S.-T. Yau College Student Mathematics Contests 2025

Probability and Statistics

(4 problems)

Problem 1.

Consider a single observation $X \sim N(\mu, \sigma^2)$ with both μ and σ^2 unknown. We are interested in constructing confidence interval (CI) for μ.

1. Consider the following CI:

$$\mathrm{CI}(X; c_1, c_2) = \begin{cases} c_1 X \le \mu \le c_2 X & \text{if } X > 0 \\ c_2 X \le \mu \le c_1 X & \text{if } X < 0 \end{cases}.$$

   What is the coverage probability of this CI?

2. Let $F(x; b_1, b_2)$ be a function that is measurable in x and is a distribution function on $b_1 < b_2$, and which generates a randomized confidence interval whose lower and upper endpoints are random variables $b_1 < b_2$ with distribution function $F(x; b_1, b_2)$. Then, there is a randomized confidence interval given by a distribution function $G(c_1, c_2)$ randomly choosing a interval of the form $\mathrm{CI}(x; c_1, c_2)$ with $(C_1, C_2) \sim G$ and satisfying the following "minimax" property:

$$\inf_{\mu,\sigma} E_{\mu,\sigma} P_G\{\mu \in \mathrm{CI}(X; c_1, c_2)\} \ge \sup_F \inf_{\mu,\sigma} E_{\mu,\sigma} P_{F(X;\cdot)}\{\mu \in (b_1, b_2)\}$$

$$E_{\mu,\sigma} E_G\left[(c_2 - c_1)|X|\right] = E_{\mu,\sigma} E_{F(X;\cdot)}\left[(b_2 - b_1)\right].$$

Problem 2.

Consider a joint distribution of the random variable pair $\{X, Y\}$, and let P_1 and P_2 be the two marginal distributions on the real line (that is, for X and Y, respectively). Let $p_1(x)$ and $p_2(y)$ be the respective densities for P_1 and P_2 with respect to a dominating measure μ. Let $\nu(\cdot)$ be any function such that $\nu(X)$ and $\nu(Y)$ are not bounded and their expectations $E_1(\nu(X))$ and $E_2(\nu(Y))$ are both finite under P_1 and P_2, respectively. Suppose a single observation r comes from population 1 with probability 1/2 and from population 2 with probability 1/2. The problem is to choose the population, i, for which the expectation is greater, based on the single observation r.

Consider a randomized decision rule (or selection rule) to be given by a function

$$\phi : \mathcal{R} \to [0, 1]$$

such that $\phi(r)$ is the probability of choosing population 1 based on the observation r. Assume $\phi(r)$ is a one-to-one function.

For any given decision function $\phi(r)$, for any two densities $p_1(x)$ and $p_2(y)$ (with respect to the Lebesgue measure) such that $E_1(\nu(X)) > E_2(\nu(Y))$, is the probability of correctly selecting population 1 always greater than or equal to 1/2? If so, prove it. If not, for any given decision function $\phi(r)$, construct a $p_1(x)$ and a $p_2(y)$ such that $E_1(\nu(X)) > E_2(\nu(Y))$ but the probability of correctly selecting population 1 is less than 1/2.

Problem 3.

Let $n \geq 2$ be a natural number. Prove that there exist independent random variables $X_1, X_2, \ldots, X_n = X_0$ such that

$$P\big(X_{k-1} < X_k\big) \;=\; 1 \;-\; \frac{1}{4\cos^2\!\big(\frac{\pi}{n+2}\big)}, \quad \text{for all } 1 \leq k \leq n.$$

Problem 4.

Consider a one-way ANOVA with p cells and n/p observations per cell:

$$Y_{ij} = \beta_j + R_{ij}, \quad i = 1, \ldots, n/p, \quad j = 1, \ldots, p,$$

where $\{R_{ij}\}$ are i.i.d. Assume the true values $\beta_j = 0$.

Let ψ be a given function and let $\hat{\beta}_j$ denote the solution of the equation

$$0 = \sum_{i=1}^{n/p} \psi\big(Y_{ij} - \hat{\beta}_j\big).$$

Let ψ be a bounded function such that ψ' is bounded and continuous near zero. Suppose $E\psi(R) = 0$ and $E\psi'(R) = d \neq 0$. Assume $p(\log p)/n \to 0$. Prove that there are solutions $\{\hat{\beta}_j\}$ of the equation and a constant $B > 0$ such that

$$P\left\{\max_j |\hat{\beta}_j| \geq \left(\frac{1}{B}\frac{p\log n}{n}\right)^{1/2}\right\} \leq \frac{2p}{n} \to 0.$$