2016 Second-Year Exam
Statistics 220A
Statistics 220B

September 15, 2016 10:00 – 2:00

Instructions: This is a CLOSED BOOK and CLOSED NOTES exam. You should attempt 2 of the 3 problems for 220A and 2 of the 3 problems for 220B. You MAY NOT turn in any more than 2 solutions for each exam.

Statistics 220A 2016 Second-Year Exam

September 15, 2016

Instructions: This is a CLOSED BOOK and CLOSED NOTES exam. You should attempt 2 of the 3 problems only. You MAY NOT turn in 3 solutions Do Well!!

1 220A Questions

1. Let $X_1, X_2, ..., X_n \mid G(\cdot) \overset{iid}{\sim} P^X(\cdot) \equiv G(\cdot)$, where $G(\cdot)$ is an unknown probability measure defined on $(\mathcal{R}, \mathcal{B})$. So $(\mathcal{R}, \mathcal{B}, G)$ is a probability space. The X_is are Borel measurable. We assume a partial prior specification for $G(\cdot)$. Let $(A_1, A_2, ..., A_k)$ be a partition of $\mathcal{R}$. Let $\alpha(\cdot)$ be a measure on $(\mathcal{R}, \mathcal{B})$ that is absolutely continuous with respect to Lebesgue measure. Then assume that $(G(A_1), G(A_2), ..., G(A_k)) \sim$ Dirichlet$((\alpha(A_1), \alpha(A_2), ..., \alpha(A_k))$. We assume this same result holds for all possible partitions of all possible dimensions. These are called finite dimensional distributions. Let $\alpha(\mathcal{R}) \equiv w$ and recall that $G(A) \sim$ Beta$(\alpha(A), \alpha(A^c))$; thus $E(G(A)) = \alpha(A)/w \equiv \bar{\alpha}(A)$, $\forall A \in \mathcal{B}$. Observe that $\bar{\alpha}(\cdot)$ is a probability measure and recall that we assume $\bar{\alpha} \ll$ Lebesgue Measure, so $\bar{\alpha}$ is zero on sets of LM zero.

   Due to Ferguson (1973), there exists a unique probability measure on the space of probability distributions (properly defined with a suitable metric) that has these finite dimensional distributions. We say that

   $$G \sim DP(w, \alpha)$$

   where DP means Dirichlet Process. Let $\mathcal{P}(\cdot)$ be this unique probability measure. Then $\mathcal{P}((G(A_1), ...G(A_k)) \in B) = \int_B f_k(x)dx$ where $f_k(x)$ is the pdf of Dirichlet$(\alpha(A_1), ..., \alpha(A_k))$, and $B \in \mathcal{B}(C_k)$ where C_k is the set of all vectors of probabilities of dimension k that add to one, namely the $k-1$ dimensional simplex. In particular, with $k = 2$, $\mathcal{P}(G(A) \in B) = \int_B f_1(x)dx$, where f_1 is a Beta$(\alpha(A), \alpha(A^c))$ pdf.

   (a) Prove that the marginal distribution of X_1 is $\bar{\alpha}(\cdot)$, namely that

   $$Pr(X_1 \in A) = \bar{\alpha}(A)$$

   for all $A \in \mathcal{B}$. By marginal probability, we mean that we have integrated G out appropriately.

   Now assume that it is known that

   $$(*)\, G(A) \mid X_1 = x_1 \sim \text{Beta}(\alpha(A) + \delta_{x_1}(A), \alpha(A^c) + \delta_{x_1}(A^c))$$

where $\delta_x(A)$ is one if $x \in A$ and zero otherwise. This is called the dirac delta function.

(b) Using (*) and part (a), what is the marginal distribution of $X_2 \mid X_1 = x_1$. Describe (characterize) the distribution carefully. Be sure to give the precise formula for

$$Pr(X_2 \in A \mid X_1 = x_1)$$

(c) Obtain the marginal probability $Pr(X_1 = X_2) \equiv q$, which will be a function of w. Hint: First obtain $Pr(X_1 = X_2 \mid X_1 = x_1)$

(d) Now derive the precise formula for the marginal probability

$$Pr(X_1 \in B_1, X_2 \in B_2)$$

(e) Using the results in (c-d), give and characterize the conditional distributions for $(X_1, X_2) \mid \{X_1 \neq X_2\}$ and for $(X_1, X_2) \mid \{X_1 = X_2\}$.

Let $\nu = \nu_1 + \nu_2$ where ν_1 is Lebesgue Measure on $\mathcal{R}^2$ and ν_2 is LM on the set $\{(x_1, x_2) \in \mathcal{R}^2 : x_1 = x_2, \}$.

(f) Argue that the joint marginal probability measure for (X_1, X_2) is absolutely continuous with respect to ν.

(g) Obtain the RN derivative of this joint distribution with respect to ν.

(h) Give the integral equation that must hold (using product sets) for this to be the RN derivative.

2. Let $(S, \mathcal{A}, P)$ be a probability space on which variables below are appropriately defined.

   (a) State and prove the First Borel-Cantelli Lemma in this context.

   (b) Suppose it is known that

   $$Pr\left(|X_n| > \frac{1}{n}\right) < \frac{1}{2^n} \qquad \forall n.$$

   What can be said about convergence properties of X_n? Justify Carefully.

   Let $X : (S, \mathcal{A}) \to (\mathcal{R}, \mathcal{B})$ and let $Y = g(X)$ where $g : (\mathcal{R}, \mathcal{B}) \to (\mathcal{R}, \mathcal{B})$.

   (c) Prove that $Y : (S, \mathcal{A}) \to (\mathcal{R}, \mathcal{B})$, that it, Y is measurable with respect to the original σ-field. Hint: Regard $Y(\omega)$ as say $h(\omega) = g(X(\omega))$.

   (d) (i) Prove that, for any $B \in \mathcal{B}$, $h^{-1}(B)$ and $h^{-1}(B^c)$ are non-overlapping sets in $\mathcal{B}$. (ii) Prove that $h^{-1}(\bigcup_{i=1}^{\infty} B_i) = \bigcup_{i=1}^{\infty} h^{-1}(B_i)$ for $B_i \in \mathcal{B}$. (iii) Prove that $h^{-1}(\mathcal{B})$ is a sub-$\sigma-$field of $\mathcal{A}$. (iv) Give an example of when it would be a proper sub-σ-field. Justify.

   (e) Prove that $\int g(x(\omega))dP(\omega) = \int g(x)dP^X(x)$ for $g \geq 0$.

   Now let $\mathcal{C}$ be a sub-$\sigma-$field of $\mathcal{A}$.

   (f) Explain precisely how to define $P(A \mid \mathcal{C})$ and justify its existence.

   (g) Prove the law of total probability for obtaining $P(A)$ by conditioning on $\mathcal{C}$.

3. Let $(S, \mathcal{A}, P)$ be a probability space on which all objects below are well defined.

   (a) Explain what it means that P is a probability measure, eg. what are the defining properties that make P a p.m.?

   (b) Using the fact that P is a p.m., prove that $P(A_1 \bigcup A_2) \leq P(A_1) + P(A_2)$.

   (c) Prove that $P(\bigcup_{i=1}^{n} A_i) \leq \sum_{i=1}^{n} P(A_i)$, for all n.

   (d) Using first principles, prove that if $B_n \nearrow B$, then $P(B_n) \to P(B)$ as $n \to \infty$. You are asked here to prove continuity of probability using basic properties of probability.

   (e) Prove that $P(\bigcup_{i=1}^{\infty} A_i) \leq \sum_{i=1}^{\infty} P(A_i)$

   (f) Consider a sequence of RVs $\{X_1, X_2, ...\}$. Then prove that if, for any subsequence of $\{1, 2, ...\}$, say $\{k_1, k_2, ...\}$, there exists a further subsequence $\{r_1, r_2, ...\}$ such that $X_{r_j} \stackrel{a.s.}{\to} X$ as $j \to \infty$, then $X_j \stackrel{P}{\to} X$ as $j \to \infty$.

   (g) Prove that if $X_n \stackrel{P}{\to} X$ as $n \to \infty$, then $g(X_n) \stackrel{P}{\to} g(X)$ as $n \to \infty$ for continuous $g(\cdot)$.

   (h) Let $Y_n \stackrel{P}{\to} 0$, and assume that X_n is bounded in probability. Then using first principles, prove that $X_n Y_n \stackrel{P}{\to} 0$ as $n \to \infty$. You may not use Slutsky's Theorem, you are asked to prove the simplest form of it starting with the definition of convergence in probability and using the bounded in probability property that is specified.

Statistics 220B 2016 Second-Year Exam

September 15, 2016

Instructions: This is a CLOSED BOOK and CLOSED NOTES exam. You should attempt 2 of the 3 problems only. You MAY NOT turn in 3 solutions Do Well!!

220B: Do only 2 of the following 3 problems.

1. Asymptotic theory

   (a) True or False: for each statement, either prove it or give a counter-example and justify.

      i. Maximum likelihood estimator (MLE) always exists.

      ii. Characteristic function (CF) always exists.

      iii. For any sequence of i.i.d. random variables, law of large numbers is applicable.

      iv. If ϕ is a valid CF, so is $|\phi|^2$.

   (b) Consider a one-way random effects model

   $$Y_{ij} = \mu + \alpha_i + \epsilon_{ij}, \quad \alpha_i \overset{\text{i.i.d}}{\sim} N(0, \sigma_\alpha^2), \; \epsilon_{ij} \overset{\text{i.i.d}}{\sim} N(0, \sigma^2), \quad \alpha_i \perp \epsilon_{ij}.$$

   Now consider testing $H_0 : \sigma_\alpha^2 = 0$. Explain why the standard chi-square approximation result for likelihood ratio test does not directly apply.

   For the rest of this question, consider a random sample $X_1, \ldots, X_n$ from Bernoulli(p), where $p \in (0, 1)$ and $n \geq 2$. Let $\theta = p(1-p)$. It's obvious (you don't have to prove it) that the MLE of θ is given by $T_n = \bar{X}_n(1 - \bar{X}_n)$, where $\bar{X}_n = \sum_{i=1}^n X_i/n$.

   (c) Assume $p \neq 1/2$. Find the limiting distribution of properly normalized T_n. In other words, find a sequence of numbers $a_n > 0$ and some function $h(\cdot)$ such that $a_n(T_n - h(p))$ converges weakly to a non-degenerate distribution.

   (d) Now do part (c) for $p = 1/2$.

   (e) Find the uniformly minimum-variance unbiased estimator (UMVUE) for θ (you don't need to prove it). Denote it by S_n.

   (f) Assume $p \neq 1/2$. Show that $a_n(S_n - T_n) \xrightarrow{L} 0$ for the same a_n in part (c). Use this result to find the limiting distribution of properly normalized S_n.

   (g) Assume $p = 1/2$. Find the limiting distribution of properly normalized S_n. Does the argument in part (f) still work?

2. Decision theory

(a) Consider a random sample $X_1, \ldots, X_n$ from some distribution P_θ, where θ is a parameter, with mean θ and finite second moment. Let $\bar{X} = \sum_{i=1}^n X_i/n$. Consider an estimator $\delta = a\bar{X} + b$, where a and b are two constants. Find the (freq) risk of δ assuming squared error loss.

(b) Show that δ is inadmissible if
(i) $a > 1$, (ii) $a < 0$, (iii) $a = 1, b \neq 0$.

(c) Let $n = 1$ and P_θ be the normal distribution $N(\theta, 1)$. Assume $a \in (0, 1)$. Show that δ is a Bayes estimator of θ with respect to some prior Π. Find Π. The following fact maybe helpful:

Consider a normal model with fixed variance $X \sim N(\mu, \sigma^2)$, and a prior $\mu \sim N(\mu_0, \sigma_0^2)$, the posterior of $\mu \mid X = x$ is

$$N\left(\frac{\mu_0\sigma_0^{-2} + x}{\sigma^{-2} + \sigma_0^{-2}}, \frac{1}{\sigma^{-2} + \sigma_0^{-2}}\right)$$

(d) Find the Bayes risk of δ under the prior Π in part (c).

(e) Consider $a = 0$. Find the values of b such that $\delta = b$ is an admissible estimator of θ.

(f) Now characterize the class of all admissible linear estimators of the form $aX + b$ for θ.

(g) Is $\delta = X$ minimax? Explain. (You can use theorems that we covered in class/notes.)

(h) The following statements are incorrect. Give a counter-example and briefly explain.

i. Bayes rule is admissible.

ii. Let $\boldsymbol{X} \sim N(\boldsymbol{\mu}, \boldsymbol{I}_{d\times d})$ for $d \geq 1$. Then $\boldsymbol{X}$ is admissible for $\boldsymbol{\mu}$ under squared error loss.

3. MCMC theory and nonparametrics

For parts (a)–(d), consider a Markov Chain $\{\theta^l, l \in \mathrm{N}\}$, a transition kernel $q(\cdot \mid \cdot) : \Theta \times \Theta \to [0, \infty)$, a finite measure μ, and an invariant density satisfying

$$p(\theta) = \int_\Theta p(\theta')q(\theta \mid \theta')d\mu(\theta').$$

Recall the definition of detailed balance condition (DBC) for q with invariant function p,

$$q(x|y)p(y) = q(y|x)p(x), \qquad \forall x, y \in \Theta.$$

Recall the definition of L_1-distance between two functions p_1 and p_2,

$$\|p_1 - p_2\|_1 = \int_\Theta |p_1(\theta) - p_2(\theta)|\, d\mu(\theta).$$

(a) Define a transition kernel at step k by

$$q_k(\theta \mid \theta') = \int_\Theta q(\theta \mid \tilde{\theta})q_{k-1}(\tilde{\theta} \mid \theta')d\mu(\tilde{\theta}),\ k = 1, 2, \ldots,$$

with $q_0(\tilde{\theta} \mid \theta') = I(\theta' = \tilde{\theta})/\mu(\Theta)$ and $q_1 = q$. Show that q_k has the invariant density p for $k = 1, 2, \ldots$.

(b) Show that for every $\theta \in \Theta$, $\|q_k(\cdot \mid \theta) - p(\cdot)\|_1$ is decreasing in k.

(c) Show that if q satisfies DBC with p, so does q_k for $k = 1, 2, \ldots$.

(d) What does ergodic theorem and DBC mean in MCMC context? Why would a Bayesian care about them? Explain in words.

(e) Give three examples of U-statistics. (you don't need to explain)

(f) Given i.i.d $X_1, \ldots, X_n$ from a continuous density function f on the real line. Consider a kernel estimator

$$\hat{f}_n(x) = \frac{1}{n}\sum_{i=1}^n \frac{1}{h}K\left(\frac{x - X_i}{h}\right),$$

where K is the standard normal kernel and h is the bandwidth. Assume that $nh \to \infty$ as $n \to \infty$. Use convolution (you can use other methods) to show that if $\hat{f}_n(x) \xrightarrow{p} f(x)$ for every $x \in \mathrm{R}$, then $h \to 0$.