Introduction to Empirical Processes and Semiparametric Inference Lecture 09: Stochastic Convergence, Continued

Michael R. Kosorok, Ph.D.

Professor and Chair of Biostatistics

Professor of Statistics and Operations Research

University of North Carolina-Chapel Hill

Stochastic Convergence

Today, we will discuss the following basic concepts:

- Spaces of Bounded Functions
- Other Modes of Convergence
- Continuous Mapping Revisited
- Outer Almost Sure Representation Theorem
- Example: The Cramér-von Mises Statistic

Spaces of Bounded Functions

Now we consider stochastic processes X_n with index set T.

The natural metric space for weak convergence in this setting is $\ell^\infty(T)$.

A nice feature of this setting is the fact that asymptotic measurability of X_n follows from asymptotic measurability of $X_n(t)$ for each $t \in T$:

Lemma 7.16. Let the sequence of maps X_n in $\ell^\infty(T)$ be asymptotically tight. Then X_n is asymptotically measurable if and only if $X_n(t)$ is asymptotically measurable for each $t \in T$.

Theorem 7.17. The sequence X_n converges to a tight limit in $\ell^\infty(T)$ if and only if X_n is asymptotically tight and all finite-dimensional marginals converge weakly to limits.

Moreover, if X_n is asymptotically tight and all of its finite-dimensional marginals

$$(X_n(t_1), \ldots, X_n(t_k))$$

converge weakly to the marginals

$$(X(t_1), \ldots, X(t_k))$$

of a stochastic process X, then there is a version of X such that $X_n \rightsquigarrow X$ and X resides in $UC(T, \rho)$ for some semimetric ρ making T totally bounded.

Recall from Chapter 2:

Theorem 2.1. X_n *converges weakly to a tight* X *in* $\ell^\infty(T)$ *if and only if:*

(i) For all finite $\{t_1, \ldots, t_k\} \subset T$*, the multivariate distribution of* $\{X_n(t_1), \ldots, X_n(t_k)\}$ *converges to that of* $\{X(t_1), \ldots, X(t_k)\}$*.*

(ii) There exists a semimetric ρ *for which* T *is totally bounded and*

$$\lim_{\delta \downarrow 0} \limsup_{n \to \infty} P^* \left\{ \sup_{s,t \in T \text{ with } \rho(s,t) < \delta} |X_n(s) - X_n(t)| > \epsilon \right\} = 0,$$

for all $\epsilon > 0$*.*

The proof of Theorem 2.1 (which we omit) shows that whenever $X_n \rightsquigarrow X$ and X is tight, any semimetric ρ defining a σ-compact set $UC(T, \rho)$ with

$$\text{pr}(X \in UC(T, \rho)) = 1$$

will also result in X_n being uniformly ρ-equicontinuous in probability.

What is not clear at this point is the converse:

- that any semimetric ρ_* which enables uniform asymptotic equicontinuity
- will also define a σ-compact set $UC(T, \rho_*)$ wherein X resides with probability 1.

The following theorem settles this question:

Theorem 7.19. Assume $X_n \rightsquigarrow X$ *in* $\ell^\infty(T)$*, and let* ρ *be a semimetric making* (T, ρ) *totally bounded. TFAE:*

(i) X_n *is asymptotically uniformly* ρ*-equicontinuous in probability.*

(ii) pr$(X \in UC(T, \rho)) = 1$.

An interesting consequence of Theorems 2.1 and 7.19, in conjunction with Lemma 7.4, happens when $X_n \rightsquigarrow X$ in $\ell^\infty(T)$ and X is a tight Gaussian process.

Recall from Section 7.1 the semimetric

$$\rho_p(s,t) \equiv (E|X(s) - X(t)|^p)^{1/(p\vee 1)},$$

for any $p \in (0, \infty)$.

Then for any $p \in (0, \infty)$,

- (T, ρ_p) is totally bounded,
- the sample paths of X are ρ_p-continuous, and
- X_n is asymptotically uniformly ρ_p-equicontinuous in probability.

While any value of $p \in (0, \infty)$ will work, the choice $p = 2$ (the “standard deviation” metric) is often the most convenient to work with.

We now point out an equivalent—but sometimes easier to work with—condition for X_n to be asymptotically uniformly ρ-equicontinuous in probability:

Lemma 7.20. Let X_n be a sequence of stochastic processes indexed by T. TFAE:

(i) *There exists a semimetric ρ making T totally bounded and for which X_n is uniformly ρ-equicontinuous in probability.*

(ii) *For every $\epsilon, \eta > 0$, there exists a finite partition $T = \cup_{i=1}^k T_i$ such that*

$$\limsup_{n\to\infty} \mathsf{P}^* \left(\sup_{1\le i\le k} \sup_{s,t\in T_i} |X_n(s) - X_n(t)| > \epsilon \right) \quad < \quad \eta.$$

Other Modes of Convergence

Recall from Chapter 2 the following modes of convergence:

convergence in probability: $X_n \xrightarrow{\text{P}} X$ if $P\{d(X_n, X)^* > \epsilon\} \to 0$ for every $\epsilon > 0$.

outer almost sure convergence: $X_n \xrightarrow{\text{as}*} X$ if there exists a sequence Δ_n of measurable random variables such that

- $d(X_n, X) \le \Delta_n$ and
- $P\{\limsup_{n\to\infty} \Delta_n = 0\} = 1$.

We now introduce two additional modes of convergence which can be useful in some settings:

almost uniform convergence: X_n *converges almost uniformly* to X if,

for every $\epsilon > 0$, there exists a measurable set A such that

$$\text{pr}(A) \geq 1 - \epsilon$$

and $d(X_n, X) \to 0$ uniformly on A.

almost sure convergence: X_n *converges almost surely* to X if

$$\mathsf{P}_*(\lim_{n\to\infty} d(X_n, X) = 0) = 1.$$

Note that an important distinction between almost sure and outer almost sure convergence is that, in the latter mode, there must exist a measurable majorant of $d(X_n, X)$ which goes to zero.

This distinction is quite important because almost sure convergence does not in general imply convergence in probability when $d(X_n, X)$ is not measurable.

For this reason, we do generally not use the almost sure convergence mode except rarely.

Here are key relationships among the three remaining modes:

Lemma 7.21. Let $X_n, X : \Omega \mapsto \mathbb{D}$ be maps with X Borel measurable. Then

(i) $X_n \stackrel{\text{as}*}{\to} X$ implies $X_n \stackrel{\text{P}}{\to} X$.

(ii) $X_n \stackrel{\text{P}}{\to} X$ if and only if every subsequence $X_{n'}$ has a further subsequence $X_{n''}$ such that $X_{n''} \stackrel{\text{as}*}{\to} X$.

(iii) $X_n \stackrel{\text{as}*}{\to} X$ if and only if X_n converges almost uniformly to X if and only if $\sup_{m\geq n} d(X_m, X) \stackrel{\text{P}}{\to} 0$.

Since almost uniform convergence and outer almost sure convergence are equivalent for sequences, we will not use the almost uniform mode much.

The next lemma describes several important relationships between weak convergence and convergence in probability.

Before presenting it, we need to extend the definition of convergence in probability—in the setting where the limit is a constant—to allow the probability spaces involved to change with n as is already permitted for weak convergence.

We denote this modified convergence $X_n \stackrel{\mathsf{P}}{\to} c$, and distinguish it from the previous form of convergence in probability only by context.

Lemma 7.23. Let $X_n, Y_n : \Omega_n \mapsto \mathbb{D}$ *be maps,* $X : \Omega \mapsto \mathbb{D}$ *be Borel measurable, and* $c \in \mathbb{D}$ *be a constant. Then*

(i) If $X_n \rightsquigarrow X$ *and* $d(X_n, Y_n) \overset{\mathsf{P}}{\rightarrow} 0$*, then* $Y_n \rightsquigarrow X$*.*

(ii) $X_n \overset{\mathsf{P}}{\rightarrow} X$ *implies* $X_n \rightsquigarrow X$*.*

(iii) $X_n \overset{\mathsf{P}}{\rightarrow} c$ *if and only if* $X_n \rightsquigarrow c$*.*

Proof. We first prove (i).

Let $F \subset \mathbb{D}$ be closed, and fix $\epsilon > 0$. Then

$$\begin{aligned}\limsup_{n\to\infty} \mathsf{P}^*(Y_n \in F) &= \limsup_{n\to\infty} \mathsf{P}^*(Y_n \in F,\, d(X_n, Y_n)^* \leq \epsilon) \\ &\leq \limsup_{n\to\infty} \mathsf{P}^*(X_n \in \overline{F^\epsilon}) \\ &\leq P(X \in \overline{F^\epsilon}).\end{aligned}$$

The result follows by letting $\epsilon \downarrow 0$.

Now assume $X_n \stackrel{\mathsf{P}}{\rightarrow} X$.

Since $X \rightsquigarrow X$, $d(X, X_n) \stackrel{\mathsf{P}}{\rightarrow} 0$ implies $X_n \rightsquigarrow X$ by (i).

This follows by taking in (i) $X_n = X$ and $Y_n = X_n$.

Thus (ii) follows.

We now prove (iii).

$X_n \xrightarrow{\text{P}} c$ implies $X_n \rightsquigarrow c$ by (ii).

Now assume $X_n \rightsquigarrow c$, and fix $\epsilon > 0$.

Note that

$$\mathsf{P}^*(d(X_n, c) \geq \epsilon) = \mathsf{P}^*(X_n \notin B(c, \epsilon)),$$

where $B(c, \epsilon)$ is the open ϵ-ball around $c \in \mathbb{D}$.

By the portmanteau theorem,

$$\limsup_{n \to \infty} \mathsf{P}^*(X_n \notin B(c, \epsilon)) \leq \mathsf{pr}(X \notin B(c, \epsilon)) = 0.$$

Thus $X_n \xrightarrow{\mathsf{P}} c$ since ϵ is arbitrary, and (iii) follows.$\square$

Continuous Mapping Revisited

We now present a generalized continuous mapping theorem for sequences of maps g_n which converge to g.

For example, suppose T is high dimensional.

The computational burden of computing the supremum of $X_n(t)$ over T may be reduced by choosing a finite mesh T_n which closely approximates T.

Theorem 7.24 (Extended continuous mapping). Let $\mathbb{D}_n \subset \mathbb{D}$ *and* $g_n : \mathbb{D}_n \mapsto \mathbb{E}$ *satisfy the following: if* $x_n \to x$ *with* $x_n \in \mathbb{D}_n$ *for all* $n \geq 1$ *and* $x \in \mathbb{D}_0$*, then*

$$g_n(x_n) \to g(x),$$

where $\mathbb{D}_0 \subset \mathbb{D}$ *and* $g : \mathbb{D}_0 \mapsto \mathbb{E}$*.*

Let X_n *be maps taking values in* $\mathbb{D}_n$*, and let* X *be Borel measurable and separable with* $\mathsf{P}_*(X \in \mathbb{D}_0) = 1$*. Then*

(i) $X_n \rightsquigarrow X$ *implies* $g_n(X_n) \rightsquigarrow g(X)$*.*

(ii) $X_n \stackrel{\mathsf{P}}{\to} X$ *implies* $g_n(X_n) \stackrel{\mathsf{P}}{\to} g(X)$*.*

(iii) $X_n \stackrel{\mathsf{as}*}{\to} X$ *implies* $g_n(X_n) \stackrel{\mathsf{as}*}{\to} g(X)$*.*

The following alternative theorem does not require separability of X and is a supplement to Theorem 7.7:

Theorem 7.25. Let $g : \mathbb{D} \mapsto \mathbb{E}$ *be continuous at all points in* $\mathbb{D}_0 \subset \mathbb{D}$*, and let* X *be Borel measurable with* $\mathrm{P}_*(X \in \mathbb{D}_0) = 1$*. Then*

(i) $X_n \overset{\mathrm{P}}{\to} X$ *implies* $g(X_n) \overset{\mathrm{P}}{\to} g(X)$*.*

(ii) $X_n \overset{\mathrm{as}*}{\to} X$ *implies* $g(X_n) \overset{\mathrm{as}*}{\to} g(X)$*.*

Outer Almost Sure Representation Theorem

We now present a useful outer almost sure representation result for weak convergence.

Such representations allow the transformation of certain weak convergence problems into problems about convergence of fixed sequences.

We give an illustration of this approach in the sketch of the proof of Proposition 7.27 below.

Theorem 7.26. Let $X_n : \Omega_n \mapsto \mathbb{D}$ *be a sequence of maps, and let* X_∞ *be Borel measurable and separable.*

If $X_n \rightsquigarrow X_\infty$*, then there exists a probability space* $(\tilde{\Omega}, \tilde{\mathcal{A}}, \tilde{P})$ *and maps* $\tilde{X}_n : \tilde{\Omega} \mapsto \mathbb{D}$ *with*

(i) $\tilde{X}_n \stackrel{\text{as}*}{\to} \tilde{X}_\infty$*;*

(i) $\mathsf{E}^* f(\tilde{X}_n) = \mathsf{E}^* f(X_n)$*, for every bounded* $f : \mathbb{D} \mapsto \mathbb{R}$ *and all* $1 \leq n \leq \infty$*.*

Moreover, $\tilde{X}_n$ *can be chosen to be equal to* $X_n \circ \phi_n$*, for all* $1 \leq n \leq \infty$*, where the* $\phi_n : \tilde{\Omega} \mapsto \Omega_n$ *are measurable and perfect maps and* $P_n = \tilde{P} \circ \phi_n^{-1}$*.*

Recall our previous discussion about perfect maps.

In the setting of the above theorem, if the $\tilde{X}_n$ are constructed from the perfect maps ϕ_n, then

$$\left[f(\tilde{X}_n)\right]^* = [f(X_n)]^* \circ \phi_n$$

for all bounded $f : \mathbb{D} \mapsto \mathbb{R}$.

Thus the equivalence between $\tilde{X}_n$ and X_n can be made much stronger than simply equivalence in law.

The following proposition can be useful in studying weak convergence of certain statistics which can be expressed as stochastic integrals.

For example, the Wilcoxon and Cramér-von Mises statistics can be expressed in this way.

Proposition 7.27. Let $X_n, G_n \in D[a, b]$ *be stochastic processes with* $X_n \rightsquigarrow X$ *and* $G_n \xrightarrow{\mathrm{P}} G$ *in* $D[a, b]$*, where* X *has continuous sample paths,* G *is fixed, and* G_n *and* G *have total variation bounded by* $K < \infty$*. Then*

$$\int_a^{(\cdot)} X_n(s) dG_n(s) \rightsquigarrow \int_a^{(\cdot)} X(s) dG(s)$$

in $D[a, b]$*.*

Sketch of Proof. First, Slutsky's theorem and Lemma 7.23 establish that $(X_n, G_n) \rightsquigarrow (X, G)$.

Next, Theorem 7.26 tells us that there exists a new probability space and processes $\tilde{X}_n$, $\tilde{X}$, $\tilde{G}_n$ and $\tilde{G}$ which have the same outer integrals for bounded functions as X_n, X, G_n and G, respectively, but which also satisfy

$$(\tilde{X}_n, \tilde{G}_n) \overset{\text{as}*}{\to} (\tilde{X}, \tilde{G}).$$

Fix $\epsilon > 0$.

By the continuity of the sample paths of $\tilde{X}$ over the compact $[a, b]$ (which can include extended reals), we have that there exists an integer $1 \leq m < \infty$ and a partition of $[a, b]$, $a = t_0 < t_1 < \cdots < t_m = b$, where

$$\max_{1 \leq j \leq m} \sup_{s,t \in (t_{j-1}, t_j]} |\tilde{X}(s) - \tilde{X}(t)| \leq \epsilon.$$

Define $\tilde{X}_m \in D[a,b]$ such that $\tilde{X}_m(a) = \tilde{X}(a)$ and

$$\tilde{X}_m(t) \equiv \sum_{j=1}^{m} \mathbf{1}\{t_{j-1} < t \leq t_j\}\tilde{X}(t_j),$$

for $t \in (a,b]$.

Note that the integral of interest evaluated at $t = a$ is zero since the set $(a,a]$ is empty.

We now have, for any $t \in (a, b]$, that

$$\begin{aligned}
&\left|\int_a^t \tilde{X}_n(s)d\tilde{G}_n(s) - \int_a^t \tilde{X}(s)d\tilde{G}(s)\right| \\
&\leq \int_a^b \left|\tilde{X}_n(s) - \tilde{X}(s)\right| \times |d\tilde{G}_n(s)| \\
&\quad + \int_a^b \left|\tilde{X}_m(s) - \tilde{X}(s)\right| \times |d\tilde{G}_n(s)| \\
&\quad + \left|\int_a^t \tilde{X}_m(s)\left\{d\tilde{G}_n(s) - d\tilde{G}(s)\right\}\right| \\
&\quad + \int_a^t \left|\tilde{X}_m(s) - \tilde{X}(s)\right| \times |d\tilde{G}(s)|
\end{aligned}$$

Note that the first term $\rightarrow 0$, while the second and fourth terms are bounded by $K\epsilon$.

By definition of $\tilde{X}_m$, the third term equals

$$\begin{aligned}\left|\sum_{j=1}^{m}\tilde{X}(t_j)\int_{(t_{j-1},t_j]\cap(a,t]}\left\{d\tilde{G}_n(s)-d\tilde{G}(s)\right\}\right| &\leq m\|\tilde{X}\|_\infty\|\tilde{G}_n-\tilde{G}\|_\infty \\ &\to 0.\end{aligned}$$

Thus all four parts summed are asymptotically bounded by $2K\epsilon$.

Since ϵ was arbitrary, we conclude that

$$\int_a^{(\cdot)} \tilde{X}_n(s)d\tilde{G}_n(s) \overset{\text{as}*}{\to} \int_a^{(\cdot)} \tilde{X}(s)d\tilde{G}(s).$$

Now part (ii) of Lemma 7.23 yields that we can replace the outer almost sure convergence with week convergence.

This now implies weak convergence in the original probability space, and the proof is complete.$\square$

Example: The Cramér-von Mises Statistic

Let $X_1, \ldots, X_n$ be i.i.d. real random variables with continuous distribution F and corresponding empirical process $\hat{F}_n$.

The Cramér-von Mises Statistic for testing the null hypothesis $H_0 : F = F_0$ has the form:

$$T_n = n \int_{-\infty}^{\infty} \left(\hat{F}_n(t) - F_0(t) \right)^2 d\hat{F}_n(t).$$

In this case, $X_n = n\left(\hat{F}_n - F_0\right)^2$, $G_n = \hat{F}_n$, and $[a, b] = [-\infty, \infty]$.

Standard empirical process results yield that when $F = F_0$ (i.e., under the null), $X_n \rightsquigarrow X$ and $G_n \xrightarrow{\mathsf{P}} G$ in $D[a, b]$, where

$$X(t) = \mathbb{B}^2(F_0(t)),$$

$\mathbb{B}$ is a standard Brownian bridage, and $G = F_0$.

Note that X has continuous sample paths and the total variation of G_n and G is at most 1.

Thus all of the conditions of Proposition 7.27 are met, and thus

$$T_n \rightsquigarrow \int_0^1 \mathbb{B}^2(t)dt,$$

which is a pivotal limit.

The distribution of this pivotal can be written as an infinite series which is not difficult to compute (see, e.g., Tolmatz, 2001, *Annals of Probability*): the critical value for $\alpha = 0.05$ is approximately 0.220 (I think).