S.-T. Yau College Student Mathematics Contests 2026

Probability and Statistics

(6 problems)

Problem 1. A *copula* is a multivariate cumulative distribution function with uniform marginals on $[0,1]$. By Sklar's Theorem, any joint distribution can be decomposed into its marginal distributions and a copula that captures the dependence structure between variables. That is, for random variables U and V with marginal CDFs F_U and F_V, their joint CDF can be written as $H(u,v) = C\{F_U(u), F_V(v)\}$ for some copula C. The copula parameter ρ indexes the strength and shape of dependence, independently of the marginals.

Let $Y_a \in \{0, 1, \ldots, L-1\}$ be an ordinal outcome observed in treatment group $a \in \{0,1\}$.

(1). Define $\psi = \Pr(Y_1 > Y_0)$ as the probability that the outcome in the treatment group exceeds that in the control group. Suppose the joint CDF satisfies $\Pr(Y_1 \leq k,\, Y_0 \leq j) = C\{F_1(k), F_0(j)\}$ with the convention $F_a(-1) = 0$, where C is a pre-specified copula. Derive a closed-form expression for ψ in terms of C, F_1, and F_0.

(2). Suppose Y_a arises from a latent continuous variable Y_a^* through the threshold model

$$Y_a^* = \mu_a + \varepsilon_a, \qquad Y_a = \ell \iff \tau_{\ell-1} < Y_a^* \leq \tau_\ell,$$

with shared thresholds $-\infty = \tau_{-1} < \tau_0 < \cdots < \tau_{L-2} < \tau_{L-1} = +\infty$ and μ_a a constant for group a. Also, suppose the joint distribution of the latent residuals satisfies

$$\Pr(\varepsilon_1 \leq e_1,\, \varepsilon_0 \leq e_0) = C\{F_{\varepsilon_1}(e_1),\, F_{\varepsilon_0}(e_0)\},$$

where C is a pre-specified copula and $F_{\varepsilon_a}(e) = \Pr(\varepsilon_a \leq e)$. Prove that this implies

$$\Pr(Y_1 \leq k,\, Y_0 \leq j) = C\{F_1(k),\, F_0(j)\}.$$

Problem 2. Consider the partitioned linear regression model

$$Y = X_1\beta_1 + X_2\beta_2 + \varepsilon,$$

where $Y \in \mathbb{R}^n$, $X_1 \in \mathbb{R}^{n\times k_1}$, $X_2 \in \mathbb{R}^{n\times k_2}$, $k_1, k_2 \geq 1$, and $[X_1\ X_2]$ have full column rank. Define the annihilator matrix

$$M_1 = I_n - X_1(X_1^\top X_1)^{-1}X_1^\top,$$

which projects onto the orthogonal complement of the column space of X_1. Recall that for a generic regression of a response $\tilde{Y}$ on a predictor matrix $\tilde{X}$ with full column rank, the OLS estimator is

$$\hat{\beta} = (\tilde{X}^\top\tilde{X})^{-1}\tilde{X}^\top\tilde{Y}.$$

Prove: the OLS estimator $\hat{\beta}_2$ obtained from the full regression of Y on $[X_1\ X_2]$ is identical to the OLS estimator obtained from regressing M_1Y on M_1X_2.

Problem 3. Let ϕ and Φ be the density and distribution functions of the standard normal, and $a > 0$ is a constant.

(a) Show that $f(x) = 2\phi(x)\Phi(ax)$ is the density of some random variable (denoted by Y).

(b) Calculate $\mathbb{E}(Y)$.

Problem 4. Let $\{B_t, t \geq 0\}$ be a standard Brownian motion. For $a > 0$, define the first exit time from the interval $(-a, a)$:

$$\tau_a = \inf\{t \geq 0 : |B_t| = a\}.$$

Compute $\mathbb{E}[\tau_a^2]$.

Problem 5. Let $X_1, X_2, \dots$ be i.i.d. random variables. Suppose that for some integer $n \geq 2$, there exist constants $a > 0$ and $b \in \mathbb{R}$ such that

$$X_1 + \cdots + X_n \stackrel{d}{=} aX_1 + b. \tag{U1}$$

Let $\alpha = \frac{\ln n}{\ln a}$.

(a). **Prove:** In the case $\alpha = 1$ and $b = 0$, the characteristic function of X_k is

$$\phi(t) = \exp\{i\mu t - \gamma|t|\}.$$

(b). **Prove:** If we additionally assume $\mathbb{E}|X_1| < \infty$, prove that excluding the degenerate case, it is impossible to satisfy (U1) and $\alpha \leq 1$.

Problem 6. Let ξ and η be independent random variables. If the sum $S = \xi + \eta$ and the difference $D = \xi - \eta$ are also independent, then ξ and η must follow normal distributions.