S.-T. Yau College Student Mathematics Contests 2023

Probability and Statistics

Solve every problem.

1. The sequence $(X_1, X_2, \ldots, X_n, \ldots)$ is a Dirichlet process with base distribution G_0 and concentration parameter $\alpha_0 > 0$ if G_0 is a probability distribution on R and satisfies:

   - $X_1 \sim G_0$
   - Conditional on $X_1, X_2, \ldots, X_n$, the distribution of X_{n+1} is $\alpha_0 G_0 + \sum_{i=1}^n \delta_{X_i}$, appropriately normalized, where δ_x is the Dirac measure with probability 1 on singleton $\{x\}$.

   Assume that G_0 has finite first and second moments.

   (a) Derive the distribution of $X_n, n \geq 1$.

   (b) Let $Y_n = I(X_n > 0)$. Prove or disprove $(Y_n)_{n\geq 1}$ forms a Dirichlet Process. If $(Y_n)_{n\geq 1}$ forms a Dirichlet Process, determine its concentration parameter and its base distribution.

2. Suppose X and Y are non-negative random variables on a probability space $(\Omega, \mathcal{F}, P)$. Let $H(x, y)$ be a function on $[0, \infty)^2$ such that $E(|H(X, Y)|) < \infty$. Define function $\phi(u) = u/(1 + u)$ for $u \geq 0$. For integer $n = 0, 1, 2, \ldots$, let

$$U_n = \sum_{j=1}^{2^n} \frac{j-1}{2^n} I\Big(\frac{j-1}{2^n} \leq \phi(X) < \frac{j}{2^n}\Big), \quad V_n = E(H(X, Y)|U_n).$$

   Prove or disprove that there exists a random variable Z such that as $n \to \infty$, V_n converges to Z almost surely. If there exists such a Z, show the expression of Z (in the sense of almost surely).

3. Let $X_1, X_2, \ldots, X_n$ be a sequence of i.i.d. random variables with a uniform distribution on $(0, 1)$. Define the events $A_1, A_2, \ldots$ by

$$A_n = \{X_n = \max(X_1, \ldots, X_n)\}.$$

   Define $R_n = \sum_{k=1}^n I(A_k)$. Let (m_n) be a sequence of positive numbers such that $\lim_{n\to\infty} m_n = \infty$. Compute the following limit

$$\lim_{n\to\infty} P\Big(|R_n - \log n| > m_n \sqrt{\log n}\Big)$$

4. Denote by $\mathcal{B}$ the Borel sigma field on the real line R, and let $(\Omega, \mathcal{F})$ be a measurable space. Define a mapping $Q(t, A)$ for $t \in R$ and $A \in \mathcal{F}$ such that $Q(t, \cdot)$ is a probability measure on $(\Omega, \mathcal{F})$ for each $t \in R$, and $Q(\cdot, A)$ is a Borel function for each $A \in \mathcal{F}$.

   Denote by Π a probability measure on $(R, \mathcal{B})$, P a probability measure on $(\Omega, \mathcal{F})$, and T a random variable on $(\Omega, \mathcal{F}, P)$. Assume that Π, P and T satisfy $\Pi = P \circ T^{-1}$, and for $A \in F$,

$$P(A) = \int_R Q(t, A)\Pi(dt),$$

   where $P \circ T^{-1}$ denotes the induced probability measure by T. Prove or disprove the following statement:

$$\text{For any } A \in \mathcal{F}, \quad P(A|T) = Q(T, A) \text{ almost surely.}$$

5. Four statisticians I, II, III and IV play a sequence of games. For each game, the winning probabilities of I, II, III and IV are $(1-\theta)/2$, $(1-\theta)/2$, $\theta/2$ and $\theta/2$, respectively, where $0 < \theta < 1$. There is only one winner in each game and no tie is allowed. Assume that outcomes of games are independent of each other. For a fixed integer $r \geq 2$, the stopping rule is to terminate as soon as one of the following conditions hold: (1) I and II together win r games; (2) III and IV together win $r+1$ games. At the time of termination, let X_1, X_2, X_3 and X_4 denote the numbers of games won by I, II, III and IV, respectively.

   (a) Prove or disprove the statistic $T = (X_1 + X_2, X_3 + X_4)$ is complete.

   (b) Find a uniformly minimum variance unbiased estimator of θ.

6. A system of interest involves three random variables, X, Y, and S, where S has a Poisson distribution with mean 2λ, for a parameter $\lambda > 0$, and where X and Y are conditionally independent Bernoulli variables, given $S = s$, with

$$P(X = 1|S = s) = \frac{1}{2^{s+1}} \text{ and } P(Y = 1|S = s) = \frac{\theta}{2^s},$$

   where $\theta \in (0, 1)$ is a second parameter. The random variable S is unobservable.

   We have n i.i.d. copies (X_i, Y_i) of (X, Y).

   (a) An intuitive estimator for θ is $\hat{\theta}_n = \bar{Y}_n/(2\bar{X}_n)$, where $\bar{X}_n = \frac{1}{n}\sum_{i=1}^n X_i$ and $\bar{Y}_n = \frac{1}{n}\sum_{i=1}^n Y_i$. Derive its asymptotic distribution.

   (b) Consider the hypothesis testing problem

$$H_0 : \theta = 1/2 \quad \text{versus} \quad H_1 : \theta \neq 1/2.$$

   Construct an exact test statistic. As θ moves away from $1/2$, describe all sources of increasing power that you can think of.