SECTION 9

Convergence in Distribution and Almost Sure Representation

Classical limit theorems for sums of independent random vectors will often suggest a standardization for the partial-sum process, S_n, so that its finite dimensional projections have a limiting distribution. It is then natural to ask whether the standardized stochastic process also has a limiting distribution, in some appropriate sense. The traditional sense has been that of a functional limit theorem. One identifies some metric space of real-valued functions on T that contains all the standardized sample paths, and then one invokes a general theory for convergence in distribution of random elements of a metric space (or weak convergence of probability measures on the space).

For example, if $T = [0, 1]$ and the sample paths of S_n have only simple discontinuities, the theory of weak convergence for $D[0, 1]$ might apply.

Unfortunately, even for such simple processes as the empirical distribution function for samples from the Uniform$[0, 1]$ distribution, awkward measurability complications arise. With $D[0, 1]$ either one skirts the issue by adopting a Skorohod metric, or one retains the uniform metric at the cost of some measure theoretic modification of the definition of convergence in distribution.

For index sets more complicated than $[0, 1]$ there is usually no adequate generalization of the Skorohod metric. The measurability complications cannot be defined away. One must face the possibility that the expectations appearing in plausible definitions for convergence in distribution need not be well defined. Of the numerous general theories proposed to handle this problem, the one introduced by Hoffmann-Jørgensen (and developed further by Dudley 1985) is undoubtedly the best. It substitutes outer expectations for expectations. It succeeds where other theories fail because it supports an almost sure representation theorem; with suitable reinterpretation, most of the useful results from the classical theory carry over to the new theory.

The theory concerns sequences of maps $\{X_n\}$ from a probability space $(\Omega, \mathcal{A}, \mathbb{P})$ into a metric space $\mathcal{X}$. If each X_n is measurable with respect to the Borel σ-field $\mathcal{B}(\mathcal{X})$, convergence in distribution to a probability measure P on $\mathcal{B}(\mathcal{X})$ can conveniently be defined to mean

$$\mathbb{P}f(X_n) \to Pf \qquad \text{for every } f \text{ in } \mathcal{U}(\mathcal{X}),$$

where $\mathcal{U}(\mathcal{X})$ stands for the class of all bounded, uniformly continuous, real functions on $\mathcal{X}$. If X_n has no particular measurability properties, $f(X_n)$ need not be measurable; the expectation $\mathbb{P}f(X_n)$ need not be well defined. But the outer (or inner) expectation is defined: for each bounded, real-valued H on Ω,

$$\mathbb{P}^*H = \inf\{\mathbb{P}h : H \leq h \text{ and } h \text{ integrable}\}.$$

The inner expectation $\mathbb{P}_*H$ is defined analogously. The new definition of convergence in distribution replaces $\mathbb{P}$ by $\mathbb{P}^*$, while retaining some measure theoretic regularity for the limit P in order to exclude some unpleasant cases.

(9.1) Definition. If $\{X_n\}$ is a sequence of (not necessarily Borel measurable) maps from Ω into a metric space $\mathcal{X}$, and if P is a probability measure on the Borel σ-field $\mathcal{B}(\mathcal{X})$, then $X_n \rightsquigarrow P$ (read as "X_n converges in distribution to P") is defined to mean $\mathbb{P}^*f(X_n) \to Pf$ for every f in $\mathcal{U}(\mathcal{X})$.

The equality $\mathbb{P}_*f(X_n) = -\mathbb{P}^*[-f(X_n)]$ shows that the definition could be stated, equivalently, in terms of convergence of inner expectations. It could also be stated in terms of convergence to a Borel measurable random element X: one replaces Pf by $\mathbb{P}f(X)$.

In requiring convergence only for f in $\mathcal{U}(\mathcal{X})$ my definition departs slightly from the Hoffmann-Jørgensen and Dudley definitions, where f runs over all bounded, continuous functions. The departure makes it slightly easier to prove some basic facts without changing the meaning of the concept in important cases.

(9.2) Example. Here is a result that shows the convenience of requiring uniform continuity for f in Definition 9.1. *If $\{Y_n\}$ is a sequence of random elements of a metric space $(\mathcal{Y}, e)$ which converges in probability to a constant y, that is, $\mathbb{P}^*\{e(Y_n, y) > \delta\} \to 0$ for each $\delta > 0$, and if $X_n \rightsquigarrow X$, then $(X_n, Y_n) \rightsquigarrow (X, y)$.* For if f is a uniformly continuous function on $\mathcal{X} \otimes \mathcal{Y}$, bounded in absolute value by a constant M, then, for an appropriate choice of δ,

$$f(X_n, Y_n) \leq f(X_n, y) + \epsilon + 2M\{e(Y_n, y) > \delta\}.$$

Taking outer expectations of both sides then letting $n \to \infty$, we get

$$\limsup \mathbb{P}^*f(X_n, Y_n) \leq \limsup \mathbb{P}^*f(X_n, y) + \epsilon.$$

Uniform continuity of $f(\cdot, y)$ ensures that the right-hand side equals $\mathbb{P}f(X, y) + \epsilon$. Replacement of f by $-f$ would give the companion lower bound needed to establish the required convergence. □

It has long been recognized (see Pyke 1969, for example) that many arguments involving convergence in distribution are greatly simplified by use of a technical device known as *almost sure representation.* Such a representation usually asserts something like:

> *If $X_n \rightsquigarrow P$ then there exist $\widetilde{X}_n$ and $\widetilde{X}$ such that $\widetilde{X}_n$ and X_n have the same distribution, $\widetilde{X}$ has distribution P, and $\widetilde{X}_n \to \widetilde{X}$ almost surely.*

The random elements $\widetilde{X}_n$ and $\widetilde{X}$ are defined on a new probability space $(\widetilde{\Omega}, \widetilde{\mathcal{A}}, \widetilde{\mathbb{P}})$. For Borel measurable X_n, "the same distribution" is interpreted to mean that

$$\mathbb{P}g(X_n) = \widetilde{\mathbb{P}}g(\widetilde{X}_n) \qquad \text{for all bounded, Borel measurable } g.$$

Without the measurability, it would seem natural to require equality of outer expectations. Dudley's (1985) form of the representation theorem achieves this in a particularly strong form.

With the Dudley representation, $\widetilde{\mathcal{A}}\backslash\mathcal{A}$-measurable maps ϕ_n from $\widetilde{\Omega}$ into Ω are constructed to be *perfect* in the sense that not only is $\mathbb{P}$ the image of $\widetilde{\mathbb{P}}$ under each ϕ_n, but also

$$\mathbb{P}^*H = \widetilde{\mathbb{P}}^*H \circ \phi_n \qquad \text{for every bounded } H \text{ on } \Omega.$$

The representing random elements $\widetilde{X}_n$ are defined by

$$\widetilde{X}_n(\widetilde{\omega}) = X_n(\phi_n(\widetilde{\omega})) \qquad \text{for each } \widetilde{\omega} \text{ in } \widetilde{\Omega}.$$

Thus $\mathbb{P}^*g(X_n) = \widetilde{\mathbb{P}}^*g(\widetilde{X}_n)$ for every bounded g on $\mathcal{X}$, regardless of its measurability properties. In general the outer integrals satisfy only an inequality,

$$\mathbb{P}^*H \geq \widetilde{\mathbb{P}}^*(H \circ \phi_n) \qquad \text{for every bounded } H \text{ on } \Omega,$$

because $h \circ \phi_n \geq H \circ \phi_n$ whenever $h \geq H$. To establish that ϕ_n is perfect it is therefore enough to prove that

$$\text{(9.3)} \qquad \mathbb{P}^*H \leq \widetilde{\mathbb{P}}g \qquad \text{for all } \widetilde{\mathcal{A}}\text{-measurable } g \geq H \circ \phi_n.$$

We then get the companion inequality by taking the infimum over all such g.

The Dudley representation also strengthens the sense in which the representing sequence converges. For possibly nonmeasurable random elements mere pointwise convergence would not suffice for applications.

(9.4) Representation Theorem. *If $X_n \rightsquigarrow P$ in the sense of Definition 9.1, and if the limit distribution P concentrates on a separable Borel subset $\mathcal{X}_0$ of $\mathcal{X}$, then there exists a probability space $(\widetilde{\Omega}, \widetilde{\mathcal{A}}, \widetilde{\mathbb{P}})$ supporting $\widetilde{\mathcal{A}}\backslash\mathcal{A}$-measurable maps ϕ_n into Ω and an $\widetilde{\mathcal{A}}\backslash\mathcal{B}(\mathcal{X})$-measurable map $\widetilde{X}$ into $\mathcal{X}_0$, such that:*

- *(i) each ϕ_n is a perfect map, in the sense that $\mathbb{P}^*H = \widetilde{\mathbb{P}}^*(H \circ \phi_n)$ for every bounded H on Ω;*
- *(ii) $\widetilde{\mathbb{P}}\widetilde{X}^{-1} = P$, as measures on $\mathcal{B}(\mathcal{X})$;*
- *(iii) there is a sequence of $\widetilde{\mathcal{A}}\backslash\mathcal{B}[0,\infty]$-measurable, extended-real-valued random variables $\{\delta_n\}$ on $\widetilde{\Omega}$ for which $d\big(\widetilde{X}_n(\widetilde{\omega}), \widetilde{X}(\widetilde{\omega})\big) \leq \delta_n(\widetilde{\omega}) \to 0$ for almost every $\widetilde{\omega}$, where $\widetilde{X}_n(\widetilde{\omega}) = X_n(\phi_n(\widetilde{\omega}))$.*

It is easy to show that if (i), (ii), and (iii) hold then $X_n \rightsquigarrow P$; the assertions of the theorem are more natural than they might appear at first glance.

A sketch of Dudley's construction will close out this section. But first an example—a revamped Continuous Mapping Theorem—to show how perfectness compensates for the lack of measurability. In my opinion, an unencumbered form of Continuous Mapping Theorem is essential for any general theory of convergence in distribution.

(9.5) EXAMPLE. Suppose $X_n \rightsquigarrow P$ with P concentrated on a separable Borel subset $\mathcal{X}_0$ of $\mathcal{X}$. Suppose τ is a map into another metric space $\mathcal{Y}$ such that

(i) the restriction of τ to $\mathcal{X}_0$ is Borel measurable,

(ii) τ is continuous at P almost all points of $\mathcal{X}_0$.

Then we can deduce from the Representation Theorem that $\tau(X_n)$ converges in distribution to the image measure $P\tau^{-1}$.

Fix an f in $\mathcal{U}(\mathcal{Y})$. Define $h = f \circ \tau$. We need to verify that $\mathbb{P}^* h(X_n) \to Ph$. With no loss of generality we may suppose $0 \le h \le 1$. Fix an $\epsilon > 0$. For each positive integer k define G_k to be the open set of all points x in $\mathcal{X}$ for which h oscillates by $> \epsilon$ within the open ball of radius $1/k$ and center x. [That is, there are points y and z with $|h(y) - h(z)| > \epsilon$ and $d(x,y) < 1/k$ and $d(x,z) < 1/k$. The same y and z will provide oscillation $> \epsilon$ for every center close enough to x.]

As $k \to \infty$ the set G_k shrinks down to a set that excludes all continuity points of τ, and thereby has zero P measure. We can therefore find a k such that $PG_k < \epsilon$.

The definition of G_k ensures that if $\widetilde{X}(\widetilde{\omega}) \notin G_k$ and if $\delta_n(\widetilde{\omega}) < 1/k$ then

$$|h(\widetilde{X}_n(\widetilde{\omega})) - h(\widetilde{X}(\widetilde{\omega}))| \le \epsilon.$$

Consequently,

$$h(\widetilde{X}_n) \le (\epsilon + h(\widetilde{X}))\{\widetilde{X} \notin G_k, \delta_n < 1/k\} + \{\widetilde{X} \in G_k\} + \{\delta_n \ge 1/k\}.$$

The expression on the right-hand side is measurable; it is one of the measurable functions that enters into the definition of the outer expectation of $h(\widetilde{X}_n)$. It follows that

$$\widetilde{\mathbb{P}}^* h(\widetilde{X}_n) \le \epsilon + \widetilde{\mathbb{P}} h(\widetilde{X}) + \widetilde{\mathbb{P}}\{\widetilde{X} \in G_k\} + \widetilde{\mathbb{P}}\{\delta_n \ge 1/k\}.$$

Measurability of δ_n and dominated convergence ensure that the last probability tends to zero. And the perfectness property lets us equate the left-hand side with $\mathbb{P}^* h(X_n)$. Passing to the limit we deduce

$$\limsup \mathbb{P}^* h(X_n) \le Ph.$$

An analogous argument with h replaced by $1 - h$ gives the companion lower bound needed to establish the desired convergence. □

OUTLINE OF A PROOF OF THE REPRESENTATION THEOREM

Step 1. The indicator function of a closed ball with zero P measure on its boundary can be sandwiched between two functions from $\mathcal{U}(\mathcal{X})$ whose expectations are arbitrarily close. If B is an intersection of finitely many such balls, the approximating functions can be combined to construct f_1 and f_2 in $\mathcal{U}(\mathcal{X})$ such that

$P(f_1 - f_2) < \epsilon$ and

$$f_1(X_n) \geq \{X_n \in B\} \geq f_2(X_n).$$

Taking outer and inner expectations, then passing to the limit, we deduce that

$$\mathbb{P}^*\{X_n \in B\} \to PB,$$
$$\mathbb{P}_*\{X_n \in B\} \to PB,$$

for every such B. (We will need this result only for sets B constructed from balls with centers in $\mathcal{X}_0$.)

Step 2. If π is a partition of $\mathcal{X}$ generated by a finite collection of closed balls, each with zero P measure on its boundary, then

$$\mathbb{P}_*\{X_n \in B\} \to PB \qquad \text{for each } B \text{ in } \pi.$$

This follows from Step 1, because the sets in π are proper differences of intersections of finitely many closed balls.

Step 3. For each positive integer k, cover $\mathcal{X}_0$ by closed balls of diameter less than $1/k$, with zero P measure on their boundaries. Use separability of $\mathcal{X}_0$ to extract a countable subcover, then use countable additivity of P to find a subcollection that covers all of $\mathcal{X}_0$ except for a piece with P measure less than 2^{-k}. Generate a finite partition $\pi(k)$ of $\mathcal{X}$ from this collection. All except one of the sets in $\pi(k)$ has diameter less than $1/k$, and that one has P measure less than 2^{-k}. The convergence property from Step 2 gives an $n(k)$ such that

$$\mathbb{P}_*\{X_n \in B\} \geq (1 - 2^{-k})PB \qquad \text{for all } B \text{ in } \pi(k), \text{ all } n \geq n(k).$$

Step 4. Assuming that $1 = n(0) < n(1) < \cdots$, define $\gamma(n)$ to equal the k for which $n(k) \leq n < n(k+1)$. For $\gamma(n) = k$ and each B_i in $\pi(k)$, find measurable A_{ni} with $A_{ni} \subseteq X_n^{-1}B_i$ and

$$\mathbb{P}A_{ni} = \mathbb{P}_*\{X_n \in B_i\}.$$

Define a probability measure μ_n on $\mathcal{A}$ by

$$2^{-\gamma(n)}\mu_n(\cdot) + \left(1 - 2^{-\gamma(n)}\right)\sum_i PB_i\,\mathbb{P}(\cdot \mid A_{ni}) = \mathbb{P}(\cdot).$$

The inequality from Step 3, and the inequality

$$\mathbb{P}A \geq \sum_i \mathbb{P}(A \mid A_{ni})\mathbb{P}A_{ni} \qquad \text{for measurable } A,$$

ensure that μ_n is nonnegative. For each t in $[0,1]$ and each x in $\mathcal{X}$ define a probability measure $K_n(t, x, \cdot)$ on $\mathcal{A}$ by

$$\text{(9.6)} \qquad K_n(t,x,\cdot) = \begin{cases} \mathbb{P}(\cdot\,|A_{ni}) & \text{if } \ t \leq 1 - 2^{-\gamma(n)} \quad \text{and} \quad x \in B_i \in \pi(\gamma(n)), \\ \mu_n(\cdot) & \text{if } \ t > 1 - 2^{-\gamma(n)}. \end{cases}$$

The kernel K_n will provide a randomization mechanism for generating $\mathbb{P}$, starting from a t distributed uniformly on $[0,1]$ independently of an x distributed according to P. Specifically, if λ denotes Lebesgue measure on $[0,1]$, then

$$\mathbb{P}A = \iint K_n(t,x,A)\lambda(dt)P(dx)$$

for each A in $\mathcal{A}$.

Step 5. Define $\widetilde{\Omega}$ as the product space $[0,1] \otimes \mathcal{X} \otimes \Omega^{\mathbb{N}}$, where $\mathbb{N} = \{1,2,\dots\}$. Equip it with its product σ-field. For t in $[0,1]$ and x in $\mathcal{X}$ define the probability measure $K(t,x,\cdot)$ on the product σ-field of $\Omega^{\mathbb{N}}$ as a product

$$K(t,x,\cdot) = \prod_n K_n(t,x,\cdot).$$

With λ denoting Lebesgue measure on $[0,1]$, define $\widetilde{\mathbb{P}}$ on the product σ-field of $\widetilde{\Omega}$ by

$$\widetilde{\mathbb{P}}(\cdot) = \lambda \otimes P \otimes K.$$

That is, for $I \in \mathcal{B}[0,1]$ and $B \in \mathcal{B}(\mathcal{X})$ and C in the product σ-field of $\Omega^{\mathbb{N}}$,

$$\widetilde{\mathbb{P}}(I \otimes B \otimes C) = \iint \{t \in I, x \in B\} K(t,x,C)\lambda(dt)P(dx).$$

Some measurability details must be checked to ensure that $\widetilde{\mathbb{P}}$ is well defined.

Step 6. Define maps ϕ_n (from $\widetilde{\Omega}$ into Ω), and $\widetilde{X}$ (from $\widetilde{\Omega}$ into $\mathcal{X}$), and $\widetilde{X}_n$ (from $\widetilde{\Omega}$ into $\mathcal{X}$) by

$$\begin{aligned} \phi_n(t,x,\omega_1,\omega_2,\dots) &= \omega_n, \\ \widetilde{X}(t,x,\omega_1,\omega_2,\dots) &= x, \\ \widetilde{X}_n(t,x,\omega_1,\omega_2,\dots) &= X_n(\omega_n). \end{aligned}$$

Use the representations from Steps 4 and 5 to verify that $\widetilde{\mathbb{P}}\phi_n^{-1} = \mathbb{P}$ and $\widetilde{\mathbb{P}}\widetilde{X}^{-1} = P$.

Step 7. Temporarily fix a value of k. Let B_0 be the member of $\pi(k)$ that might have diameter greater than $1/k$. Define the subset $\widetilde{\Omega}_k$ of $\widetilde{\Omega}$ to consist of all those $\widetilde{\omega}$ for which $t \le 1 - 2^{-k}$ and $x \in B_i$ for some $i \ge 1$ and $\omega_n \in A_{ni}$ for that same i, for all n in the range $n(k) \le n < n(k+1)$. By the construction of $\pi(k)$,

$$d(\widetilde{X}_n(\widetilde{\omega}), \widetilde{X}(\widetilde{\omega})) \le 1/k \qquad \text{for } n(k) \le n < n(k+1) \text{ and } \widetilde{\omega} \text{ in } \widetilde{\Omega}_k.$$

If $n(k) \le n < n(k+1)$, define $\delta_n(\widetilde{\omega})$ to equal $1/k$ on $\widetilde{\Omega}_k$ and ∞ elsewhere. By the Borel-Cantelli lemma, the δ_n sequence converges to zero almost surely, because the construction of $\widetilde{\mathbb{P}}$ ensures $\widetilde{\mathbb{P}}\widetilde{\Omega}_k^c \le 2(1/2)^k$.

Step 8. Prove that each ϕ_n is perfect. Let H be a bounded function on Ω, and let g be a bounded, measurable function on $\widetilde{\Omega}$ for which

$$g(\widetilde{\omega}) \ge H(\omega_n) \qquad \text{for all} \quad \widetilde{\omega} = (t,x,\omega_1,\omega_2,\dots).$$

Establish (9.3) by finding a measurable function g^* on Ω for which $\widetilde{\mathbb{P}}g \ge \mathbb{P}g^*$ and

$$g^*(\omega_n) \ge H(\omega_n) \qquad \text{for all } \omega_n.$$

For fixed t, x, and ω_n, let $g_n(t,x,\omega_n)$ be the measurable function obtained by integrating g with respect to the product of all those $K_i(t,x,\cdot)$ with $i \neq n$. Then $\widetilde{\mathbb{P}}g_n = \widetilde{\mathbb{P}}g$ and

$$g_n(t,x,\omega_n) \geq H(\omega_n) \qquad \text{for all } t,x,\omega_n.$$

Now comes the crucial argument. The kernel K_n depends on (t,x) in a very simple way. There is a finite partition of $[0,1] \otimes \mathcal{X}$ into measurable sets D_α, and there are probability measures m_α on Ω, such that

$$K_n(t,x,\cdot) = \sum_\alpha \{(t,x) \in D_\alpha\} m_\alpha(\cdot).$$

Define g^* by

$$g^*(\omega_n) = \min_\alpha P \otimes \lambda\left(g_n(t,x,\omega_n) \mid (t,x) \in D_\alpha\right),$$

with the minimum running over those α for which $P \otimes \lambda(D_\alpha) > 0$. Finiteness of the $\{D_\alpha\}$ partition ensures that g^* is measurable. It is easy to check that it also satisfies the desired inequalities.

REMARKS. Typically measurability is not a major concern in specific problems. Nevertheless, it is highly desirable that a general theory for convergence in distribution, free from unnatural measurability constraints, should exist. Unfortunately, Hoffmann-Jørgensen's (1984) theory was presented in a manuscript for a book that has not yet been published. However, detailed explanations of some parts of the theory have appeared in the papers of Andersen (1985a, 1985b) and Andersen and Dobrić (1987, 1988).

Many measure theoretic details have been omitted from the outline of the proof of the Representation Theorem, but otherwise it is quite similar to the version in Chapter IV of Pollard (1984), which was based on Dudley's (1968) original paper. Dudley (1985) discussed the notion of a perfect map in some detail, and also showed how slippery a concept almost sure convergence can be for nonmeasurable random processes.