Project 2

Convergence in distribution

2.1 Sample paths

A stochastic process $\{X_t : t \in T\}$ defined on $(\Omega, \mathcal{F}, \mathbb{P})$ defines an $\mathcal{F}\backslash\mathcal{F}_T$-measurable random element of $\mathbb{R}^T$ if $\mathcal{F}_T$ is the cylinder sigma-field on $\mathbb{R}^T$, the smallest sigma-field for which each coordinate projection π_t is $\mathcal{F}_T\backslash\mathcal{B}(\mathbb{R})$-measurable. Why?

If each each sample path $X(\cdot, \omega)$ is bounded then X is also a random element of $\ell^\infty(T)$, the set of all bounded real functions on T, also equipped with its cylinder sigma-field. Also equip $\ell^\infty(T)$ with its uniform metric,

$$d(x, y) = \sup_{t\in T} |x(t) - y(t)|$$

and the corresponding Borel sigma-field $\mathcal{B}$. Continuity of each π_t implies that $\mathcal{F}_T \subseteq \mathcal{B}$. Why? The process X need not be $\mathcal{F}\backslash\mathcal{B}$-measurable.

Suppose T is equipped with a metric ρ. Write $C(T)$ for the set of all ρ-uniformly continuous members of $\ell^\infty(T)$. If T is separable then the Borel sigma-field $\mathcal{B}$ is equal to the cylinder sigma-field, because the supremum in the definition of d can be taken over a countable, dense subset of T. The space $C(T)$ is a closed subset of $\ell^\infty(T)$. Why?

psseudometric enough

When $T = [0, 1]$ the space $D(T)$ is defined as the set of all ***cadlag*** real-valued functions on T. That is, $x \in D(T)$ if it is continuous from the right at each t in $[0, 1)$ and the limit from the left exists (and is finite) at each t in $(0, 1]$. The Borel sigma-field on $D[0, 1]$ is strictly larger than the cylinder sigma-field, as the next example shows.

> **Remark.** I am in the bad habit of writing $\mathcal{F}_T$ for the cylinder sigma-field on $\mathbb{R}^T$ and on $\ell^\infty(T)$ and on $C(T)$ and on $D(T)$ for $T = [0, 1]$. Also I have been writing $\mathcal{B}$ for the Borel sigma-field, generated by the open subsets for the d metric, on $\ell^\infty(T)$ and $C(T)$ and $D[0, 1]$. Would it be better to invent different symbols for these quantities?

<1> **Example.** (Compare with Billingsley 1968, page 152.) Expand the following argument.

Suppose ξ is defined on $(\Omega, \mathcal{F}, \mathbb{P})$ has a Uniform$[0, 1]$ distribution. The process $X(t, \omega) = \{\xi(\omega) \le t\}$ for $0 \le t \le 1$ has sample paths in $D[0, 1]$ and is measurable with respect to the cylinder sigma-field. It cannot be Borel measurable.

To see why, define elements $x_\theta(t) = \{\theta \le t\}$ of $D[0,1]$. Note that $X(\cdot,\omega) = x_{\xi(\omega)}$. For each $A \subseteq [0,1]$ define

$$\widetilde{A} = \{y \in D[0,1] : \inf_{\theta \in A} d(y, x_\theta) < 1/3\}.$$

Each $\widetilde{A}$ is an open subset of $D[0,1]$. Also $X(\cdot,\omega) \in \widetilde{A}$ if and only if $\xi(\omega) \in A$.

If ξ were $\mathcal{F}\backslash\mathcal{B}$-measurable then

$$\nu(A) := \mathbb{P}\{\omega \in \Omega : \xi(\omega) \in \widetilde{A}\}$$

would define a probability measure that extends Lebesgue measure to all subsets of $[0,1]$, which is not possible if we accept the Axiom of Choice.

□

Exercise Suppose $\xi_1, \dots, \xi_n$ are independent Uniform$[0,1]$ distributed random variables. The empirical distribution function is defined as

$$U_n(t,\omega) = n^{-1} \sum\nolimits_{i \le n} \{\xi_i(\omega) \le t\} \qquad \text{for } 0 \le t \le 1.$$

Show that U_n is cylinder measurable but not Borel measurable, as a map from Ω into $D[0,1]$. (I'd be happy with just the case $n = 2$.) It might help to think of the ξ_i's as coordinate maps on $[0,1]^n$ equipped with n-dimensional Lebesgue measure.

□

The failure of Borel measurability for the empirical distribution function was one motivation for the creation of different metrics (Skorohod 1956) for $D[0,1]$. The Borel sigma-field for the best known of the Skorohod metrics coincides with the cylinder sigma-field. For more general empirical processes indexed by sets of functions there was no obvious way to generalize the Skorohod approach. Instead Hoffmann-Jørgensen (1984) developed a theory based on outer expectations, which handled the measurability problem in a most elegant way. Subsequent work by Dudley (1985) established the H-J approach as the new standard.

2.2 Convergence in distribution

Let $(\mathcal{X}, d)$ be a metric space equipped with its Borel sigma-field $\mathcal{B}(\mathcal{X})$. Suppose $P_\infty, P_1, P_2, \dots$ are probability measures on $\mathcal{B}(\mathcal{X})$ and $X_\infty, X_1, X_2, \dots$ are $\mathcal{B}(\mathcal{X})$-measurable random elements of $\mathcal{X}$, all defined on the same $\big(\Omega, \mathcal{F}, \mathbb{P}\big)$.

In the classical theory of weak convergence/convergence in distribution (Pollard 2001, Section 7.1), if X_n has distribution P_n and X_∞ has distribution P_∞ then each of the following assertions means the same thing:

$$X_n \rightsquigarrow X_\infty \qquad X_n \rightsquigarrow P_\infty \qquad P_n \rightsquigarrow P_\infty \qquad P_n \rightsquigarrow X_\infty$$

namely,

$$P_n f = \mathbb{P} f(X_n) \to P_\infty f = \mathbb{P} f(X_\infty) \qquad \text{for all } f \text{ in } \mathrm{BL}(\mathcal{X}).$$

Here $\mathrm{BL}(\mathcal{X})$ is the set of all bounded, Lipschitz, real functions on $\mathcal{X}$. That is, $\mathrm{BL}(\mathcal{X})$ consists of all real-valued functions f on $\mathcal{X}$ for which both

$$\begin{aligned} \|f\|_\infty &:= \sup\{|f(x)| : x \in \mathcal{X}\} \\ \|f\|_{\mathrm{Lip}} &:= \inf\{C : |f(x) - f(y)| \le C d(x, y) \text{ for all } x,\ y \text{ in } \mathcal{X}\}. \end{aligned}$$

Equivalently, $\|f\|_{\mathrm{BL}} := \max\left(\|f\|_\infty, \|f\|_{\mathrm{Lip}}\right)$ is finite.

If we wish to allow the possibility that the X_n's, for $1 \le n < \infty$, might not be $\mathcal{B}(\mathcal{X})$-measurable then we cannot think of X_n as having a distribution in the sense that the image of $\mathbb{P}$ underX_n is a probability measure on $\mathcal{B}(\mathcal{X})$. (Does it then make sense to talk of convergence in distribution?) Instead, work with inner and outer expectations.

<3> **Definition.** *For a bounded real-valued random variable Z on Ω define*

$$\mathbb{P}^* Z = \inf\{\mathbb{P} U : Z \le U \text{ and } U \text{ is measurable }\}$$

and

$$\mathbb{P}_* Z = \sup\{\mathbb{P} L : Z \ge L \text{ and } L \text{ is measurable }\}.$$

Notice that $\mathbb{P}^*(-Z) = -\mathbb{P}_* Z$, so we don't really need to work with lower expectations, although it is helpful to note that $\mathbb{P}^* Z \ge \mathbb{P}_* Z$ with equality if and only if Z is measurable with respect to the $\mathbb{P}$-completion of $\mathcal{F}$.

<4> **Definition.** *For (possibly nonmeasurable) maps $X_n : \Omega \to \mathcal{X}$ and a probability measure P on $\mathcal{B}(\mathcal{X})$ define $X_n \rightsquigarrow P$ to mean $\mathbb{P}^* f(X_n) \to P f$ for each f in $\mathrm{BL}(\mathcal{X})$. If X_∞ is a $\mathcal{B}(\mathcal{X})$ measurable random element of $\mathcal{X}$ with distribution P, define $X_n \rightsquigarrow X_\infty$ to mean the same as $X_n \rightsquigarrow P$. Equivalently, $\mathbb{P}^* f(X_n) \to \mathbb{P} f(X_\infty)$ for each f in $\mathrm{BL}(\mathcal{X})$.*

You should convince yourself that $\mathbb{P}^*$ could be replaced by $\mathbb{P}_*$ without changing the meaning of $X_n \rightsquigarrow P$. Indeed you should show that the definition is equivalent to

<5> $$\limsup_{n\to\infty} \mathbb{P}^* f(X_n) \le Pf \qquad \text{for each } f \text{ in } \mathrm{BL}(\mathcal{X})$$

and to

<6> $$\liminf_{n\to\infty} \mathbb{P}_* f(X_n) \ge Pf \qquad \text{for each } f \text{ in } \mathrm{BL}(\mathcal{X}).$$

In fact, we could even replace $\mathrm{BL}(\mathcal{X})$ by the set of all upper semicontinuous functions that are bounded from above in <5> and by the set of all lower semicontinuous functions that are bounded from below in <6>. These facts are sometimes collected together as part of the so-called ***Portmanteau Theorem*** (van der Vaart and Wellner 1996, Section 1.3).

The H-J theory parallels the classical theory exposited by Billingsley (1968). For our purposes it will be enough to understand the following slight generalization of the method explained by Pollard (1984, page 92).

<7> **Lemma.** *Let $\{X_n : n \in \mathbb{N}\}$ be a sequence of maps from Ω into $\mathcal{X}$ and let P be a probability measure defined on $\mathcal{B}(\mathcal{X})$. Suppose $\{\epsilon_k : k \in \mathbb{N}\}$ and $\{\delta_k : k \in \mathbb{N}\}$ are both sequences of positive numbers that converge to zero as $k \to \infty$. For each k suppose there exist maps $X_{n,k} : \Omega \to \mathcal{X}$ and probability measures P_k on $\mathcal{B}(\mathcal{X})$ for which*

(i) $X_{n,k} \rightsquigarrow P_k$ *as $n \to \infty$, for each fixed k*

(ii) $P_k \rightsquigarrow P$ *as $k \to \infty$*

(iii) $\limsup_{n\to\infty} \mathbb{P}^*\{d(X_n, X_{n,k}) > \delta_k\} < \epsilon_k$ *for each fixed k*

Then $X_n \rightsquigarrow P$.

Proof In class I gave a proof by first principles, without using facts about outer expectations. (I sneaked the "without loss of generality $f \ge 0$" past you.) This time I present a more concise proof. You should justify all the implicit appeals to facts about outer expectations in what follows.

Write $A_{n,k}$ for $\{d(X_n, X_{n,k}) > \delta_k\}$. Consider an f in $\mathrm{BL}(\mathcal{X})$ with $\|f\|_{\mathrm{BL}} = C$. Without loss of generality, $f \ge 0$. Use the $\|\cdot\|_\infty$ bound for $\omega \in A_{n,k}$ and the $\|\cdot\|_{\mathrm{Lip}}$ bound for $\omega \in A_{n,k}^c$ to get

$$f(X_n) \le f(X_{n,k}) + C\delta_k + CA_{n,k}.$$

(Where did I use the assumption $f \geq 0$?) Take outer expectations then let n tend to infinity to deduce

$$\begin{aligned}\limsup_{n\to\infty} & \mathbb{P}^* f(X_n) \\ & \leq \limsup_{n\to\infty} \mathbb{P}^* f(X_{n,k}) + C\delta_k + C \limsup_{n\to\infty} \mathbb{P}^* A_{n,k} \\ & \leq P_k f + C\delta_k + C\epsilon_k\end{aligned}$$

The upper bound tends to Pf as k tends to infinity.

□

2.3 Partial-sum processes

Let $\xi_1, \xi_2, \ldots$ be iid random variables, defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, with $\mathbb{P}\xi_i = 0$ and $\mathbb{P}\xi_i^2 = 1$. Define $S_n := \sum_{i \leq n} \xi_i$ and

$$X_n(t, \omega) = n^{-1/2} S_{\lfloor nt \rfloor} \qquad \text{for } 0 \leq t \leq 1.$$

Consider X_n as a map into $\mathcal{X} = D[0,1]$ equipped with its uniform metric. Let B denote a Brownian motion with continuous sample paths.

For each finite $J \subseteq [0,1]$ with $0, 1 \in J$ write π_J for the map that projects $D[0,1]$ into $\mathbb{R}^J$ and let A_J denote the map from $\mathbb{R}^J$ into $D[0,1]$ defined by

$$A_J(z) := z(t_{k+1})\{t = 1\} + \sum\nolimits_{i=0}^{k} z(t_i)\{t_i \leq t < t_{i+1}\} \qquad \text{for } 0 \leq t \leq 1$$
$$\text{if } z \in \mathbb{R}^J \text{ and } J = \{t_0, \ldots, t_{k+1}\} \text{ with } 0 = t_0 < \cdots < t_{k+1} = 1.$$

Define $L_J = A_J \circ \pi_J$, a map from $D[0,1]$ into $D[0,1]$.

(i) For each fixed J show that π_J and A_J and L_J are continuous functions, for each fixed J.

(ii) Suppose $J = \{t_0, \ldots, t_{k+1}\}$ with $0 = t_0 < \cdots < t_{k+1} = 1$. Define $I_j := [t_j, t_{j+1}]$. Show that

$$d(x, L_J x) \leq \Delta_J(x) := \max\nolimits_{j=0}^{k} \sup\nolimits_{t \in I_j} |x(t) - x(t_j)|$$

(iii) Explain why X_n is actually $\mathcal{B}(\mathcal{X})$-measurable. Hint: Consider $\psi(X_n)$ for a continuous $\psi : \mathcal{X} \to \mathbb{R}$.

(iv) Show that $\sup_t |\text{var}(X_n(t) - t| \leq n^{-1}$.

(v) Invoke a multivariate central limit theorem to show that $\pi_J X_n \rightsquigarrow \pi_J B$ for each finite $J \subseteq [0,1]$.

(vi) For each k in $\mathbb{N}$ define $J(k) := \{j/k : j = 0, 1, \ldots, k\}$. Define $X_{n,k} = L_{J(k)}(X_n)$ and $B_k = L_{J(k)}B$.

(vii) Show that $B_k \rightsquigarrow B$ as $k \to \infty$.

(viii) Show that $X_{n,k} \rightsquigarrow B_k$ as $n \to \infty$, for each fixed k.

(ix) For each $\delta > 0$ show that there exists a $k = k_\delta$ for which, when $n \geq k$,

$$\begin{aligned} &\mathbb{P}\{d(X_n, X_{n,k}) > \delta\} \\ &\quad \leq \mathbb{P}\{\Delta_{J(k)}(X_n) > \delta\} \\ &\quad \leq \sum\nolimits_{j=0}^{k} 2\mathbb{P}\{|X_n(t_{j+1}) - X_n(t_j)| > \delta/2\} \qquad \text{where } t_j = j/k \\ &\quad \to \sum\nolimits_{j=0}^{k} 2\mathbb{P}\{|B(t_{j+1}) - B(t_j)| > \delta/2\} \qquad \text{as } n \to \infty \\ &\quad \leq 2k \exp(-k\delta^2/8). \end{aligned}$$

For the second inequality you might find Pollard (2001, Inequality 6.38) useful.

(x) Figure out choices for δ_k and ϵ_k so that Lemma 7 leads you to the conclusion: $X_n \rightsquigarrow B$.

EXERCISE Show that $\max_{j \leq n} |S_j|/\sqrt{n}$ converges in distribution to $\sup_{0 \leq t \leq 1} |B_t|$.

□

References

Billingsley, P. (1968). *Convergence of Probability Measures.* New York: Wiley.

Dudley, R. M. (1985). An extended Wichura theorem, definitions of Donsker classes, and weighted empirical distributions. *Springer Lecture Notes in Mathematics 1153*, 141–178. Springer, New York.

Hoffmann-Jørgensen, J. (1984). Stochastic Processes on Polish Spaces. Unpublished manuscript, Aarhus University, Denmark.

Pollard, D. (1984). *Convergence of Stochastic Processes.* New York: Springer.

Pollard, D. (2001). *A User's Guide to Measure Theoretic Probability.* Cambridge University Press.

Skorohod, A. V. (1956). Limit theorems for stochastic processes. *Theory Probability and Its Applications 1*, 261–290.

van der Vaart, A. W. and J. A. Wellner (1996). *Weak Convergence and Empirical Process: With Applications to Statistics*. Springer-Verlag.