2022 春信息论 B 第一次习题课补充材料

考察范围：<u>重要结论和技巧</u>　课程助教：<u>高源</u>

1. Jensen 不等式。对于给定的凸函数 f 和随机变量 X, 有

$$E[f(X)] \geqslant f(E[X])$$

- 等号成立当且仅当 X 为常量。

2. 利用 Jensen 不等式证明相对熵非负。

$$\begin{aligned}
D(p\|q) &= \sum_A p(x)\log\frac{p(x)}{q(x)} \\
&= -\sum_A p(x)\log\frac{q(x)}{p(x)} \\
&\geqslant -\log\sum_A p(x)\log\frac{q(x)}{p(x)} \\
&= -\log\sum_A q(x) \\
&\geqslant 0
\end{aligned}$$

- 等号成立当且仅当 p 和 q 同分布。
- 注意，Jensen 不等式在连续情形也是成立的，连续情形下相对熵非负证明方法相同。

3. 利用相对熵非负证明熵的极值性。

记 $\mu(x) = \dfrac{1}{|\mathcal{X}|}$ 服从均匀分布. 则有

$$\begin{aligned}
D(p\|\mu) &= \sum p(x)\log\frac{p(x)}{\mu(x)} \\
&= \sum p(x)\log p(x) + \sum p(x)\log|\mathcal{X}| \\
&= \log|\mathcal{X}| - H(X) \\
&\geqslant 0
\end{aligned}$$

- 等号成立当且仅当 X 服从均匀分布。

4. 利用相对熵非负证明功率受限条件下正态分布取得最大微分熵。

$$
\begin{aligned}
0 &\leqslant D\left(g \| \phi_k\right) \\
&= \int_{-\infty}^{+\infty} g \ln \frac{g}{\phi_k} dx \\
&= -h(g) - \int_{-\infty}^{+\infty} g\left(c - \frac{x^2}{2\sigma^2}\right) dx \\
&= -h(g) - \left(\int_{-\infty}^{+\infty} (g \cdot c) dx - \frac{E\left[X^2\right]}{2\sigma^2}\right) \\
&= -h(g) - \left(\int_{-\infty}^{+\infty} \left(\phi_k \cdot c\right) dx - \frac{E\left[X^2\right]}{2\sigma^2}\right) \\
&= -h(g) + h\left(\phi_k\right)
\end{aligned}
$$

- 等号成立当且仅当 g 服从正态分布。

5. 利用相对熵非负证明均值受限（取值非负）条件下指数分布取得最大微分熵。

$$
\begin{aligned}
0 &\leqslant D\left(g \| \phi_\mu\right) \\
&= \int_{0}^{+\infty} g \ln \frac{g}{\phi_\mu} dx \\
&= -h(g) - \int_{0}^{+\infty} g\left(c - \lambda x\right) dx \\
&= -h(g) - \left(\int_{0}^{+\infty} (g \cdot c) dx - \lambda E[X]\right) \\
&= -h(g) - \left(\int_{0}^{+\infty} \left(\phi_\mu \cdot c\right) dx - \lambda E[X]\right) \\
&= -h(g) + h\left(\phi_\mu\right)
\end{aligned}
$$

- 等号成立当且仅当 g 服从指数分布。

6. 利用相对熵非负证明峰值受限条件下均匀分布取得最大微分熵。

$$
\begin{aligned}
0 &\leqslant D\left(g \| u\right) \\
&= \int_{a}^{b} g \ln \frac{g}{u} dx \\
&= -h(g) - \int_{a}^{b} g \cdot c dx \\
&= -h(g) - \int_{a}^{b} g dx \cdot c \\
&= -h(g) - \int_{a}^{b} u dx \cdot c \\
&= -h(g) + h\left(u\right)
\end{aligned}
$$

- 等号成立当且仅当 g 服从均匀分布。

7. 数据处理不等式。若 $X \to Y \to Z$，则有 $I(X;Y) \geqslant I(X;Z)$.

$$
\begin{aligned}
I(X;Y,Z) &= I(X;Y) + I(X;Z \mid Y) \\
&= I(X;Z) + I(X;Y \mid Z)
\end{aligned}
$$

- 等号成立当且仅当 $I(X;Y \mid Z) = 0$。
- 构造一个信息量的不同展开式是一个重要的证明技巧。

8. 利用构造一个信息量的不同展开式方法证明。随机变量 $X_1\ X_2\ X_3\ X_4$ 构成了马尔科夫链 $X_1 \to X_2 \to X_3 \to X_4$. 证明：

$$
I\left(X_1;X_3\right) + I\left(X_2;X_4\right) \leqslant I\left(X_1;X_4\right) + I\left(X_2;X_3\right)
$$

9. 费诺不等式。对于满足 $X \to Y \to \hat{X}$ 的估计量 $\hat{X}$, 记 $P_e = p\{X \neq \hat{X}\}$, 有

$$
H\left(P_e\right) + P_e \log|\mathcal{X}| \geqslant H(X \mid \hat{X}) \geqslant H(X \mid Y)
$$

为了构造一个信息量的不同展开式，引入示性变量（按照 P_e 的定义提供的提示）

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

构造展开式

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

由示性变量的定义可知, $H(E \mid X,\hat{X}) = 0$，又由于

$$
H(E \mid \hat{X}) \leqslant H(E) = H\left(P_e\right)
$$

和

$$
\begin{aligned}
H(X \mid E,\hat{X}) &= p(E=1)H(X \mid E=1,\hat{X}) + p(E=0)H(X \mid E=0,\hat{X}) \\
&\leqslant p(E=1)\log|\mathcal{X}| \\
&= P_e \log|\mathcal{X}|
\end{aligned}
$$

可得

$$
H\left(P_e\right) + P_e \log|\mathcal{X}| \geqslant H(X \mid \hat{X})
$$

- 等号成立条件？

- 引入示性变量结合链式法则是证明和求解问题的重要技巧。

10. 利用示性变量求解并联信道的容量。信道模型

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

对于这种输入的随机变量按照一定概率分布组合在一起的问题，一般的处理技巧是引入示性变量。记

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

引入示性变量的目的是通过构造互信息的分解，简化 $I(X;Y)$ 的表达，从而完成最大值求解。

$$\begin{aligned} I(X;Y,\theta) &= I(X;\theta)+I(X;Y\mid\theta) \\ &= I(X;Y)+I(X;\theta\mid Y) \end{aligned}$$

考虑到 $X\to Y\to\theta$ 于是有

$$I(X;\theta\mid Y)=0$$

进而有

$$\begin{aligned} I(X;Y) &= I(X;\theta)+I(X;Y\mid\theta) \\ &= H(\theta)-H(\theta\mid X)+\alpha I\left(X_1;Y_1\right)+(1-\alpha)I\left(X_2;Y_2\right) \\ &= H(\alpha)+\alpha I\left(X_1;Y_1\right)+(1-\alpha)I\left(X_2;Y_2\right) \end{aligned}$$

通过对 α 求导得到

$$\alpha=\frac{2^{C_1}}{2^{C_1}+2^{C_2}}$$

于是有

$$C=H\left(\frac{2^{C_1}}{2^{C_1}+2^{C_2}}\right)+\frac{2^{C_1}}{2^{C_1}+2^{C_2}}C_1+\frac{2^{C_2}}{2^{C_1}+2^{C_2}}C_2=\log\left(2^{C_1}+2^{C_2}\right)$$

达到信道容量时的输入概率分布为

$$X=\begin{cases} X_1 & 概率\frac{2^{C_1}}{2^{C_1}+2^{C_2}} \\ X_2 & 概率\frac{2^{C_2}}{2^{C_1}+2^{C_2}} \end{cases}$$

进一步可以推广得到

$$2^C=\sum_{i=1}^{n}2^{C_i}$$