

第 4 讲 费诺不等式, 更多例子

费诺不等式

例题

费诺不等式

例题

- ▶ T. M. Cover and J. A. Thomas, Elements of Information Theory, 2.8.
- ▶ R. W. Yeung, Information and Network Coding, 2.8.
- ▶ F. Alajaji, P. Chen, An Introduction to Single-User Information Theory, 2.5.
- ▶ Y. Polyanskiy, Y. Wu, Information Theory: from Coding to Learning, 3.6.

费诺不等式

例题

费诺不等式

例题

罗伯特·马利欧·费诺 (Robert Mario Fano, 1917 年 11 月 11 日—2016 年 7 月 13 日) 是一名意裔美籍计算机科学家, 为麻省理工学院电机工程与计算机科学教授. 专长于信息论, 曾与克劳德·香农共同开发出香农-费诺编码, 并曾提出费诺不等式. 1976 年获得香农奖.



- ▶ 我们观察与 X 相关的随机变量 Y , 相应的条件分布为 $p(y|x)$, 通过 Y 计算函数 $g(Y) = \hat{X}$, 其中 $\hat{X}$ 是对 X 的估计, 取值空间为 $\hat{\mathcal{X}}$.
- ▶ 我们并不要求 $\hat{X}$ 必须与 $\mathcal{X}$ 相同, 也允许函数 $g(Y)$ 是随机的.
- ▶ 我们希望得到事件 $\{\hat{X} \neq X\}$ 的概率的一个界.
- ▶ 我们记

$$P_e = \Pr\{\hat{X} \neq X\}.$$

- ▶ 注意到 $X \rightarrow Y \rightarrow \hat{X}$ 构成一个马尔可夫链.

定理 1.1: 费诺不等式

对任何满足 $X \rightarrow Y \rightarrow \hat{X}$ 的估计量 $\hat{X}$, 设 $P_e = P\{X \neq \hat{X}\}$, 有

$$H(P_e) + P_e \log |\mathcal{X}| \geq H(X|\hat{X}) \geq H(X|Y)$$

或

$$P_e \geq \frac{H(X|Y) - 1}{\log |\mathcal{X}|}.$$

证明.

我们首先定义一个随机变量

$$E = \begin{cases} 1 & \text{如果 } \hat{X} \neq X \\ 0 & \text{如果 } \hat{X} = X \end{cases}$$

费诺不等式

例题

利用熵的链式法则将 $H(E, X|\hat{X})$ 以两种不同方式展开, 有

$$\begin{aligned} H(E, X|\hat{X}) &= H(X|\hat{X}) + H(E|X, \hat{X}) \\ &= H(E|\hat{X}) + H(X|E, \hat{X}) \end{aligned}$$

因为 E 是 X 和 $\hat{X}$ 的函数, 所以, 条件熵 $H(E|X, \hat{X})$ 等于 0. 由于条件作用使熵减少, 可知 $H(E|\hat{X}) \leq H(E) = H(P_e)$ (因为 E 是二值随机变量, 故 $H(E) = H(P_e)$). 对于剩余项 $H(X|E, \hat{X})$ 可以界定如下:

$$\begin{aligned} H(X|E, \hat{X}) &= P(E=0)H(X|\hat{X}, E=0) + P(E=1)H(X|\hat{X}, E=1) \\ &\leq (1 - P_e)0 + P_e \log |\mathcal{X}|. \end{aligned}$$

上述不等式成立是因为当 $E=0$ 时, $X = \hat{X}$; 当 $E=1$ 时, 条件熵的上界为 X 的可能取值数目的对数值.

证明.

综合这些结果, 可得

$$H(P_e) + P_e \log |\mathcal{X}| \geq H(X|\hat{X}).$$

因为 $X \rightarrow Y \rightarrow \hat{X}$ 构成马尔可夫链, 由数据处理不等式可知

$$I(X; \hat{X}) \leq I(X; Y),$$

从而 $H(X|\hat{X}) \geq H(X|Y)$. 于是, 有

$$H(P_e) + P_e \log |\mathcal{X}| \geq H(X|\hat{X}) \geq H(X|Y).$$

□

推论 1.2

对任意两个随机变量 X 和 Y , 设 $p = P(X \neq Y)$,

$$H(p) + p \log |\mathcal{X}| \geq H(X|\hat{X}) \geq H(X|Y).$$

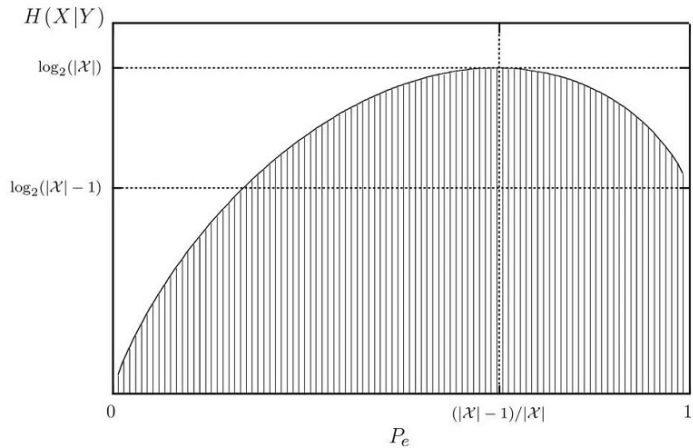
推论 1.3

设 $P_e = \Pr(X \neq \hat{X})$, $\hat{X} : \mathcal{Y} \rightarrow \mathcal{X}$, 则

$$H(P_e) + P_e \log(|\mathcal{X}| - 1) \geq H(X|Y).$$

费诺不等式

例题



注记 1.4

假定没有任何关于 Y 的知识, 只能在毫无信息的情况下对 X 进行推测. 设 $X \in \{1, 2, \dots, m\}$ 且 $p_1 \geq p_2 \geq \dots \geq p_m$, 则对 X 的最佳估计是 $\hat{X} = 1$, 而此时误差概率 $P_e = 1 - p_1$. 费诺不等式变为

$$H(P_e) + P_e \log(m-1) \geq H(X)$$

且概率密度函数

$$(p_1, p_2, \dots, p_m) = (1 - P_e, \frac{P_e}{m-1}, \dots, \frac{P_e}{m-1})$$

可以达到等号成立的界. 因此, 费诺不等式是精确的.

设 X 和 X' 为两个独立同分布的随机变量, 则它们有相同的熵 $H(X)$, 那么 $X = X'$ 的概率为

$$P(X = X') = \sum_x p^2(x).$$

由此我们可以得到如下的不等式:

引理 1.5

如果 X 和 X' 为独立同分布, 则它们有相同的熵 $H(X)$, 我们有

$$P(X = X') \geq 2^{-H(X)}.$$

当且仅当 X 服从均匀分布, 等号成立.

证明.

设 $X \sim p(x)$. 由 Jensen 不等式, 我们有

$$2^{E \log p(X)} \leq E 2^{\log p(X)},$$

从而有

$$2^{-H(X)} = 2^{\sum p(x) \log p(x)} \leq \sum p(x) 2^{\log p(x)} = \sum p^2(x). \quad \square$$

推论 1.6

设 X 和 Y 相互独立, 且 $X \sim p(x)$, $X' \sim r(x)$, $x, x' \in \mathcal{X}$, 那么

$$P(X = X') \geq 2^{-H(p) - D(p\|r)}$$

$$P(X = X') \geq 2^{-H(r) - D(r\|p)}$$

证明.
我们有

$$\begin{aligned} 2^{-H(p)-D(p\|r)} &= 2^{\sum p(x) \log p(x) + \sum p(x) \log \frac{r(x)}{p(x)}} \\ &= 2^{\sum p(x) \log r(x)} \\ &\leq \sum p(x) 2^{\log r(x)} \\ &= \sum p(x) r(x) \\ &= P(X = X'), \end{aligned}$$

其中上面的不等号成立是由函数 $f(y) = 2^y$ 的凸性和 Jensen 不等式得到. $\square$

费诺不等式

例题

例 2.1: 课本习题 2.10

设离散随机变量 X_1 和 X_2 的概率质量函数分别为 $p_1(\cdot)$ 和 $p_2(\cdot)$, 字母表分别为 $\mathcal{X}_1 = \{1, 2, \dots, m\}$, $\mathcal{X}_2 = \{m+1, \dots, n\}$. 设

$$X = \begin{cases} X_1 & \text{概率为 } \alpha \\ X_2 & \text{概率为 } 1 - \alpha. \end{cases}$$

- (1) 试求 $H(X)$ 关于 $H(X_1)$ 、 $H(X_2)$ 和 α 的表达式.
- (2) 试关于 α 最大化 $H(X)$, 证明 $2^{H(X)} \leq 2^{H_1(X)} + 2^{H(X_2)}$.

证明.
注意到

$$X = \begin{cases} X_1 & \text{概率为 } \alpha \\ X_2 & \text{概率为 } 1 - \alpha, \end{cases}$$

我们定义

$$\theta = f(X) = \begin{cases} 1 & \text{若 } X = X_1, \\ 2 & \text{若 } X = X_2, \end{cases}$$

那么我们有

$$\begin{aligned} H(X) &= H(X, f(X)) = H(\theta) + H(X|\theta) \\ &= H(\theta) + p(\theta = 1)H(X|\theta = 1) + p(\theta = 2)H(X|\theta = 2) \\ &= H(\alpha) + \alpha H(X_1) + (1 - \alpha)H(X_2) \end{aligned}$$

其中 $H(\alpha) = -\alpha \log \alpha = (1 - \alpha) \log(1 - \alpha)$.

证明.

记 $f(\alpha) = H(\alpha) + \alpha H(X_1) + (1 - \alpha)H(X_2)$, 则

$$f'(\alpha) = \log\left(\frac{1 - \alpha}{\alpha}\right) + H(X_1) - H(X_2).$$

解方程 $f'(\alpha_0) = 0$ 可得 $\alpha_0 = \frac{2^{H(X_1)}}{2^{H(X_1)} + 2^{H(X_2)}}$. 不难看出 f 在 α_0 处取到最大值, 于是

$$\begin{aligned} H(X) &\leq -\frac{2^{H(X_1)}}{2^{H_1(X)} + 2^{H(X_2)}} \log \frac{2^{H(X_2)}}{2^{H_1(X)} + 2^{H(X_2)}} - \frac{2^{H(X_2)}}{2^{H_1(X)} + 2^{H(X_2)}} \log \frac{2^{H(X_1)}}{2^{H_1(X)} + 2^{H(X_2)}} \\ &\quad + \frac{2^{H(X_2)}}{2^{H_1(X)} + 2^{H(X_2)}} H(X_1) + \frac{2^{H(X_1)}}{2^{H_1(X)} + 2^{H(X_2)}} H(X_2) \\ &= \log(2^{H(X_1)} + 2^{H(X_2)}). \end{aligned}$$

从而 (2) 得证. □

例 2.2: 课本习题 2.39

设 X, Y 和 Z 为三个服从 $\text{Bernoulli}(\frac{1}{2})$ 的二元随机变量, 且两两独立. 在上述约束条件下, $H(X, Y, Z)$ 的最小值是多少?

证明.
我们有

$$\begin{aligned} H(X, Y, Z) &= H(X, Y) + H(Z|X, Y) \\ &\geq H(X, Y) \\ &= 2 \text{ 比特}. \end{aligned}$$

下面我们验证这个下界可以达到, 从而 $H(X, Y, Z)$ 的最小值为 2 比特.
设 $X, Y \sim \text{Bernoulli}(1/2)$, $Z = X \oplus Y$, 其中 $\oplus$ 表示模 2 加法.

□

例 2.3

设 $Z_1, Z_2, \dots$ 为独立同分布的随机变量, 服从 $\{0, 1\}$ 上的均匀分布. 对于 $1 \leq i \leq n$, 设

$$X_i = \sum_{j=1}^i Z_j.$$

求 $I(X_1; X_2, X_3, \dots, X_n)$.

证明.

首先注意到 $X_1 \rightarrow X_2 \rightarrow \cdots \rightarrow X_n$ 构成一个马尔可夫链. (思考: 为什么?)

由互信息的链式法则, 我们有

$$\begin{aligned} I(X_1; X_2, X_3, \cdots, X_n) &= \sum_{i=2}^n I(X_1; X_i | X_2, \cdots, X_{i-1}) \\ &= I(X_1; X_2). \\ &= I(Z_1; Z_1 + Z_2) \\ &= H(Z_1 + Z_2) - H(Z_1 + Z_2 | Z_1) \\ &= \frac{3}{2} - 1 = 1/2 \text{ 比特}. \end{aligned}$$

□

例 2.4

设 $X_1 \rightarrow X_2 \rightarrow X_3 \rightarrow X_4$ 构成一个马尔科夫链. 证明:

$$I(X_1; X_3) + I(X_2; X_4) \leq I(X_1; X_4) + I(X_2; X_3).$$

证明.
注意到

$$\begin{aligned} I(X_1; X_3, X_4) &= I(X_1; X_3) + I(X_1; X_4|X_3) \\ &= I(X_1; X_4) + I(X_1; X_3|X_4) \end{aligned}$$

于是

$$I(X_1; X_4) - I(X_1; X_3) = I(X_1; X_4|X_3) - I(X_1; X_3|X_4).$$

类似地,

$$I(X_2; X_3) - I(X_2; X_4) = I(X_2; X_3|X_4) - I(X_2; X_4|X_3).$$

从而我们有

$$\begin{aligned} &I(X_1; X_4) + I(X_2; X_3) - I(X_1; X_3) + I(X_2; X_4) \\ &= I(X_1; X_4|X_3) - I(X_1; X_3|X_4) + I(X_2; X_3|X_4) - I(X_2; X_4|X_3). \end{aligned}$$

费诺不等式

例题

证明.
又由

$$\begin{aligned} I(X_1; X_2; X_4|X_3) &= I(X_1; X_4|X_3) + I(X_2; X_4|X_1, X_3) \\ &\quad I(X_2; X_4|X_3) + I(X_1; X_4|X_2, X_3), \end{aligned}$$

以及

$$\begin{aligned} I(X_1, X_2; X_3|X_4) &= I(X_1; X_3|X_4) + I(X_2; X_3|X_1, X_4) \\ &= I(X_2; X_3|X_4) + I(X_1; X_3|X_2, X_4), \end{aligned}$$

我们知

$$\begin{aligned} &I(X_1; X_4) + I(X_2; X_3) - I(X_1; X_3) + I(X_2; X_4) \\ &= I(X_1; X_4|X_2, X_3) - I(X_2; X_4|X_1, X_3) + I(X_2; X_3|X_1, X_4) - I(X_1; X_3|X_2, X_4) \\ &= I(X_2; X_3|X_1, X_4) \geq 0, \end{aligned}$$

从而得证.

例 2.5

记 $\pi(n)$ 为不超过 n 的素数个数. 注意到对于任一正整数 n , 其有唯一的形如下的素因子分解

$$n = \prod_{i=1}^{\pi(n)} p_i^{X_i}.$$

其中 $p_1, p_2, p_3, \dots$ 为素数. 即 $p_1 = 2, p_2 = 3, p_3 = 5, \dots$. $X_i = X_i(n)$ 为一个非负整数, 其表示 n 的素因子分解中 p_i 的重数. 设 N 为在 $\{1, 2, \dots, n\}$ 上均匀分布的一个随机变量.

(1) 证明 $X_i(N)$ 为一个整数值随机变量, 满足

$$0 \leq X_i(N) \leq \log n.$$

(2) 证明 $\log n = H(N) \leq \pi(n) \log(\log n + 1)$. 从而 $\pi(n) \geq \frac{\log n}{\log(\log n + 1)}$.
特别地当 $n \rightarrow \infty$ 时有 $\pi(n) \rightarrow \infty$.

证明.

- (1) 由于 $p_i^{X_i} | N$, 所以 $2^{X_i(N)} \leq p_i^{X_i(N)} \leq N \leq n$. 从而 $X_i(N) \leq \log n$.
- (2) 注意到对于每个 N , 其对应的 $(X_1(N), \dots, X_{\pi(n)}(N))$ 是唯一的, 所以

$$\begin{aligned}\log n &= H(N) \\ &= H(X_1(N), X_2(N), \dots, X_{\pi(n)}(N)) \\ &\leq H(X_1(N)) + \dots + H(X_{\pi(n)}(N)) \\ &\leq \pi(n) \log(\log(n+1)),\end{aligned}$$

从而 $\pi(n) \geq \frac{\log n}{\log(\log n+1)}$. 特别地当 $n \rightarrow \infty$ 时有 $\pi(n) \rightarrow \infty$. □

前述结论的改进

设 $N \sim \text{Uniform}\{1, 2, \dots, n\}$, 其可以唯一表示为

$$N = M^2 \cdot p_1^{Y_1} \cdot p_2^{Y_2} \cdot \dots \cdot p_{\pi(n)}^{Y_{\pi(n)}},$$

其中 M 是使得 $M^2|N$ 的最大整数. 这样随机变量 Y_i 取值范围是 $\{0, 1\}$. 由于 $M^2|N$, $1 \leq M \leq \sqrt{n}$. 从而我们有

$$\begin{aligned} \log n &= H(N) \\ &= H(M, Y_1, \dots, Y_{\pi(n)}) \\ &\leq H(M) + H(Y_1) + H(Y_2) + \dots + H(Y_{\pi(n)}) \\ &\leq \frac{1}{2} \log n + \pi(n). \end{aligned}$$

从而 $\pi(n) \geq \frac{1}{2} \log n$.

费诺不等式

例题