应用信息论基础——
第五、六章作业

2022年12月14日

第五章重点总结

- 率失真
  - 失真函数
  - 信息率失真及其性质：非增、凸性

信道容量： 信道确定，评估容量
率失真： 信源确定，对“变换”的最低要求

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

- **率失真定理：** 失真条件下对码字数的要求

- **率失真函数的计算**
  - 伯努利、高斯：**引入反向信道；** **两种展开：** $I(X;\hat{X})=H(X)-H(X|\hat{X})=H(\hat{X})-H(\hat{X}|X)$
  - 多元高斯：**反注水法**

- **信源信道分离定理：** 建立率失真与信道容量的联系 $\boxed{R(D)\le C}$

- 率失真与信道容量的对偶关系：迭代计算

1：$R(0)=H(X)$的失真矩阵性质

1. 设离散无记忆信源X熵为$H(X)$，失真矩阵为$[d(k,j)]_{K\times J}$，率失真函数为$R(D)$，试证明$R(0)=H(X)$的充要条件是失真矩阵的每一行至少有一个0元素，而每一列至多有一个0元素。

1：$R(0)=H(X)$的失真矩阵性质

Tsinghua University

- **证明**
  - **充分性：若失真矩阵的每一行至少有一个0元素，而每一列至多有一个0元素，则$\boldsymbol{R(0)=H(X)}$。**

在此失真矩阵的条件下，可以找到这样的试验信道，使输出端只传输到使得$d(x,\hat{x})=0$的对应符号$\hat{x}$，在此试验信道中必有$\mathrm{E}\{d(x,\hat{x})\}=0$满足；

又由于失真矩阵每列最多一个0，则**对于每个收端符号$\hat{x}$，有0个或唯一的x与之对应**，即满足条件的信道为无损信道，$H(X|\hat{X})=0$，即$R(0)=H(X)$。

  - **必要性：若$\boldsymbol{R(0)=H(X)}$，则失真矩阵中每一行至少有一个零元素，且每一列至多有一个零元素。**

由于$R(0)=H(X)$，即失真可以取到0，则对于每个x至少有一个$\hat{x}$在失真度量下无失真，即失真矩阵每行至少有一个0；

如果失真矩阵第j列有多于1个0，**对于$\hat{x}_j$，可以有多个x与之对应，都能保证总失真为0，此时$\boldsymbol{H(X|\hat{X})>0}$**，即存在试验信道使得$I(X;\hat{X})=H(X)-H(X|\hat{X})<H(X)$，则$R(0)<H(X)$，矛盾，因此每列至多有一个0。

1：$R(0)=H(X)$的失真矩阵性质

- **直观理解**

$R(0)=H(X)$：无失真的情况下想要用$H(X)$的码率描述X，失真矩阵的定义不能过分严苛，也不能太宽松。

1）**定义的失真矩阵至少能实现0失真**（每行至少一个0）；

2）且$\boldsymbol{X}$**可被唯一对应还原**（每列至多一个0），互信息不可低于X的信息量。

暗含$K\leq J$：可能存在$\hat{x}$用不到

2. 令X和$\tilde{X}$为离散随机变量，均取值于集合$\{0,1\}$，X的取值分布为$\{0.5,0.5\}$。现针对离散无记忆信源进行限失真压缩编码，编码器输入和输出分别为X和$\tilde{X}$，定义非对称失真度量矩阵为$\begin{bmatrix} 0 & \infty \\ 1 & 0 \end{bmatrix}$。设失真度量为$d(x,\tilde{x})$，$R(D)$为$X$的率失真函数。

1）求$R(0)$；

2）求使$R(D_0)=0$的最小D_0；

3）对于$D<D_0$，在$\mathrm{E}[d(x,\tilde{x})]\le D$条件下，求使$I(X;\tilde{X})$最小的条件分布$p(\tilde{x}|x)$；

4）给出率失真函数$R(D)$，$0\le D\le D_0$.

Tsinghua University 清華大學

- 1) 对于此失真矩阵则信道为无损信道
  $H(X|\hat{X})=0,\ R(0)=H(X)=1$
- 2) $R(D_0)=0$：X与$\hat{X}$独立。
- 3) 有限失真全部由$p(x=1,\hat{x}=0)$带来
  可得联合分布，进而得到条件分布。
  $p(x,\hat{x})$中$\frac{1}{2}$是由于$p(x=0,\hat{x}=1)=0$。
- 4) 带入定义即可。

目标是求解以下优化问题

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

由于 $d(0,1)=\infty$，对有限的失真 D 一定有 $p(\tilde{x}=1|x=0)=0$。因此失真 $D=p(\tilde{x}=0,x=1)$。因此对$(X,\tilde{X})$有以下的联合分布（假设$D\le\frac{1}{2}$）

$$\boxed{p(x,\tilde{x})=\begin{bmatrix}\frac{1}{2} & 0\\ D & \frac{1}{2}-D\end{bmatrix}}$$

由此联合分布可以得到互信息

$$R(D)=I(X;\tilde{X})=H(X)-H(X|\tilde{X})$$
$$=H\left(\frac{1}{2}\right)-\left(\frac{1}{2}+D\right)H\left(\frac{D}{\frac{1}{2}+D},\frac{\frac{1}{2}}{\frac{1}{2}+D}\right)$$
$$=1+\frac{1}{2}\log_2\frac{\frac{1}{2}}{\frac{1}{2}+D}+D\log_2\frac{D}{\frac{1}{2}+D}$$

(1) 令$D=0$，将其代入$R(D)$表达式可得$R(0)=1$。

(2) 当$D=\frac{1}{2}$时，$R(D)=0$。而$R(D)$是关于失真D的减函数，因此使得$R(D_0)=0$的最小$D_0=\frac{1}{2}$。

(3) 由以上分析知

$$p(\tilde{x}|x)=\begin{bmatrix}1 & 0\\ 2D & 1-2D\end{bmatrix}$$

(4) 率失真函数为

$$R(D)=1+\frac{1}{2}\log_2\frac{\frac{1}{2}}{\frac{1}{2}+D}+D\log_2\frac{D}{\frac{1}{2}+D}$$

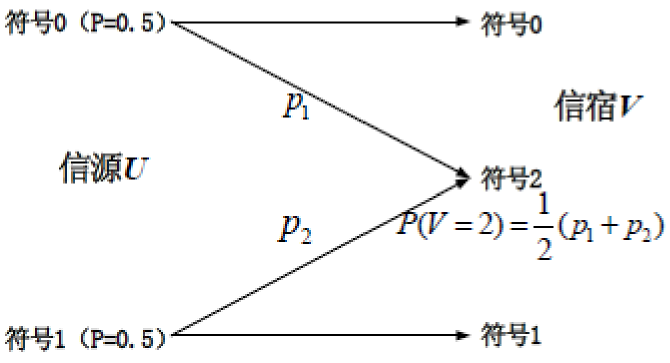
3-2对3的率失真计算

3. 已知信源$U=\{0,1\}$，信宿$V=\{0,1,2\}$。设信源输入为等概分布，失真矩阵为$D=\begin{bmatrix} 0 & \infty & 1 \\ \infty & 0 & 1 \end{bmatrix}$，求此信源的率失真函数。

3-2对3的率失真计算

- 将失真为∞对应的转移概率置为0。

$$\frac{1}{2}(p_1+p_2)H\left(\frac{p_1}{p_1+p_2}\right) \le D\cdot H\left(\frac{1}{2}\right)=D$$



$$R(D)=\min_{\frac{1}{2}(p_1+p_2)\le D} I(U;V)=\min_{\frac{1}{2}(p_1+p_2)\le D}[H(U)-H(U|V)]$$

$$=1-\boxed{\max_{\frac{1}{2}(p_1+p_2)\le D}\frac{1}{2}(p_1+p_2)H\left(\frac{p_1}{p_1+p_2}\right)}$$

$D\le 1$时，为取得极值，需令$p_1=p_2=D$，求得：

$$R(D)=\begin{cases}1-D & 0\le D\le 1\\ 0 & D>1\end{cases}$$

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

4. 设已知离散无记忆信源在给定失真量度$d(k,j)$，$k=1,2,\cdots,K$，$j=1,2,\cdots,J$下的率失真函数为$R(D)$。现定义新的失真度量$d'(k,j)=d(k,j)-g_k$，试证：在新的失真度量下率失真函数$R'(D)=R(D+G)$，其中$G=\sum_k p_k g_k$。

4-失真度量变换后的率失真

- 考虑不改变测试信道的情况下失真的变换关系。

对新的失真度量，当失真为 D 时，失真函数对应的信道为$p(a_k,b_j)$，有

$$D=\sum_{k=1}^{K}\sum_{j=1}^{J}p(a_k,b_j)d'(k,j)=\sum_{k=1}^{K}\sum_{j=1}^{J}p(a_k,b_j)d(k,j)-\sum_{k=1}^{K}g_k\sum_{j=1}^{J}p(a_k,b_j)$$

$$=\sum_{k=1}^{K}\sum_{j=1}^{J}p(a_k,b_j)d(k,j)-\sum_{k=1}^{K}p(a_k)g_k$$

即

$$\sum_{k=1}^{K}\sum_{j=1}^{J}p(a_k,b_j)d(k,j)=D+\sum_{k=1}^{K}p(a_k)g_k=D+G$$

对于原失真函数，当失真为$D+G$时，失真函数对应的信道也为$p(a_k,b_j)$。对同一个信道，其最小互信息是唯一确定的，因此有$R'(D)=R(D+G)$

不改变率失真函数的形状，仅进行搬移
对于此类失真矩阵的变换，无需重新求解

5. 设有离散无记忆信源X经编码后输出Y，失真矩阵的所有列是集合$\{d_1, \cdots, d_m\}$的某一置换。

定义函数：$\Phi(D) = \max_{P: \sum_{i=1}^{m} P_i d_i \leq D} H(P)$，试证明：

1）$\Phi(D)$是D的上凸函数；

2）$I(X;Y) \geq H(X) - \Phi(D)$；

3）$R(D) \geq H(X) - \Phi(D)$；

4）若信源服从均匀分布，且失真矩阵的所有行互为置换，则$R(D) = H(X) - \Phi(D)$.

5-率失真函数的香农下界

- 1) 验证上凸，构造达到失真的输入分布。

a) 设 $\mathbf{P}_1$ 是达到 $\Phi(D_1)$ 的输入分布，$\mathbf{P}_2$ 是达到 $\Phi(D_2)$ 的输入分布，即 $H(\mathbf{P}_1)=\Phi(D_1)$，

$H(\mathbf{P}_2)=\Phi(D_2)$。设 $\lambda_1+\lambda_2=1,(\lambda_1,\lambda_2$非负) 且 $\lambda_1 D_1+\lambda_2 D_2=D$。令 $\mathbf{P}=\lambda_1\mathbf{P}_1+\lambda_2\mathbf{P}_2$，则：

$$\sum_{i=1}^{m} P_i d_i=\sum_{i=1}^{m}(\lambda_1 P_{1i}+\lambda_2 P_{2i})d_i=\lambda_1 D_1+\lambda_2 D_2=D$$

能达到的失真也是D_1, D_2的线性组合

由 $H(\mathbf{P})$ 的上凸性可知：

$$\Phi(\lambda_1 D_1+\lambda_2 D_2)=\Phi(D)\geq H(\mathbf{P})=H(\lambda_1\mathbf{P}_1+\lambda_2\mathbf{P}_2)\geq\lambda_1 H(\mathbf{P}_1)+\lambda_2 H(\mathbf{P}_2)=\lambda_1\Phi(D_1)+\lambda_2\Phi(D_2)$$

得证 $\Phi(D)$ 是 D 的上凸函数。另外，易证 $\Phi(D)$ 是 D 非减函数。

Tsinghua University

- 2) $\Phi(D) \geq H(X|Y) = H(X) - I(X;Y)$：借助$\Phi(D)$的定义和上凸性质。

将$p(X|Y=y)$看作$\Phi(D)$定义中的$\{P_i\}$。

b) $$D = E[d(x,y)] = \sum_{i=1}^{m}\sum_{j=1}^{m} P(x_i y_j) d(x_i, y_j) = \sum_{j=1}^{m}\sum_{i=1}^{m} P(y_j) P(x_i \mid y_j) d(x_i, y_j)$$

反向信道

令$\boxed{D_j = \sum_{i=1}^{m} P(x_i \mid y_j) d(x_i, y_j) = \sum_{i=1}^{m} q_i^j d(x_i, y_j)}$，于是有$D = E[d(x,y)] = \sum_{j=1}^{m} P(y_j) D_j$

失真矩阵所有的列是集合$\{d_1, d_2, \cdots, d_m\}$的某一置换，$D_j$也可看成是是概率分布

$\begin{bmatrix} P(x_1 \mid y_j) & P(x_2 \mid y_j) & \cdots & P(x_m \mid y_j) \end{bmatrix} = \begin{bmatrix} q_1^j & q_2^j & \cdots & q_m^j \end{bmatrix} = \mathbf{q}^j$ 的某一置换和

$\begin{bmatrix} d_1 & d_2 & \cdots & d_m \end{bmatrix}$的内积。定义函数$\Phi(D_j) = \max\limits_{\mathbf{q}^j : \sum q_i^j d_i \leq D_j} H(\mathbf{q}^j)$。由$\Phi(D)$的凸性可得：

$$\Phi(D) = \Phi\left(\sum_{j=1}^{m} P(y_j) D_j\right) \geq \sum_{j=1}^{m} P(y_j) \Phi(D_j) \geq \sum_{j=1}^{m} P(y_j) H(\mathbf{q}^j)$$

$$= -\sum_{j=1}^{m}\sum_{i=1}^{m} P(y_j) P(x_i \mid y_j) \log P(x_i \mid y_j) = H(X \mid Y) = H(X) - I(X;Y)$$

5-率失真函数的香农下界

- 3) $R(D) \geq H(X) - \Phi(X;Y)$：率失真的下界

$$R(D) = \min_{P_{j|i}:\sum_i\sum_j P_{i,j}d_{i,j}\leq D} I(X;Y) \geq H(X) - \Phi(D)$$

- 4) **构造反向信道，使得3)中下界能达到**

令$p(x|y)$每行也为P^*的一个置换，

验证：

①X总满足均匀分布（行置换）；

②总失真为D （列置换）；

③$H(X|Y) = H(P^*) = \Phi(D)$。

设概率分布 $\mathbf{P}^* = \{P_1^*, P_2^*, \cdots, P_m^*\}$ 满足 $H(\mathbf{P}^*) = \Phi(D)$，失真矩阵为$[d_{i,j}]$，若 $d_{i,j} = d_k$ 时令 $\mathbf{P}_{i|j} = \mathbf{P}_k^*$。设 Y 服从均匀分布，即 $P_j = \frac{1}{m}, (j = 1, 2, \dots, m)$，则：

$$P_i = \sum_j P_{i,j} = \sum_j P_j P_{i|j} = \frac{1}{m}\sum_j P_{i|j} = \frac{1}{m}\sum_k P_k^* = \frac{1}{m}$$（因为失真矩阵的所有行互为置换）

即信源服从均匀分布。此时：

$$\sum_i\sum_j p_{i,j}d_{i,j} = \sum_j \frac{1}{m}\sum_i p_{i|j}d_{i,j} = \sum_j \frac{1}{m}\sum_k P_k^* d_k = \sum_j \frac{1}{m} D = D$$

$$\begin{aligned} I(X;Y) &= H(X) - H(X|Y) \\ &= H(X) - \sum_i\sum_j P_{i,j}\log P_{i|j} \\ &= H(X) - \sum_j \frac{1}{m}\sum_j P_{i|j}\log P_{i|j} \\ &= H(X) - \sum_j \frac{1}{m}\sum_k P_k^*\log P_k^* \\ &= H(X) - H(\mathbf{P}^*) \\ &= H(X) - \Phi(D) \end{aligned}$$

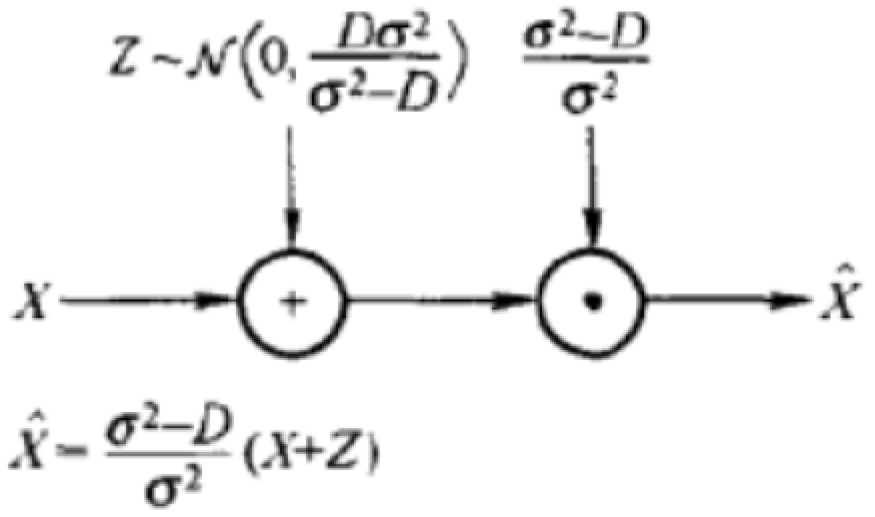
目标：构造出$H(X|Y) = H(P^*)$

清華大學 Tsinghua University

6. 考虑连续型随机变量X，其均值为0，方差为σ^2，失真度量是平方误差的失真度量，试证明：

$$h(X)-\frac{1}{2}\log(2\pi eD)\le R(D)\le\frac{1}{2}\log\frac{\sigma^2}{D}.$$

（可考虑下图所示系统）



两种互信息展开方式
分别求正项和负项的熵的最大值

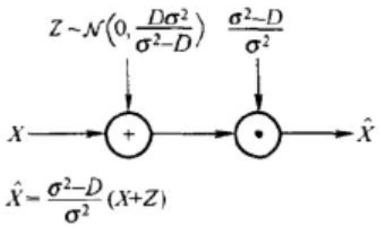
- 下界

常用性质： 对于方差受限的连续随机变量，高斯分布下熵最大。

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

- 上界

考虑$\hat{X}=\frac{\sigma^2-D}{\sigma^2}(X+Z)$



可得，

$$
\begin{aligned}
E(X-\hat{X})^2 &= E\left(\frac{D}{\sigma^2}X-\frac{\sigma^2-D}{\sigma^2}Z\right)^2\\
&=\left(\frac{D}{\sigma^2}\right)^2EX^2+\left(\frac{\sigma^2-D}{\sigma^2}\right)^2EZ^2\\
&=\left(\frac{D}{\sigma^2}\right)^2\sigma^2+\left(\frac{\sigma^2-D}{\sigma^2}\right)^2\frac{D\sigma^2}{\sigma^2-D}\\
&=D
\end{aligned}
$$

同时，互信息为：

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

则上界得证

找到$\hat{X}$的方差可得本项上界

$$
\begin{aligned}
E\hat{X}^2 &= \left(\frac{\sigma^2-D}{\sigma^2}\right)^2E(X+Z)^2\\
&=\left(\frac{\sigma^2-D}{\sigma^2}\right)^2\left(EX^2+EZ^2\right)\\
&=\left(\frac{\sigma^2-D}{\sigma^2}\right)^2\left(\sigma^2+\frac{D\sigma^2}{\sigma^2-D}\right)\\
&=\sigma^2-D.
\end{aligned}
$$

7. 设信源为N长随机向量：$X = X_1 X_2 \cdots X_N$，各分量统计独立，且$X_n \sim \mathcal{N}(0, \sigma_n^2)$。

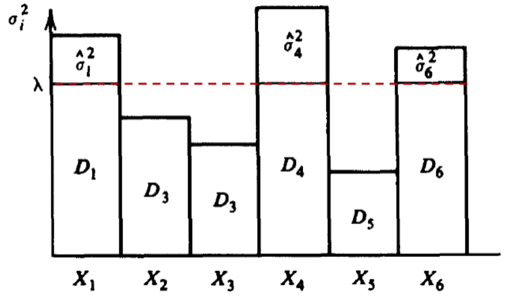
1）若对该信源进行熵压缩编码，则在平方和失真准则下，如果允许的均方误差为D，试求压缩后的信源信息速率$R(D)$。（给出表达式即可）

2）设存在一无记忆加性噪声信道，当噪声功率限定为P_N，而输入信号的功率限制在P_S以下时，若要保证该信道能够全部传送压缩后的上述信源信息，试求此时信道噪声的最大微分熵。

- 1) 反注水法：优化问题求解中σ_n^2/D_i非负对应的KKT条件导致。

参考课件P27或Cover的书P179

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



由于各分量独立，则有

$$R(D) = \sum_{n=1}^{N} R_n(D_n) = \sum_{n=1}^{N} \frac{1}{2}\log\frac{\sigma_n^2}{D_n(\lambda)}$$

其中λ的取值需满足$\sum D_n(\lambda) = D$。

- 2) 加性噪声信道：$\hat{X} = X + Z$，最大接收功率为$P_S + P_N$。

保证该信道能够传输压缩后的信息，由信源信道分离定理可得

$$R(D) \leq C = \max\{I(X;\hat{X})\} = \max\{h(\hat{X}) - h(\hat{X}|X)\}$$
$$= \max\{h(\hat{X}) - h(Z)\} \leq \frac{1}{2}\log(2\pi e(P_S + P_N)) - h(Z)$$

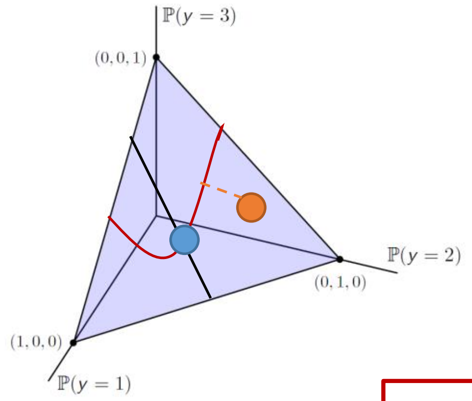
则$h(Z) \leq \frac{1}{2}\log(2\pi e(P_S + P_N)) - R(D)$。

$$C = \frac{1}{2}\log(1 + SNR)$$

适用于加性高斯白噪信道

第六章重点总结

- 信息几何：直观理解
  - 概率单纯形simplex：n维概率空间
  - 信息投影：KLD；线性指数族



- **型type：（经验）分布**
  - 型类的性质：数目与熵/等概
  - 大数定律：远离真实分布的型类的概率以指数衰减

典型集与型类
序列概率/各元素频率
宏观/围观

- 大偏差理论：经验偏离期望
  - Sanov定理：（一阶指数意义下）偏差概率—投影+KLD
  - 条件极限定理：**落在集合内的型大多落在投影点处**

- 假设检验：误差估计
  - 两类误差：Chernoff-Stein引理；误差指数（对称）：Chernoff Information
  - 均方误差：Fisher Information & Cramer-Rao Bound

1. **信息投影**：分布q在概率空间$\mathcal{P}$上的投影定义为

$$p^* = \underset{p\in\mathcal{P}}{\operatorname{argmin}}\ D(p\|q)$$

设在元素集$\mathcal{X}$上存在一分布Q，两个线性族$\mathcal{P}_1$、$\mathcal{P}_2$，$Q \notin \mathcal{P}_1 \cup \mathcal{P}_2$。证明将其先投影到$\mathcal{P}_1$再投影到$\mathcal{P}_1 \cap \mathcal{P}_2$，得到的投影与直接投影到$\mathcal{P}_1 \cap \mathcal{P}_2$的相同。
其中$\mathcal{P}_1$满足：

$$\sum_x p(x) = 1$$

$$\sum_x p(x)h_i(x) \geq \alpha_i,\ i = 1, 2, \cdots, r$$

$\mathcal{P}_2$满足：

$$\sum_x p(x) = 1$$

$$\sum_x p(x)g_j(x) \geq \beta_j,\ j = 1, 2, \cdots, s$$

清華大學 Tsinghua University

常用结论：线性约束的分布为指数分布族（通过解带约束的优化问题得到）

参考课件P51或Cover的书11.5

$$J(P)=\sum_x P(x)\log\frac{P(x)}{Q(x)}+\sum_i\lambda_i\sum_x P(x)g_i(x)+\nu\sum_x P(x)$$

由拉格朗日乘子法，可求得

$$\begin{aligned}p*(x) &= \arg\min_{p\in\mathcal{P}_1} D(p\|q)\\ &= c_1 q(x)\boxed{e^{\sum_{i=1}^{r}\lambda_i h_i(x)}}\\ r*(x) &= \arg\min_{p\in\mathcal{P}_1\cap\mathcal{P}_2} D(p\|q)\\ &= c_2 q(x)\boxed{e^{\sum_{i=1}^{r}\lambda_i' h_i(x)+\sum_{j=1}^{s}\nu_j' g_j(x)}}\end{aligned}$$

把p^*投影到$\mathcal{P}_1\cap\mathcal{P}_2$上，会得到

c_1, c_2, c_3：归一项

$$\begin{aligned}p^{**}(x) &= \arg\min_{p\in\mathcal{P}_1\cap\mathcal{P}_2} D(p\|p^*)\\ &= c_3 p^*(x) e^{\sum \nu_j g_j(x)}\\ &= c_3 c_1 q(x) e^{\sum \nu_i g_i(x)+\sum\lambda_i h_i(x)}\end{aligned}$$

和r^*的形式相同。因为约束条件是相同的，所以常数的值也是相同的，故：

$$r^*(x)=p^{**}(x)\ \forall x$$

因此r^*是满足$r\in\mathcal{P}_1\cap\mathcal{P}_2$上，$D(r\|p^*)$的最小值

2-无偏/有偏估计

Tsinghua University

2. 抽自$\mathcal{N}(\mu,\sigma^2)$的$n$个iid数据样本，考虑对分布参数的估计问题。

   1）证明$\bar{X}$为μ的无偏估计。

   2）证明估计量

$$S_n^2 := \frac{1}{n}\sum_{i=1}^{n}(X_i-\bar{X})^2$$

是σ^2的有偏估计，而估计量

$$S_{n-1}^2 := \frac{1}{n-1}\sum_{i=1}^{n}(X_i-\bar{X})^2$$

是无偏的。

2-无偏/有偏估计

1) $E\,\bar{X} = E\frac{1}{n}\sum X_i = \frac{1}{n}E\,X_i = \mu$

2) n-1由自相关项带来

$$E[(X_i-\mu)(\bar{X}_n-\mu)] = E[(X_i-\mu)(\frac{1}{n}\sum_j (X_j-\mu))]$$
$$= \frac{1}{n}E(x_i-\mu)^2 + \frac{1}{n}\sum_{j\neq i} E[(X_i-\mu)(X_j-\mu)]$$
$$= \frac{\sigma^2}{n}$$

同理.

$$E(\bar{X}_n-\mu)^2 = \frac{\sigma^2}{n}$$

令$W = \sum_i (X_i-\bar{X}_n)^2$ 故:

$$E(W) = E[\sum_i ((X_i-\mu)-(\bar{X}_n-\mu))^2]$$
$$= \sum_i E(X_i-\mu)^2 - 2\sum_i E(X_i-\mu)(\bar{X}_n-\mu) + nE(\bar{X}_n-\mu)^2$$
$$= n\sigma^2 - 2n\frac{\sigma^2}{n} + n\frac{\sigma^2}{n}$$
$$= (n-1)\sigma^2$$

解释统计中的$n-1$

因此:

$$E(S_n^2) = \frac{n-1}{n}\sigma^2$$
$$E(S_{n-1}^2) = \sigma^2$$

S_n有更小的MSE（课后题11.6）

3. $X_1, X_2, \cdots, X_n \sim$ i.i.d. $p(x)$，考虑假设检验$H_1: p = p_1$与$H_2: p = p_2$，其中

$$p_1(x)) = \begin{cases} \frac{1}{2}, & x = -1 \\ \frac{1}{4}, & x = 0 \\ \frac{1}{4}, & x = 1 \end{cases} \qquad p_1(x)) = \begin{cases} \frac{1}{4}, & x = -1 \\ \frac{1}{2}, & x = 0 \\ \frac{1}{4}, & x = 1 \end{cases}$$

在约束条件$\Pr\{\text{判定}H_1|H_2\text{真}\} \leq \frac{1}{2}$之下，求出$\Pr\{\text{判定}H_2|H_1\text{真}\}$关于$H_1$与$H_2$的假设检验的最佳误差指数。

Chernoff-Stein引理：两类假设，给定一个误差概率范围，另一个以指数渐进0。

由Chernoff-Stein引理，可得

$$D(P_2||P_1)=\frac{1}{4}\log\frac{1/4}{1/2}+\frac{1}{4}\log\frac{1/4}{1/4}+\frac{1}{2}\log\frac{1/2}{1/4}=0.25$$

因此误差概率会随$2^{-\frac{n}{4}}$趋向0

4. 令$X_1, X_2, \cdots, X_n$为服从几何分布的独立同分布随机变量：

$$\Pr\{X=k\}=p^{k-1}(1-p),\quad k=1,2,\cdots$$

针对下面的情形，找出（在一阶指数意义下）好的估计：

1）$\Pr\{\frac{1}{n}\sum_{i}^{n} X_i \ge \alpha\}$；

2）$\Pr\{X_1=k|\frac{1}{n}\sum_{i}^{n} X_i \ge \alpha\}$；

3）当$p=1/2$，$\alpha=4$时，计算1）和2）。

4-大偏差

1) Sanov定理：解带约束的优化问题+KLD
2) 条件极限定理：$\Pr\{X_1 = k | P_{X^n} \in E\} \to r_k$

定理条件本身就是一阶指数意义下最优

假设r是在整数1, 2...上的整数，其跟几何分布的相对熵距离为：

$$D(r||p) = \sum r_i \log \frac{r_i}{p^{i-1}(1-p)}$$

在$\sum r_i = 1$和$\sum ir_i = \alpha$的约束下，最小化$D(r||p)$。令

$$J(r) = \sum r_i \log \frac{r_i}{p^{i-1}(1-p)} + \lambda_1 \sum r_i + \lambda_2 \sum ir_i$$

令其微分为0，则：

$$\log r_i - \log(p^{i-1}(1-p)) + \lambda_1 + \lambda_2 i = 0$$

可解得：

$$\boxed{r_i = p^{i-1}(1-p)c_1 c_2^i}$$

r也是几何分布。由于$\sum ir_i = \alpha$，故其参数为$1-\frac{1}{\alpha}$，因此

$$r_i = (1-\frac{1}{\alpha})^{i-1}\frac{1}{\alpha}$$

$$\begin{aligned} D(r||p) &= \sum_i r_i \log \frac{r_i}{p^{i-1}(1-p)} \\ &= \sum_i r_i \log \frac{p}{1-p}\frac{1-\frac{1}{\alpha}}{\frac{1}{\alpha}}(\frac{1}{\alpha p})^i \\ &= \log \frac{p}{(1-p)(\alpha-1)} + \alpha \log \frac{\alpha-1}{\alpha p} \end{aligned}$$

几何分布向linear family上投影得到的仍为几何分布
几何分布属于指数族

(a)

$$\frac{1}{n}\log \Pr\{\frac{1}{n}\sum_{i=1}^n X_i \geq \alpha\} = D(r||p) = \log \frac{p}{(1-p)(\alpha-1)} + \alpha \log \frac{\alpha-1}{\alpha p}$$

(b)

$$\Pr\{X_1 = k | \frac{1}{n}\sum_{i=1}^n X_i \geq \alpha\} = r_k = (1-\frac{1}{\alpha})^{k-1}\frac{1}{\alpha}$$

(c) 代入数值可得：

$$\Pr\{X_1 = k | \frac{1}{n}\sum_{i=1}^n X_i \geq \alpha\} = r_k = 0.75^{k-1}0.25$$