$

这表示不允许有任何失真。

- 问题：D能否达到其下限值0?
- 回答：与单个符号的失真函数有关。

寻找最小平均失真度D_{min}

- 方法：在失真矩阵的每一行找出一个最小的$d(x_i, y_j)$，各行的最小$d(x_i, y_j)$值都不同。对所有这些最小值求数学期望，就是信源的最小平均失真度。

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

- 显然，只有当失真矩阵的每一行至少有一个0元素时，平均失真度才能达到下限值0。
- 此时，信源不允许任何失真存在，信息率至少应等于信源输出的平均信息量（信源熵），即$R(0)=H(X)$。

分析信源最大平均失真度D_{max}

- 必须传输的信息率R越小，容忍的失真D就越大。
- 当$R(D)$等于0时，对应的平均失真最大，也就是函数R(D)定义域的上界值D_{max}。
- 信息率失真函数是平均互信息的极小值：
  - 当$R(D)=0$时，即平均互信息的极小值等于0；
  - 这相当于输入X和输出Y统计独立。
  - 意味着在接收端收不到信源发送的任何信息，与信源不发送任何信息等效。或者说传送信源符号的信息率可以压缩至0。

$$R(D) \triangleq \min_{p(y_j/x_i)\in P_D} I(X;Y)$$

计算D_{max}的值

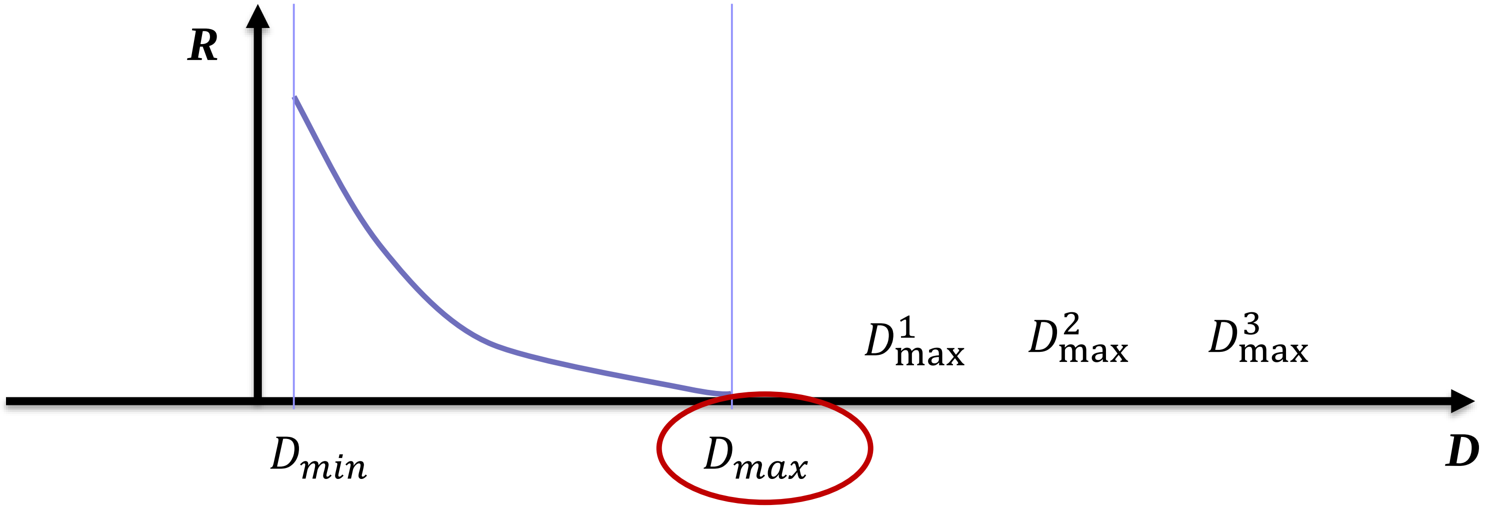
令试验信道特性$p(y_j|x_i)=p(y_j)$，这时X和Y相互独立，等效于通信中断，因此$I(X,Y)=0$，即$R(D)=0$。

满足上式的试验信道有许多，这些试验信道对应的R都为0

令 P_0 为满足上述独立要求的全体转移概率集合，相应地可求出许多平均失真值，$D_{\max}^1$，$D_{\max}^2$，$D_{\max}^3$…

从中选取最小的一个，就是这类平均失真值的下确界 D_{max} 。

$$D_{max}=\min_{p(y|x)\in P_0} E[d(x,y)]$$



计算D_{max}的值

令试验信道特性$p(y_j|x_i) = p(y_j)$，这时X和Y相互独立，等效于通信中断，因此$I(X,Y) = 0$，即$R(D)=0$。

满足上式的试验信道有许多，这些试验信道对应的R都为0

令 P_0 为满足上述独立要求的全体转移概率集合，相应地可求出许多平均失真值，$D_{\max}^1$，$D_{\max}^2$，$D_{\max}^3$…

从中选取最小的一个，就是这类平均失真值的**下确界** D_{max}。

$$D_{max} = \min_{p(y|x)\in P_0} E[d(x,y)]$$

由于，X和Y相互独立，故有：

$$D_{\max} = \min_{p(y_j)} \sum_j p(y_j) \sum_i p(x_i) d(x_i, y_j) = \min_{p(y_j)} \sum_j p(y_j) D_j$$

$$\begin{bmatrix} p(x_1) & p(x_2) & \cdots & p(x_n) \end{bmatrix}$$

$$\begin{bmatrix} d_{11} & d_{12} & \cdots & d_{1m} \\ d_{21} & d_{22} & \cdots & d_{2m} \\ \vdots & \vdots & d_{ij} & \vdots \\ d_{n1} & d_{n2} & \cdots & d_{nm} \end{bmatrix} = [D]$$

计算D_{max}的值

$$D_{max} = \min_{p(y_j)} \sum_j p(y_j) \sum_i p(x_i) d(x_i, y_j) = \min_{p(y_j)} \sum_j p(y_j) D_j$$

信源分布 失真函数 已经给定

上式是用不同的概率分布$p(y_j)$ 对D_j求数学期望，取数学期望当中最小的一个作为D_{max}

若其中最小D_j的分布选取为$p(y_i)$=1，而其他非最小D_j时的分布选取为$p(y_i)$=0，此时数学期望必然最小,有：

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

R(D)的定义域例子

例: 二元信源$\begin{Bmatrix} x_1 & x_2 \\ 0.4 & 0.6 \end{Bmatrix}$, $[D]=\begin{bmatrix} \alpha & 0 \\ 0 & \alpha \end{bmatrix}$,求$R(D)$的定义域和值域。

$$\boldsymbol{D}_{\max}=\min_{p(y_j)}\sum_j p(y_j)\sum_i p(x_i)d(x_i,y_j)$$

$$=\min_{p(y_j)}\sum_j p(y_j)D_j$$

解： 由定义：$D_{\min}=0$

$$D_1=p(x_1)d(x_1,\ y_1)+p(x_2)d(x_2,\ y_1)=0.4\alpha+0.6*0=0.4\alpha$$

$$D_2=p(x_1)d(x_1,\ y_2)+p(x_2)d(x_2,\ y_2)=0.4*0+0.6*\alpha=0.6\alpha$$

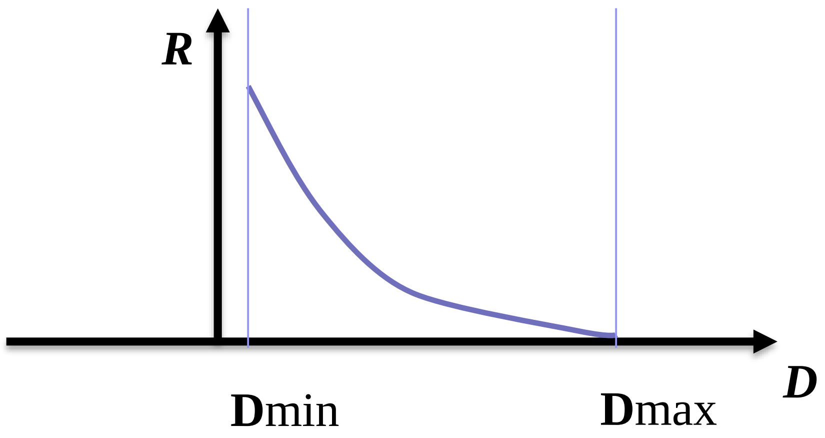
$$D_{\max}=\min(D_1,D_2)=0.4\alpha$$

当D=$D_{\min}=0$时，$R(D)=H(X)$，无失真

当$D\geq D_{\max}$时，$R(D)=0$

信息率失真函数的性质

- 信息率失真函数*R*(*D*)是*D*的函数
- 问题：*R*随*D*的变化规律是怎样的?



1. D的有效取值范围是多少?

2. R(D)是关于D的形态是上凸还是下凸，还是其他?

3. R(D)在区间是递增还是递减?

回答：率失真函数性质

- **R(D)的定义域(0,Dmax)**
- **R(D)是关于D的下凸函数**
- **R(D)在区间(0,Dmax)上是严格递减函数**

2. R(D)是关于D的下凸函数

假定D_1和D_2是两个失真度，$p_1(y|x)$和$p_2(y|x)$是满足保真度准则D_1和D_2前提下使$I(X,Y)$达到极小的信道，即

$$R(D_1) = \min_{p(y|x) \in P_{D_1}} I[p(y|x)] = I[p_1(y|x)]$$

$$R(D_2) = \min_{p(y|x) \in P_{D_2}} I[p(y|x)] = I[p_2(y|x)]$$

令 $D_0 = \alpha D_1 + (1-\alpha) D_2$，$0 < \alpha < 1$，**新失真度**

$p_0(y|x) = \alpha p_1(y|x) + (1-\alpha) p_2(y|x)$ **新信道**

可以证明$p_0(y|x)$是满足保真度准则D_0的信道。

进一步可证： **下凸函数得证！**

$$\boxed{R[\alpha D_1 + (1-\alpha) D_2] \le \alpha R(D_1) + (1-\alpha) R(D_2)}$$

3. R(D)在区间(0,Dmax)上是严格递减函数

$I[X,Y]$是$p(y|x)$的连续函数，由$R(D)$的定义可知

$R(D)$是连续函数。

设有两个失真度D_2和D_1，对应的满足保真度D_2和D_1的试验信道集合为P_{D_2}和P_{D_1}

若$D_2 > D_1$，则有$P_{D_2} \supset P_{D_1}$

在一个较大范围内求极小值，一定不大于在其中一个小范围内求极小值：

$$\min_{p(y|x)\in P_{D_2}} I[p(y|x)] \le \min_{p(y|x)\in P_{D_1}} I[p(y|x)]$$

有 $R(D_2) \le R(D_1)$，即$R(D)$是非增函数。

可以证明上述不等式中的等号不成立（略），

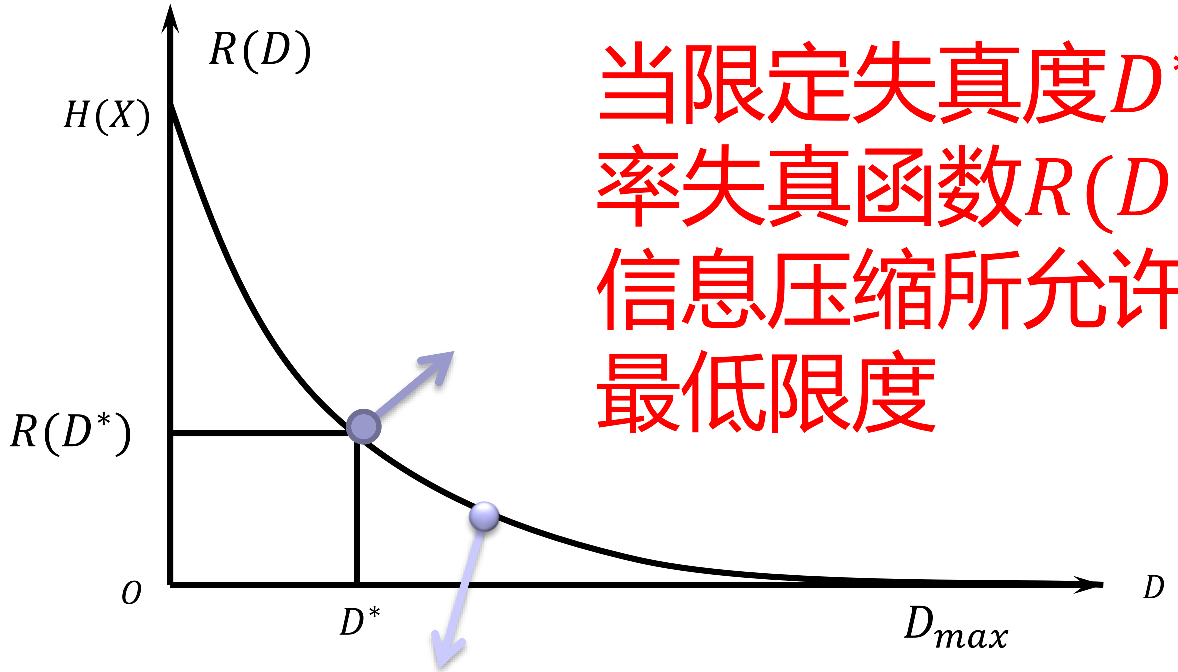
所以，$R(D)$是严格递减函数。

R(D)三点结论（三个性质）

1. $R(D)$是非负函数，其定义域为$0 \sim D_{max}$，其值域为$0 \sim H(X)$，当$D > D_{max}$时，$R(D) = 0$；
2. $R(D)$是关于失真度D的下凸函数；
3. $R(D)$是关于失真度D的严格递减函数。

根据上述性质可以画出离散信源信息率失真函数$R(D)$的一般曲线图形。

离散信源信息率失真函数R(D)的一般曲线



当限定失真度D^*时，率失真函数$R(D^*)$是信息压缩所允许的最低限度

若信息率压缩至$R(D) < R(D^*)$

则失真度D必大于限定失真度D^*；

结论：信息率失真函数给出了限失真条件下信息压缩允许的下界。

总结

- 由以上分析可见，当规定了允许的失真D，又找到了适当的失真函数d_{ij}，就可以找到该失真条件下的最小信息率$R(D)$，这个最小信息率是一个极限值。
- 在满足保真度准则的前提下，用不同方法进行数据压缩时，$R(D)$函数就是压缩程度的衡量，由$R(D)$可知是否还有压缩的潜力，潜力有多大，因此对它的研究很有实际意义。

谢谢！

黑晓军
华中科技大学
电子信息与通信学院
Email: heixj@hust.edu.cn
网址：http://eic.hust.edu.cn/aprofessor/heixiaojun

参考资料

- 陈运，信息论与编码，第3版，第7章7.1节，电子工业出版社，2015