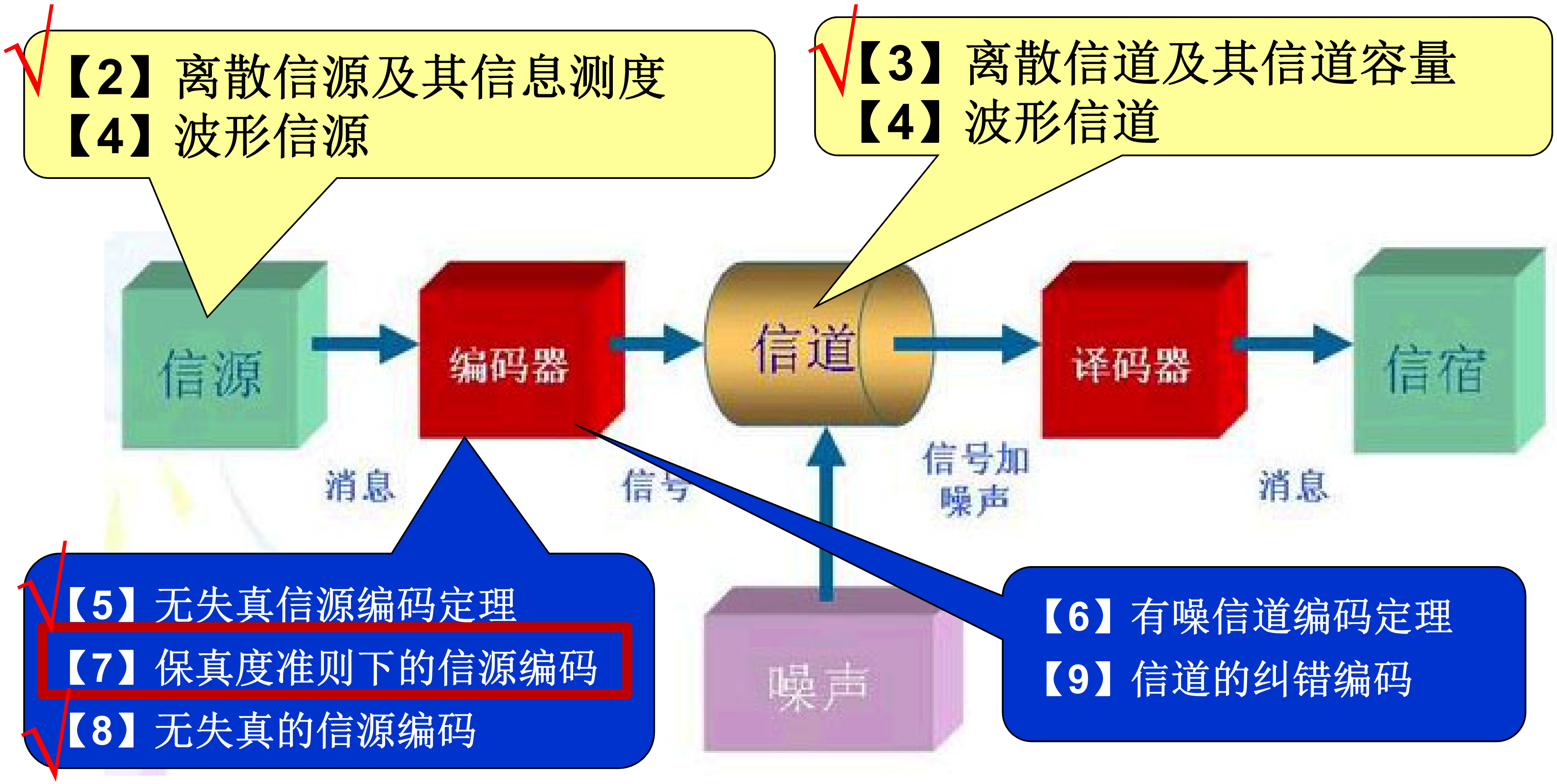
第七章 保真度准则下的信源编码

2022/5/24

信息论的旅程

- 本章将着重讨论允许一定失真的条件下可把信源信息压缩到什么程度。



主要内容

- 基本概念
  1. 概述
  2. 系统模型
- 失真测度
- 信息率失真函数
- 限失真信源编码定理

1.1 概述

- 失真产生的原因
  - 信道噪声的干扰使得信息传输过程会产生差错；
  - 当信息传输率超过信道容量时，必然产生差错；
  - 信源熵（$H(S)$）是信源无失真压缩（$\bar{l}$）的极限，若再继续压缩则会带来失真。 $\bar{l}\downarrow$ $\quad \bar{l} \geq \frac{H(S)}{\log r}$ $\quad H_r(S) \longrightarrow$ $r=2$时 $\bar{l} \geq H(S)$
- 失真存在的合理性
  - 信宿的灵敏度和分辨率是有限的，不要求绝对无失真；
  - 允许失真的存在，可以提高信息传输率，从而降低通信成本。

1.1 概述（续）

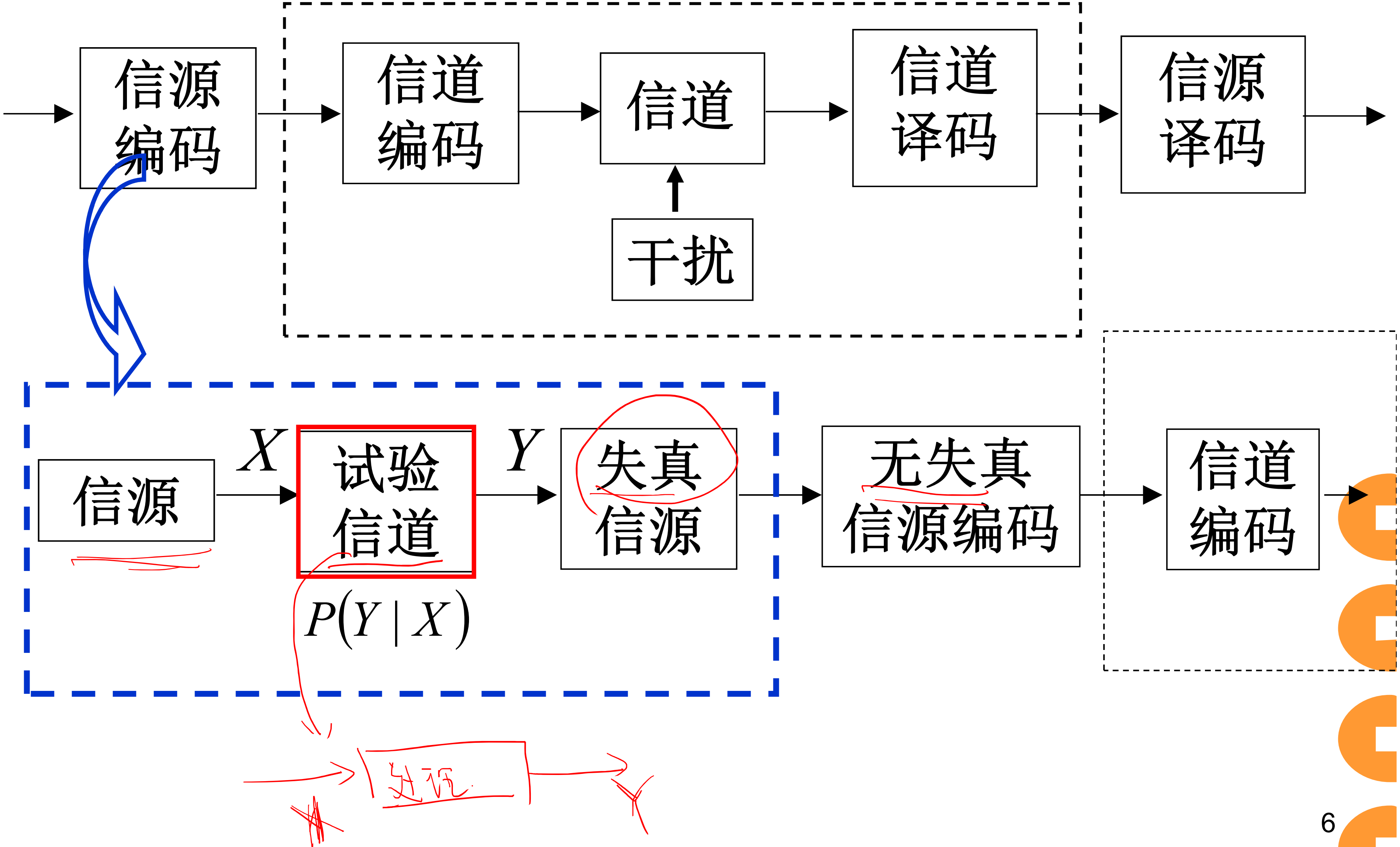
- 无失真信源压缩的极限：信源的信息熵
- 本章的研究内容
  - 在允许一定程度失真的条件下，能够把信源信息压缩到什么程度，即，最少需要多少比特才能描述信源。
- 研究方法
  - 用研究信道的方法，来研究有失真信源压缩问题。

$I(X;Y) = f(P_X, P_{Y|X})$　　$X \to \square \to Y$

$[\cdots]_{r\times s}$　$P_{Y|X}$,　$I(X;Y)$　P_X　　$C = \max_{P_X} I(X;Y)$ ✓

$\begin{bmatrix} X \\ P_X \end{bmatrix} = [\quad]$　P_X　$I(X;Y)$　$P_{Y|X}$　　$\min_{P_{Y|X}} I(X;Y)$

1.2 系统模型 – 只讨论信源编码问题★



主要内容

- 基本概念
- 失真度和平均失真度
  1. 失真函数
  2. 平均失真度
- 信息率失真函数
- 限失真信源编码定理

2.1 失真度 – 失真函数

- 失真函数 $d(x,y)$
  - 非负函数；函数形式可根据需要定义
  - 定量描述输入/输出符号之间的差异/失真

$$\begin{bmatrix} X \\ P \end{bmatrix} = \begin{bmatrix} x_1 & x_2 & \cdots & x_n \\ p(x_1) & p(x_2) & \cdots & p(x_n) \end{bmatrix} \quad Y:\{y_1, y_2, \cdots, y_m\}$$

失真矩阵 $\boldsymbol{D} = \begin{bmatrix} d(x_1,y_1) & d(x_1,y_2) & \cdots & d(x_1,y_m) \\ d(x_2,y_1) & d(x_2,y_2) & \cdots & d(x_2,y_m) \\ \vdots & \vdots & & \vdots \\ d(x_n,y_1) & d(x_n,y_2) & \cdots & d(x_n,y_m) \end{bmatrix}_{n \times m}$

2.1 失真度 – 失真函数（续）

- 常用的失真函数有：

（1） 汉明失真

$$d(x_i, y_j) = \begin{cases} 0 & x_i = y_j \\ 1 & x_i \neq y_j \end{cases}$$

（2） 平方误差失真函数

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

- 失真函数是根据人们的实际需要和失真引起的损失、风险大小等人为规定的。

2.1 失真度 – 失真函数 – 例题

例7.1 设信道输入 $X=\{0,1\}$，输出 $Y=\{0,?,1\}$，规定失真函数

$d(x_i, y_j)$

$$\boldsymbol{d(0,0)=d(1,1)=0,}$$

$$\boldsymbol{d(0,1)=d(1,0)=1,}$$

$\boldsymbol{d(0,?)=d(1,?)=0.5}$，求 失真矩阵 。

解:

$$[d]=\begin{array}{c c} & \begin{array}{ccc} 0 & ? & 1 \end{array} \\ \begin{array}{c} 0 \\ 1 \end{array} & \begin{bmatrix} d(0,0) & d(0,?) & d(0,1) \\ d(1,0) & d(1,?) & d(1,1) \end{bmatrix} \end{array} = \begin{array}{c c} & \begin{array}{ccc} 0 & ? & 1 \end{array} \\ \begin{array}{c} 0 \\ 1 \end{array} & \begin{bmatrix} 0 & 0.5 & 1 \\ 1 & 0.5 & 0 \end{bmatrix} \end{array}$$

2×3

2.2 失真度 – 平均失真度

有损压缩

$X \rightarrow \boxed{\quad} \rightarrow Y$

保真度准则

信源编码定理

- 平均失真度：★

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

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

- 平均失真度：信源P_X在给定信道$P_{Y|X}$中传输时所引起的失真的总体量度。

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

2.2 失真度 － 平均失真度 － 例题7.2

$\bar{D} \le \frac{1}{2}$

P_X

$P_{Y|X}$

$d(x.y)$

- 设信源符号集 $A=\{a_1,a_2,\cdots,a_{2r}\}$，概率分布

  $$p(a_i)=1/2r \qquad (i=1,2,\cdots,2r)$$

  失真函数定义为

  $$d(x_i,y_j)=\begin{cases}0 & (x_i=y_j)\\ 1 & (x_i\neq y_j)\end{cases}$$

  假定允许的失真限度为 $D^*=1/2$

  分析在给定失真限度条件下信源压缩的程度。

- 考虑：无失真编码时每个符号需要多少个码元？

$\bar{L} \geq H_r(S) = H(S) = \log 2r$

$r=2.$

2.2 失真度 – 平均失真度 – 例题7.2(续)

1. 假定输入等概，则信源熵 $H(X)=\log 2r$ ，即无失真编码中，每个信源符号需要 $\log 2r$ **bit**
2. 给出一种编码方案，将**2*r***个输入信源编码对应到***r***个输出符号中。编码方法对应如右所示的试验信道。
3. 平均失真为：

$$P_{Y|X}=\begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & 1 \\ 0 & 0 & \cdots & 1 \\ 0 & 0 & \cdots & 1 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix}$$

$$\overline{D}=E[d]=\sum_{i=1}^{2r}\sum_{j=1}^{r}p(x_i)p(y_j\mid x_i)d(x_i,y_j)$$

$$=\sum_{i=r+1}^{2r}p(x_i)=\frac{1}{2}$$

2.2 失真度 – 平均失真度 – 例题7.2(续)

- 该信道为无噪信道，则$H(Y|X)=0$。
- 信道的输出熵 $\begin{bmatrix} Y \\ P \end{bmatrix} = \begin{bmatrix} y_1 & y_2 & \cdots & y_{r-1} & y_r \\ 1/2r & 1/2r & \cdots & 1/2r & r+1/2r \end{bmatrix}$

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

$$= H\left(\frac{1}{2r},\frac{1}{2r},\cdots,\frac{1}{2r},\frac{r+1}{2r}\right)$$

$$= \log 2r - \frac{r+1}{2r}\log(r+1)$$

$$P_{Y|X} = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 \\ 0 & 0 & \cdots & 1 \\ 0 & 0 & \cdots & 1 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix}$$

信源信息压缩：

$$\boxed{H(X)-H(Y)=\frac{r+1}{2r}\log(r+1)}$$

有无压缩下界？

主要内容

- 基本概念
- 失真测度
- 信息率失真函数
  1. 保真度准则
  2. 信息率失真函数$R(D)$
  3. $R(D)$的性质
- 限失真信源编码定理

3.1 信息率失真函数 –★保真度准则

- 保真度准则：若预先规定的平均失真度为D，则称信源压缩后的平均失真度$\overline{D}$不大于D的准则为保真度准则，即$\overline{D} \le D$。

- ***D*允许的试验信道**：将满足保真度准则的所有信道称为失真度***D***允许信道（也称***D***允许的试验信道），记为

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

- 对于确定信源和失真函数，不同编码对应不同的试验信道。

$[P_{Y|X}]$ $r \times s$

$d(x,y)$

P_X

$P_{Y|X}$

$\overline{D} \le D$

$X >$

3.2 信息率失真函数

$B_D = \{ P_{Y|X} : \bar{D} \le D \}$

- 在满足失真要求的信道中，寻找一种信道，使给定的信源经过此信道传输时，其信息传输率 $I(X;Y)$ 达到最小，定义这个最小值为信息率失真函数：

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

★
$$R(D) = \min_{p(y_j|x_i) \in B_D} I(X;Y)$$

$$= \min_{p(y_j|x_i) \in B_D} \sum_{i=1}^{n} \sum_{j=1}^{m} p(x_i) p(y_j | x_i) \log \frac{p(y_j | x_i)}{p(y_j)}$$

注：此时满足 $I(X;Y) = H(Y) - H(Y \mid X) = H(Y)$

$\min I$

信道容量和信息率失真函数的比较

1. 信道容量：$C = \max\limits_{p(x_i)} I(X;Y)$

$P_{Y|X}$

是在信道固定前提下，选择一种信源概率分布使信息传输率最大（求极大值）。它反映了信道传输信息的能力，是信道可靠传输的最大信息传输率。信道容量与信源无关，是反映信道特性的参量，不同的信道其信道容量不同。

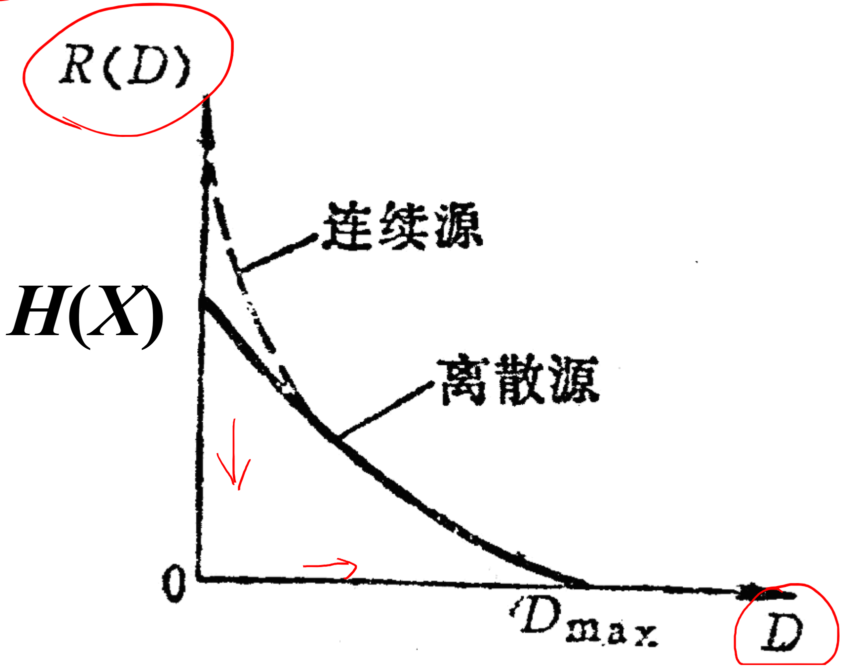
2. 率失真函数：$R(D) = \min\limits_{p(y_j|x_i):\bar{D} \le D} I(X;Y)$

P_X

是在信源固定，满足保真度准则的条件下的信息传输率的最小值。反映了满足一定失真度的条件下信源可以压缩的程度，也就是满足失真要求而再现信源消息所必须获得的最少平均信息量。

3.3 率失真函数的性质

- $R(D)$的定义域和值域
  - 平均失真度D的取值范围：$[0,\ D_{max}]$；
  - 率失真函数的取值：$[H(X),\ 0]$
- $R(D)$是关于D的下凸函数。
- $R(D)$在定义区间是严格递减函数。



主要内容

- 基本概念
- 失真测度
- 信息率失真函数
- 限失真信源编码定理 – 香农第三定理

香农第一 -

变长无失真 - - - -

4 限失真信源编码定理 – 香农第三定理

- 设离散无记忆平稳信源的信息率失真函数为 $R(D)$，只要满足 $R \geq R(D)$，当信源序列长度 N 足够大时，一定存在一种编码方法，其译码失真小于或等于 $D+\varepsilon$，其中 ε 是任意小的正数。
- 反之，若 $R < R(D)$，则无论采用什么样的编码方法，其译码失真必大于 D。

R(D)

定理说明：在允许失真 D 的条件下，信源最小的、可达的信息传输率是信源的 $R(D)$。

4 限失真信源编码说明

- 该定理是对离散无记忆信源给出的，对于连续无记忆平稳信源有类似结论。
- 存在性定理 。
  - 正定理：$R \geq R(D)$时，译码失真小于或等于***D+ε***的码肯定存在，但定理本身并未告知码的具体构造方法。
  - 逆定理：$R < R(D)$ 时，译码失真必大于***D***，肯定找不到满足条件的码，因此用不着浪费时间和精力。
- 香农信息论的三个基本概念——信源熵、信道容量和信息率失真函数，都是临界值，是从理论上衡量通信能否满足要求的重要界限。

本章总结

- 概论
  - 失真产生的原因和存在的合理性。
  - 率失真通过信道形式来研究信源的压缩问题。
- 失真函数、平均失真；失真测度分析。
- 信息率失真函数
  - 保真度准则；
  - 率失真函数$R(D)$、含义、及其性质；
- 香农第三定理的理解

作业题

1. 设输入符号集为X={0, 1}且等概分布，输出符号集为Y={0, 1}。定义失真函数为$d(0,0)=d(1,1)=0$；$d(0,1)=d(1,0)=2$

   求解：失真矩阵D；

2. 简述$R(D)$的性质，画出一般$R(D)$的曲线并说明其物理意义。