概率, 散度

请在 11 月 27 日课前提交纸质作业.

1. (6 分) 用 Poisson(λ) 表示期望为 λ 的泊松分布. 随机变量 X 服从柏松分布, 记为 $X \sim \text{Poisson}(\lambda)$, 当且仅当 $\forall k \in \mathbb{N}, \Pr[X = k] = \frac{e^{-\lambda}\lambda^k}{k!}$.
   (1) 如果独立的随机变量 X, Y 分别服从 $X \sim \text{Poisson}(\lambda_1)$, $Y \sim \text{Poisson}(\lambda_2)$, 定义另一个随机变量 $Z = X + Y$. 证明 $Z \sim \text{Poisson}(\lambda_1 + \lambda_2)$.
   (2) 随机变量 X, Y, Z 的定义和上一文相同. 求已知 Z 时, X 的条件分布. 具体来说, 对任意 $n, k \in \mathbb{N}$, 计算 $\Pr[X = k | Z = n]$.

2. (4 分) 有一族相关事件 $E_1, \ldots, E_{2024}$, 满足 $\Pr[E_i] = \frac{1}{2}, \Pr[E_i \wedge E_j] = \frac{1}{3}, \ldots$. 更一般地, 对任意 $S \subseteq \{1, 2, \ldots, 2024\}$, $\Pr[\bigwedge_{i \in S} E_i] = \frac{1}{|S|+1}$. 计算 $\Pr[\bigvee_{i=1}^{2024} E_i]$.

3. (10 分) 令 $X_1, X_2, \cdots \in \{H, T\}$ 表示反复独立的投掷一枚均匀硬币的结果. 用 T_{HHH} (resp. $T_{HHT}, T_{HTH}, \ldots$) 表示首次出现连续的 HHH (resp. $HHT, HTH, \ldots$) 的投掷次数. 求出 $\mathbb{E}[T_{HHH}], \mathbb{E}[T_{HHT}], \mathbb{E}[T_{HTH}]$.

   Remark: 对于计算密集的步骤, 不妨使用计算机辅助.

4. (5 分) 证明信息熵是凹函数. 也就是对任意分布 P, Q 和 $\lambda \in (0, 1)$

$$H[\lambda P + (1-\lambda)Q] \geq \lambda H[P] + (1-\lambda)H[Q].$$

5. (5 分) 令 P_{XY} 表示 X, Y 的联合分布. 证明

$$I(X;Y) = D(P_{XY} \| P_X P_Y).$$

6. (5 分) 证明 $d(p\|q) = D(\text{Bern}(p) \| \text{Bern}(q)) \geq 2 \log e \cdot (p-q)^2$.

   Remark: 不妨两边都除以 $\log e$, 这等价于使用 e 作底数.

7. (5 分) 证明散度的 data-processing 不等式. 对任意 P_X, Q_X 和 kernel $P_{Y|X}$, 令 $P_Y = P_X \circ P_{Y|X}$, $Q_Y = Q_X \circ P_{Y|X}$ (也就是说, P_Y, Q_Y 分别是 $P_{XY} = P_X P_{Y|X}, Q_{XY} = Q_X P_{Y|X}$ 的边缘分布). 证明

$$D(P_X \| Q_X) \geq D(P_Y \| Q_Y).$$

   Remark: 互信息的 data-processing 不等式可以由散度的 data-processing 不等式推出. 如果 X, Y, Z 的依赖关系可以用有向图 $X \to Y \to Z$ 表示 (即 $P_{XYZ} = P_X P_{Y|X} P_{Z|Y}$), 注意到

$$I(X;Y) = D(P_{XY} \| P_X P_Y), \qquad I(X;Z) = D(P_{XZ} \| P_X P_Z).$$

   只需定义一个合适的 kernel 便证明了互信息的 data-processing 不等式 $I(X;Y) \geq I(X;Z)$.

8. (5 分) 使用 data-processing 不等式, 证明

$$\sqrt{\frac{1}{2\log e}D(P\|Q)} \geq \Delta(P,Q)$$

这里 $\Delta(P,Q)$ 表示 P,Q 之间的统计距离（statistical distance, 也可以更精确地称为 total variation distance）

$$\Delta(P,Q) = \frac{1}{2}\sum_{x}|P(x)-Q(x)| = \max_{\text{事件 } E}(P(E)-Q(E)).$$

9. (5 分) 证明对于服从任意联合分布的随机变量 X,Y,Z,

$$2H[X,Y,Z] \leq H[X,Y]+H[X,Z]+H[Y,Z].$$

据此证明 Shearer 引理: 令 Ω 是 $\mathbb{R}^3$ 上 n 个点组成的集合, Ω 向三个坐标平面投影分别有 n_1,n_2,n_3 个像, 那么 $n^2 \leq n_1n_2n_3$. 并说明何时可以取到等号.