S.-T. Yau College Student Mathematics Contests 2022

Probability and Statistics

Solve every problem.

Problem 1. Let $\{X_n\}$ be a sequence of Gaussian random variables. Suppose that X is a random variable such that X_n converges to X in distribution as $n \to \infty$. Show that X is also a (possibly degenerate, *i.e.*, variance zero) Gaussian random variable.

Problem 2. For two probability measures μ and ν on the real line $\mathbf{R}$, the total variation distance $\|\mu - \nu\|_{TV}$ is defined as

$$\|\mu - \nu\|_{TV} = \sup\{\mu(C) - \nu(C) : C \in \mathcal{B}(\mathbf{R})\},$$

where $\mathcal{B}(\mathbf{R})$ is the σ-algebra of Borel sets on $\mathbf{R}$. Let $\mathcal{C}(\mu,\nu)$ be the space of couplings of the probability measures μ and ν, *i.e.*, the space of $\mathbf{R}^2$ valued random variables (X,Y) defined on some (not necessarily same) probability space $(\Omega, \mathcal{F}, \mathbb{P})$ such that the marginal distributions of X and Y are μ and ν, respectively. Show that

$$\|\mu - \nu\|_{TV} = \inf\{\mathbb{P}(X \neq Y) : (X,Y) \in \mathcal{C}(\mu,\nu)\}.$$

For simplicity you may assume that μ and ν are absolutely continuous with respect to the Lebesgue measure on $\mathbf{R}$.

Problem 3. We throw a fair die repeatedly and independently. Let τ_{11} be the first time the pattern 11 (two consecutive 1's) appears and τ_{12} the first time the pattern 12 (1 followed by 2) appears.

(a) Calculate the expected value $\mathbb{E}\tau_{11}$.

(b) Which is larger, $\mathbb{E}\tau_{11}$ or $\mathbb{E}\tau_{12}$? It is sufficient to give an intuitive argument to justify your answer. You can also calculate $\mathbb{E}\tau_{12}$ if you wish.

Problem 4. Let $\{X_n\}$ be a Markov chain on a discrete state space S with transition function $p(x,y)$, $x,y \in S$. Suppose that there is a state $y_0 \in S$ and a positive number θ such that $p(x,y_0) \geq \theta$ for all $x \in S$.

(a) Show that is a positive constant $\lambda < 1$ such that for any two initial distribution μ and ν,

$$\sum_{y \in S} \left|\mathbb{P}_\mu\{X_1 = y\} - \mathbb{P}_\nu\{X_1 = y\}\right| \leq \lambda \sum_{y \in S} |\mu(y) - \nu(y)|.$$

(b) Show that the Markov chain has a unique stationary distribution π and

$$\sum_{y \in S} \left|\mathbb{P}_\mu\{X_n = y\} - \pi(y)\right| \leq 2\lambda^n.$$

Problem 5. Consider a linear regression model with p predictors and n observations:

$$\mathbf{Y} = X\beta + \mathbf{e},$$

where $X_{n\times p}$ is the design matrix, β is the unknown coefficient vector, and the random error vector $\mathbf{e}$ has a multivariate normal distribution with mean zero and $\mathrm{Var}(\mathbf{e}) = \sigma^2 I_n$ ($\sigma^2 > 0$ unknown and I_n is the identity matrix).

Here $\text{rank}(X) = k \le p$, p may or may not be greater than n, but we assume $n - k > 1$. Let $\mathbf{x}_1 = (x_{1,1}, \dots, x_{1,p})$ be the first row of X and define

$$\gamma = \frac{\mathbf{x}_1 \beta}{\sigma}.$$

Find the uniformly minimum variance unbiased estimator (UMVUE) of γ or prove it does not exist.

Problem 6. Let $X_1, \dots, X_{2022}$ be independent random variables with $X_i \sim N(\theta_i, i^2)$, $1 \le i \le 2022$. For estimating the unknown mean vector $\theta \in R^{2022}$, consider the loss function $L(\theta, \mathbf{d}) = \sum_{i=1}^{2022} (d_i - \theta_i)^2 / i^2$. Prove that $\mathbf{X} = (X_1, \dots, X_{2022})$ is a minimax estimator of θ.

Recall: *If* $Y|\mu \sim N(\mu, \sigma^2)$ *and* $\mu \sim N(\mu_0, \sigma_0^2)$ *then* $\mu | Y = y \sim N\left(\frac{\mu_0/\sigma_0^2 + y/\sigma^2}{1/\sigma_0^2 + 1/\sigma^2}, \frac{1}{1/\sigma_0^2 + 1/\sigma^2}\right)$.