信息论第二单元复习问题集

概率不等式

1. 请复述**马尔可夫不等式 (Markov's Inequality)** 并给出证明。

答案：对于非负随机变量 $X \geq 0$ 和常数 $a > 0$:

$$\Pr(X \geq a) \leq \frac{\mathbb{E}[X]}{a}$$

证明：

$$\begin{aligned}
\mathbb{E}[X] &= \int_0^\infty x f(x) dx \\
&\geq \int_a^\infty x f(x) dx \\
&\geq a \int_a^\infty f(x) dx \\
&= a \Pr(X \geq a) \\
\Rightarrow \Pr(X \geq a) &\leq \frac{\mathbb{E}[X]}{a} \quad \square
\end{aligned}$$

2. 请复述**切比雪夫不等式 (Chebyshev's Inequality)** 并给出证明。

答案：对于随机变量 X 有有限方差 $\mathrm{Var}(X)$，则：

$$\Pr(|X-\mu| \geq k\sigma) \leq \frac{1}{k^2}$$

其中 $\mu = \mathbb{E}[X]$, $\sigma^2 = \mathrm{Var}(X)$, $k > 0$。
证明：对 $|X-\mu|^2$ 应用马尔可夫不等式：

$$\begin{aligned}\Pr(|X-\mu| \geq k\sigma) &= \Pr((X-\mu)^2 \geq k^2\sigma^2) \\ &\leq \frac{\mathbb{E}[(X-\mu)^2]}{k^2\sigma^2} \\ &= \frac{\sigma^2}{k^2\sigma^2} = \frac{1}{k^2} \quad \square\end{aligned}$$

3. 请复述**弱大数定律 (Weak Law of Large Numbers, WLLN)** 并给出证明概要。

答案： 设 $X_1, X_2, \ldots$ 是独立同分布随机变量，$\mathbb{E}[X_i] = \mu$，$\mathrm{Var}(X_i) = \sigma^2 < \infty$，则：

$$\lim_{n\to\infty} \Pr\left(\left|\frac{1}{n}\sum_{i=1}^{n} X_i - \mu\right| > \varepsilon\right) = 0$$

对任意 $\varepsilon > 0$。

证明概要：

(a) 定义样本均值 $S_n = \frac{1}{n}\sum_{i=1}^{n} X_i$

(b) 由线性性：$\mathbb{E}[S_n] = \mu$

(c) 由独立性：$\mathrm{Var}(S_n) = \frac{\sigma^2}{n}$

(d) 对 S_n 应用切比雪夫不等式：

$$\Pr(|S_n - \mu| \geq \varepsilon) \leq \frac{\sigma^2}{n\varepsilon^2}$$

(e) 当 $n \to \infty$ 时，右边趋于 0 $\square$

Shearer 引理

4. 请严格复述 **Shearer 引理 (Shearer's Lemma)** 并解释其中符号含义。

答案：设 $X_1, \dots, X_n$ 是随机变量，$\mathcal{F}$ 是子集集合覆盖 $[n]$，使得每个 X_i 出现在至少 k 个子集中，则：

$$H(X_1, \dots, X_n) \leq \frac{1}{k} \sum_{F \in \mathcal{F}} H(X_F)$$

其中：

- $\mathcal{F} = \{F_1, \dots, F_m\}$ 是 $[n] = \{1, 2, \dots, n\}$ 的子集族
- $k = \min_i |\{F \in \mathcal{F} : i \in F\}|$（每个变量被覆盖的最小次数）
- $X_F = \{X_i : i \in F\}$（子集 F 对应的随机变量子集）

5. 请证明 **Shearer 引理**。

证明：使用熵的链式法则和条件熵的非负性：

$$\begin{aligned}
\sum_{F \in \mathcal{F}} H(X_F) &= \sum_{F \in \mathcal{F}} H\left(X_F \mid X_{F \cap \{1,\dots,i-1\}}\right) \quad \text{（链式法则）} \\
&\geq \sum_{F \in \mathcal{F}} \sum_{i \in F} H\left(X_i \mid X_1, \dots, X_{i-1}\right) \\
&= \sum_{i=1}^{n} H\left(X_i \mid X_1, \dots, X_{i-1}\right) \cdot |\{F \in \mathcal{F} : i \in F\}| \\
&\geq k \sum_{i=1}^{n} H\left(X_i \mid X_1, \dots, X_{i-1}\right) \\
&= kH(X_1, \dots, X_n) \quad \text{（链式法则）} \\
\Rightarrow H(X_1, \dots, X_n) &\leq \frac{1}{k} \sum_{F \in \mathcal{F}} H(X_F) \quad \square
\end{aligned}$$

渐近均分性与典型集

6. 请解释**渐近均分性 (Asymptotic Equipartition Property, AEP)** 的含义。

答案： AEP 描述 i.i.d. 序列的典型性质：

- 对于 i.i.d. 序列 $X_1, \ldots, X_n \sim p(x)$
- $-\frac{1}{n} \log p(X_1, \ldots, X_n)$ 以高概率趋近于 $H(X)$
- 当 n 很大时，序列分为两类：
  - **典型序列**：概率接近 $2^{-nH(X)}$
  - **非典型序列**：总概率可忽略

7. 请定义**典型集 (Typical Set)** $A_\varepsilon^{(n)}$。

答案：熵典型集定义为：

$$A_\varepsilon^{(n)} = \left\{ x^n \in \mathcal{X}^n : \left| -\frac{1}{n} \log p(x^n) - H(X) \right| \leq \varepsilon \right\}$$

其中 $p(x^n) = \prod_{i=1}^n p(x_i)$。

8. 请证明以下**典型集性质**：

   (a) 概率集中性：$\lim_{n\to\infty} \Pr(X^n \in A_\varepsilon^{(n)}) = 1$

证明：

- 定义 $Y_i = -\log p(X_i)$，则 $\mathbb{E}[Y_i] = H(X)$
- $-\frac{1}{n}\log p(X^n) = \frac{1}{n}\sum_{i=1}^n Y_i$
- 由弱大数定律：

$$\Pr\left(\left|\frac{1}{n}\sum_{i=1}^{n} Y_i - H(X)\right| > \varepsilon\right) \to 0$$

- 故 $\Pr(X^n \in A_\varepsilon^{(n)}) \to 1$ □

(b) 基数上界：$|A_\varepsilon^{(n)}| \le 2^{n(H(X)+\varepsilon)}$

证明：

$$\begin{aligned} 1 \ge \sum_{x^n \in A_\varepsilon^{(n)}} p(x^n) &\ge \sum_{x^n \in A_\varepsilon^{(n)}} 2^{-n(H(X)+\varepsilon)} \\ &= |A_\varepsilon^{(n)}| \cdot 2^{-n(H(X)+\varepsilon)} \\ \Rightarrow |A_\varepsilon^{(n)}| &\le 2^{n(H(X)+\varepsilon)} \quad \square \end{aligned}$$

(c) 基数下界（大 n）：$|A_\varepsilon^{(n)}| \ge (1-\varepsilon)2^{n(H(X)-\varepsilon)}$

证明：

$$\begin{aligned}
1-\varepsilon &< \Pr(A_\varepsilon^{(n)}) \\
&= \sum_{x^n \in A_\varepsilon^{(n)}} p(x^n) \\
&\le |A_\varepsilon^{(n)}| \cdot \max_{x^n \in A_\varepsilon^{(n)}} p(x^n) \\
&\le |A_\varepsilon^{(n)}| \cdot 2^{-n(H(X)-\varepsilon)} \\
\Rightarrow |A_\varepsilon^{(n)}| &\ge (1-\varepsilon) 2^{n(H(X)-\varepsilon)} \quad \square
\end{aligned}$$

(d) 典型集占整个序列空间的比例趋于 0：$\frac{|A_\varepsilon^{(n)}|}{|\mathcal{X}^n|} \to 0 \ (n \to \infty)$

证明：

$$\begin{aligned}
\frac{|A_\varepsilon^{(n)}|}{|\mathcal{X}^n|} &\le \frac{2^{n(H(X)+\varepsilon)}}{|\mathcal{X}|^n} \\
&= 2^{n(H(X)+\varepsilon)} \cdot 2^{-n \log |\mathcal{X}|} \\
&= 2^{-n(\log |\mathcal{X}| - H(X) - \varepsilon)}
\end{aligned}$$

当 $H(X) < \log |\mathcal{X}|$ 且 ε 足够小时：

$$\log |\mathcal{X}| - H(X) - \varepsilon > 0$$

指数衰减导致比例趋近于 0 $\quad \square$

9. 为什么说"$H(A_\varepsilon^{(n)}) \neq \log |\mathcal{X}^n|$"？请解释典型集的熵特性。

答案：

- $\log|\mathcal{X}^n| = n\log|\mathcal{X}|$ 是序列空间的最大可能熵
- 典型集的熵满足：

$$\log[(1-\varepsilon)2^{n(H(X)-\varepsilon)}] \leq H(A_\varepsilon^{(n)}) \leq \log[2^{n(H(X)+\varepsilon)}]$$

- 即 $n(H(X)-\varepsilon)+o(n) \leq H(A_\varepsilon^{(n)}) \leq n(H(X)+\varepsilon)$
- 当 $H(X) < \log|\mathcal{X}|$ 时，$H(A_\varepsilon^{(n)}) \ll n\log|\mathcal{X}|$
- 这意味着典型序列集相比整个空间有指数级压缩