第5章 率失真理论

授课教师：樊平毅教授
清华大学电子工程系

2022年11月9日

率失真理论的出发点

Tsinghua University

目标：连续信源的数字化处理与数字编码

- 思考
  - 连续随机变量的有限表示如何优化？
  - 对于一个给定的信源分布与失真的度量，在特定码率下，能达到的最小期望失真是多少？
  - 满足一定失真限制的条件下，描述符号所需的最小码率为多少？

目录

清華大學
Tsinghua University

本章学习逻辑线条：

失真率定义 → 失真率的性质 → 失真率计算的例子

↓

是否存在编码模式满足给定的失真率？

↓

一般情况失真率如何计算：计算方法是什么

⇒

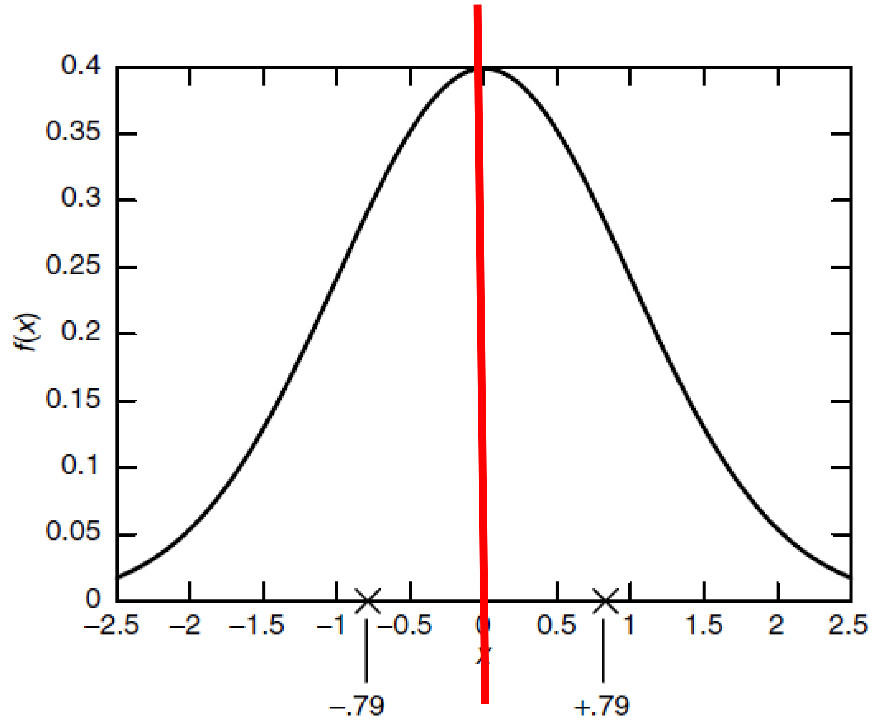
新发现：
失真率与信道容量的对偶关系

量化

- 例子：高斯随机变量的1比特量化

设$X \sim \mathcal{N}(0,\sigma^2)$，考虑平方误差，仅给定1比特表示$X$，显然，此1比特将用来区分$X>0$与否，为使平方误差最小，表示函数$\hat{X}(X)$应为所在区域上$X$的条件均值，于是

$$\boxed{\hat{X}(x)=\begin{cases}\sqrt{\dfrac{2}{\pi}}\sigma & \text{if } x\ge 0,\\ -\sqrt{\dfrac{2}{\pi}}\sigma & \text{if } x<0.\end{cases}}$$

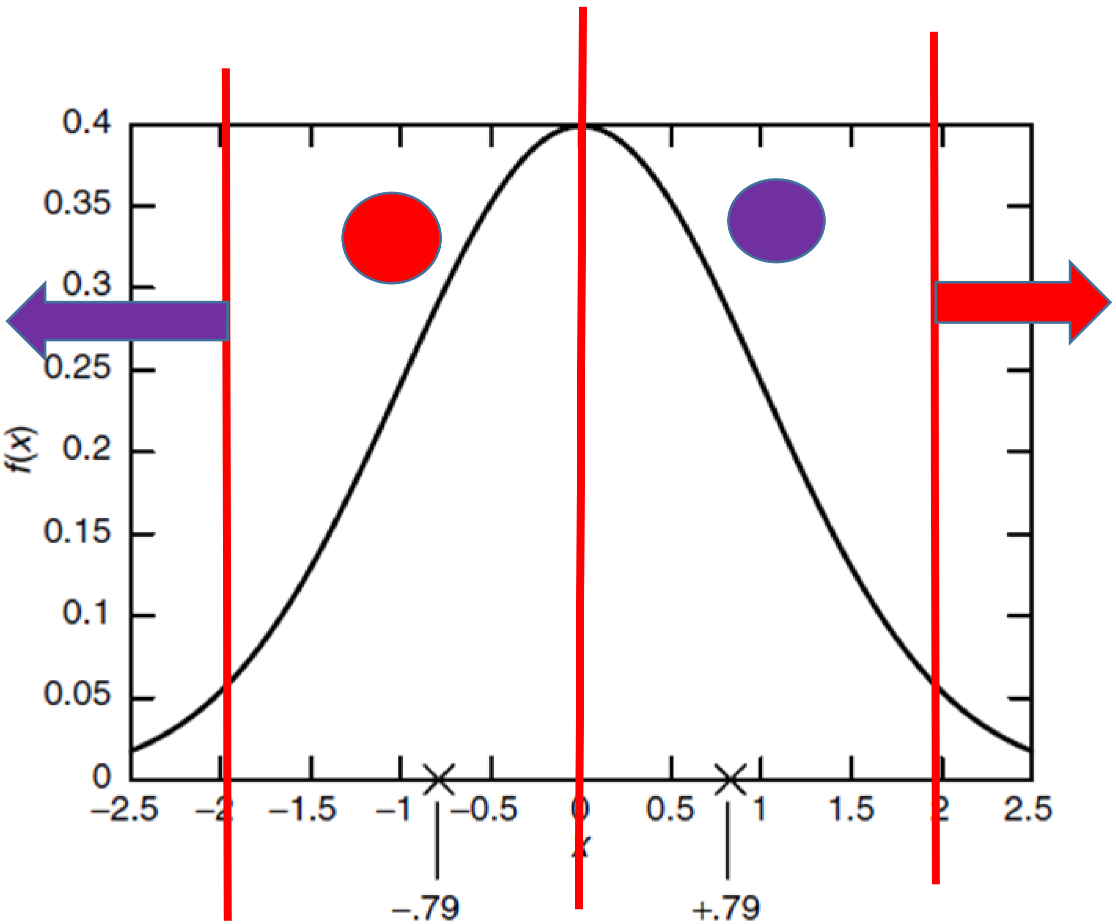


两比特表示

两比特的表示

（1）选择区域；

（2）单个区域内代表元选择，
原则上条件均值
保证均方误差最小



简单推导

- 计算平均值（区间），采用条件概率模式，实际计算的是条件平均；

- 在偏差计算中，二阶统计偏差等于方差+位移偏差的平方，

当位移为零，即位移和均值相同时，二阶统计偏差最小。

量化

- 信源中单个样本的表示问题
  - 设X为随机变量，其表示记为$\hat{X}(X)$
  - 用R比特表示X，则$\hat{X}$至多有2^R个取值（指标集）
  - 寻找$\hat{X}$的最优取值（再生点/码点）集合及与所对应的原像区域

- 最优区域划分及再生点性质

  - 再生点集合$\{\hat{X}(\omega)\}$给定时，可通过将信源随机变量X映射为再生点集合中最接近它的表示$\hat{X}(\omega)$，使失真最小化。该映射定义一个χ的区域构成的集合，称为由再生点定义的Voronoi划分或Dirichlet划分。

  - 再生点应该在各自划分到的区域上使条件期望失真最小化。

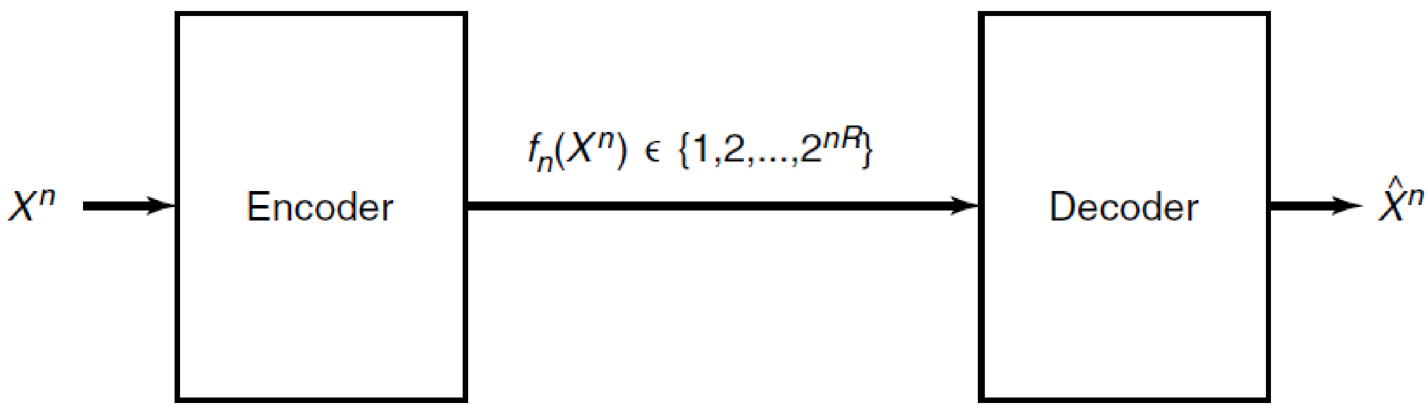
失真函数：距离定义

- 失真函数（失真度量）$d(x,\hat{x})$

  E(d)= $\mathbf{Pr}(X \neq \hat{X}) = D$

  - 汉明失真：$d(x,\hat{x}) = \begin{cases} 0 & x = \hat{x} \\ 1 & x \neq \hat{x} \end{cases}$ 也被称为错误概率失真
  - 平方误差失真：$d(x,\hat{x}) = (\mathrm{x} - \hat{x})^2$ 可以定义l_p距离，p>0 的任意实数

- 序列的失真度量
  - $d(x^n,\hat{x}^n) = \frac{1}{n}\sum d(x_i,\hat{x}_i)$



率失真 ---连续信源数字化

Tsinghua University

- 率失真码：一个$(2^{nR}, n)$率失真码包括编码函数

$$f_n : \mathcal{X}^n \to \{1, 2, \ldots, 2^{nR}\}$$

和译码函数

$$g_n : \{1, 2, \ldots, 2^{nR}\} \to \hat{\mathcal{X}}^n$$

注意：此处只关心连续信源的编解码处理

则其失真定义为失真函数的期望：

$$D = \mathrm{E}d(X^n, g_n(f_n(X^n))) = \sum_{x^n} p(x^n) d(x^n, g_n(f_n(x^n)))$$

- 率失真的可达：若存在一个$(2^{nR}, n)$率失真码序列(f_n, g_n)，满足

$$\lim_{n\to\infty} \mathrm{E}d(X^n, g_n(f_n(X^n))) \le D$$

则称率失真对(R,D)是可达的。

基本定义

- 率失真区域：全体可达率失真对(R,D)形成的集合闭包称为信源的**率失真区域**。

- **率失真函数**：给定失真D，满足(R,D)包含于信源的率失真区域中所有码率R的**下确界**，$R(D)$。
  （**最小比特的充分表示**）

- **失真率函数：**
- 给定码率R，满足(R,D)包含于信源的率失真区域中所有失真D的**下确界**，$D(R)$。
  （可以观测到的**最小扭曲程度刻画**）

信息率失真函数

- **定义：** 设信源X的失真度量为$d(x,\hat{x})$，定义其信息率失真函数$R^{(I)}(D)$为

$$R^{(I)}(D)=\min_{p(\hat{x}|x):\sum_{(x,\hat{x})}p(x)p(\hat{x}|x)d(x,\hat{x})\leq D} I(X;\hat{X})$$

联合分布需满足期望失真限制。

- **率失真定理：** 对于独立同分布信源X，若公共分布为$p(x)$且失真函数$d(x,\hat{x})$有界，则其率失真函数与对应的信息率失真函数相等，即在失真D下的最小可达码率。

$$R(D)=R^{(I)}(D)=\min_{p(\hat{x}|x):\sum_{(x,\hat{x})}p(x)p(\hat{x}|x)d(x,\hat{x})\leq D} I(X;\hat{X})$$

  - 率失真函数与信息率失真函数等价，统一用$R(D)$表示。

信息率失真函数的性质

- $R(D)$的凸性：上述给出的率失真函数$R(D)$是关于D的非增凸函数。

证明：

由于当D增大时，$R(D)$是随之增大的集合上的互信息的最小值，因此$R(D)$关于D非增。

下面证明凸性，

考虑率失真曲线上两点$(R_1, D_1), (R_2, D_2)$，

记达到率失真的联合分布为

$p_1(x, \hat{x}) = p(x)p_1(\hat{x}|x), p_2(x, \hat{x}) = p(x)p_2(\hat{x}|x)$。

考虑新的联合分布 $p_\lambda = \lambda p_1 + (1-\lambda)p_2$

由于失真是关于分布的线性函数，

则有$D_\lambda = \lambda D_1 + (1-\lambda)D_2$

$$D = \mathrm{E}d(X^n, g_n(f_n(X^n))) = \sum_{x^n} p(x^n) d(x^n, g_n(f_n(x^n)))$$

信息率失真函数

- **$R(D)$的凸性：** 上述给出的率失真函数$R(D)$是关于D的非增凸函数。

证明：

另一方面，互信息为条件分布的下凸函数，于是

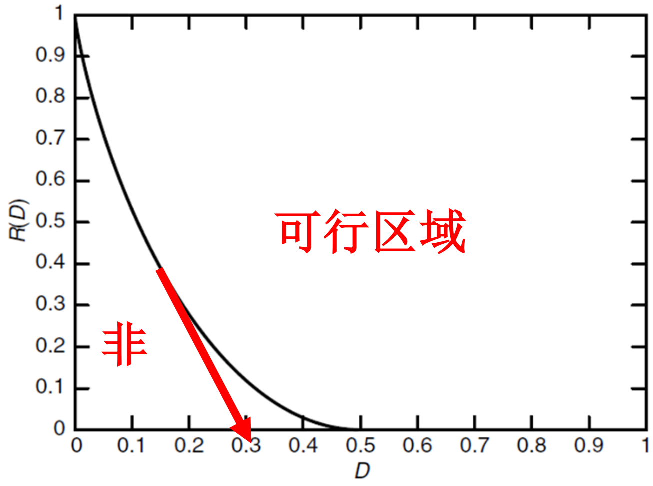
$$I_{p_\lambda}(X;\hat{X}) \le \lambda I_{p_1}(X;\hat{X}) + (1-\lambda) I_{p_2}(X;\hat{X})$$

由率失真函数的定义，

$$\begin{aligned} R(D_\lambda) &\le I_{p_\lambda}(X;\hat{X}) \\ &\le \lambda I_{p_1}(X;\hat{X}) + (1-\lambda) I_{p_2}(X;\hat{X}) \\ &= \lambda R(D_1) + (1-\lambda) R(D_2), \end{aligned}$$

即证明了其凸性。

图像可预测性



注解：（1）下凸性可以用于寻找最小值，或者说，（2）寻找其切线交叉点，说明系统的实现难以按照D的变化线性快速下降；（3）下降速度受限

率失真函数的计算

采用Hamming 距离

- 二元信源
  - Bernoulli(p)信源在汉明失真度量下的率失真函数为：

$$R(D)=\begin{cases} H(p)-H(D), & 0\le D\le \min\{p,1-p\}, \\ 0, & D>\min\{p,1-p\}. \end{cases}$$

证明：先找到互信息下界，再证明其可达。

（存在性）考虑二元信源$X\sim Bernoulli(p)$，不失一般性，假定$p<\frac{1}{2}$。用$\oplus$表示模2加法运算，则$X\oplus\hat{X}=1$等价于$X\neq\hat{X}$，有

$$\begin{aligned} I(X;\hat{X}) &= H(X)-H(X|\hat{X}) \\ &= H(p)-H(X\oplus\hat{X}|\hat{X}) \\ &\ge H(p)-H(X\oplus\hat{X}) \\ &\ge H(p)-H(D), \end{aligned}$$

由于$\Pr(X\neq\hat{X})\le D$且$H(D)$在$D\le\frac{1}{2}$时单增，则有

$$R(D)\ge H(p)-H(D).$$

下界选取是前提

率失真函数计算

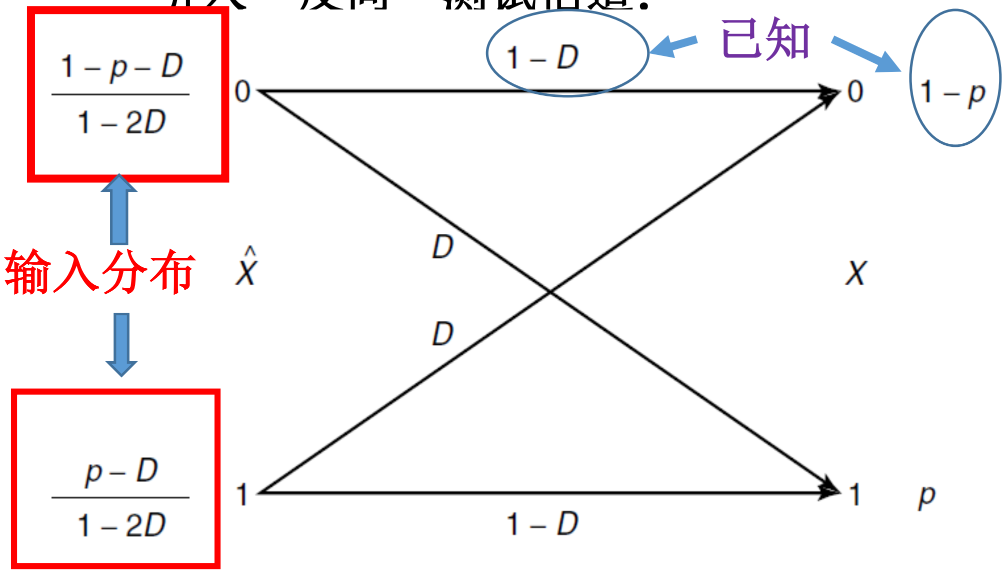
- 二元信源
  - Bernoulli(p)信源在汉明失真度量下的率失真函数为：

• 二元信源
• Bernoulli(p)信源在汉明失真度量下的率失真函数为：
证明：先找到互信息下界，再证明其可达。
（可达性）找到满足失真限制且$I(X,\hat{X}) = R(D)$的联合分布。
引入“反向”测试信道：

证明：先找到互信息下界，再证明其可达。

（可达性）找到满足失真限制且$I\left(X,\hat{X}\right) = R(D)$的联合分布。

引入“反向”测试信道：



反向测试信道：

- 为求得达到下界时的条件分布$p(\hat{x}|x)$，转为考虑更为简便的反向条件分布$p(x|\hat{x})$。
- 失真限制可直接满足。
- 强调了率失真与信道容量的对偶性。

率失真函数的计算

- 二元信源
  - Bernoulli(p)信源在汉明失真度量下的率失真函数为：

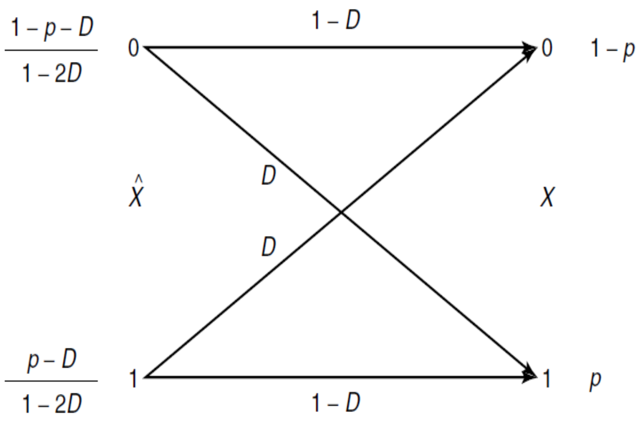
$$R(D)=\begin{cases} H(p)-H(D), & 0\le D\le \min\{p,1-p\},\\ 0, & D>\min\{p,1-p\}.\end{cases}$$

证明：先找到互信息下界，再证明其可达。

（可达性）找到满足失真限制且$I(X,\hat{X})=R(D)$的联合分布。

引入“反向”测试信道：

利用 互信息位等价交换



强调率失真与信道容量的对偶性
考虑反向条件密度$f(x|\hat{x})$

显然，其期望失真为$\Pr(X\neq\hat{X})=D$。

➢若$D\le p<\frac{1}{2}$，则$\Pr(\hat{X}=1)\ge 0, Pr(\hat{X}=0)\ge 0$，

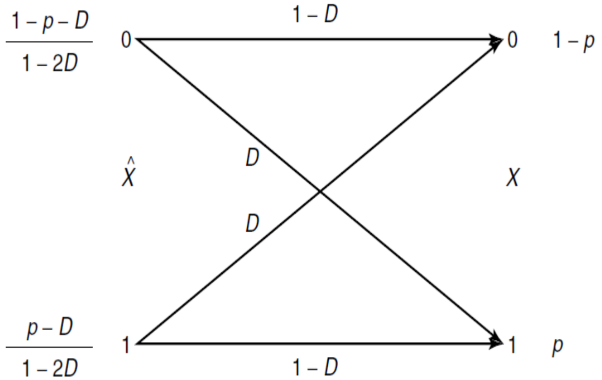
➢进而有$I(X;\hat{X})=H(X)-H(X|\hat{X})=H(p)-H(D)$；

■

- 二元信源
  - Bernoulli(p)信源在汉明失真度量下的率失真函数为：

$$R(D)=\begin{cases}H(p)-H(D), & 0\le D\le \min\{p,1-p\},\\ 0, & D>\min\{p,1-p\}.\end{cases}$$

证明： 其他条件情况



➢若$\boldsymbol{D\ge p}$，则通过可令$\boldsymbol{Pr(\hat{X}=0)=1}$
而到达码率$\boldsymbol{R(D)=0}$，
此时$\boldsymbol{I(X;\hat{X})=0}$，期望失真$\boldsymbol{D=P}$；

➢若$\boldsymbol{D\ge 1-p}$，类似的可令$\boldsymbol{Pr(\hat{X}=1)=1}$而到达码率$\boldsymbol{R(D)=0}$。
综上，有

$$R(D)=\begin{cases}H(p)-H(D), & 0\le D\le \min\{p,1-p\},\\ 0, & D>\min\{p,1-p\}.\end{cases}$$

提示

- 利用 互信息 $I(X,\hat{X})$ 关于互换位置不变的特征，引入反向信道；
- 选取反向信道的输入，满足基本的 **Distoration** 形式；
- 选取合适的输入分布，得到的可达速率等于互信息的下界；

从而说明是可达的（至少有一个可达到的）。

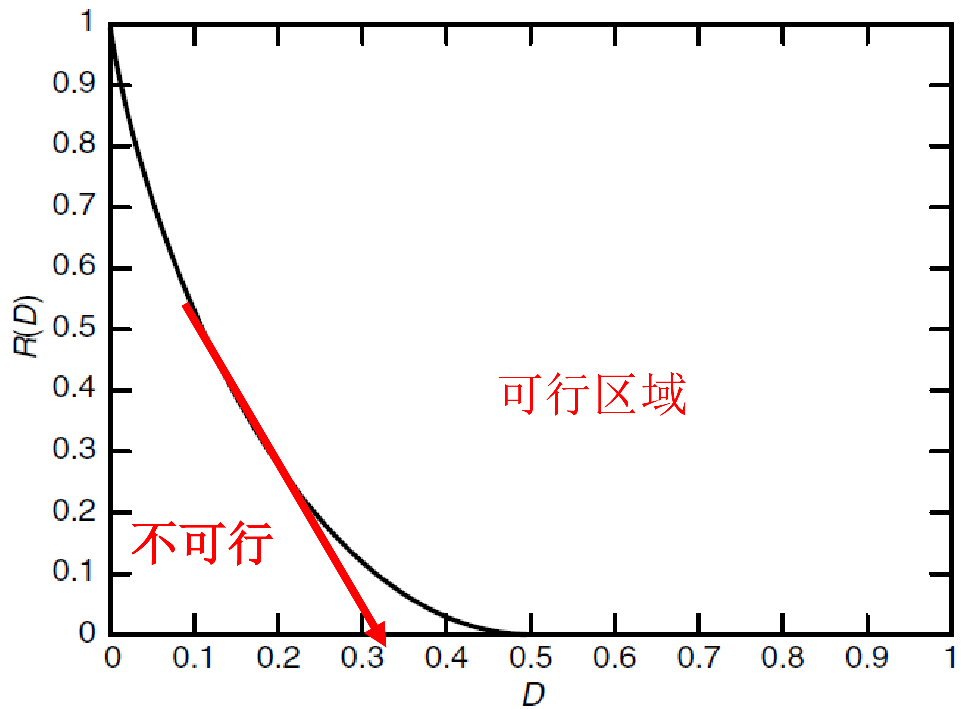
率失真函数的计算

- 二元信源
  - Bernoulli(p)信源在汉明失真度量下的率失真函数为：

$$R(D)=\begin{cases} H(p)-H(D), & 0\le D\le \min\{p,1-p\},\\ 0, & D>\min\{p,1-p\}.\end{cases}$$

  - $p=0.5$的情形

图例说明：



率失真函数的计算

- **高斯信源**
  - 一个$\mathcal{N}(0,\sigma^2)$信源在平方误差失真度量下的率失真函数为：

$$R(D)=\begin{cases}\frac{1}{2}\log\frac{\sigma^2}{D}, & 0\le D\le\sigma^2,\\ 0, & D>\sigma^2.\end{cases}$$

证明： 同样的思路。（存在性）由于$E(X-\hat{X})^2\le D$，有 **寻找下界**

$$\begin{aligned}I(X;\hat{X}) &= h(X)-h(X|\hat{X})\\ &=\frac{1}{2}\log(2\pi e)\sigma^2-h(X-\hat{X}|\hat{X})\\ &\ge\frac{1}{2}\log(2\pi e)\sigma^2-h(X-\hat{X})\\ &\ge\frac{1}{2}\log(2\pi e)\sigma^2-h(\mathcal{N}(0,E(X-\hat{X})^2))\\ &=\frac{1}{2}\log(2\pi e)\sigma^2-\frac{1}{2}\log(2\pi e)E(X-\hat{X})^2\\ &\ge\frac{1}{2}\log(2\pi e)\sigma^2-\frac{1}{2}\log(2\pi e)D\\ &=\frac{1}{2}\log\frac{\sigma^2}{D},\end{aligned}$$

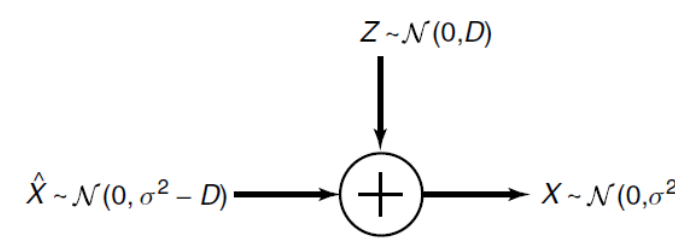
连续变量给定二阶矩正态分布使熵最大化

即 $R(D)\ge\frac{1}{2}\log\frac{\sigma^2}{D}$

率失真函数的计算

- 高斯信源
  - 一个$\mathcal{N}(0,\sigma^2)$信源在平方误差失真度量下的率失真函数

$$R(D)=\begin{cases}\frac{1}{2}\log\frac{\sigma^2}{D}, & 0\le D\le\sigma^2,\\ 0, & D>\sigma^2.\end{cases}$$



证明：同样的思路。（可达性）构建使得等号成立的反向测试信道$f(x|\hat{x})$：

构造反向信道模式，对偶模式（信道容量与率失真）

➢若$D\le\sigma^2$，取

$$X=\hat{X}+Z,\quad \hat{X}\sim\mathcal{N}(0,\sigma^2-D),\quad Z\sim\mathcal{N}(0,D)$$

➢其中$\hat{X}$与Z独立，对于该联合分布，则有 以及$E(X-\hat{X})^2=D$，则率失真下界可达；

$$I(X;\hat{X})=\frac{1}{2}\log\frac{\sigma^2}{D}$$

■

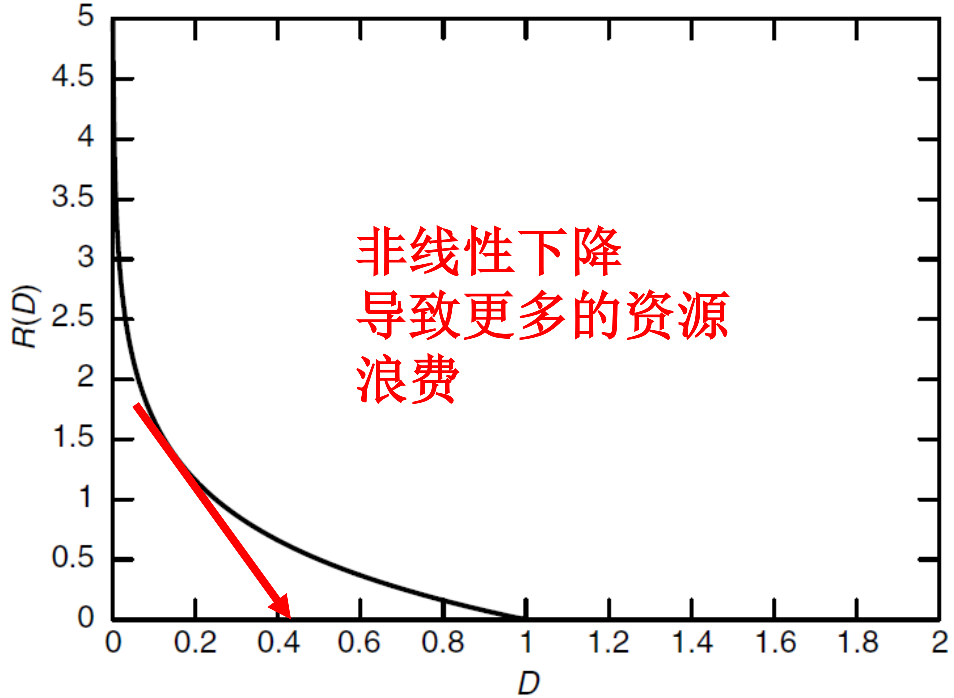
➢若$D>\sigma^2$，令以概率1选取$\hat{X}=0$则$R(D)=0$。

率失真函数的计算

- 高斯信源
  - 一个$\mathcal{N}(0,\sigma^2)$信源在平方误差失真度量下的率失真函数为：

$$R(D)=\begin{cases}\frac{1}{2}\log\frac{\sigma^2}{D}, & 0\le D\le\sigma^2,\\ 0, & D>\sigma^2.\end{cases}$$

  - 标准高斯信源：



率失真函数的计算

清華大學 Tsinghua University

- 独立高斯随机变量的同步描述（向量形式的推广）
  - 对于m个独立但不同分布高斯随机信源$X_1, \cdots X_m$的表示问题，其中$X_i \sim \mathcal{N}(0, \sigma_i^2)$，考虑平方失真度量$d(x^n, \hat{x}^n) = \sum (x_i - \hat{x}_i)^2$ 。
  - 用R bits来表示这个随机向量，如何分配这些比特到各个成员，使得总失真最小？将率失真函数推广至向量情形：

$$R(D) = \min_{f(\hat{x}^m|x^m):\mathrm{E}d(X^m,\hat{X}^m)\le D} I(X^m; \hat{X}^m)$$

由前面的例子，有

$$\begin{aligned}
I(X^m; \hat{X}^m) &= h(X^m) - h(X^m|\hat{X}^m) \\
&= \sum_{i=1}^{m} h(X_i) - \sum_{i=1}^{m} h(X_i|X^{i-1}, \hat{X}^m) \\
&\ge \sum_{i=1}^{m} h(X_i) - \sum_{i=1}^{m} h(X_i|\hat{X}_i) \\
&= \sum_{i=1}^{m} I(X_i; \hat{X}_i) \\
&\ge \sum_{i=1}^{m} R(D_i) \\
&= \sum_{i=1}^{m} \left(\frac{1}{2}\log\frac{\sigma_i^2}{D_i}\right)^+,
\end{aligned}$$

$$\boxed{D_i = E\left(X - \hat{X}\right)^2}$$

核心部分是率失真是否受限

思考题

- 当编码速率为1比特时， 高斯信源（例1中）的选取代表元模式是否可以达到规定的失真率？
- 当编码速率为2时，高斯信源（例1中）的选取代表元模式是什么？是否达到规定的失真率？
- 为什么书中没有给出多离散信源的率失真函数？
- 为什么书中没有讨论离散Markov离散信源的率失真函数？
- 有何建议或新思考？ 理论是否存在缺陷？它们与机器学习的关系是什么？能否发现在机器学习理论中的新应用？

思考题

➢为什么书中没有给出多离散信源的率失真函数？
各个信源的错误概率的非一致性（多重要求带来的复杂性）

➢为什么书中没有讨论离散Markov离散信源的率失真函数？
这个问题目前有结论，可以参考有关论文

➢有何建议或新思考？ 理论是否存在缺陷？它们与机器学习的关系是什么？能否发现在机器学习理论中的新应用？

（信息表示(隐形的表示模式）是机器学习的重要部分）

率失真函数的计算

- **独立高斯随机变量的同步描述**
  - 对于m个独立但不同分布高斯随机信源$X_1, \cdots X_m$的表示问题，其中$X_i \sim \mathcal{N}(0, \sigma_i^2)$，考虑平方失真度量$d(x^n, \hat{x}^n) = \sum (x_i - \hat{x}_i)^2$ 。
  - 求解上述率失真函数的问题可简化为如下优化问题：

$$R(D) = \min_{\sum D_i = D} \sum_{i=1}^{m} \max\left\{\frac{1}{2}\ln\frac{\sigma_i^2}{D_i}, 0\right\}$$

构建拉格朗日函数 $$J(D) = \sum_{i=1}^{m} \frac{1}{2}\ln\frac{\sigma_i^2}{D_i} + \lambda \sum_{i=1}^{m} D_i$$

利用KKT条件可得并联高斯信源的率失真定理：

$$R(D) = \sum_{i=1}^{m} \frac{1}{2}\log\frac{\sigma_i^2}{D_i}$$

λ的选取满足$\Sigma D_i = D$

其中 $$D_i = \begin{cases} \lambda & if\ \lambda < \sigma_i^2 \\ \sigma_i^2 & if\ \lambda \ge \sigma_i^2 \end{cases}$$

常数和非负性导致注水策略

率失真函数的计算

Tsinghua University

- 独立高斯随机变量的同步描述
  - 并联高斯信源的率失真定理：

$$\boxed{R(D) = \sum_{i=1}^{m} \frac{1}{2} \log \frac{\sigma_i^2}{D_i}}$$

其中 $D_i = \begin{cases} \lambda & \text{if } \lambda < \sigma_i^2 \\ \sigma_i^2 & \text{if } \lambda \geq \sigma_i^2 \end{cases}$

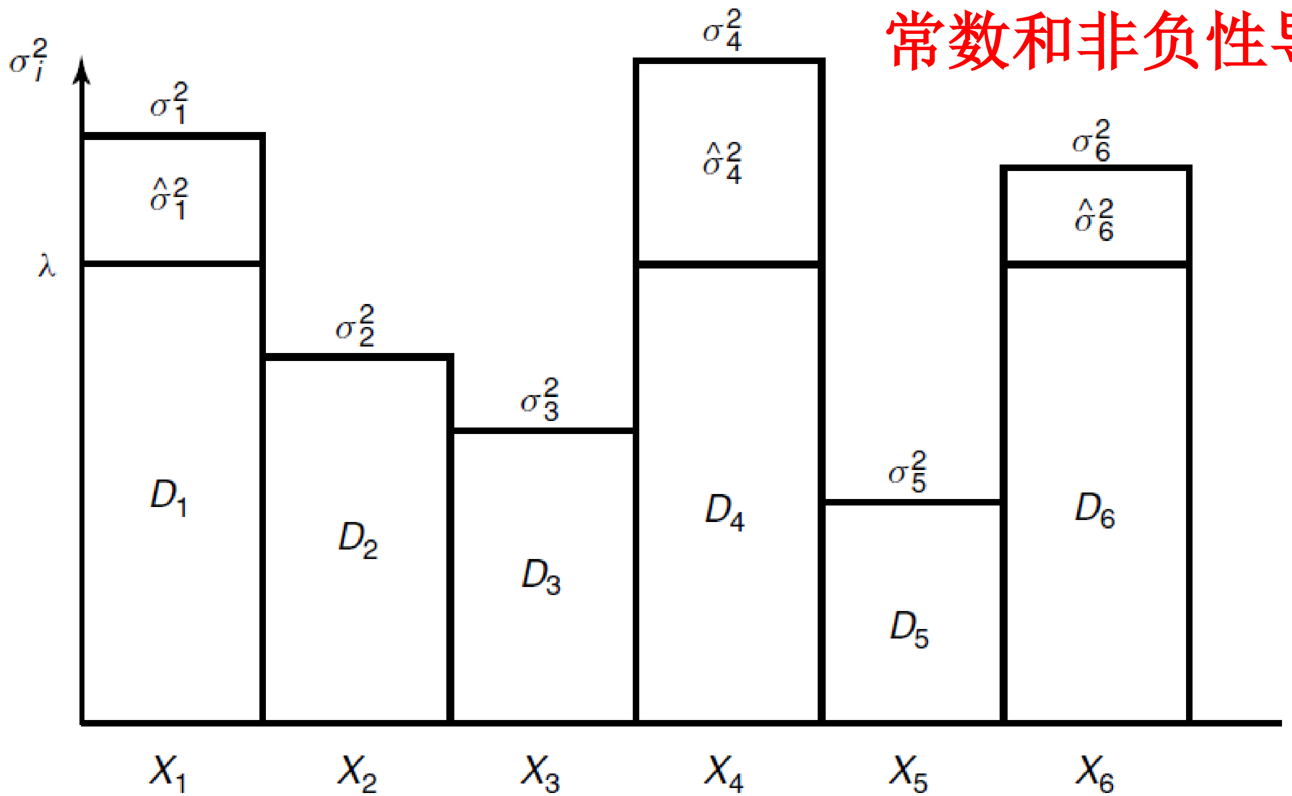
  - 对于失真度量为均方误差的多元高斯信源，其率失真函数可由反注水法依据协方差矩阵的特征值给出。

$$X \sim \mathcal{N}(0, \begin{bmatrix} \sigma_1^2 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \sigma_m^2 \end{bmatrix}), \qquad \hat{X} \sim \mathcal{N}(0, \begin{bmatrix} \hat{\sigma}_1^2 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \hat{\sigma}_m^2 \end{bmatrix})$$

率失真函数的计算

- 独立高斯随机变量的同步描述
  - 反注水法：选定常量λ，只描述方差比λ大的随机变量，方差小的不用比特描述。

常数和非负性导致注水策略



本章学习逻辑线条：

失真率定义

失真率的性质

失真率计算的例子

是否存在编码模式满足给定的失真率？

一般情况失真率如何计算：计算方法是什么

新发现：失真率与信道容量的对偶关系

科学问题：

是否满足失真率的编码都需要满足规定的编码速率界R（D）？

可行的编码其编码方案是什么？
是否可以举一个例子说明？
（Shannon的回答）

率失真定理的逆定理

- **率失真定理的逆定理：**

对于失真度量$d(x,\hat{x})$，且i.i.d.服从$p(x)$的任何信源X，以及失真$\le D$的任何一个$(2^{nR},n)$率失真码，该编码的码率必定满足$R \ge R(D)$。（充要条件）

证明：设$\hat{X}^n$为信源X^n的再生序列，则有：

$$
\begin{aligned}
nR &\overset{(a)}{\ge} H(f_n(X^n)) && \boxed{f_n\text{的值域最多为}2^{nR}} \\
&\overset{(b)}{\ge} H(f_n(X^n)) - H(f_n(X^n)|X^n) && \boxed{H(f_n(\hat{X}^n)|X^n) \ge 0} \\
&= I(X^n; f_n(X^n)) \\
&\overset{(c)}{\ge} I(X^n; \hat{X}^n) && \boxed{\text{数据处理不等式}} \\
&= H(X^n) - H(X^n|\hat{X}^n) \\
&\overset{(d)}{=} \sum_{i=1}^{n} H(X_i) - H(X^n|\hat{X}^n) && \boxed{X_i\text{相互独立}} \\
&\overset{(e)}{=} \sum_{i=1}^{n} H(X_i) - \sum_{i=1}^{n} H(X_i|\hat{X}^n, X_{i-1}, \ldots, X_1) && \boxed{\text{熵的链式法则}} \\
&\overset{(f)}{\ge} \sum_{i=1}^{n} H(X_i) - \sum_{i=1}^{n} H(X_i|\hat{X}_i) && \boxed{\text{加入条件熵减}}
\end{aligned}
$$

率失真定理的逆定理

- **率失真定理的逆定理：**

对于失真度量$d(x,\hat{x})$，且i.i.d.服从$p(x)$的任何信源X，以及失真$\leq D$的任何一个$(2^{nR},n)$率失真码，该编码的码率必定满足$R \geq R(D)$。（充要条件）

证明：（续）

$$= \sum_{i=1}^{n} I(X_i;\hat{X}_i)$$

$$\overset{(g)}{\geq} \sum_{i=1}^{n} R(Ed(X_i,\hat{X}_i))$$ 率失真函数定义

$$= n\left(\frac{1}{n}\sum_{i=1}^{n} R(Ed(X_i,\hat{X}_i))\right)$$

$$\overset{(h)}{\geq} nR\left(\frac{1}{n}\sum_{i=1}^{n} Ed(X_i,\hat{X}_i)\right)$$ 率失真函数的凸性及Jensen不等式

$$\overset{(i)}{=} nR(Ed(X^n,\hat{X}^n))$$ 分组长度为n的失真函数的定义

$$\overset{(j)}{=} nR(D),$$ $R(D)$非增且$Ed(X_n,\hat{X}^n) \leq D$

即**任意率失真码的码率R都要比在失真水平$D = Ed(X^n,\hat{X}^n)$下的率失真函数$R(D)$要大。**

率失真定理的逆定理

- 信道传输角度：带失真的信源信道分离定理

令$\boldsymbol{V_1,\cdots V_n}$为有限个独立同分布字母表的信源，编码为容量$\boldsymbol{C}$的离散无记忆信道中的$\boldsymbol{n}$个输入字符序列$\boldsymbol{X^n}$。而信道输出$\boldsymbol{Y^n}$映射为重构字母表$\boldsymbol{\hat{V}^n = g(Y^n)}$。令$\boldsymbol{D = Ed(V^n,\hat{V}^n) = \frac{1}{n}\Sigma d(V_i,\hat{V}_i)}$为该组合信源和信道编码方案构成的平均失真。

该失真$\boldsymbol{D}$可达当且仅当$\boldsymbol{C > R(D)}$成立。实时通信

率失真函数的可达性

Shannon第三定理：保真度准则下的信源编码定理

- 设$X_1, \cdots X_n$为i.i.d.~ $p(x)$，该信源失真度量有界，率失真函数为$R(D)$，则对任意的D及任意的$R > R(D)$，必定存在具有码率R和渐进失真D的率失真码序列，满足编码后平均失真小于等于$D+\delta$，$\delta > 0$。（即率失真对(R, D)可达）

- 关键解释：存在信源编码（编码器），其平均失真可以任意接近给定的D，其码率
- 可以任意接近$R(D)$。

- 证明思路：
  - 考虑联合AEP修正情形，在给定失真度量下，增加条件为考虑的序列对是典型的。
  - 选定$p(\hat{x}|x)$使得$R(D) = I(X; \hat{X})$，计算$p(\hat{x})$，
  - 证明码率为R且失真不大于$D+\delta$的率失真码的存在性。

率失真函数的可达性

- 失真典型序列
  - 设$p(x,\hat{x})$为$\mathcal{X}\times\hat{\mathcal{X}}$上的联合分布， $d(x,\hat{x})$为失真度量。若对任意$\epsilon>0$，序列对$(x^n,\hat{x}^n)$满足：

$$\left|-\frac{1}{n}\log p(x^n)-H(X)\right|<\epsilon$$

$$\left|-\frac{1}{n}\log p(\hat{x}^n)-H(\hat{X})\right|<\epsilon$$

$$\left|-\frac{1}{n}\log p(x^n,\hat{x}^n)-H(X,\hat{X})\right|<\epsilon$$

$$|d(x^n,\hat{x}^n)-Ed(X,\hat{X})|<\epsilon$$ 新增条件

则称其为失真典型的，所有失真典型序列构成的集合称为失真典型集$A_{d,\epsilon}^{(n)}$。

- 注1：上述定义即附加失真接近期望值限制条件下的联合典型集定义，因此失真典型集是联合典型集的子集。
- 注2：由大数定律，序列对的平均失真$d(X^n,\hat{X}^n)=\frac{1}{n}\sum d(X_i,\hat{X}_i)$依概率收敛于期望失真。

率失真函数的可达性

- 证明中所用到的引理
  - 引理1：

设$(X_i, \hat{X}_i)$为独立同分布的序列$\sim p(x, \hat{x})$，那么当$n \to \infty$时，$\Pr\left(A_{d,\epsilon}^{(n)}\right) \to 1$。

证明：

由定义$A_{d,\epsilon}^{(n)}$中的4个条件求和具有i.i.d.随机变量的标准化求和形式，由大数

定律，这些求和值均将以概率1收敛与各自的期望值。

于是，当 $n \to \infty$ 时，满足4个条件的所有序列构成的集合的概率将趋近于1。

率失真函数的可达性

- **证明中所用到的引理**
  - **引理2：**

对任意$(\boldsymbol{x}^n,\hat{\boldsymbol{x}}^n)\in A_{d,\varepsilon}^{(n)}$，有 $p(\hat{x}^n)\geq p(\hat{x}^n|x^n)2^{-n(I(X;\hat{X})+3\epsilon)}$

证明：

由$A_{d,\epsilon}^{(n)}$的定义，可以对任意典型序列$(\boldsymbol{x}^n,\hat{\boldsymbol{x}}^n)\in A_{d,\epsilon}^{(n)}$的概率值$p(\boldsymbol{x}^n),p(\hat{\boldsymbol{x}}^n)$和$p(\boldsymbol{x},\hat{\boldsymbol{x}})$做出界估计：

$$
\begin{aligned}
p(\hat{x}^n|x^n) &= \frac{p(x^n,\hat{x}^n)}{p(x^n)} \\
&= p(\hat{x}^n)\frac{p(x^n,\hat{x}^n)}{p(x^n)p(\hat{x}^n)} \\
&\leq p(\hat{x}^n)\frac{2^{-n(H(X,\hat{X})-\epsilon)}}{2^{-n(H(X)+\epsilon)}2^{-n(H(\hat{X})+\epsilon)}} \\
&= p(\hat{x}^n)2^{n(I(X;\hat{X})+3\epsilon)},
\end{aligned}
$$

则引理成立。

率失真函数的可达性

- 证明中所用到的引理
  - 引理3：

对$0 \le x, y \le 1$，$n > 0$，有如下不等式成立

$$\boxed{(1-xy)^n \le 1 - x + e^{-yn}.}$$

证明：设$f(y) = e^{-y} - 1 + y$，则有$f(0) = 0$，当$y > 0$时，$f'(y) = -e^{-y} + 1 > 0$，

可得$f(y) > 0$。于是对于$0 \le y \le 1$，有$1 - y \le e^{-y}$，

则$(1-y)^n \le e^{-yn}$，即$x = 1$时引理成立。检验可得$x = 0$也成立。通过求导容易得

到$g_y(x) = (1-xy)^n$是x的凸函数，则对$0 \le x \le 1$，有

$$\begin{aligned}
(1-xy)^n &= g_y(x) \\
&\le (1-x)g_y(0) + xg_y(1) \\
&= (1-x)1 + x(1-y)^n \\
&\le 1 - x + xe^{-yn} \\
&\le 1 - x + e^{-yn}. \quad \square
\end{aligned}$$

连续用了两次
凸函数性质
第二次很巧妙

率失真函数的可达性

- 保真度准则下的信源编码定理（4条基本规定步骤）

  - 设$X_1,\cdots X_n$为i.i.d.~ $p(x)$，该信源失真度量有界，率失真函数为$R(D)$，则对任意的D及任意的$R > R(D)$，必定存在具有码率R和渐进失真D的率失真码序列，满足编码后平均失真小于等于$D+\delta$，$\delta > 0$。

1. 码簿的生成

随机生成由2^{nR}个i.i.d. $\sim \prod p(\hat{x}_i)$的序列$\hat{X}_n$组成的率失真码簿$\mathcal{C}$。为这些码字做下标$\omega \in \{1, 2, \cdots, 2^{nR}\}$，该码簿呈现在编码器和译码器两端。

率失真函数的可达性

2. 编码方法

- 若码簿中存在一个标号ω的码字，使得$\left(X^n, \hat{X}^n(\omega)\right) \in A_{d,\epsilon}^{(n)}$，则将$X^n$编码为$\omega$；
- 若码簿中存在多个标号ω的码字，使得$\left(X^n, \hat{X}^n(\omega)\right) \in A_{d,\epsilon}^{(n)}$，则选取最小的$\omega$；
- 若码簿中不存在这样的码字，则令$\omega = 1$。

由此，将X空间所有序列编码为码簿中的码字，且nR比特足以描述联合典型码字的下标ω。

3. 译码方法

译码端输出的再生序列即为$\hat{X}^n(\omega)$。

率失真函数的可达性

4. 平均失真计算

计算在所有随机选取的码簿$\mathcal{C}$上的期望失真：$\overline{D} = E_{X^n,\mathcal{C}} d(X^n, \hat{X}^n)$

所取的期望是针对码簿的随机选取和X^n而言

对于选定的码簿$\mathcal{C}$与$\epsilon > 0$，将所有序列x^n分为两类：

- 存在一个码字$\hat{X}^n(\omega)$与序列x^n是失真典型，即$d\left(x^n, \hat{x}^n(\omega)\right) < D + \epsilon$。由于这些序列的总概率至多为1，则这些序列对期望失真的贡献不会超过$D + \epsilon$；
- 不存在上述要求的码字$\hat{X}^n(\omega)$的序列x^n。记P_e为这类序列的总概率，单个序列失真的上界为d_{max}，则其对期望失真的贡献最多为$P_e d_{max}$。

因此，可将总失真定界如下：

$$Ed(X^n, \hat{X}^n(X^n)) \leq D + \epsilon + P_e d_{\max}$$

若能证明P_e足够小，则当适当选取ε使得上式左边小于$D + \delta$。

即对于随机选取的码簿和信源序列，估计不存在与该信源序列失真典型的码字的概率的界。

率失真函数的可达性

Tsinghua University

P_e的计算

记$J(\mathcal{C})=\left\{x^n:\left(x^n,\hat{x}^n(\omega)\right)\in A_{d,\epsilon}^{(n)},\omega\in\{1,2,\cdots,2^{nR}\}\right\}$，为满足$\mathcal{C}$中至少存在一个码字与$x^n$构成失真典型的序列$x^n$构成的集合。$P_e$是由不在集合内的信源序列引起的，于是

$$P_e=\sum_{\mathcal{C}}P(\mathcal{C})\sum_{x^n:x^n\notin J(\mathcal{C})}p(x^n)$$

所有码本都考虑

改变求和顺序，即选取的码簿不能很好表示信源序列 x^n的概率，则有

$$P_e=\sum_{x^n}p(x^n)\sum_{\mathcal{C}:x^n\notin J(\mathcal{C})}p(\mathcal{C})$$

定义

$$K(x^n,\hat{x}^n)=\begin{cases}1 & \text{if } (x^n,\hat{x}^n)\in A_{d,\epsilon}^{(n)}\\ 0 & \text{if } (x^n,\hat{x}^n)\notin A_{d,\epsilon}^{(n)}\end{cases}$$

则选取码字$\hat{X}^n$不能与信源序列x^n构成失真典型的概率为：

$$\Pr((x^n,\hat{X}^n)\notin A_{d,\epsilon}^{(n)})=\Pr(K(x^n,\hat{X}^n)=0)=1-\sum_{\hat{x}^n}p(\hat{x}^n)K(x^n,\hat{x}^n)$$

率失真函数的可达性

P_e的计算

$$P_e = \sum_{x^n} p(x^n) \sum_{\mathcal{C}: x^n \notin J(\mathcal{C})} p(\mathcal{C})$$

进而得到

$$= \sum_{x^n} p(x^n) \left[1 - \sum_{\hat{x}^n} p(\hat{x}^n) K(x^n, \hat{x}^n)\right]^{2^{nR}}$$

利用引理2估计括号中求和项的界，可得

$$\sum_{\hat{x}^n} p(\hat{x}^n) K(x^n, \hat{x}^n) \geq \sum_{\hat{x}^n} p(\hat{x}^n|x^n) 2^{-n(I(X;\hat{X})+3\epsilon)} K(x^n, \hat{x}^n).$$

则有 $$P_e \leq \sum_{x^n} p(x^n) \left(1 - 2^{-n(I(X;\hat{X})+3\epsilon)} \sum_{\hat{x}^n} p(\hat{x}^n|x^n) K(x^n, \hat{x}^n)\right)^{2^{nR}}$$

再利用引理3估计

上式右侧项的界：

$$\left(1 - 2^{-n(I(X;\hat{X})+3\epsilon)} \sum_{\hat{x}^n} p(\hat{x}^n|x^n) K(x^n, \hat{x}^n)\right)^{2^{nR}}$$

$$\leq 1 - \sum_{\hat{x}^n} p(\hat{x}^n|x^n) K(x^n, \hat{x}^n) + e^{-(2^{-n(I(X;\hat{X})+3\epsilon)} 2^{nR})}$$

率失真函数的可达性

P_e的计算

代入估计的界可得

$$P_e \le 1 - \sum_{x^n} \sum_{\hat{x}^n} p(x^n) p(\hat{x}^n | x^n) K(x^n, \hat{x}^n) + e^{-2^{-n(I(X;\hat{X})+3\epsilon)} 2^{nR}}$$

其中最后一项也等于

$$e^{-2^{n(R-I(X;\hat{X})-3\epsilon)}}$$

当$R > I(X;\hat{X}) + 3\epsilon$时，它随$n$以指数级快速衰减至0。因此，选取$p(\hat{x}|x)$为达到率失真函数最小值时的条件分布，则$R > R(D)$意味着$R > I(X;\hat{X})$，并且只要选取足够小的$\varepsilon$，即可让最后一项趋于0。

率失真函数的可达性

P_e的计算

$$P_e \le 1 - \sum_{x^n}\sum_{\hat{x}^n} p(x^n)p(\hat{x}^n|x^n)K(x^n,\hat{x}^n) + e^{-2^{-n(I(X;\hat{X})+3\epsilon)}2^{nR}}$$

对于前两项，由引理1可得

$$1 - \sum_{x^n}\sum_{\hat{x}^n} p(x^n)p(\hat{x}^n|x^n)K(x^n,\hat{x}^n) = \Pr\left(\left(X^n,\hat{X}^n(\omega)\right) \notin A_{d,\epsilon}^{(n)}\right) < \epsilon$$

因此，适当的选取ϵ和n，能使P_e任意小。

率失真函数的可达性

上述证明表明，对于任意选取的$\delta > 0$，存在ϵ和n，对于分组长度为n且码率为R的所有随机选取的编码，期望失真小于$D+\delta$。因此必定存在一个具有该码率和分组长度的编码，其平均失真不大于$D+\delta$，由于δ是任意的，于是证明了$R > R(D)$时(R, D)是可达的。■

小结：

- 提供了和前面信道容量可达性的类似证明；
- 典型集就是最主要的工具
- 提供了一个特殊的不等式，需要重点关注

信源失真压缩与信道容量对偶规则的发现

率失真函数的可达性

Tsinghua University

- 率失真定理的随机编码证明 ⟷ 信道编码定理的随机编码证明

- 以高斯分布为例讨论二者之间的相似性

信道传输： $Y=X+N;$

信源失真编码： $X=\hat{X}+N$

率失真函数的可达性

清華大學 Tsinghua University

- **高斯信道的信道编码**

考虑高斯信道$Y_i = X_i + Z_i$，其中$Z_i \sim \mathcal{N}(0, N)$ i.i.d.，且该信道在传输码字上的单符号功率限制为P，考虑一个n长的传输序列。功率限制使传输序列限制在$\mathcal{R}^n$中半径为$\sqrt{nP}$的球内。编码问题等价于在该球内找一个由2^{nR}个序列构成的集合，使其中任意序列被误认为其他序列的概率尽可能地小，即以每个序列为中心半径为$\sqrt{nN}$的球体几乎不相交。这相当于用$\sqrt{nN}$的小球去填塞半径为$\sqrt{n(P+N)}$的球。期望容纳球的最大数量为体积之比，也为半径之比的n次幂。记M为能有效传输的码字的数量，则有

$$M \le \frac{(\sqrt{n(P+N)})^n}{(\sqrt{nN})^n} = \left(\frac{P+N}{N}\right)^{\frac{n}{2}}$$

信道编码定理说明，当n很大时，大约可以找到

$$\boxed{2^{nC} = \left(\frac{P+N}{N}\right)^{\frac{n}{2}}}$$

接收端信号分离观点

个码字，使以它们为中心的有噪声球邻域几乎不相交（相交的总体积任意小）。

率失真函数的可达性

- 高斯信源的率失真

方差为σ^2的高斯信源，具有失真D的某$(2^{nR}, n)$率失真码为$\mathcal{R}^n$中2^{nR}个序列组成的集合，其中大多数长度为n的信源序列（即所有位于半径$\sqrt{n\sigma^2}$球内的信源序列）在某个码字$\sqrt{nD}$邻域内。同样使用填球模型法，可得最少所需码字数量为

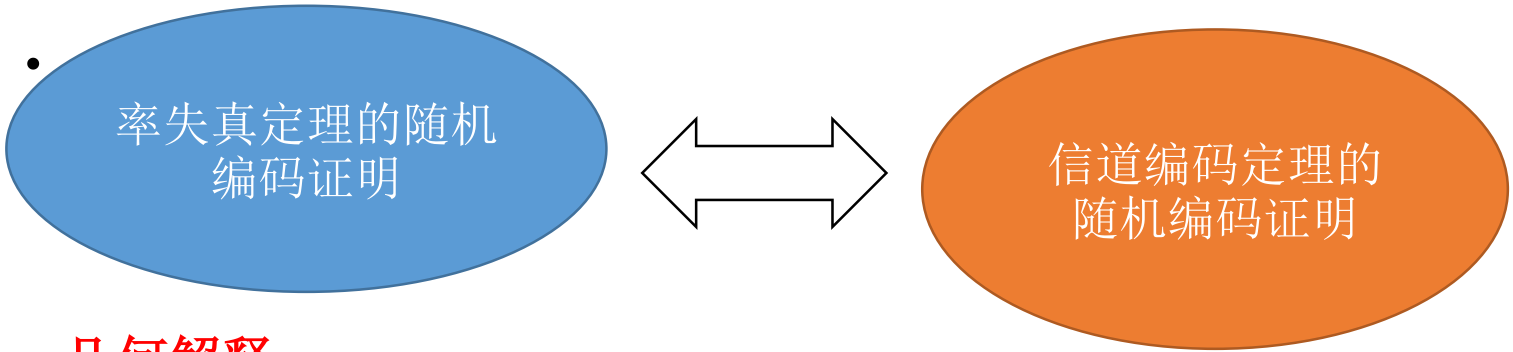
发端覆盖方法可以包含所有信源情况

$$\boxed{2^{nR(D)} = \left(\frac{\sigma^2}{D}\right)^{\frac{n}{2}}}$$

率失真定理说明此最小码率是渐进可达的，即存在一族半径为$\sqrt{nD}$的球，能够覆盖除去其概率可以任意小的一个集合之外的空间。

$$\boxed{2^{nC} = \left(\frac{P+N}{N}\right)^{\frac{n}{2}}}$$

率失真函数的可达性



- 几何解释
  - 信道编码： 填球模型 （追求最多）
  - 率失真编码：球覆盖模型 （追求最少）
- 信道传输码与率失真码的转换
  - 满足其一填球模型的界，则对另一个情形也满足填球模型的界。
  - 高斯情形：对于率失真编码和信道编码，码字为高斯且有适当方差的都为渐进最佳

科学问题：

◆对于一个给定的离散信源，率失真函数是否有显示表达式？

◆如果没有显示表达式，如何计算？

◆是否可以给一个具体的例子？

◆这个算例的通用性如何？

强典型序列与率失真

- **强典型性定义**
  - **强典型序列** （计数策略）

    称序列$x^n \in \mathcal{X}^n$关于$\mathcal{X}$上的分布$p(x)$是ϵ-强典型的，满足：

    1）对于任意$a \in \mathcal{X}$，且$p(a) > 0$，有

$$\left|\frac{1}{n}N(a|x^n) - p(a)\right| < \frac{\epsilon}{|\mathcal{X}|}$$

    2）对任意$a \in \mathcal{X}$，且$p(a) = 0$，有$N(a|x^n) = 0$。

    其中$N(a|x^n)$表示字符a在序列x^n中出现的次数。

  - 这是针对随机变量的局部处理技术，加强关注每一个序列的类型

  - 离散随机变量序列的处理模式？

强典型序列与率失真

- **强典型性定义**
  - **强典型序列对**：

称序列对$(\boldsymbol{x}^n, \boldsymbol{y}^n) \in \mathcal{X}^n \times \mathcal{Y}^n$关于$\mathcal{X} \times \mathcal{Y}$上的分布$\boldsymbol{p}(\boldsymbol{x}, \boldsymbol{y})$是$\epsilon$-强典型的，若满足：

1）对于任意$(\boldsymbol{a}, \boldsymbol{b}) \in \mathcal{X} \times \mathcal{Y}$，且$\boldsymbol{p}(\boldsymbol{a}, \boldsymbol{b}) > \mathbf{0}$，有

$$\left|\frac{1}{n}N(a, b|x^n, y^n) - p(a, b)\right| < \frac{\epsilon}{|\mathcal{X}||\mathcal{Y}|}$$

2）对任意$(\boldsymbol{a}, \boldsymbol{b}) \in \mathcal{X} \times \mathcal{Y}$，且$\boldsymbol{p}(\boldsymbol{a}, \boldsymbol{b}) = \mathbf{0}$，有$\boldsymbol{N}(\boldsymbol{a}, \boldsymbol{b}|\boldsymbol{x}^n, \boldsymbol{y}^n) = \mathbf{0}$。

其中$\boldsymbol{N}(\boldsymbol{a}, \boldsymbol{b}|\boldsymbol{x}^n, \boldsymbol{y}^n)$表示字符对$(\boldsymbol{a}, \boldsymbol{b})$在序列对$(\boldsymbol{x}^n, \boldsymbol{y}^n)$中出现的次数。

强典型序列与率失真

- **强典型序列率失真函数的可达性**

  期望失真接近于$\boldsymbol{D}$，且能够找到一个码字使其与给定序列间的失真依概率小于$\boldsymbol{D}+\boldsymbol{\delta}$。

- 证明思路同上，使用强典型序列

1. 码簿的生成 | 2. 编码方法 | 3. 译码方法 | 4. 平均失真计算 | P_e的计算

估计界时考虑三类信源序列

对偶性理论

- **观察信道容量与率失真之间的关系？**

- 信道容量强调**输入分布**的优化问题，选择一个KL-divergence的上界其物理意义？

最好信息转移水平。

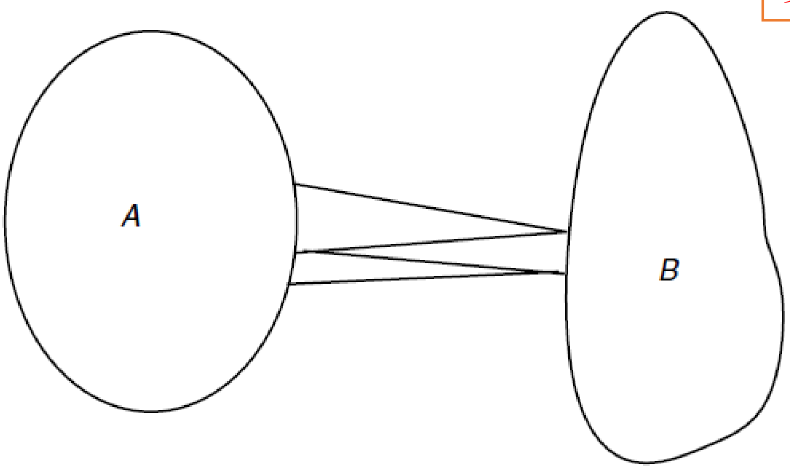
- 率失真强调**转移概率**的优化问题，选择一个**KL-Divergence 的下界其物理意义？**

最坏情况下的信息表示能力。

信道容量与率失真函数的计算

- 交替迭代式计算凸集间的最小距离
  - 交替固定，找最优值。
  - 若两个集合为概率分布集，距离度量为KL距离，则算法收敛至两个分布集之间的最小相对熵。

类比**EM**算法



- 信道容量与通用数据压缩的对偶性

- **信道容量与通用数据压缩的对偶性**
  - 改写为相对熵的形式

引理：设$p(x)p(y|x)$为给定联合分布，则使相对熵$D\big(p(x,y)\|p(x)r(y)\big)$最小化的分布$r(y)$是对应于$p(y|x)$的边际分布$r^*(y)$，即：

$$D(p(x)p(y|x)\|p(x)r^*(y)) = \min_{r(y)} D(p(x)p(y|x)\|p(x)r(y))$$

其中

$$r^*(y) = \Sigma\, p(x)p(y|x)。$$

同时

$$\max_{r(x|y)} \sum_{x,y} p(x)p(y|x)\log\frac{r(x|y)}{p(x)} = \sum_{x,y} p(x)p(y|x)\log\frac{r^*(x|y)}{p(x)}$$

其中

$$\boxed{r^*(x|y) = \frac{p(x)p(y|x)}{\sum_x p(x)p(y|x)}.}$$

- 信道容量与通用数据压缩的对偶性
  - 率失真函数的计算

利用上述引理，可将率失真函数的计算改写为双重最小化问题：

$$R(D)=\min_{r(\hat{x})}\min_{q(\hat{x}|x):\sum p(x)q(\hat{x}|x)d(x,\hat{x})\le D}\sum_{x}\sum_{\hat{x}}p(x)q(\hat{x}|x)\log\frac{q(\hat{x}|x)}{r(\hat{x})}$$

令集合A为其边际分布$\boldsymbol{p}(\boldsymbol{x})$满足失真限制的所有联合分布构成的集合，B为乘积分布$\boldsymbol{p}(\boldsymbol{x})\boldsymbol{r}(\hat{\boldsymbol{x}})$全体构成的集合，其中$\boldsymbol{r}(\hat{\boldsymbol{x}})$为任意的，则有下述优化问题：

$$R(D)=\min_{q\in B}\min_{p\in A}D(p||q)$$

信道容量与率失真函数的计算

Tsinghua University

- 信道容量与通用数据压缩的对偶性
  - 率失真函数的计算

$$R(D) = \min_{q \in B} \min_{p \in A} D(p||q)$$

Double Min 的处理技术及其内涵？

  - 可通过交替最小化算法，Blahut-Arimoto算法求解。

先选定某个λ以及初始分布$r(\hat{x})$，计算在失真限制下使互信息最小的$q(\hat{x}|x)$。

（1）由拉格朗日乘子法：

$$q(\hat{x}|x) = \frac{r(\hat{x}) e^{-\lambda d(x,\hat{x})}}{\sum_{\hat{x}} r(\hat{x}) e^{-\lambda d(x,\hat{x})}}$$

（2）固定此条件分布，由引理得到使互信息最小的输出分布：

$$r(\hat{x}) = \sum_{x} p(x) q(\hat{x}|x)$$

以此输出分布作为下次迭代的起点，每一步先关于$q(\hat{x}|x)$最小化，再关于$r(\hat{x})$最小化。

信道容量与率失真函数的计算

- 信道容量与通用数据压缩的对偶性
  - 信道容量的计算

类似地，由信道容量定义

$$C = \max_{r(x)} I(X;Y) = \max_{r(x)} \sum_x \sum_y r(x)p(y|x) \log \frac{r(x)p(y|x)}{r(x)\sum_{x'} r(x')p(y|x')}$$

可将信道容量的计算改写为双重最大化问题：

$$C = \max_{q(x|y)} \max_{r(x)} \sum_x \sum_y r(x)p(y|x) \log \frac{q(x|y)}{r(x)}$$

Double Max 的处理技术及其内涵？

信道容量与率失真函数的计算

- 信道容量与通用数据压缩的对偶性
  - 信道容量的计算

$$C = \max_{q(x|y)} \max_{r(x)} \sum_x \sum_y r(x) p(y|x) \log \frac{q(x|y)}{r(x)}$$

  - 可通过交替最大化算法，Csiszár-Tusnády算法求解。

先猜测最大化分布$r(\hat{x})$，然后求出最佳的条件分布。
（1）由引理知条件分布即

$$q(x|y) = \frac{r(x)p(y|x)}{\sum_x r(x)p(y|x)}$$

（2）固定此条件分布，利用拉格朗日乘子法求解带约束的最大化问题，得到最优的输入分布：

$$r(x) = \frac{\prod_y (q(x|y))^{p(y|x)}}{\sum_x \prod_y (q(x|y))^{p(y|x)}}$$

以此作为起点进行下次迭代。

总结

- 率失真理论：

- 与信道容量分析的对偶处理技术；

- 最优性；
- 存在性；
- 可达性；

- 信道容量和率失真的迭代计算 **double max or double min**

中间用到一个重要的不等式