Paper 4, Section II

24J Principles of Statistics

Given independent and identically distributed observations $X_1, \ldots, X_n$ with finite mean $E(X_1) = \mu$ and variance $\mathrm{Var}(X_1) = \sigma^2$, explain the notion of a *bootstrap sample* $X_1^b, \ldots, X_n^b$, and discuss how you can use it to construct a confidence interval C_n for μ.

Suppose you can operate a random number generator that can simulate independent uniform random variables $U_1, \ldots, U_n$ on $[0, 1]$. How can you use such a random number generator to simulate a bootstrap sample?

Suppose that $(F_n : n \in \mathbb{N})$ and F are cumulative probability distribution functions defined on the real line, that $F_n(t) \to F(t)$ as $n \to \infty$ for every $t \in \mathbb{R}$, and that F is continuous on $\mathbb{R}$. Show that, as $n \to \infty$,

$$\sup_{t \in \mathbb{R}} |F_n(t) - F(t)| \to 0.$$

State (without proof) the theorem about the consistency of the bootstrap of the mean, and use it to give an asymptotic justification of the confidence interval C_n. That is, prove that as $n \to \infty$, $P^{\mathbb{N}}(\mu \in C_n) \to 1 - \alpha$ where $P^{\mathbb{N}}$ is the joint distribution of $X_1, X_2, \ldots$.

[You may use standard facts of stochastic convergence and the Central Limit Theorem without proof.]

Paper 3, Section II

24J Principles of Statistics

Define what it means for an estimator $\hat{\theta}$ of an unknown parameter θ to be *consistent*.

Let S_n be a sequence of random real-valued continuous functions defined on $\mathbb{R}$ such that, as $n \to \infty$, $S_n(\theta)$ converges to $S(\theta)$ in probability for every $\theta \in \mathbb{R}$, where $S : \mathbb{R} \to \mathbb{R}$ is non-random. Suppose that for some $\theta_0 \in \mathbb{R}$ and every $\varepsilon > 0$ we have

$$S(\theta_0 - \varepsilon) < 0 < S(\theta_0 + \varepsilon),$$

and that S_n has exactly one zero $\hat{\theta}_n$ for every $n \in \mathbb{N}$. Show that $\hat{\theta}_n \to^P \theta_0$ as $n \to \infty$, and deduce from this that the maximum likelihood estimator (MLE) based on observations $X_1, \ldots, X_n$ from a $N(\theta, 1)$, $\theta \in \mathbb{R}$ model is consistent.

Now consider independent observations $\mathbf{X}_1, \ldots, \mathbf{X}_n$ of bivariate normal random vectors

$$\mathbf{X}_i = (X_{1i}, X_{2i})^T \sim N_2\left[(\mu_i, \mu_i)^T, \sigma^2 I_2\right], \quad i = 1, \ldots, n,$$

where $\mu_i \in \mathbb{R}$, $\sigma > 0$ and I_2 is the 2×2 identity matrix. Find the MLE $\hat{\mu} = (\hat{\mu}_1, \ldots, \hat{\mu}_n)^T$ of $\mu = (\mu_1, \ldots, \mu_n)^T$ and show that the MLE of σ^2 equals

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n} s_i^2, \quad s_i^2 = \frac{1}{2}[(X_{1i} - \hat{\mu}_i)^2 + (X_{2i} - \hat{\mu}_i)^2].$$

Show that $\hat{\sigma}^2$ is *not* consistent for estimating σ^2. Explain briefly why the MLE fails in this model.

[You may use the Law of Large Numbers without proof.]

Paper 2, Section II

25J Principles of Statistics

Consider a random variable X arising from the binomial distribution $\mathrm{Bin}(n,\theta)$, $\theta \in \Theta = [0,1]$. Find the maximum likelihood estimator $\hat{\theta}_{MLE}$ and the Fisher information $I(\theta)$ for $\theta \in \Theta$.

Now consider the following priors on Θ:

(i) a uniform $U([0,1])$ prior on $[0,1]$,

(ii) a prior with density $\pi(\theta)$ proportional to $\sqrt{I(\theta)}$,

(iii) a $\mathrm{Beta}(\sqrt{n}/2, \sqrt{n}/2)$ prior.

Find the means $E[\theta|X]$ and modes $m_\theta|X$ of the posterior distributions corresponding to the prior distributions (i)–(iii). Which of these posterior decision rules coincide with $\hat{\theta}_{MLE}$? Which one is minimax for quadratic risk? Justify your answers.

[You may use the following properties of the $\mathrm{Beta}(a,b)$ $(a>0,\ b>0)$ distribution. Its density $f(x;a,b)$, $x \in [0,1]$, is proportional to $x^{a-1}(1-x)^{b-1}$, its mean is equal to $a/(a+b)$, and its mode is equal to

$$\frac{\max(a-1,0)}{\max(a,1)+\max(b,1)-2}$$

provided either $a>1$ or $b>1$.

You may further use the fact that a unique Bayes rule of constant risk is a unique minimax rule for that risk.]

Paper 1, Section II

25J Principles of Statistics

Consider a normally distributed random vector $X \in \mathbb{R}^p$ modelled as $X \sim N(\theta, I_p)$ where $\theta \in \mathbb{R}^p$, I_p is the $p \times p$ identity matrix, and where $p \geqslant 3$. Define the *Stein estimator* $\hat{\theta}_{STEIN}$ of θ.

Prove that $\hat{\theta}_{STEIN}$ dominates the estimator $\tilde{\theta} = X$ for the risk function induced by quadratic loss

$$\ell(a, \theta) = \sum_{i=1}^{p} (a_i - \theta_i)^2, \quad a \in \mathbb{R}^p.$$

Show however that the worst case risks coincide, that is, show that

$$\sup_{\theta \in \mathbb{R}^p} E_\theta \, \ell(X, \theta) = \sup_{\theta \in \mathbb{R}^p} E_\theta \, \ell(\hat{\theta}_{STEIN}, \theta).$$

[You may use Stein's lemma without proof, provided it is clearly stated.]