Second Year Exam - 220AB (Theory)
2019

This is a closed book and notes examination. You are to answer exactly 5 of the following 6 questions. Use your time wisely. Clearly justify each step. The 5 questions you choose to answer will be worth equal credit. If you attempt all 6 questions, your grade will be based on your 5 highest scores. Please write only on one side of each page.

1. Let $(\Omega, \mathcal{F}, P)$ be a probability space.

   (a) Let $X : \Omega \to \mathbb{R}$ be a function. Show that $\{\omega : X(\omega) < a\} \in \mathcal{F}$ for all $a \in \mathbb{R}$ is equivalent to $\{\omega : a < X(\omega) \leq b\} \in \mathcal{F}$ for all $a < b$.

   (b) Suppose $X : \Omega \to \mathbb{R}$ is measurable. Construct a sequence of simple functions X_n such that $X_n \to X$ almost surely. If you cannot find such a sequence then at least describe how simple functions are defined in a measure-theoretic setting.

   (c) State Monotone Convergence Theorem and Fatou's Lemma.

   (d) Deduce Fatou's Lemma from Monotone Convergence Theorem.

2. Let $H(x) = I(x \geq 0)(1 + x + x^2)$ where I is the indicator function.

(a) Verify that $H(x)$ is a generalized distribution function. That is, $H(x)$ is real-valued, right continuous and non-decreasing.

(b) Let $g(x) = -1$ for $x \in [0, 2]$ and $g(x) = 1$ for $x \in (2, 3]$ and $g(x) = 0$ otherwise. Compute $\int_{\mathbb{R}} g(x)dH(x)$ using the definition.

(c) Let G be non-decreasing and continuous. Use Fubini's theorem to prove the following integration by parts formula:

$$\int_{(a,b]} G(x)dH(x) = G(b)H(b) - G(a)H(a) - \int_{(a,b]} H(x)dG(x), \quad -\infty < a < b < \infty \tag{1}$$

Justify all steps.

(d) How would you revise your formula (1) if we do not assume G is continuous? Justify.

(e) Use your revised formula in part (d) to compute $\int_{(-1,2]} H(x)dH(x)$.

3. Let $Y_1, \ldots, Y_n, \ldots$ be identically distributed with mean μ, variance $\sigma^2 > 0$ and finite $E(Y_j^4)$. The sample variance is defined as $S_n = \sum_{i=1}^n (Y_i - \bar{Y}_n)^2/(n-1)$ where $\bar{Y}_n = \sum_{i=1}^n Y_i/n$. In parts (b) -(d) we assume the Y_j's are iid.

   (a) Show that $(n-1)S_n = \sum_{i=1}^n Y_i^2 - n\bar{Y}_n^2$

   (b) Show that $S_n \to \sigma^2$ almost surely as $n \to \infty$. You must explain why σ^2 is finite.

   (c) State the joint asymptotic distribution of $(\sum_{i=1}^n Y_i^2/n, \sum_{i=1}^n Y_i/n)$

   (d) Use the Delta method to find the asymptotic distribution of S_n, suitably standardized.

   (e) Instead of assuming Y_j's are iid, let us assume they arise from a moving average process. Specifically, suppose $Y_j = (Z_j + Z_{j+1})/2$ with Z_j's being iid. Show that $S_n \to c$ almost surely and identify the constant c.

4. **An example on convergence**

Consider the Borel sigma field on $\Omega = [0, 1)$ and the Lebesgue measure. For every $\omega \in \Omega$, define a sequence of random variables $X_n : \omega \to X_n(\omega)$ as $X_1(\omega) = \mathrm{I}(0 \le \omega < 1) = \mathrm{I}[0, 1)$, $X_2 = \mathrm{I}[0, 1/2)$, $X_3 = \mathrm{I}[1/2, 1)$, $X_4 = \mathrm{I}[0, 1/4)$, $X_5 = \mathrm{I}[1/4, 1/2), \dots$ In other words,

$$X_n = \mathrm{I}\big[\frac{k}{2^m}, \frac{k+1}{2^m}\big), \quad m = \lfloor \log_2 n \rfloor, \; k = n - 2^m, \quad n = 1, 2, \dots$$

where $\lfloor \cdot \rfloor$ is the floor function (e.g., $\lfloor 5.8 \rfloor = 5$).

(a) Show that $X_n \overset{p}{\to} 0$.

(b) Explain why X_n does not converge to 0 almost surely. Hence conclude that convergence in probability does not necessarily imply convergence almost surely.

(c) Recall this fact: if $Y_n \overset{p}{\to} Y$, then there exists a sub-sequence of Y_n that converges to Y almost surely. Illustrate this fact using the example here.

(d) Recall Skorokhod's representation theorem: if $Y_n \overset{d}{\to} Y$, then there exists a sequence of random variables $\{Y_n^*\}$ and Y^* on the same probability space such that $Y_n^* \overset{d}{=} Y_n$, $Y^* \overset{d}{=} Y$, and $Y_n^* \overset{a.s.}{\to} Y^*$. Illustrate this theorem using the example here. (Here we write $X \overset{d}{=} Y$ if X and Y have the same distribution)

(e) Show that $X_n \to 0$ in L_r-norm for any $r > 0$. (Recall $X_n \overset{L_r}{\to} 0$ if $E|X_n|^r \to 0$ for $r > 0$.)

(f) Give an example where L_1-convergence holds while L_2-convergence does not and explain. In other words, give an example of Y_n and Y such that $Y_n \overset{L_1}{\to} Y$ but $Y_n \overset{L_2}{\not\to} Y$.

5. **U-statistics and projection**

Given a random sample $X_1, \dots, X_n \sim F$. Define a U-statistic as

$$U = \binom{n}{r}^{-1} \sum_{1 \le j_1 < \cdots < j_r \le n} h(X_{j_1}, \dots, X_{j_r}), \tag{2}$$

which can be viewed as the sample average of $h(X_{j_1}, \dots, X_{j_r})$ over all possible permutations of indexes $1 \le j_1 < \cdots < j_r \le n$. The parameter of interest is $\theta = \mathrm{E}h(X_1, \dots, X_r)$, where h is called a kernel (a symmetric function over its inputs), and $r \in \mathbb{N}$ is called the order of the kernel. Assume $\mathrm{E}h^2(X_1, \dots, X_r) < \infty$.

(a) Choose $r = 1$ and $h(x_1) = x_1^2$. What is U and θ then?

(b) Show that U defined in (2) can be written as

$$U = \mathrm{E}\{h(X_1, \dots, X_r) \mid X_{(1)}, \dots, X_{(n)}\},$$

where $X_{(1)}, \dots, X_{(n)}$ are order statistics for $X_1, \dots, X_n$.

(c) Switch the topic to projection now - consider a general setting, where T is a random variable, $\mathcal{S}$ is a linear space of random variables (including constants, with finite second moment), and we say $\hat{S}$ is the projection of T on the space $\mathcal{S}$ if
(i) $\hat{S} \in \mathcal{S}$, (ii) $E(T) = \mathrm{E}\hat{S}$ and (iii) $\mathrm{Cov}(T - \hat{S}, S) = 0$ for every $S \in \mathcal{S}$.

Now show that if $\hat{S}$ is the projection of T on the space $\mathcal{S}$, then $\mathrm{Var}(T) \ge \mathrm{Var}(\hat{S})$. In other words, doing projection decreases the variance.

(d) Given two random variables X and Y on the same probability space. Explain why the conditional expectation $\mathrm{E}(X \mid Y)$ can be viewed as a projection based on part (c), in other words, verify the three properties of projection (i) – (iii) in part (c).

(e) Now back to part (b), we can view U as a projection of $h(X_1, \dots, X_r)$. Why is U a better estimator for θ than $h(X_1, \dots, X_r)$? Explain.

6. **Decision theory**

Consider a single observation X from a Bernoulli distribution, $\text{Bin}(1, p)$, with parameter $p \in [0, 1]$. We will use the squared error loss throughout this problem.

(a) Consider a prior $p \sim \text{Beta}(\alpha, \beta)$, then we know the Bayes estimator is

$$\delta = \frac{\alpha + X}{\alpha + \beta + 1}, \quad \alpha > 0, \beta > 0. \tag{3}$$

Find the (frequentist) risk for δ.

(b) Prove this general result: let θ be a k-dimensional parameter of interest, $\Theta \subset \mathbb{R}^k$ be the parameter space and π be a prior fully supported on Θ. If δ_π is the Bayes estimator for π with finite Bayes risk, then δ_π is admissible. (Hint: proof by contradiction)

(c) Use the result in part (b) to explain why δ defined in (3) is admissible.

(d) Is X an extended (limiting) Bayes estimator in this example? Explain.

(e) Show that when $\alpha = \beta = 1/2$, δ becomes an equalizer estimator (an estimator with constant risk, i.e., the risk does not depend on the parameter).

(f) Now prove this result: if an equalizer rule is admissible, then it is minimax.

(g) Find a minimax estimator for p.