TOPICS IN STATISTICAL THEORY Part III.

Some necessary measure theory YZ/Lent 2020

Comments and corrections to zemel@statslab.cam.ac.uk.

The material in this document is non-examinable.

1 Portmanteau

Theorem 1.1. *Let X_n and X be random vectors on $\mathbb{R}^d$ with distribution functions F_n, F. The following are equivalent.*

1. $F_n(x) \to F(x)$ *whenever F is continuous at x. That is, $X_n \to X$ in distribution.*
2. $\mathbb{E}g(X_n) \to \mathbb{E}g(X)$ *for any bounded continuous function $g : \mathbb{R}^d \to \mathbb{R}$.*
3. $\mathbb{E}g(X_n) \to \mathbb{E}g(X)$ *for any bounded 1-Lipschitz function $g : \mathbb{R}^d \to \mathbb{R}$.*
4. $P(X_n \in A) \to P(X \in A)$ *for any Borel set A such that $P(X \in \partial A) = 0$. Here $\partial A = \overline{A} \setminus \mathrm{int(A)}$ is the boundary of A.*

Proof. (1→2). Let $g : \mathbb{R}^d \to \mathbb{R}$ be continuous with $M = \sup|g| < \infty$, fix $\epsilon > 0$ and denote D the set of discontinuity points of F. Since D is countable the set of "discontinuity coordinates"

$$B = \{d_i : d \in D\} \subset \mathbb{R}$$

is countable. Call x a strict continuity point of F if $x_i \notin B$ for all $i = 1, \dots, d$; such an x is also a continuity point of F. Let $x_1 \leq x_2$ be strict continuity points of F such that $F(x_1) \leq \epsilon$ and $F(x_2) \geq 1-\epsilon$. Since g is uniformly continuous on $K = [x_1, x_2] = \{x \in \mathbb{R}^d : x_1 \leq x \leq x_2\}$, we can split K into smaller rectangles $R_j = [y_j, z_j]$, $j = 1, \dots, k$ such that y_j, z_j are strict continuity points of F and

$$M_j - m_j \leq \epsilon, \qquad M_j = \sup_{x \in R_j} g(x), \qquad m_j = \inf_{x \in R_j} g(x).$$

Then

$$\mathbb{E}g(X) \leq \mathbb{E}g(X)[1\{X \leq x_1\} + 1\{X \not\leq x_2\}] + \sum_{j=1}^k \mathbb{E}g(X)1\{X \in R_j\} \leq 2M\epsilon + \sum_{j=1}^k M_j P(X \in R_j)$$

and similarly

$$\mathbb{E}g(X) \geq \mathbb{E}g(X)[1\{X \leq x_1\} + 1\{X \not\leq x_2\}] + \sum_{j=1}^k \mathbb{E}g(X)1\{X \in \mathrm{int}(R_j)\} \geq -2M\epsilon + \sum_{j=1}^k m_j P(X \in \mathrm{int}(R_j))$$

Notice however that $P(X \in \text{int}(R_j)) = P(X \in R_j)$ since y_j and z_j are continuity points of F. Now use the same decomposition to obtain

$$\limsup_{n\to\infty} \mathbb{E}g(X_n) \leq \limsup_{n\to\infty} M[P(X_n \leq x_1) + P(X_n \nleq x_2)] + \sum_{j=1}^{k} M_j P(X_n \in R_j)$$

and

$$\liminf_{n\to\infty} \mathbb{E}g(X_n) \geq \liminf_{n\to\infty} -M[P(X_n \leq x_1) + P(X_n \nleq x_2)] + \sum_{j=1}^{k} m_j P(X_n \in \text{int}(R_j)).$$

We now claim that $P(X_n \in R_j) \to P(X \in R_j)$ and $P(X_n \in \text{int}(R_j)) \to P(X \in \text{int}(R_j))$. Admitting this momentarily, we conclude

$$\limsup_{n\to\infty} \mathbb{E}g(X_n) - \mathbb{E}g(X) \leq 4M\epsilon + \sum_{j=1}^{k} [M_j - m_j] P(X \in R_j) \leq 4M\epsilon + \epsilon$$

and

$$\liminf_{n\to\infty} \mathbb{E}g(X_n) - \mathbb{E}g(X) \geq -4M\epsilon - \sum_{j=1}^{k} [M_j - m_j] P(X \in \text{int}(R_j)) \geq -4M\epsilon - \epsilon.$$

Since $\epsilon > 0$ is arbitrary conclude that $\mathbb{E}g(X_n) \to \mathbb{E}g(X)$. It remains to show the claim. Let $R = [y, z]$ where y and z are strict continuity points of F. The inclusion-exclusion formula yields

$$P(X \in R) = \sum_{I \subseteq \{1,\dots,d\}} (-1)^{|I|} F(v_I), \qquad (v_I)_i = \begin{cases} y_i & i \in I \\ z_i & i \notin I. \end{cases} \tag{1}$$

Since y and z are strict continuity points of F this gives

$$P(X_n \in R) \to P(X \in R).$$

For the interior of R, assume that $z_i > y_i$ for all i (otherwise the interior is empty). First notice that

$$\limsup P(X_n \in \text{int}R)) \leq \limsup P(X_n \in R) = P(X \in R).$$

There exist sequences $y_k \searrow y$ and $z_k \nearrow z$ such that $y_k > y$ coordinatewise, $z_k < z$ coordinatewise and both y_k and z_k are strict continuity points of F. Letting $R_k = [y_k, z_k]$ gives

$$\liminf P(X_n \in \text{int}R)) \geq \liminf P(X_n \in R_k) = P(X \in R_k)$$

for all k. As $k \to \infty$, formula (**??**) shows that $P(X \in R_k) \to P(X \in R)$ because y and z are strict continuity points of F. The last two inequalities combined yield

$$P(X_n \in \text{int}R) \to P(X \in \text{int}(R)) = P(X \in R)$$

for any $R = [y, z]$ with y, z strict continuity points of F. This completes the proof.

(2→3) is obvious.

(3→4) By linearity $\mathbb{E}g(X_n) \to \mathbb{E}g(X)$ for any $L > 0$ and any L-Lipschitz function g. Let $\epsilon > 0$ and A as in 4. Denote $d(x, B) = \inf_{y \in B} \|x - y\|$ for any nonempty set $B \subseteq \mathbb{R}^d$ and any $x \in \mathbb{R}^d$. If $A = \emptyset$ or $A = \mathbb{R}^d$ then the assertion holds trivially, so assume $\emptyset \subset A \subset \mathbb{R}^d$. Define $g_k(y) = 1 - \min(1, kd(x, A))$ and $h_k(y) = \min(1, kd(x, \mathbb{R}^d \setminus A))$. These functions are bounded and k-Lipschitz on $\mathbb{R}^d$. Observe that g_k equals 1 on A and vanishes on the closed set

$$B_k = \left\{ x : d(x, A) \geq \frac{1}{k} \right\}.$$

and h_k vanishes on $\mathbb{R}^d \setminus A$ and equals 1 on the closed set

$$C_k = \left\{ x : d(x, \mathbb{R}^d \setminus A) \geq \frac{1}{k} \right\}.$$

Hence B_k and C_k are measurable and

$$P(X_n \in C_k) \leq \mathbb{E}h_k(X_n) \leq P(X_n \in A) \leq \mathbb{E}g_k(X_n) \leq P(X_n \notin B_k).$$

The same holds with X_n replaced by X. Thus

$$\limsup_{n\to\infty} P(X_n \in A) \leq \limsup_{n\to\infty} \mathbb{E}g_k(X_n) = \mathbb{E}g_k(X) \leq P(X \notin B_k)$$

and

$$\liminf_{n\to\infty} P(X_n \in A) \geq \liminf_{n\to\infty} \mathbb{E}h_k(X_n) = \mathbb{E}h_k(X) \geq P(X \in C_k).$$

Since k is arbitrary and $B_k \subset B_{k+1}$,

$$\limsup_{n\to\infty} P(X_n \in A) \leq P(X \notin \cup B_k) = P(d(X, A) = 0) = P(X \in \overline{A}),$$

and since $C_k \subseteq C_{k+1}$,

$$\liminf_{n\to\infty} P(X_n \in A) \geq P(X \in \cap C_k) = P(d(X, \mathbb{R}^d \setminus A) = 0) = P(X \in \text{int}(A)).$$

The assumption that $P(X \in \partial A) = 0$ now gives $P(X_n \in A) \to P(X \in A)$.

(4→1). Define $A = (-\infty, x]$ for x continuity point of F. Then

$$P(X \in \text{int}(A)) = \lim_{k\to\infty} P(X \leq x - 1/k) = \lim_{k\to\infty} F(x - 1/k) = F(x) = P(X \in A).$$

Since A is closed this implies that $P(X \in \partial A) = 0$ and consequently

$$F_n(x) = P(X_n \in A) \to P(X \in A) = F(x)$$

as desired. □

2 Gluing lemma

Let $(X_1, Y_1) \in \mathbb{R}^{d_X} \times \mathbb{R}^{d_Y}$ have joint distribution P_{XY} and $(Y_2, Z_2) \in \mathbb{R}^{d_Y} \times \mathbb{R}^{d_Z}$ have joint distribution P_{YZ}. Suppose that Y_1 and Y_2 have the same distribution. The gluing lemma asserts that we can "glue" the two joint distributions using the common marginal distribution (of Y_1 and Y_2).

Lemma 2.1. *There exists a random vector $(X, Y, Z) \in \mathbb{R}^{d_X} \times \mathbb{R}^{d_Y} \times \mathbb{R}^{d_Z}$ such that $(X, Y) \sim P_{XY}$ and $(Y, Z) \sim P_{YZ}$.*

Proof. (sketch) Let A, B, C be Borel subsets of $\mathbb{R}^{d_X}$, $\mathbb{R}^{d_Y}$ and $\mathbb{R}^{d_Z}$ respectively. Define

$$f_A(y) = \mathbb{P}[X_1 \in A | Y_1 = y], \qquad g_C(y) = \mathbb{P}[Z_2 \in C | Y_2 = y].$$

Then

$$P_{XY}(A \times B) = E[1_B(Y) f_A(Y)], \qquad P_{YZ}(B \times C) = E[1_B(Y) g_C(Y)].$$

Define

$$P_{XYZ}(A \times B \times C) = E[1_B(Y) f_A(Y) g_C(Y)].$$

Since $f_{\mathbb{R}^{d_X}}(y) = 1 = g_{\mathbb{R}^{d_Z}}(y)$ we have $P_{XYZ}(A \times B) = P_{XY}(A \times B)$ and $P_{XYZ}(B \times C) = P_{YZ}(B \times C)$. Therefore P_{XYZ} has the correct marginals P_{XY} and P_{YZ}. The rigorous mathematical justification for this construction comes from the disintegration theorem, as given in Theorem 5.4 in Kallenberg's book [Foundations of Modern Probability]. □

3 Prokhorov theorem

This theorem establishes a useful relation between distributional convergence and tightness. A sequence of random vectors (X_n) is tight if for any $\epsilon > 0$ there exists $M > 0$ such that

$$\sup_n P(\|X_n\| > M) < \epsilon.$$

Theorem 3.1 (Prokhorov 1956)**.** *If $X_n \to X$ in distribution, then (X_n) is tight. If (X_n) is tight, then there exists a subsequence n_k and a random variable X such that $X_{n_k} \to X$ in distribution.*

Proof. Let F_n and F denote the corresponding distribution functions. Assume $X_n \to X$ in distribution and fix $\epsilon > 0$. Since $P(\|X\| > M) \to 0$ as $M \to \infty$, there exist $x_1 \leq x_2$ such that $F(x_1) < \epsilon/4$ and $F(x_2) \geq 1 - \epsilon/4$. Replacing x_1 and x_2 if needed, we may assume that x_1 and x_2 are continuity point of F. Then there exists N such that $F_n(x_1) < \epsilon/2$ and $F_n(x_2) \geq 1 - \epsilon/2$ for all $n > N$. There exists $x_3 \leq x_1$ and $x_4 \geq x_2$ such that $P(X_i \leq x_3) \leq \epsilon/2$ and $P(X_i \leq x_4) \geq 1 - \epsilon/2$ for all $i \leq N$. These inequalities hold also hold for $i > N$.

Hence when M is large enough such that the ball of radius M centred at the origin contains $[x_3, x_4] = \{x : x_3 \leq x \leq x_4\}$,

$$\sup_n P(\|X_n\| > M) < \epsilon.$$

The converse is more difficult. By a diagonal argument one can extract a subsequence n_k such that $F_{n_k}(q) \to G(q)$ for all $q \in \mathbb{Q}^d$. Define

$$F(x) = \liminf_{q_k \searrow x, q_k \in \mathbb{Q}^d} G(q).$$

By $q_k \searrow x$ we mean that $q_{k+1} \geq q_k > x$, and the last inequality is strict in any coordinate. We shall show that $F_{n_k}(x) \to F(x)$ for all continuity points x of F and that F is a distribution function.

a) It is clear that $0 \leq F(x) \leq 1$ for all x and that $F(x_2) \geq F(x_1)$ for all $x_2 \geq x_1$. This implies that F has at most countably many discontinuity points.

b) F is right-continuous. Fix $x \in \mathbb{R}^d$ and $\epsilon > 0$. There exists $q > x$ rational such that $F(x) \leq G(q) \leq F(x) + \epsilon$. Then for all $x \leq y < q$ we have $F(x) \leq F(y) \leq F(x + \epsilon)$ implying right-continuity.

c) Let $\epsilon > 0$. Since (X_n) is tight there exists $q \in \mathbb{Q}^d$ such that $F_n(q) > 1 - \epsilon$ for all n. This implies that $F(q) \geq 1 - \epsilon$. Thus $f(x) \to 1$ as $x \to \infty$ on each coordinate. Now tightness of (X_n) also implies tightness of each coordinate projection $(e_i^t X_n)$ where $e_i \in \mathbb{R}^d$ is the i-th unit vector. Thus there exists $q_i \in \mathbb{Q}$ such that $F_n^i(q_i) < \epsilon$ for all n, where F^i denotes the i-th marginal distribution. Consequently $F_n(x) < \epsilon$ whenever $x_i \leq q_i$. This implies that $F(x) \leq \epsilon$ whenever $x_i < q_i$. Since i is arbitrary we conclude that $F(x) \to 0$ if $\min_i x_i \to -\infty$ (note that it is enough if merely *one* coordinate of x goes to $-\infty$.)

d) Conditions a)–c) suffice for F being a distribution function, so that there exists a random vector $X \sim F$. See Theorem 2.25 in Kallenberg's book [Foundations of Modern Probability].

e) Suppose that F is continuous at x. Let $q > x$ such that $G(q) \leq F(x) + \epsilon$. Then

$$\limsup F_{n_k}(x) \leq \limsup F_{n_k}(q) = G(q) \leq F(x) + \epsilon.$$

Now let $q < x$ such that $F(q) > F(x) - \epsilon$. Pick $q < q' < x$. Then

$$\liminf F_{n_k}(x) \geq \liminf F_{n_k}(q') = G(q') \geq F(q) > F(x) - \epsilon.$$

Since ϵ is arbitrary we have $F_{n_k}(x) \to F(x)$. Thus $X_{n_k} \to X$ in distribution as required. □