Prohorov's theorem

Bengt Ringnér[*][†]

October 26, 2008

1 The theorem

Definition 1 *A set* Π *of probability measures defined on the Borel sets of a topological space is called tight if, for each* $\varepsilon > 0$*, there is a compact set* K *such that*

$$\boldsymbol{P}(K) > 1 - \varepsilon$$

for all $\boldsymbol{P} \in \Pi$

Theorem 1 *A tight set,* Π*, of probability measures on the Borel sets of a metric topological space,* $\mathfrak{X}$*, is relatively compact in the sense that for each sequence,* $\boldsymbol{P}_1, \boldsymbol{P}_2, \ldots$ *in* Π *there exists a subsequence that converges to a probability measure* $\boldsymbol{P}$*, not necessarily in* Π*, in the sense that*

$$\int g \, d\boldsymbol{P}_{n_j} \to \int g \, d\boldsymbol{P}$$

for all bounded continuous integrands. Conversely, if the metric space is separable and complete, then each relatively compact set is tight.

This is a generalisation of the Helly selection and the Helly-Bray theorems in which $\mathfrak{X}$ is the real line. If $F_1, F_2, \ldots$ is a sequence of right-continuous cumulative distribution functions, $F_n(x) = \mathbb{P}(X_n \leq x)$, then there is a subsequence and a limiting F such that, as $j \to \infty$,

$$F_{n_j}(x) \to F(x) \quad \text{if} \quad F \quad \text{is continuous at} \quad x$$

[*]Centre for Mathematical Sciences, Lund University, Lund, Sweden.
[†]Homepage: http://www.maths.lth.se/~bengtr

This is the same as, for all g continuous with *compact support*,

$$\int g(x)\,dF_{n_j}(x) \to \int g(x)\,dF(x),$$

in other words,

$$\int g\,d\boldsymbol{P}_{n_j} \to \int g\,d\boldsymbol{P}$$

where $\boldsymbol{P}_n(A) = \mathbb{P}(X_n \in A)$ for Borel sets A. The limiting F can be chosen as right continuous and nondecreasing, but is not neccessarily a distribution function, since one has $\lim_{x\to\infty} F(x) \le 1$ and $\lim_{x\to-\infty} F(x) \ge 0$, but not necessarily equality. Equality is equivalent to the original sequence being tight. In this case pointwise convergence at continuity points is the same as $\int g\,dF_{n_j}$ converges for all *bounded* continuous integrands.

Note also that working with continuous functions with compact support is only possible in locally compact spaces, a concept which is too restrictive in probability theory. Here $\mathfrak{X}$ may contain trajectories of random processes, and one might be interested in

$$\boldsymbol{P}_n(\{x : \sup_t x(t) \le M\}) = \mathbb{P}(\sup_t X_n(t) \le M).$$

2 Proof of the direct part

The idea the same as on the real line. First prove convergence on a denumerable dense subset, then use equicontinuity. To get a dense subset of integrands, we need a lemma.

Lemma 1 *Let K be a compact set in a space $\mathfrak{X}$ equipped with a metric topology. Then the space of bounded real valued functions on $\mathfrak{X}$ is separable in the sense that there exist a countable subset, $\varphi_1, \varphi_2, \ldots$ such that, for any bounded continuous g,*

$$\sup_{x\in K} |g(x) - \varphi_k(x)|$$

can be made arbitrarily small, and

$$\sup_{x\in\mathfrak{X}} |g(x) - \varphi_k(x)| \le 3 \sup_{x\in\mathfrak{X}} |g(x)|.$$

Proof Being compact in a metric space, K has a countable dense subset, $x_1, x_2, \ldots$. From the (proof of the) Stone-Weierstrass approximation theorem it follows that any continuous function f can be approximated in the sup-norm on K with functions of the form

$$\varphi = \min_{0<j\le j_0} \max_{0<k\le k_0} \psi_{j\,k}$$

where the ψ are of the form

$$\psi(x) = a - b\rho(x, x_i)$$

and ρ is any metric generating the topology. Since K is bounded, we may let a and b be rational. Finally, make the φ bounded by replacing them with

$$\min(\max(\varphi, \inf_K \varphi), \sup_K \varphi).$$

□

To prove the theorem, choose compact $K_1, K_2, \ldots$ such that $\boldsymbol{P}(K_m) < 1 - 1/m$ for all $\boldsymbol{P} \in \Pi$. For each such K_m the lemma gives a dense subset of functions. Let $\varphi_1, \varphi_2, \ldots$ be an enumeration of all these. For a given bounded continuous g and $\varepsilon > 0$, choose $m > 1/\varepsilon$ and φ_k with $|g - \varphi_k| < \varepsilon$ on K_m. Then

$$\int |g - \varphi_k| \, d\boldsymbol{P} = \int_{K_m} \underbrace{|g - \varphi_k|}_{<\varepsilon} \, d\boldsymbol{P} + \int_{K_m^c} \underbrace{|g - \varphi_k|}_{\leq 3 \sup |f|} \, d\boldsymbol{P} \leq (1 + 3 \sup_{\mathfrak{X}} |g|)\varepsilon.$$

Therefore, for any given g and $\varepsilon > 0$ there is a φ_k such that

$$\sup_{\boldsymbol{P} \in \Pi} \int |g - \varphi_k| \, d\boldsymbol{P} < \varepsilon.$$

By the Bolzano-Weierstrass theorem combined with Cantor's diagonal method, there is a subsequence such that, for each k, $\int \varphi_k \, d\boldsymbol{P}_{n_j}$ converges as $j \to \infty$. This gives

$$\limsup_{j \to \infty} \int g \, d\boldsymbol{P}_{n_j} - \liminf_{j \to \infty} \int g \, d\boldsymbol{P}_{n_j} < 2\varepsilon.$$

Since this holds for arbitrarily small $\varepsilon > 0$, the limit exists, so we can define a functional $\boldsymbol{I}$ by

$$\boldsymbol{I}(g) = \lim_{j \to \infty} \int g \, d\boldsymbol{P}_{n_j}.$$

Clearly, $\boldsymbol{I}$ is a linear functional and $\boldsymbol{I}(g) \geq 0$ if $g \geq 0$. To prove that it can be represented with a probability measure, we use the Stone-Daniell representation theorem. Let $g_k \searrow 0$ pointwise. For a given $\varepsilon > 0$ choose a compact K such that $\boldsymbol{P}(K^c) < \varepsilon$ for all $\boldsymbol{P}$ in Π. Then

$$\int g_k \, d\boldsymbol{P}_{n_j} \leq \int_K g_k \, d\boldsymbol{P}_{n_j} + \int_{K^c} g_k \, d\boldsymbol{P}_{n_j} \leq \sup_{x \in K} g_k(x) + \varepsilon \sup_x g_1(x),$$

so

$$\boldsymbol{I}(g_k) \leq \sup_{x\in K} g_k(x) + \varepsilon \sup_x g_1(x),$$

By Dini's theorem, $g_k \to 0$ uniformly on K, so the first term tends to zero as $k \to \infty$, giving $\limsup_{k\to\infty} \boldsymbol{I}(g_k) \leq \varepsilon \sup_x g_1(x)$ for all $\varepsilon > 0$, which gives

$$\boldsymbol{I}(g_k) \to 0$$

Since $|g|$ and $\min(1, g)$ are bounded continuous if g is so, $\boldsymbol{I}$ is a Daniell integral on a Stone lattice. Therefore, and since $\boldsymbol{I}(1) = 1$, it can be represented as an integral with respect to a probability measure, so

$$\lim_{j\to\infty} \int g\, d\boldsymbol{P}_{n_j} = \boldsymbol{I}(g) = \int g\, d\boldsymbol{P}$$

which proves the first part of the theorem.

3 Proof of the converse

All sets of the form

$$K = \cap_{i=1}^{\infty} \cup_{i=1}^{k_j} \overline{B(x_i, 1/j)}$$

in which $B(x_i, 1/j) = \{x : \rho(x, x_i) < 1/j\}$ are compact, since $\mathfrak{X}$ is complete and K is closed and totally bounded. In order to make K large enough, we use separability and choose $x_1, \ldots$ as a dense subset. We shall use the fact that, if $\boldsymbol{P}_k \to \boldsymbol{P}$ is the sense of the theorem, then[1]

$$\liminf_{k\to\infty} \boldsymbol{P}_k(U) \geq \boldsymbol{P}(U)$$

for all open sets U.

Let $\varepsilon > 0$. There is a K with $\boldsymbol{P}(K) > 1 - \varepsilon$ for all $\boldsymbol{P} \in \Pi$

- if we can prove that for each j there is a k_j such that

$$\boldsymbol{P}(\cup_{i=1}^{k_j} B(x_i, 1/j)) > 1 - \varepsilon/2^j$$

  for all $\boldsymbol{P}$ in Π.

[1]This is a part of the so called Portemanteau Theorem. It follows by considering $g_n(x) = \min(1, n\rho(x, U^c))$. Then g_n increases towards the indicator function of U, and

$$\boldsymbol{P}_k(U) \geq \int g_n\, d\boldsymbol{P}_k \to \int g_n\, d\boldsymbol{P}, \quad k \to \infty,$$

which gives

$$\liminf_{k\to\infty} \boldsymbol{P}_k(U) \geq \int g_n\, d\boldsymbol{P} \to \boldsymbol{P}(U), \quad n \to \infty.$$

If this does not hold, there is a j_0 such that, for each k there is a $\boldsymbol{P}_k$ with

$$\boldsymbol{P}_k(\cup_{i=1}^k B(x_i, 1/j_0)) \leq 1 - \varepsilon/2^{j_0}.$$

Of course, this also holds with $\cup_{i=1}^{k'}$ if $k' \leq k$. By assumption, there is a converging subsequence, so

$$\boldsymbol{P}(\cup_{i=1}^k B(x_i, 1/j_0)) \leq \liminf_{n\to\infty} \boldsymbol{P}_{k_n}(\cup_{i=1}^k B(x_i, 1/j_0)) \leq 1 - \varepsilon/2^{j_0}.$$

But this would give

$$1 = \boldsymbol{P}(\mathfrak{X}) = \lim_{k\to\infty} \boldsymbol{P}(\cup_{i=1}^k B(x_i, 1/j_0)) \leq 1 - \varepsilon/2^{j_0}.$$

Therefore • holds and the proof is complete.

□