Paper 1, Section II

28J Principles of Statistics

The distribution of a random variable X is obtained from the binomial distribution $\mathcal{B}(n;\Pi)$ by conditioning on $X>0$; here $\Pi \in (0,1)$ is an unknown probability parameter and n is known. Show that the distributions of X form an exponential family and identify the natural sufficient statistic T, natural parameter Φ, and cumulant function $k(\phi)$. Using general properties of the cumulant function, compute the mean and variance of X when $\Pi=\pi$. Write down an equation for the maximum likelihood estimate $\widehat{\Pi}$ of Π and explain why, when $\Pi=\pi$, the distribution of $\widehat{\Pi}$ is approximately normal $\mathcal{N}(\pi, \pi(1-\pi)/n)$ for large n.

Suppose we observe $X=1$. It is suggested that, since the condition $X>0$ is then automatically satisfied, general principles of inference require that the inference to be drawn should be the same as if the distribution of X had been $\mathcal{B}(n;\Pi)$ and we had observed $X=1$. Comment briefly on this suggestion.

Paper 2, Section II

28J Principles of Statistics

Define the *Kolmogorov–Smirnov* statistic for testing the null hypothesis that real random variables $X_1,\ldots,X_n$ are independently and identically distributed with specified continuous, strictly increasing distribution function F, and show that its null distribution does not depend on F.

A composite hypothesis H_0 specifies that, when the unknown positive parameter Θ takes value θ, the random variables $X_1,\ldots,X_n$ arise independently from the uniform distribution $\mathrm{U}[0,\theta]$. Letting $J := \arg\max_{1\leqslant i\leqslant n} X_i$, show that, under H_0, the statistic (J, X_J) is sufficient for Θ. Show further that, given $\{J=j,\ X_j=\xi\}$, the random variables $(X_i : i\neq j)$ are independent and have the $\mathrm{U}[0,\xi]$ distribution. How might you apply the Kolmogorov–Smirnov test to test the hypothesis H_0?

Paper 3, Section II

27J Principles of Statistics

Define the *normal* and *extensive* form solutions of a Bayesian statistical decision problem involving parameter Θ, random variable X, and loss function $L(\theta, a)$. How are they related? Let $R_0 = R_0(\Pi)$ be the Bayes loss of the optimal act when $\Theta \sim \Pi$ and no data can be observed. Express the Bayes risk R_1 of the optimal statistical decision rule in terms of R_0 and the joint distribution of (Θ, X).

The real parameter Θ has distribution Π, having probability density function $\pi(\cdot)$. Consider the problem of specifying a set $S \subseteq \mathbb{R}$ such that the loss when $\Theta = \theta$ is $L(\theta, S) = c\,|S| - \mathbf{1}_S(\theta)$, where $\mathbf{1}_S$ is the indicator function of S, where $c > 0$, and where $|S| = \int_S dx$. Show that the "highest density" region $S^* := \{\theta : \pi(\theta) \geqslant c\}$ supplies a Bayes act for this decision problem, and explain why $R_0(\Pi) \leqslant 0$.

For the case $\Theta \sim \mathcal{N}(\mu, \sigma^2)$, find an expression for R_0 in terms of the standard normal distribution function Φ.

Suppose now that $c = 0.5$, that $\Theta \sim \mathcal{N}(0,1)$ and that $X|\Theta \sim \mathcal{N}(\Theta, 1/9)$. Show that $R_1 < R_0$.

Paper 4, Section II

27J Principles of Statistics

Define *completeness* and *bounded completeness* of a statistic T in a statistical experiment.

Random variables X_1, X_2, X_3 are generated as $X_i = \Theta^{1/2}\, Z + (1-\Theta)^{1/2}\, Y_i$, where Z, Y_1, Y_2, Y_3 are independently standard normal $\mathcal{N}(0,1)$, and the parameter Θ takes values in $(0,1)$. What is the joint distribution of (X_1, X_2, X_3) when $\Theta = \theta$? Write down its density function, and show that a minimal sufficient statistic for Θ based on (X_1, X_2, X_3) is $T = (T_1, T_2) := (\sum_{i=1}^3 X_i^2, (\sum_{i=1}^3 X_i)^2)$.

[*Hint: You may use that if I is the $n \times n$ identity matrix and J is the $n \times n$ matrix all of whose entries are 1, then $aI + bJ$ has determinant $a^{n-1}(a+nb)$, and inverse $cI + dJ$ with $c = 1/a$, $d = -b/(a(a+nb))$.*]

What is $\mathbb{E}_\theta(T_1)$? Is T complete for Θ?

Let $S := \text{Prob}(X_1^2 \leqslant 1 \mid T)$. Show that $\mathbb{E}_\theta(S)$ is a positive constant c which does not depend on θ, but that S is not identically equal to c. Is T boundedly complete for Θ?