Stats 310A Session 8

December 10, 2019

In this session, we will go through some practice problems. These problems fall in the scope of Stats 310A and involves a lot of what we learned comprehensively.

Problem 1 Let $\mathcal{X}$ be a set, $\mathcal{B}$ be a countably generated σ-algebra of subsets of $\mathcal{X}$. Let $\mathcal{P}(\mathcal{X}, \mathcal{B})$ be the set of all probability measures on $(\mathcal{X}, \mathcal{B})$. Make $\mathcal{P}(\mathcal{X}, \mathcal{B})$ into a measurable space by declaring that the map $P \mapsto P(A)$ is Borel measurable for each $A \in \mathcal{B}$. Call the associated σ-algebra $\mathcal{B}^*$.

(a) Show that $\mathcal{B}^*$ is countably generated.

(b) For $\mu \in \mathcal{P}(\mathcal{X}, \mathcal{B})$, show that $\{\mu\} \in \mathcal{B}^*$.

(c) For $\mu, \nu \in \mathcal{P}(\mathcal{X}, \mathcal{B})$, let

$$\|\mu - \nu\| = \sup_{A \in \mathcal{B}} |\mu(A) - \nu(A)|.$$

Show that the map $(\mu, \nu) \mapsto \|\mu - \nu\|$ is $\mathcal{B}^* \times \mathcal{B}^*$ measurable.

Solution

(a) We have by the definition of $\mathcal{B}^*$ that

$$\mathcal{B}^* = \sigma\left(\{\{P \in \mathcal{P}(\mathcal{X}, \mathcal{B}) : P(A) \leq p\} : A \in \mathcal{B}, p \in [0,1]\}\right).$$

As $\mathcal{B}$ is countably generated, there exists some countable $\mathcal{B}_0$ such that $\mathcal{B} = \sigma(\mathcal{B}_0)$. Without loss of generality, we can let $\mathcal{B}_0$ be an algebra (if not, consider the smallest algebra containing $\mathcal{B}_0$: this also generates $\sigma(\mathcal{B}_0)$ and is a countable set, see Exercise 1.1.29(b) in Dembo's Notes).

We now define

$$\mathcal{B}^* = \sigma\left(\{\{P \in \mathcal{P}(\mathcal{X}, \mathcal{B}) : P(A) \leq p\} : A \in \mathcal{B}_0, p \in [0,1] \cap \mathbb{Q}\}\right),$$

which is the smallest σ-algebra that makes $f_A := P \mapsto P(A)$ measurable for all $A \in \mathcal{B}_0$.

Recalling that the total variation distance makes $\mathcal{P}(\mathcal{X}, \mathcal{B})$ into a metric space, let $\mathcal{B}^*_{\text{tv}}$ be the corresponding Borel σ-algebra. We will now show that

$$\mathcal{B}^*_0 \subseteq \mathcal{B}^* \subseteq \mathcal{B}^*_{\text{tv}} \subseteq \mathcal{B}^*_0$$

The first inclusion is trivial and the second follows from the fact that

$$|f_A(P) - f_A(P')| = |P(A) - P'(A)| \leq \|P - P'\|_{\text{tv}},$$

rendering each f_A Lipschitz, so continuous, and thus $\mathcal{B}^*_{\text{tv}}$-measurable.

For the last inclusion, a slight modification of Exercise 1.2.15(a) in Dembo's notes yields that for any $P, P' \in \mathcal{P}$ and $A \in \mathcal{B}$,

$$\inf_{B \in \mathcal{B}_0} (P(A\Delta B) \vee P'(A\Delta B)) = 0.$$

In particular, for any $B \in \mathcal{B}$ and $\epsilon > 0$, we can take $A_\epsilon \in \mathcal{B}_0$ such that $P(A_\epsilon \Delta A), P'(A_\epsilon \Delta A) < \epsilon$, which renders

$$\begin{aligned} |P(A) - P(A_\epsilon)| &= |P(A \cup A_\epsilon) - P(A_\epsilon) - P(A \cup A_\epsilon) + P(A)| \\ &\le P(A \setminus A_\epsilon) + P(A_\epsilon \setminus A) \\ &\le 2\epsilon, \end{aligned}$$

and similarly for P'. We thus have that, for any $\epsilon > 0$,

$$\begin{aligned} \sup_{A \in \mathcal{B}_0} |P(A) - P'(A)| &\le \sup_{A \in \mathcal{B}} |P(A) - P'(A)| \\ &\le \sup_{A \in \mathcal{B}} |P(A_\epsilon) - P'(A_\epsilon)| + 4\epsilon \\ &\le \sup_{A \in \mathcal{B}_0} |P(A) - P'(A)| + 4\epsilon, \end{aligned}$$

from which we have that

$$\sup_{A \in \mathcal{B}} |P(A) - P'(A)| = \sup_{A \in \mathcal{B}_0} |P(A) - P'(A)|.$$

But now, we can write any TV-open ball as

$$\begin{aligned} \{P : \|P - P_0\|_{\text{tv}} < r\} &= \bigcup_{q<r, q \in \mathbb{Q}} \{P : \sup_{A \in \mathcal{B}} |P(A) - P_0(A)| \le q\} \\ &= \bigcup_{q<r, q \in \mathbb{Q}} \{P : \sup_{A \in \mathcal{B}_0} |P(A) - P_0(A)| \le q\} \\ &= \bigcup_{q<r, q \in \mathbb{Q}} \bigcap_{r=1}^{\infty} \bigcap_{A \in \mathcal{B}_0} \{P : |P(A) - P_0(A)| < q + 1/r\} \\ &\in \mathcal{B}_0^*. \end{aligned}$$

We thus have our chain of set inclusions and in particular, $\mathcal{B}^* = \mathcal{B}_0^*$, which is finitely-generated.

(b) Given any $\mu \in \mathcal{P}(\mathcal{X}, \mathcal{B})$, we clearly have

$$\{\mu\} \subseteq \{P : P(A) = \mu(A), \text{ for all } A \in \mathcal{B}_0\} = \bigcap_{A \in \mathcal{B}_0} \{P : P(A) = \mu(A)\}.$$

Our goal is to show the converse direction, thereby showing that $\{\mu\}$ is the intersection of countably many generating sets and thus $\{\mu\} \in \mathcal{B}^*$. This is to say that any two measures that coincide on the generating set $\mathcal{B}_0$ has to coincide on $\mathcal{B}$, which is guaranteed by the uniqueness of the Caratheodory extension.

(c) From the working in part (a), it suffices to show that for any $t \in \mathbb{R}$,

$$\{(\mu,\nu) : \|\mu-\nu\|_{\text{tv}} \le t\} = \bigcap_{A\in\mathcal{B}_0} \{(\mu,\nu) : |\mu(A)-\nu(A)| \le t\}$$

is a measurable subset of $\mathcal{B}^* \times \mathcal{B}^*$. But note that each set on the RHS is $\mathcal{B}^* \times \mathcal{B}^*$-measuable as the function $(\mu,\nu) \to |\mu(A)-\nu(A)|$ is measurable for all A, so the result follows.

□

Problem 2 Let $\{X_n\}_n$ be iid symmetric random variables such that

$$\lim_{y\to\infty} \frac{y^2 \Pr(|X_1| > y)}{\mathbb{E}(X_1^2; |X_1| < y)} = 0. \tag{1}$$

Show that there exists a sequence $\{b_n\}_n$ of positive constants such that

$$\frac{1}{b_n}\sum_{k=1}^n X_k \xrightarrow{d} \mathcal{N}(0,1). \tag{2}$$

Solution We will be truncating the random variables at some $c_n \to \infty$, but we will leave the specification of this sequence for later. Define

$$\sigma_n^2 = \mathbb{E}(X_1^2; |X_1| < c_n) \tag{3}$$

so that $\sigma_n^2 \to \mathbb{E}X_1^2 \in (0,\infty]$. Next, we define the truncations

$$\tilde{X}_{n,k} = \frac{1}{\sigma_n\sqrt{n}} X_k \mathbf{1}_{\{|X_k|\le c_n\}}. \tag{4}$$

Defining

$$S_n = \frac{1}{\sigma_n\sqrt{n}}\sum_{k=1}^n X_k, \tag{5}$$

$$\tilde{S}_n = \sum_{k=1}^n \tilde{X_{n,k}}, \tag{6}$$

we verify the Lindeberg condition, so for any $\epsilon > 0$, we have

$$\mathbb{E}(\tilde{X}_{n,k}^2; |\tilde{X}_{n,k} \ge \epsilon) = \frac{1}{n\sigma_n^2}\mathbb{E}(X_1^2; \epsilon\sigma_n\sqrt{n} \le |X_1| < c_n), \tag{7}$$

$$g_n(\epsilon) = \frac{1}{\sigma_n^2}\mathbb{E}(X_1^2; \epsilon\sigma_n\sqrt{n} \le |X_1| < c_n). \tag{8}$$

Notice that if $c_n \ll \sigma_n\sqrt{n}$, then for large enough n, the condition in the expectation fails and $g_n(\epsilon) = 0$. The Lindeberg CLT thus yields that $\tilde{S}_n \xrightarrow{d} \mathcal{N}(0,1)$.

Next, we use the usual truncation trick to write

$$\Pr(S_n \neq \tilde{S}_n) \leq \sum_{k=1}^{n} \Pr(|X_n| > c_n) \tag{9}$$

$$= n \Pr(|X_1| > c_n) \tag{10}$$

$$= \frac{n}{c_n^2} \cdot c_n^2 \Pr(|X_1| > c_n) \tag{11}$$

$$= \frac{n}{c_n^2} \mathbb{E}(X_1^2; |X_1| < c_n) f(c_n), \tag{12}$$

where $f(y)$ is the function tending to 0 defined in eq. (3).

Suppose now that $c_n \geq \sigma_n n^{1/4}$ and define

$$\bar{f}(x) = \sup_{y \geq x} f(y). \tag{13}$$

This function dominates f, is decreasing and tends to 0 as $x \to \infty$. In particular, we can define

$$a_n = \bar{f}(\sigma_n n^{1/4}) \geq f(c_n), \tag{14}$$

so that

$$\Pr(S_n \neq \tilde{S}_n) \leq \frac{n\sigma_n^2 a_n}{c_n^2}, \tag{15}$$

which converges to 0 as long as

$$a_n \sigma_n \sqrt{n} \ll c_n. \tag{16}$$

At last, we can define

$$c_n = \sigma_n(\sqrt{a_n} \vee n^{-1/4})\sqrt{n} \tag{17}$$

Notice that this satisfies $c_n \to \infty$, $c_n \geq \sigma_n n^{1/4}$ and $c_n \gg a_n \sigma_n \sqrt{n}$ so that $\Pr(S_n \neq \tilde{S}_n) \to 0$, and also satisfies $c_n \ll \sigma_n \sqrt{n}$ so that $\tilde{S}_n \xrightarrow{d} \mathcal{N}(0,1)$. Therefore, we conclude that $S_n \xrightarrow{d} \mathcal{N}(0,1)$. □

Problem 3 Recall that given two measures μ, ν on $(\mathbb{R}, \mathcal{B}_\mathbb{R})$, a coupling of μ and ν is any probability measure γ on $(\mathbb{R}^2, \mathcal{B}_{\mathbb{R}^2})$ such that, for any Borel set A, we have $\gamma(A \times \mathbb{R}) = \mu(A)$, $\gamma(\mathbb{R} \times A) = \nu(A)$. (In words, the one-dimensional marginals of γ are –respectively– μ and ν.) We denote by $\Gamma(\mu, \nu)$ the set of couplings of μ and ν. For $p \geq 1$, let $\mathcal{P}_p$ be the space of probability measures μ such that $\int |x|^p \, \mu(\mathrm{d}x) < \infty$. For $\mu, \nu \in \mathcal{P}_p$, their p-th Wasserstein distance is

$$W_p(\mu, \nu) = \left\{ \inf_{\gamma \in \Gamma(\mu,\nu)} \int_{\mathbb{R} \times \mathbb{R}} |x - y|^p \, \gamma(\mathrm{d}x, \mathrm{d}y) \right\}^{1/p} \tag{18}$$

1. For $\mu = \mathcal{N}(0,1)$ and $\nu = \mathcal{N}(a,1)$, prove that $W_2(\mu, \nu) = |a|$.
2. For $\mu = \mathcal{N}(0,1)$ and $\nu = \mathcal{N}(0,v)$, $v > 1$, prove that $W_2(\mu, \nu) = \sqrt{v} - 1$.
3. Prove that $\Gamma(\mu, \nu)$ is uniformly tight.
4. Fix $p \geq 1$. Prove that there exists a sequence of probability measures $\{\gamma_n\}_{n \in \mathbb{R}} \subseteq \Gamma(\mu, \nu)$ amd $\gamma \in \Gamma(\mu, \nu)$ such that $\gamma_n \stackrel{w}{\Rightarrow} \gamma$, and

$$\lim_{n \to \infty} \int_{\mathbb{R} \times \mathbb{R}} |x - y|^p \, \gamma_n(\mathrm{d}x, \mathrm{d}y) = W_p(\mu, \nu)^p \,. \tag{19}$$

5. Prove that (for $\{\gamma_n\}_{n\in\mathbb{N}}$, γ constructed as in the previous point)

$$\liminf_{n\to\infty} \int_{\mathbb{R}\times\mathbb{R}} |x-y|^p \,\gamma_n(\mathrm{d}x,\mathrm{d}y) \geq \int_{\mathbb{R}\times\mathbb{R}} |x-y|^p \,\gamma(\mathrm{d}x,\mathrm{d}y)\,, \tag{20}$$

and deduce that

$$W_p(\mu,\nu) = \left\{ \int_{\mathbb{R}\times\mathbb{R}} |x-y|^p \,\gamma(\mathrm{d}x,\mathrm{d}y) \right\}^{1/p} \tag{21}$$

Solution

1. For any $X \sim \mu$ and $Y \sim \nu$, we have by Jensen's inequality that

$$\sqrt{\mathbb{E}(X-Y)^2} \geq |\mathbb{E}X - \mathbb{E}Y| = |a|, \tag{22}$$

and this lower bound is achieved by taking $Z \sim \mathcal{N}(0,1)$ and $X = Z, Y = Z + a$ so that

$$\sqrt{\mathbb{E}(X-Y)^2} = \sqrt{\mathbb{E}(Z-(Z+a))^2} = |a|. \tag{23}$$

Hence, $W_2(\mu,\nu) = |a|$.

2. For any $X \sim \mu$ and $Y \sim \nu$, we have by Cauchy-Schwarz that

$$\mathbb{E}(X-Y)^2 = \mathbb{E}X^2 - 2\mathbb{E}XY + \mathbb{E}Y^2 \geq v - 2\sqrt{v} + 1 = (\sqrt{v}-1)^2 \tag{24}$$

and this lower bound is achieved by taking $Z \sim \mathcal{N}(0,1)$ and $X = Z, Y = \sqrt{v}Z$ so that

$$\sqrt{\mathbb{E}(X-Y)^2} = \sqrt{\mathbb{E}[(\sqrt{v}-1)^2 Z]} = \sqrt{v} - 1. \tag{25}$$

Hence, $W_2(\mu,\nu) = \sqrt{v} - 1$.

3. Let $\epsilon > 0$ and let K_ϵ be such that $\mu([-K,K]^v), \nu([-K,K]^c) < \epsilon/2$. We then have that $[-K,K]^2$ is compact such that, for any $\gamma \in \Gamma$, we have that

$$\begin{aligned} \gamma(([-K,K]^2)^c) &= \Pr(\{X \notin [-K,K]\} \cup \{Y \notin [-K,K]\}) && (26)\\ &\leq \Pr(X \notin [-K,K]) + \Pr(Y \notin [-K,K]) && (27)\\ &< \epsilon. && (28) \end{aligned}$$

That is, Γ is uniformly tight.

4. Since $W_p^p(\mu,\nu) = \inf_{\gamma\in\Gamma} \int_{\mathbb{R}^2} |x-y|^p \gamma(\mathrm{d}x,\mathrm{d}y)$, choose a sequence $\gamma_n \in \Gamma$ such that

$$\lim_{n\to\infty} \int_{\mathbb{R}\times\mathbb{R}} |x-y|^p \gamma_n(\mathrm{d}x,\mathrm{d}y) = W_p^p(\mu,\nu). \tag{29}$$

By the uniform tightness of Γ, Prokhorov's theorem allows us to choose a subsequence n_k such that $\gamma_{n_k} \Rightarrow \gamma$. Joint weak convergence implying marginal weak convergence (since coordinate projections are continuous) allows us to conclude that $\gamma \in \Gamma$.

5. Since $\gamma_{n_k} \Rightarrow \gamma$, Skorokhod's representation theorem yields a sequence of random variables $(X_k, X_k') \sim \gamma_{n_k}$ converging almost surely to some $(X, X') \sim \gamma$. Fatou's lemma then gives

$$\liminf_{k\to\infty} \int_{\mathbb{R}\times\mathbb{R}} |x-y|^p \gamma_{n_k}(\mathrm{d}x, \mathrm{d}y) = \liminf_{k\to\infty} \mathbb{E}|X_k - X_k'| \tag{30}$$

$$\geq \mathbb{E}|X - X'| \tag{31}$$

$$= \int_{\mathbb{R}\times\mathbb{R}} |x-y|^p \gamma(\mathrm{d}x, \mathrm{d}y). \tag{32}$$

Since $\gamma \in \Gamma$, we combine this with the previous result to conclude

$$W_p^p(\mu,\nu) \geq \int_{\mathbb{R}\times\mathbb{R}} |x-y|^p \gamma(\mathrm{d}x, \mathrm{d}y) \geq W_p^p(\mu,\nu), \tag{33}$$

from which we conclude that the infimum in the definition of $W_p^p(\mu,\nu)$ is necessarily achieved.

□