Second Year Exam - Part I (Theory)
2018

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. Please write only on one side of each page.

1. Let $F(x) = 0.5I(x \geq 0) + 0.5I(x \geq 1) - 0.5e^{-x}I(x \geq 0)$ where I is the indicator function. Let $\mu((a,b]) = F(b) - F(a)$ for $-\infty < a < b < \infty$.

   (a) Which properties of the function F ensure that μ is a probability measure? Verify these properties.

   (b) Use Fubini's theorem to prove that for a nonnegative random variable X with distribution function G we have $EX = \int_0^\infty (1 - G(x))dx$

   (c) Use part (b) to compute the expected value of a random variable whose distribution function is given by F.

   (d) Is μ absolutely continuous with respect to Lebesgue measure? Is Lebesgue measure absolutely continuous with respect to μ? Justify.

2. Let x_j, $j = 1, 2, \ldots$, be constants such that $x_j \in [L, U]$ where $0 < L < U < \infty$. Let $Y_j = \beta x_j + \epsilon_j$ where ϵ_j are independent identically distributed random variables with mean 0 and variance $\sigma^2 \in (0, \infty)$.

(a) Denote the least squares estimator by $\hat{\beta} \equiv \sum_{j=1}^n Y_j x_j / \sum_{j=1}^n x_j^2$. Show that $\hat{\beta}$ converges to β in mean squared as $n \to \infty$.

(b) Show that $\hat{\beta}$ in part (a) is asymptotically normal by verifying Lindeberg's condition.

(c) An alternative estimator of β is $\tilde{\beta} \equiv \sum_{j=1}^n Y_j / \sum_{j=1}^n x_j$. Show that this estimator also converges to β in mean squared but for every n it has a variance at least as large as that of $\hat{\beta}$.

(d) In this part assume x_j are i.i.d. random variables bounded between L and U (rather than constants). Show that the least squares estimator $\hat{\beta}$ converges to β almost surely.

3. (a) Employing characteristic functions, show that if $\lambda > 0$ is a constant and X_n has a binomial$(n, \lambda/n)$ distribution then $X_n \xrightarrow{D} \text{Po}(\lambda)$ as $n \to \infty$. Note that the Po(λ) distribution has pmf $\lambda^k e^{-\lambda}/k!$ for $k = 0, 1, \ldots$ and its mean and variance are both λ.

(b) Give an example of a sequence of random variables X_n such that $X_n \to \text{Po}(\lambda)$ in distribution ($\lambda \in (0, \infty)$) but EX_n does not converge to λ.

(c) Prove that $nP(X > n) \to 0$ as $n \to \infty$ if X is a random variable with $E \max\{X, 0\} < \infty$.

(d) Give an example of a real-valued random variable X such that $nP(X > n)$ does not tend to zero as $n \to \infty$.

4. **Subindependence**

Consider two random variables X and Y, and define their characteristic functions (CF) by $\phi_X(t)$ and $\phi_Y(t)$, respectively. We say X and Y are subindependent if

$$\phi_{X+Y}(t) = \phi_X(t)\phi_Y(t),$$

in other words, CF of $X+Y$ equals to the product of CFs of X and Y.

(a) It is clear that independence implies subindependence. However, subindependence does not necessarily imply independence. Let X follow the standard Cauchy distribution, with CF $\phi_X(t) = \exp(-|t|)$. Show that X and $Y = X$ (itself) are subindependent but not independent.

(b) Still let X follow the standard Cauchy distribution. Show that X and $-X$ are not subindependent.

(c) Recall by taking the derivatives of CF, one can obtain the moments of a random variable. In particular, if the k-th moment of X exists, then

$$E(X^k) = (-i)^k \frac{\partial^k \phi_X(t)}{\partial t^k} |_{t=0}, \quad k = 1, 2, \ldots$$

Use this fact to show that subindependence implies no correlation, i.e., if X and Y are subindependent, then $\mathrm{Cov}(X,Y) = 0$ (assuming the second moments of X and Y exist. Hint: consider $\mathrm{E}(X+Y)^2$).

(d) Independence plays a key role in statistics and probability. For example, CLT may not necessarily hold without the independence assumption. Now consider $X_1, \ldots, X_n$ being identically distributed. If we don't assume independence, then we can choose $X_k = S$ for odd number k and $X_k = -S$ for even number k, where S follows some symmetric distribution with mean 0 and variance 1. Show that

$$\frac{\sum_{i=1}^n X_i}{\sqrt{n}} \xrightarrow{D} 0.$$

In other words, CLT does not hold here.

(e) Now explain intuitively, why CLT may still be valid if we replace the independence assumption by subindependence.

5. **Decision theory**

Suppose that we have a single observation X from a Bernoulli distribution with success probability parameter θ. Let θ only take two possible values, $\{0.3, 0.6\}$. For any estimator δ, define a loss function as $L(\theta, \delta) = \mathrm{I}(\theta \neq \delta)$.

(a) Consider three estimators,

$$\delta_1(X) = 0.3, \quad \delta_2(X) = 0.6, \quad \delta_3(X) = 0.3\mathrm{I}(X = 0) + 0.6\mathrm{I}(X = 1).$$

Show that their risks are:

$$R(\theta, \delta_1) = \mathrm{I}(\theta = 0.6), \quad R(\theta, \delta_2) = \mathrm{I}(\theta = 0.3),$$
$$R(\theta, \delta_3) = 0.3\mathrm{I}(\theta = 0.3) + 0.4\mathrm{I}(\theta = .6).$$

(b) Since the parameter space is discrete with just two points, we can plot the risk vector $(R(\theta = 0.3, \delta), R(\theta = 0.6, \delta))$ of the estimator δ and form a risk set. Note that in our case, the risk set is the triangle area with nodes $(0, 1)$, $(1, 0)$, $(0.3, 0.4)$. Explain why the risk set is always a convex set.

(c) Mark all the admissible estimators on the plot of the risk set.

(d) Find a prior distribution such that the corresponding Bayes estimator is not unique. Explain. (I will need the mathematical expression for the prior distribution, marking on the plot is not good enough.)

(e) Show that the minimax estimator has the risk vector of $(\frac{4}{11}, \frac{4}{11})$. Is it a Bayes estimator?

(f) Prove this result in general (not under the specific setting of this problem, but for general cases): if a minimax estimator is unique, then it is admissible. Then use this result to determine if the minimax estimator you find in part (e) is admissible.

(g) Now prove this result in general, if the parameter space is finite, i.e., $\Theta = \{\theta_1, \ldots, \theta_k\}$, and a prior π is positive on Θ. Then the Bayes estimator under π is admissible. Explain why the condition π being positive is required.

6. MLE

Consider i.i.d. observations $X_1, \ldots, X_n$ from some distribution with density function $f_\theta(x)$ indexed by a parameter $\theta \in \Theta \subset \mathbb{R}$. Consider an M-estimator $\hat{\theta}_n$ that maximizes a function of type

$$M_n(\theta) = \frac{1}{n}\sum_{i=1}^{n} U_\theta(X_i),$$

where U_θ is a known function of X. For example, one can choose $U_\theta(x) = \log f(x)$ and then obtain the MLE as the maximizer of $M_n(\theta)$.

(a) What are the M-estimators for the choice of $U_\theta(x) = -(x-\theta)^2$ and $-|x-\theta|$? What about $U_\theta(x) = -(1-p)(x-\theta)^- - p(x-\theta)^+$ for $0 < p < 1$? Explain.

Here is a theorem (Theorem 5.7 from "Asymptotic Statistics") that is useful for proving consistency of MLE.

Theorem: Let M_n be random functions and let M be a fixed function of θ such that for every $\epsilon > 0$,

$$\sup_{\theta \in \Theta} |M_n(\theta) - M(\theta)| \xrightarrow{P} 0, \tag{1}$$

$$\sup_{\theta : |\theta - \theta_0| \geq \epsilon} M(\theta) < M(\theta_0). \tag{2}$$

Then any sequence of estimators $\hat{\theta}_n$ with $M_n(\hat{\theta}_n) \geq M_n(\theta_0) - o_p(1)$ converges in probability to θ_0.

The following questions are based on this theorem.

(b) Write down the definition of $o_p(1)$, and prove $o_p(1) + o_p(1) = o_p(1)$ (you can use any theorems we have discussed in the class).

(c) Prove the **Theorem**. (Hint: first show $M(\theta_0) - M(\hat{\theta}_n) \leq o_p(1)$, then combine this with condition (2).)

(d) For condition (1), explain why the uniform convergence "$\sup_{\theta \in \Theta}$" is needed based on part (c). Can we obtain (1) by law of large numbers? Explain.

(e) Give an example of M where (2) is not satisfied (drawing a picture of M will be good enough). Assuming Θ is a closed interval, and $M : \Theta \to \mathbb{R}$ is a continuous function. Show that if θ_0 is the unique global maximum of M, then (2) holds. Discuss how you would relax the continuity assumption of M.