Homework Assignment #2 for Prob-Stats, Fall 2018
Due date: Monday, October 22, 2018

Topics: consistent estimators; sub-σ-fields and partial observations; Doob's theorem about sub-σ-field measurability; conditional expectations as mean-squared loss optimal estimators; the Bridge-Crossing Lemma and conditional distributions; Jensen's inequality and expected utility maximization.

Readings: Bierens: Ch. 3.1-2, 3.4-5. CB: Ch. 2; Ch. 8.1-3; and Ch. 10.1. CSZ: Ch. 6.6; Ch. 7.6; Ch. 8.4.

A. Let X be a measurable function taking values in $\mathbb{R}$. $E\,X1_{X>0} = 0$ if and only if $P(X > 0) = 0$. [Despite its simplicity, this will be a valuable factoid for working with conditional distributions.]

B. This problem will take you through several aspects and applications of the probability integral transform. The end goal of this material, in a later problem set, is Skorohod's representation theorem. A note: there are two results having that name, one involves stopping times and Brownian motions; the one we're looking at shows that you can replace a weaker kind of convergence of random variables with almost everywhere convergence.
   1. CB, Exercise 2.10.
   2. We say that the random variable X **first order stochastically dominates** the random variable Y if for all $r \in \mathbb{R}$, $P(X > r) \geq P(Y > r)$. Suppose that X FOSDs Y, let $X' : (0,1) \to \mathbb{R}$ and $Y' : (0,1) \to \mathbb{R}$ denote the probability integral transform of X and Y. Show that $X'(r) \geq Y'(r)$ for all $r \in (0,1)$.
   3. X FOSDs Y if and only if $\int u(X(\omega))\,dP(\omega) \geq \int u(Y(\omega))\,dP(\omega)$ for all bounded, non-decreasing $u : \mathbb{R} \to \mathbb{R}$.
   4. If $X_n \to X$ almost everywhere, then for all bounded continuous $f : \mathbb{R} \to \mathbb{R}$, $\int f(X_n(\omega))\,dP(\omega) \to \int f(X(\omega))\,dP(\omega)$.
   5. If for all bounded continuous $f : \mathbb{R} \to \mathbb{R}$, $\int f(X_n(\omega))\,dP(\omega) \to \int f(X(\omega))\,dP(\omega)$, then for all closed $F \subset \mathbb{R}$, $\limsup_n P(X_n \in F) \leq P(X \in F)$. [Hint: the functions $f(r) = \max\{0, 1 - n \cdot d(r, F)\}$ are continuous and bounded; what do you learn from integrating them?]
   6. If for all closed interval $F_r := (-\infty, r]$, $\limsup_n P(X_n \in F_r) \leq P(X \in F_r)$, then the probability integral transforms of the X_n converge almost everywhere to the probability integral transform of X.

C. The reading by Breiman et al. (1964) shows that a set of probabilities Π on an observation space is **strongly** 0-1 if and only if there exists a consistent sequence of estimators for the $p \in \Pi$. This problem takes you through the most basic of such situtations.

   The space of observations for this problem is $\mathfrak{X} := \{0,1\}^{\mathbb{N}}$. For each $\theta \in \Theta = [0,1]$, let p_θ be the distribution on $\mathfrak{X}$ defined by $\{\text{proj}_n(\cdot) : n \in \mathbb{N}\}$ is an iid collection of Bernoulli(θ) distributions. In case it is needed for later notational purposes, let $Y_n(x) = \text{proj}_n(x)$. Let $\mathcal{X}_n$ be the smallest σ-field on $\{0,1\}^{\mathbb{N}}$ making each proj_m measurable, $m \leq n$. Let $\mathcal{X}$ be the smallest σ-field containing all of the $\mathcal{X}_n$.

1. Show that the functions

   $x \mapsto \liminf_n \frac{1}{n} \sum_{i \leq n} \text{proj}_i(x)$ and $x \mapsto \limsup_n \frac{1}{n} \sum_{i \leq n} \text{proj}_i(x)$

   are $\mathcal{X}$-measurable.
2. Show that for each $\theta \in [0,1]$, the set

   $$E_\theta := \{x \in \mathfrak{X} : \lim_n \tfrac{1}{n} \textstyle\sum_{i \leq n} \text{proj}_i(X) = \theta$$

   is measurable.
3. Show that the set $E = \{x \in \{0,1\}^{\mathbb{N}} : \lim_n \frac{1}{n} \sum_{i \leq n} \text{proj}_i(X) \text{ exists}\}$ is measurable.
4. Show that for each $\theta \in \Theta$, $p_\theta(E_\theta) = 1$.
5. Show that $\widehat{\theta}_n(Y_1, \ldots, Y_n) = \frac{1}{n} \sum_{i \leq n} Y_i$ is a consistent sequence of estimators for θ.
6. A set of probabilities Π on a measure space of observations, $(X, \mathcal{X})$, is **strongly zero-one** if there exists a measurable $E \subset X$ and an onto $\Phi : E \to \Pi$ such that for all $p \in \Pi$, $p(\Phi^{-1}(p)) = 1$. Show that $\Pi = \{p_\theta : \theta \in \Theta\}$ is strongly zero-one. [A consistent sequence of estimators for the elements of Π exists if and only if it is strongly zero-one.]

D. [The space of observations as a metric space] On $\mathfrak{X} = \{0,1\}^{\mathbb{N}}$, consider the metric $d(x,y) = \sum_n |\text{proj}_n(x) - \text{proj}_n(y)|/2^n$. We use the σ-fields $\mathcal{X}_n$ and $\mathcal{X}$ from the previous problem.

1. Let x^α be a sequence in X. Show that $d(x^\alpha, x) \to 0$ if and only if $(\forall M \in \mathbb{N})(\exists A)(\forall \alpha \geq A)(\forall m \leq M)[\text{proj}_m(x^\alpha) = \text{proj}_m(x)]$.
2. Show that $(\mathfrak{X}, d)$ is a compact metric space, hence is both complete and separable. [A metric space is said to be separable if it has a countable dense subset.]
3. Characterize the compact subsets of $\mathfrak{X}$.
4. Show that $f : \mathfrak{X} \to \mathbb{R}$ is continuous if and only if for each $\epsilon > 0$, there exists an $M \in \mathbb{N}$ and an $\mathcal{X}_M$-measurable function g such that $\max_{x \in \mathfrak{X}} |f(x) - g(x)| < \epsilon$. [The function g is called finitely-determined, so this is saying that all continuous functions on $\mathfrak{X}$ are nearly finitely-determined.]
5. Show that $\mathcal{X}$ is the Borel σ-field on $(\mathfrak{X}, d)$.
6. Let $\mathcal{X}^\circ = \cup_n \mathcal{X}_n$. Show that this is the smallest field containing all of the $\mathcal{X}_n$, and that it is <u>not</u> a σ-field.
7. Show that every p_θ is countably additive on the field $\mathcal{X}^\circ$, that is, show that for every $E_n \downarrow \emptyset$ in $\mathcal{X}^\circ$, $p_\theta(E_n) \downarrow 0$.

Comments: Carathéodory's extension theorem tells us that if p is a countably additive probability on a field $\mathcal{F}^\circ$ and $\mathcal{F} = \sigma(\mathcal{F}^\circ)$, the p is a unique countably additive extension to $\mathcal{F}$.

E. Let (M,d) be a separable metric space with Borel σ-field $\mathcal{M}$, and let $(\Omega,\mathcal{F},P)$ be our usual probability space, i.e. $\Omega \neq \emptyset$, $\mathcal{F}$ a σ-field of subsets of Ω, and $P : \mathcal{F} \to [0,1]$ a countably additive probability.

1. Suppose that $X : \Omega \to M$ is measurable and p is the image law of X, i.e. $p(A) = P(X \in A)$ for all $A \in \mathcal{M}$. Show that p is countably additive.
2. If p is a countably additive probability on the Borel σ-field, $\mathcal{M}$, then for all $E \in \mathcal{M}$, $p(E) = \sup\{p(F) : F \subset E, F \text{ closed}\}$. [This is a good sets argument.]
3. If p and q are countably additive probabilities on the Borel σ-field, then $p = q$ if and only if $p(F) = q(F)$ for all closed F.
4. We say that a probability p is **tight** if for every $\epsilon > 0$, there exists a compact K with $p(K) > (1-\epsilon)$. If p is tight and p and q are countably additive probabilities on the Borel σ-field, then $p = q$ if and only if $p(K) = q(K)$ for all compact K.
5. If $X : \Omega \to \mathbb{R}$ is a measurable random variable and $p(A) = P(X \in A)$ is the distribution of X, then p is tight.
6. If (M,d) is a separable metric space and p and q are countably additive probabilities on the Borel σ-field, then $p = q$ if and only if $\int_M f(x)\,dp(x) = \int_M f(x)\,dq(x)$ for all bounded, Lipschitz continuous $f : M \to \mathbb{R}$.

F. [The Neyman-Pearson Lemma] Suppose that $X_1,\dots,X_n \in \{0,1\}^n$ has the Bernoulli density $f(\mathbf{x}|\theta)$ for some $\theta \in \Theta = \{\theta_0,\theta_1\}$. Ahead of time, we assign probability $\beta \in (0,1)$ to θ_0 and probability $1-\beta$ to θ_1.

After observing the data, we are going to take an action $a \in \{0,1\}$ to solve

$$\max_{a\in\{0,1\}} E\left(u(a,\theta)|X_1,\dots,X_n\right)$$

where if $\theta = \theta_0$, then $a = 0$ is the better choice and if if $\theta = \theta_1$, then $a = 1$ is the better choice. Specifically, we are going to suppose that the utilities are

$a_\downarrow\ \theta \to$	θ_0	θ_1
$a = 0$	$u(0,0)$	$u(0,1)$
$a = 1$	$u(1,0)$	$u(1,1)$

where $u(0,0) > u(1,0)$ and $u(1,1) > u(1,0)$.

1. Show that for any function $f(\theta)$ and probability q of θ_0, the set of solutions to the following two problems are the same,

   $$\max_{a\in\{0,1\}} u(a,\theta) \text{ and } \max_{a\in\{0,1\}}[u(a,\theta)+f(\theta)].$$

2. Let $f(\theta_0) = -u(1,0)$ and $f(\theta_1) = -u(0,1)$ and rewrite the payoff matrix above in the form

$a_\downarrow\ \theta \to$	θ_0	θ_1
$a = 0$	r	0
$a = 1$	0	s

   where $r, s > 0$.

3. Give the set of data realizations, $\mathbf{x} \in \{0,1\}^n$, for which the optimal decision is $a = 0$ and for which it is $a = 1$.
4. Express your regions above in terms of likelihood ratios.
5. Show that for large n, the effect of β "washes out."
6. Compare the optimal regions with $\beta = \frac{1}{2}$ to the Neyman-Pearson lemma's probability of Type I and Type II error.

G. [Bernoulli estimations with missing data] Suppose that $(Y_i, Z_i) \in \{0,1\} \times \{0,1\}$, $i = 1, \ldots, n$, are iid. Let $\theta_Y = E\,Y_i$, $\theta_1 = E\,(Y_i|Z_i = 1)$, $\theta_0 = E\,(Y_i|Z_i = 0)$, and $\theta_Z = E\,Z_i$. Assume that all four θ's belong to the open interval $(0,1)$.
1. Assuming that the entire vector (Y_i, Z_i) is observable, give consistent, unbiased estimators of the θ's and show that they are consistent and unbiased.
2. Now suppose that when $Z_i = 0$, the value of Y_i cannot be observed. That is, suppose that what is observed is X_i defined, for some constant g, as
$$X_i = \begin{cases} (Y_i, 1) & \text{if } Z_i = 1 \\ (g, 0) & \text{if } Z_i = 0. \end{cases}$$
Your previous consistent unbiased estimators of θ_1 and θ_Z are still consistent and unbiased. Supposing that there is enough data to pin down θ_1 and θ_Z. Give the range of possible values of θ_Y as a function of the other three θ's. Explain the patterns of dependence of the range.
3. Suppose now that we know that $|\theta_1 - \theta_0| \leq b$ for some $0 \leq b < 2$. Give the range of possible values of θ_Y as a function of b and the other three θ's.
4. Suppose now that the Y_i and Z_i are independent. How does this change the previous answer?

H. [L_2-best estimators] Let $L_2 = \{X \in L_0 : \int X^2(\omega)\, dP(\omega) < \infty\}$. For any $X, Y \in L_2$, define $\langle X, Y \rangle = \int X(\omega)Y(\omega)\, dP(\omega)$. For any $X \in L_2$, define $\|X\|_2 = \sqrt{\langle X, X \rangle}$, and for any $X, Y \in L_2$, define $d(X,Y) = \|X - Y\|_2$.
1. Solve $\min_r \|Y - r\|_2$.
2. If $Y = \sum_{n \leq N} \gamma_n 1_{A_n}$ and $X = \sum_{m \leq M} \beta_n 1_{B_n}$, solve $\min_g \|Y - g(X)\|_2$ where $g : \mathbb{R} \to \mathbb{R}$. The random variable $g(X)$ is called "the expectation of Y conditional on X." It is denoted $E\,(Y|X)$ or $E\,(Y|\sigma(X))$.
3. If $Y \in L_2$ and $X = \sum_{m \leq M} \beta_n 1_{B_n}$, solve $\min_g \|Y - g(X)\|_2$ where $g : \mathbb{R} \to \mathbb{R}$. The random variable $g(X)$ is called "the expectation of Y conditional on X." It is denoted $E\,(Y|X)$ or $E\,(Y|\sigma(X))$.
4. If $\|Y_n - Y\|_2 \to 0$, and $X = \sum_{m \leq M} \beta_n 1_{B_n}$, then $\|E\,(Y_n|X) - E\,(Y|X)\|_2 \to 0$.
5. For $X, Y \in L_2$ and X_n a sequence of simple functions in L_2 with $\|X_n - X\|_2 \to 0$, show that $E\,(Y|X_n)$ is a Cauchy sequence in L_2. Its limit is denoted $E\,(Y|X)$.
6. Bierens 3.6.10 (p. 83).

I. Random variables X, Y are **independent**, denoted $X \perp\!\!\!\perp Y$ if for all measurable A, B, $P((X \in A) \cap (Y \in B)) = P(X \in A) \cdot P(Y \in B)$. If X, Y are in L_2, they are **orthogonal**, denoted $X \perp Y$, if $\langle X, Y \rangle = 0$. Independence can be understood as a very strong form of independence.
   1. Suppose that $X, Y \in L_2$ and $X \perp Y$. Solve the problem

   $$\min_{\alpha,\beta \in \mathbb{R}} \|Y - (\alpha + \beta X)\|_2.$$

   2. For $X, Y \in L_2$, $\langle X - E\,X, Y - E\,Y \rangle = 0$ if and only if $E\,XY = (E\,X) \cdot (E\,Y)$.
   3. If X and Y are simple, show that $X \perp\!\!\!\perp Y$ if and only if for all bounded (measurable) $f, g : \mathbb{R} \to \mathbb{R}$, $\langle f(X) - E\,f(X), g(Y) - E\,g(Y) \rangle = 0$.
   4. If X and Y belong to L_2, show that $X \perp\!\!\!\perp Y$ if and only if for all bounded measurable $f, g : \mathbb{R} \to \mathbb{R}$, $\langle f(X) - E\,f(X), g(Y) - E\,g(Y) \rangle = 0$.
   5. If $X, Y \in L_2$ and $X \perp\!\!\!\perp Y$, solve the problem $\min_g \|Y - g(X)\|_2$ where g is a measurable function satisfying $g(X) \in L_2$.

J. Suppose that for some $(\alpha^\circ, \beta^\circ) \in \Theta = \mathbb{R}^2$ and some $X \in L_2$, $Y = \alpha^\circ + \beta^\circ X + \varepsilon$ where $E\,\varepsilon = 0$ and $\varepsilon \perp\!\!\!\perp X$.
   1. Let X_n^L be the Lebesgue sequence of simple approximations to X. Solve $\min_{\alpha,\beta} \|Y - (\alpha + \beta X_n^L)\|_2$.
   2. Show that your solutions, (α_n, β_n), to the previous problem converge to $(\alpha^\circ, \beta^\circ)$.

K. Yield per acre is a random variable Y and fertilizer per acre is a random variable X. Let $h(x) = E\,(Y|X = x)$. Suppose that $h'(x) > 0$ but that $ph'(x) > c$ where p is the price of the crop and c the cost of the fertilizer. This looks like an argument that the farmers are under-utilizing fertilizer. This problem is about omitted variable bias.

   However, suppose that yield per acre satisfies $Y = f(X, Q) + \varepsilon$, $\varepsilon \perp\!\!\!\perp X, Q$, where Q is the quality of the acre. Suppose that $\partial f / \partial x > 0$, $\partial f / \partial q > 0$, and that $\partial^2 f / \partial x \partial q > 0$, i.e. suppose that $f(\cdot, \cdot)$ is increasing and supermodular. Suppose also that farmers know their Q but that it is random to the observers/econometricians who are forming their conditional expectations. The risk-neutral farmers pick $x(Q)$ to solve $x = \arg\max_q E(pf(x, q) + \varepsilon) - cx$. Compare the true marginal product of fertilizer in a field that uses an amount $X = x$ to $h'(x)$.

L. [Doob's Theorem] Let $(\Omega, \mathcal{F})$ and let $(Y, \mathcal{Y})$ be non-empty sets and σ-fields of subsets. The most frequent class of sub-σ-fields, $\mathcal{G} \subset \mathcal{F}$, that we will encounter arise from a measurable $g : \Omega \to Y$, they are of the form $\mathcal{G} = g^{-1}(\mathcal{Y})$. Let $\mathcal{B}$ denote the Borel σ-field on $\mathbb{R}$, that is, the smallest σ-field of subsets of $\mathbb{R}$ containing the open subsets of $\mathbb{R}$. This problem shows that if $f : \Omega \to [0, 1]$ has the property that $f^{-1}(\mathcal{B}) \subset \mathcal{G}$, then there exists a measurable $h : Y \to \mathbb{R}$ such that $f(\omega) = h(g(\omega))$. Thus, the only $\mathcal{G}$-measurable functions are in fact functions of g, the function g contains everything (measurable) that one could ever get from $\mathcal{G}$.

1. Give an elementary proof of the assertion if $g(\Omega)$ is a finite set $G = \{y_1, \ldots, y_N\}$ and $\{y_n\} \in \mathcal{Y}$ for each n.
2. Show that $\mathcal{G}$ is a σ-field.
3. Show that $A_{i,n} := \{\omega : f(\omega) \in [i/2^n, (i+1)/2^n)\}$ is of the form $g^{-1}(B_{i,n})$ for some $B_{i,n} \in \mathcal{Y}$.
4. Define the functions $f_n : \Omega \to [0,1]$ and $h_n : Y \to [0,1]$ by
$$f_n = \sum_{i=1}^{2^n} \tfrac{i}{2^n} 1_{A_{i,n}} \text{ and } h_n = \sum_{i=1}^{2^n} \tfrac{i}{2^n} 1_{B_{i,n}}.$$
Show that for all ω, $f_n(\omega) = h_n(g(\omega))$.
5. Define $h(y) = \limsup_n h_n(y)$. Show that for all ω, $f(\omega) = h(g(\omega))$.
6. In the previous step, why couldn't we define $h(y) = \lim_n h_n(y)$?

M. [I'll cross that bridge when I come to it] Suppose that X and S are simple random variables taking values in $\{x_1, \ldots, x_N\}$ and $s_1, \ldots, s_M$. The time line for the problem is that the value of S is observed before X is known, and an action $a \in A$ must be picked. We suppose that A is a compact metric space, and that for each x_n, $u(\cdot, x_n)$ is a continuous function. Amongst the elements of A^S, let the function $a^*(\cdot)$ solve the problem
$$\max_{a(\cdot)} E\left(u(a^*(S), X)\right).$$
The functions $a(\cdot)$ are called, for hopefully obvious reasons, "complete contingent plans." Define $P(\cdot|S = s_m) \in \Delta(\{x_1, \ldots, x_N\})$ as the usual conditional probability.

1. Show that $a^*(\cdot)$ is a solution if and only if for all s_m with $P(S = s_m) > 0$, $a^*(s_m)$ solves the problem
$$\max_{a \in A} \sum_{x_n} u(a, x_n) P(x_n | s_m).$$
2. Check that the average conditional probability of any event $X \in A$ is the prior probability, $P(X \in A)$. That is, $P(X \in A) = \sum_m P(X \in A | S = s_m) \cdot P(S = s_m)$.

Comments: Conditional probabilities map s_m's to $\Delta(\{x_1, \ldots, x_n\})$. When S and X are more general random variables, we will want a measurable function from the range of S to $\Delta(\mathbb{R})$ that also solves these kinds of optimization problems. For example, we are often interested in seeing how educational procedures affect the performance of the median and the bottom and top quartiles. Knowing $P(\cdot|S = s)$ answers this, and many other questions.