概率, 散度

参考答案

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 $\Pr[X = k | Z = n]$.

解

(1) 我们有如下计算:

$$\begin{aligned}
\Pr[Z = k] &= \sum_{i=0}^{k} \Pr[X = i, Y = k - i] \\
&= \sum_{i=0}^{k} \Pr[X = i] \Pr[Y = k - i] \\
&= \sum_{i=0}^{k} \frac{e^{-\lambda_1}\lambda_1^i}{i!} \cdot \frac{e^{-\lambda_2}\lambda_2^{k-i}}{(k-i)!} \\
&= \frac{e^{-(\lambda_1+\lambda_2)}}{k!} \sum_{i=0}^{k} \binom{k}{i} \lambda_1^i \lambda_2^{k-i} \\
&= \frac{e^{-(\lambda_1+\lambda_2)}(\lambda_1+\lambda_2)^n}{k!}.
\end{aligned}$$

因此 $Z \sim \text{Poisson}(\lambda_1 + \lambda_2)$.

另一种解法: X 的概率生成函数是 $e^{\lambda_1(z-1)}$, Y 的概率生成函数是 $e^{\lambda_2(z-1)}$, 它们的独立和 $Z = X + Y$ 的概率生成函数是 $e^{(\lambda_1+\lambda_2)(z-1)}$, 所以 $Z \sim \text{Poisson}(\lambda_1 + \lambda_2)$.

(2) 首先, 当 $n < k$, $\Pr[X = k | Z = n] = 0$, 对于其他情况, 我们有如下计算:

$$\begin{aligned}
\Pr[X = k | Z = n] &= \frac{\Pr[X = k, Y = n - k]}{\Pr[Z = n]} \\
&= \frac{\Pr[X = k][Y = n - k]}{\Pr[Z = n]} \\
&= \frac{e^{-\lambda_1}\lambda_1^k}{k!} \cdot \frac{e^{-\lambda_2}\lambda_2^{n-k}}{(n-k)!} \cdot \left(\frac{e^{-(\lambda_1+\lambda_2)}(\lambda_1+\lambda_2)^n}{n!}\right)^{-1} \\
&= \binom{n}{k} \left(\frac{\lambda_1}{\lambda_1+\lambda_2}\right)^k \left(\frac{\lambda_2}{\lambda_1+\lambda_2}\right)^{n-k}.
\end{aligned}$$

这是一个二项分布.

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 $\Pr[\bigvee_{i=1}^{2024} E_i]$.

解 2024/2025.

对任意 $n \geq 1$, 根据容斥原理,

$$\begin{aligned}
\Pr\Big[\bigvee_{i=1}^{n} E_i\Big] &= \sum_{j=1}^{n} (-1)^{j-1} \sum_{S \subseteq [n], |S|=j} \Pr\Big[\bigwedge_{i \in S} E_i\Big] \\
&= \sum_{j=1}^{n} (-1)^{j-1} \binom{n}{j} \frac{1}{j+1} \\
&= \sum_{j=1}^{n} (-1)^{j-1} \frac{1}{n} \binom{n+1}{j+1} \\
&= \frac{1}{n+1} \left(\sum_{j=0}^{n+1} (-1)^{j} \binom{n+1}{j} - \binom{n+1}{0} + \binom{n+1}{1} \right) \\
&= \frac{1}{n+1} \left(0 - \binom{n+1}{0} + \binom{n+1}{1} \right) = \frac{n}{n+1}.
\end{aligned}$$

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.

解

(1) $\mathbb{E}[T_{HHH}] = 14$.

用 $\mathcal{S}$ 表示出现 HHH 后即刻停止抛掷, 可能得到的序列集合. 也就是所有以 HHH 结尾, 且除了末尾不包含其它 HHH 的序列的集合. 那么

$$\mathcal{S} = T\mathcal{S} + HT\mathcal{S} + HHT\mathcal{S} + HHH.$$

这里 $+$ 表示集合的不交并, 或多重集的并. 据此, 我们知道 T_{HHH} 的概率生成函数 $G(z)$ 满足

$$G(z) = \frac{z}{2} G(z) + \frac{z^2}{4} G(z) + \frac{z^3}{8} G(z) + \frac{z^3}{8}$$

进而解出

$$G(z) = \frac{\frac{z^3}{8}}{1 - \frac{z}{2} - \frac{z^2}{4} - \frac{z^3}{8}}$$

计算得到 $\mathbb{E}[T_{HHH}] = G'(1) = 14$.

(2) $\mathbb{E}[T_{HHT}] = 8$.

用 $\mathcal{S}$ 表示出现 HHT 后即刻停止抛掷, 可能得到的序列集合. 那么

$$\mathcal{S} = T\mathcal{S} + HT\mathcal{S} + HHT + HHHT + HHHHT + \dots.$$

据此, 我们知道 T_{HHH} 的概率生成函数 $G(z)$ 满足

$$G(z) = \frac{z}{2}G(z) + \frac{z^2}{4}G(z) + \frac{z^3}{8} + \frac{z^4}{16} + \frac{z^5}{32} + \dots$$

进而解出

$$G(z) = \frac{\frac{z^3}{8}/\left(1-\frac{z}{2}\right)}{1-\frac{z}{2}-\frac{z^2}{4}}$$

计算得到 $\mathbb{E}[T_{HHT}] = G'(1) = 8$.

(3) $\mathbb{E}[T_{HTH}] = 10$.

用 $\mathcal{S}$ 表示出现 HTH 后即刻停止抛掷, 可能得到的序列集合. 那么

$$\begin{aligned}\mathcal{S} &= T\mathcal{S} + HTH + HTT\mathcal{S} + HHTH + HHTT\mathcal{S} + HHHTH + HHHTT\mathcal{S} + \cdots \\ &= T\mathcal{S} + (H + HH + HHH + \dots)(TH + TT\mathcal{S})\end{aligned}$$

据此, 我们知道 T_{HTH} 的概率生成函数 $G(z)$ 满足

$$G(z) = \frac{z}{2}G(z) + \left(\frac{z}{2} + \frac{z^2}{4} + \frac{z^3}{8} + \dots\right)\left(\frac{z^2}{4} + \frac{z^2}{4}G(z)\right)$$

进而解出

$$G(z) = \frac{\frac{z^3}{8}/\left(1-\frac{z}{2}\right)}{1-\frac{z}{2}-\frac{z^3}{8}/\left(1-\frac{z}{2}\right)}$$

计算得到 $\mathbb{E}[T_{HTH}] = G'(1) = 10$.

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

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

解 令 Ω 是 P, Q 的支撑集的并集. 如果 Ω 有限, 用 U_Ω 表示 Ω 上的均匀分布, 那么直接使用熵与散度的关系

$$\begin{aligned}H[P] &= \log|\Omega| - D(P\|U_\Omega) \\ H[Q] &= \log|\Omega| - D(P\|U_\Omega) \\ H[\lambda P + (1-\lambda)Q] &= \log|\Omega| - D(\lambda P + (1-\lambda)Q\|U_\Omega)\end{aligned}$$

以及散度的凸性便证明了结论.

对于一般情况, 借鉴散度凸性的证明, 可以定义一个联合分布 P_{XZ},

$$P_Z = \mathrm{Bern}(\lambda), \qquad P_{X|Z=1} = P, \qquad P_{X|Z=0} = Q.$$

那么恰有 $P_X = \lambda P + (1-\lambda)Q$. 于是

$$H[\lambda P + (1-\lambda)Q] = H[X] \geq H[X\|Z] = \lambda H[P] + (1-\lambda)H[Q].$$

5. (5 分) 令 P_{XY} 表示 X, Y 的联合分布. 证明

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

解

$$\begin{aligned}
D(P_{XY}\|P_X P_Y) &= \mathop{\mathbb{E}}_{(X,Y)\sim P_{XY}}\left[\log \frac{P_{XY}(X,Y)}{(P_X P_Y)(X,Y)}\right] \\
&= \mathop{\mathbb{E}}_{(X,Y)\sim P_{XY}}\left[\log P_{XY}(X,Y)\right] + \mathop{\mathbb{E}}_{(X,Y)\sim P_{XY}}\left[\log \frac{1}{P_X(X)}\right] + \mathop{\mathbb{E}}_{(X,Y)\sim P_{XY}}\left[\log \frac{1}{P_Y(Y)}\right] \\
&= -\mathop{\mathbb{E}}_{(X,Y)\sim P_{XY}}\left[\log \frac{1}{P_{XY}(X,Y)}\right] + \mathop{\mathbb{E}}_{X\sim P_X}\left[\log \frac{1}{P_X(X)}\right] + \mathop{\mathbb{E}}_{Y\sim P_Y}\left[\log \frac{1}{P_Y(Y)}\right] \\
&= -H[X,Y] + H[X] + H[Y] \\
&= I(X;Y)
\end{aligned}$$

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

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

解 固定 q, 定义 $f(x) = d(x\|q)$. 那么

$$\begin{aligned}
f(p) &= p\log\frac{p}{q} + (1-p)\log\frac{1-p}{1-q} \\
f'(p) &= \log\frac{p}{q} + \log e - \log\frac{1-p}{1-q} - \log e \\
f''(p) &= \frac{1}{p}\log e + \frac{1}{1-p}\log e
\end{aligned}$$

特别地, $f(q) = f'(q) = 0$, $f''(p) \geq 4\log e$. 于是

$$d(p\|q) = f(p) \geq 2\log e \cdot (p-q)^2.$$

解 固定 p, 定义 $f(x) = d(p\|x)$. 那么

$$\begin{aligned}
f(q) &= p\log\frac{p}{q} + (1-p)\log\frac{1-p}{1-q} \\
f'(q) &= -\frac{p}{q}\log 2 + \frac{1-p}{1-q}\log e = \frac{q-p}{q(1-q)}\log e
\end{aligned}$$

当 $q > p$ 时, $f'(q) \geq 4\log e \cdot (q-p)$, 于是由中值定理

$$d(p\|q) = f(q) \geq 2\log e \cdot (q-p)^2.$$

对称地, 当 $q < p$ 时, $f'(q) \leq -4\log e \cdot (p-q)$, 于是

$$d(p\|q) = f(q) \geq 2\log e \cdot (q-p)^2.$$

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 $I(X;Y) \geq I(X;Z)$.

解 对 $D(P_{XY}\|Q_{XY})$ 使用散度的 chain rule 两次.

$$D(P_{XY}\|Q_{XY}) = D(P_X\|Q_X) + \underbrace{D(P_{Y|X}\|Q_{Y|X}|P_X)}_{=0},$$
$$D(P_{XY}\|Q_{XY}) = D(P_Y\|Q_Y) + \underbrace{D(P_{X|Y}\|Q_{X|Y}|P_Y)}_{\geq 0}.$$

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)

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

解 用 E^* 表示统计距离的定义中最能区分分布 P_X, Q_X 的事件

$$E^* = \arg\max_{\text{事件 } E}(P_X(E) - Q_X(E)).$$

定义条件概率 $P_{Z|X}$

$$P_{Z|X}(1|x) = 1 \text{ if } x \in E^* \qquad P_{Z|X}(0|x) = 1 \text{ if } x \notin E^*$$

因此 $\mathrm{Bern}(P_X(E^*)), \mathrm{Bern}(Q_X(E^*))$ 分别是 $P_X P_{Y|X}$ 和 $Q_X P_{Y|X}$ 的边缘分布. 根据 data-processing 不等式,

$$D(P_X \| Q_X) \geq D(\mathrm{Bern}(P_X(E^*)) \| \mathrm{Bern}(Q_X(E^*))) \geq 2\log e \cdot (P_X(E^*) - Q_X(E^*))^2 = 2\log e \cdot \Delta(P_X, Q_X)^2.$$

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 $n^2 \leq n_1 n_2 n_3$. 并说明何时可以取到等号.

解 先证明这个熵不等式.

$$\begin{aligned}
&H[X,Y] + H[X,Z] + H[Y,Z] - 2H[X,Y,Z] \\
&= H[X,Y] - H[Y|X,Z] - H[X|Y,Z] \\
&\geq H[X,Y] - H[Y|X] - H[X|Y,Z] \\
&= H[X] - H[X|Y,Z] \geq 0.
\end{aligned}$$

不等式取等号的必要条件是 $H[X] = H[X|Y,Z]$, 也就是 X 与 (Y,Z) 独立. 由对称性, 不等式取等号的必要条件 X, Y, Z 相互独立. 不难验证这也是充分条件.

令随机变量 (X,Y,Z) 是从 Ω 中随机选取的一个点的坐标, 那么 $H[X,Y,Z] = \log n$. 根据题目条件, $(X,Y), (X,Z), (Y,Z)$ 的支撑集大小分别为 n_1, n_2, n_3, 所以 $H[X,Y], H[X,Z], H[Y,Z]$ 分别不超过 $\log n_1, \log n_2, \log n_3$. 根据刚刚证明的熵不等式,

$$2\log n = H[X,Y,Z] \leq H[X,Y] + H[X,Z] + H[Y,Z] \leq \log n_1 + \log n_2 + \log n_3.$$

也就是题目要求的 $n^2 \leq n_1 n_2 n_3$.

不等式取等号, 首先需要熵不等式取等号. 根据之前的讨论, 必要条件是 X, Y, Z 相互独立, 这时支撑集 Ω 一定是一个笛卡尔积 $\Omega = \mathrm{Supp}(X) \times \mathrm{Supp}(Y) \times \mathrm{Supp}(Z)$. 不难验证, Ω 是笛卡尔积也是 $n^2 = n_1 n_2 n_3$ 的充分条件.