Empirical Processes: General Weak Convergence Theory

Moulinath Banerjee

May 18, 2010

1 Extended Weak Convergence

The lack of measurability of the empirical process with respect to the sigma-field generated by the 'natural' l^∞ metric, as illustrated in the previous notes, needs an extension of the standard weak convergence theory that can handle situations where the converging stochastic processes may no longer be measurable (though the limit will be a tight Borel measurable random element). Of course, an alternative solution is to use a different metric that will make the converging processes measurable, which in the scenario of Donsker's theorem, Version 1 in the previous notes is achieved by equipping the space $D[0,1]$ with the Skorohod (metric) topology. However, for empirical processes that assume values in very general (and often more complex spaces) such cute generalizations are not readily achievable and the easier way out is to keep the topology simple and tackle the measurability issues. Of course, any generalization of the notion of weak convergence must allow a powerful continuous mapping theorem.

In what follows, we briefly describe this extended weak convergence theory following the development in Section 9 of Pollard (1990). An extensive coverage is available in Chapter 1 on stochastic convergence in van der Vaart and Wellner (1996) which will be referred to sparingly. Our goal here is to develop the extended weak convergence ideas to the extent required for a proper understanding of the characterization of weak convergence of the empirical process in terms of finite-dimensional convergence and asymptotic equicontinuity. Our emphasis will be on building the tools needed to verify these conditions in different situations and as we will see, such tools typically involve quantifying the 'largeness' of the class of functions indexing the empirical process.

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

$$\mathbb{P}\, f(X_n) \to P\, f \quad \forall f \in \mathcal{U}(\mathcal{X})\,,$$

$\mathcal{U}(\mathcal{X})$ denoting the class of bounded uniformly continuous real-valued functions on $\mathcal{X}$. If X_n's are not necessarily measurable then neither is $f(X_n)$ and the above definition is inadequate. However, we can still define the *outer expectation*: for each bounded real-valued H on Ω, set:

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

Similarly, define the inner expectation:

$$\mathbb{P}_\star\, H = \sup\, \{\mathbb{P}\, h : H \geq h \;\text{ and }\; h \;\; integrable\}\,.$$

Check that $P^\star\, H = -P_\star(-H)$. We can now formally extend the definition of weak-convergence to accommodate non-measurability.

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 defined on $\mathcal{B}(\mathcal{X})$, then $X_n \to_d P$ is defined to mean: $\mathbb{P}^\star(f(X_n)) \to P\, f$ for every $f \in \mathcal{U}(\mathcal{X})$.

Note that the definition could also have been stated in terms of inner expectations. It could also have been stated in terms of a limiting Borel-measurable random element X (a measurable map from some probability space, say $(\Omega', \mathcal{A}', \mathbb{P}')$ into $(\mathcal{X}, \mathcal{B}(\mathcal{X}))$ by requiring $\mathbb{P}^\star(f(X_n)) \to \mathbb{P}'\, f(X)$ for all $f \in \mathcal{U}(\mathcal{X})$. Also note that the definition here differs slightly from the one used in van der Vaart and Wellner (1996) who require f to vary among all bounded continuous functions. However, not much is lost by changing to 'bounded uniformly continuous' instead.

An example: If $\{Y_n\}$ is a sequence of random elements assuming values in a metric space $(\mathcal{Y}, e)$ converging in probability to a constant y, i.e. $\mathbb{P}^\star(e(Y_n, y) > \delta) \to 0$ for every $\delta > 0$, and if $X_n \to_d X$, then $(X_n, Y_n) \to_d (X, y)$ in the product space $\mathcal{X} \times \mathcal{Y}$ (equipped with the product topology which is generated, say, by the metric $\rho((x_1, y_1), (x_2, y_2)) = d(x_1, x_2) + e(y_1, y_2)$). To see this, suppose that f is uniformly continuous on $\mathcal{X} \times \mathcal{Y}$ and bounded in absolute value by a constant M. Given $\epsilon > 0$, we can find $\delta > 0$ such that if two points in the domain of f are at distance less than δ apart in the ρ metric then the function values at these points differ in absolute magnitude by less than ϵ. Conclude that:

$$f(X_n, Y_n) \leq f(X_n, y) + \epsilon + 2\, M\, 1\{e(Y_n, y) > \delta\}\,.$$

Taking outer expectations gives:

$$\mathbb{P}^\star\, f(X_n, Y_n) \leq \mathbb{P}\, f(X_n, y) + 2\, M\, \mathbb{P}^\star(e(Y_n, y) > \delta) + \epsilon\,,$$

using the facts that $\mathbb{P}^\star(H_1 + H_2) \leq \mathbb{P}^\star(H_1) + \mathbb{P}^\star(H_2)$ with equality if either is measurable. Letting n go to infinity and using that Y_n converges in probability to y, we have:

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

Now, since f is arbitrary, it is certainly the case that:

$$\limsup \mathbb{P}^\star\, (-f(X_n, Y_n)) \leq \mathbb{P}\, (-f(X, y)) + \epsilon\,,$$

which shows that:

$$-\liminf \mathbb{P}_\star\, f(X_n, Y_n) \leq -\mathbb{P}\, f(X, y) + \epsilon\,.$$

Putting things together, we get:

$$\mathbb{P}\, f(X, y) - \epsilon \leq \liminf \mathbb{P}_\star\, f(X_n, Y_n) \leq \limsup \mathbb{P}^\star\, f(X_n, Y_n) \leq \mathbb{P}\, f(X, y) + \epsilon\,.$$

Since $\epsilon > 0$ is arbitrary, we conclude that $\mathbb{P}\, f(X, y) = \liminf \mathbb{P}_\star\, f(X_n, Y_n) = \limsup \mathbb{P}^\star\, f(X_n, Y_n)$. This shows that $\mathbb{P}^\star\, f(X_n, Y_n)$ converges to $\mathbb{P}\, f(X, y)$, as we sought to establish.

2 A Representation Theorem

The extended weak-convergence theory admits a nice representation theorem originally due to Dudley (1985). We next discuss an elaborate statement of this theorem and use it to formally prove a continuous mapping theorem. An almost sure representation for X_n, converging in distribution to a Borel measure P, would require the construction of a common probability space $(\tilde{\Omega}, \tilde{\mathcal{A}}, \tilde{\mathbb{P}})$ and random elements $\tilde{X}_n$ defined on this space as well as a measurable random element $\tilde{X}$ into $\mathcal{X}$ with distribution P, such that $\tilde{X}_n$ has the 'same distribution" as X_n and furthermore $\tilde{X}_n$ 'converges almost surely' to $\tilde{X}$. However, note that the term 'same distribution' is ill-defined since the X_n's are not measurable and therefore do not have a distribution in the proper sense of the term. Also, the set of points in Ω for which $\tilde{X}_n$'s have a limit may not be measurable. These difficulties can however be circumvented using the notion of a *perfect map*.

In Dudley's representation theorem, perfect maps ϕ_n are constructed from a new probability space $(\tilde{\Omega}, \tilde{\mathcal{A}}, \tilde{\mathbb{P}})$ to the space $(\Omega, \mathcal{A}, \mathbb{P})$ with the *defining property* that ϕ_n is measurable and that for every bounded map H from Ω to $\mathbb{R}$, $\mathbb{P}^\star H = \tilde{\mathbb{P}}^\star [H \circ \phi_n]$. Note that this property implies that $\tilde{\mathbb{P}} \circ \phi_n^{-1} = \mathbb{P}$ (but is much stronger than this measure preservation property). To see this, let $H = 1_A$ for $A \in \mathcal{A}$. Then $\mathbb{P}(A) = \mathbb{P}\, 1_A = \tilde{\mathbb{P}}\, [1_A \circ \phi_n] = \tilde{\mathbb{P}}\, 1_{\phi_n^{-1}(A)} = \tilde{\mathbb{P}}(\phi_n^{-1}(A))$. The versions of X_n in the almost sure representation are defined by $\tilde{X}_n(\tilde{\omega}) = X_n(\phi_n(\tilde{\omega}))$. Using the defining property of perfect maps, it follows that $\mathbb{P}^\star\, g(X_n) = \tilde{\mathbb{P}}^\star[g(\tilde{X}_n)]$, regardless of the measurability properties of X_n. For a measurable X_n, the outer expectations in this equality can be replaced by ordinary expectations and the equality itself would follow from the measure-preservation property (this is the theorem of the unconscious statistician that one encounters in probability). However, measure preservation alone would not guarantee the equality in the absence of measurability, because in general, outer expectations only satisfy an inequality:

$$\mathbb{P}^\star\, H \geq \tilde{\mathbb{P}}^\star[H \circ \phi_n] \quad \forall \ \text{bounded} \ \ H \ \ \text{on} \ \ \Omega\,.$$

This follows from the fact that for any measurable $h \geq H$, $h \circ \phi_n \geq H \circ \phi_n$ but the set of measurable majorants of $H \circ \phi_n$ could be larger than $\{h \circ \phi_n : h \geq H \ \text{ and } \ \text{measurable}\}$. To establish the perfectness of ϕ_n, one needs to show that $\mathbb{P}^\star\, H \leq \tilde{\mathbb{P}}\, g$ for all measurable real $g \geq H \circ \phi_n$ and the reverse inequality follows by taking the infimum over all such g.

The Representation Theorem: If $X_n \to_d P$ in the extended sense discussed above and if the limit distribution P concentrates on a separable Borel subset $\mathcal{X}_0$ of $\mathcal{X}$, then there exists a probability space $(\tilde{\Omega}, \tilde{\mathcal{A}}, \tilde{\mathbb{P}})$ supporting measurable maps ϕ_n into $(\Omega, \mathcal{A})$ and a measurable map $\tilde{X}$ into $(\mathcal{X}_0, \mathcal{B}_{\mathcal{X}_0})$, such that:

(a) each ϕ_n is a perfect map: i.e. $\mathbb{P}^\star H = \tilde{\mathbb{P}}^\star(H \circ \phi_n)$ for every bounded H on Ω;

(b) $\tilde{\mathbb{P}}\, \tilde{X}^{-1} = P$ as measures on $\mathcal{B}(\mathcal{X}_0)$;

(c) there is a sequence of extended-real-valued measurable random variables $\{\tilde{\delta}_n\}$ defined from $(\tilde{\Omega}, \tilde{\mathcal{A}})$ to $([0,\infty], \mathcal{B}_{[0,\infty]})$, such that $d(\tilde{X}_n(\tilde{\omega}), \tilde{X}(\tilde{\omega})) \leq \tilde{\delta}_n(\tilde{\omega}) \to 0$ for almost every $\tilde{\omega}$, where $\tilde{X}_n(\tilde{\omega}) = X_n(\phi_n(\tilde{\omega}))$.

It is easy to see that Conditions (a), (b) and (c) imply that X_n converges weakly to P. To see this, it suffices to show that for any bounded uniformly continuous g from $\mathcal{X}$ to $\mathbb{R}$, $\mathbb{P}^\star g(X_n)$, which equals $\tilde{\mathbb{P}}^\star(g(\tilde{X}_n))$ by perfectness, converges to $\mathbb{P}\, g(X)$, which equals $\tilde{\mathbb{P}}\, g(\tilde{X})$. Given any $\epsilon > 0$, there exists $\eta > 0$ such that $d(\tilde{X}_n, \tilde{X}) < \eta$ implies that $|g(\tilde{X}_n) - g(\tilde{X})| < \epsilon$. Also, let M be an upper bound on the absolute value of g. Then,

$$\begin{aligned} |\tilde{\mathbb{P}}^\star g(\tilde{X}_n) - \tilde{\mathbb{P}}\, g(\tilde{X})| &\leq \tilde{\mathbb{P}}^\star |g(\tilde{X}_n) - g(\tilde{X})| \\ &\leq \tilde{\mathbb{P}}^\star [\epsilon\, 1(d(\tilde{X}_n, \tilde{X}) \leq \eta) + 2M\, 1(d(\tilde{X}_n, \tilde{X}) > \eta) \\ &\leq \epsilon + 2M\, \tilde{\mathbb{P}}^\star\, 1(\delta_n > \eta) \\ &\leq 2\,\epsilon \ \text{ eventually,} \end{aligned}$$

using the fact that $\tilde{\mathbb{P}}^\star\, 1(\delta_n > \eta) = \tilde{\mathbb{P}}\, 1(\delta_n > \eta) \to 0$, since δ_n converges to 0 almost surely. **Continuous Mapping Theorem:** Suppose that $X_n \to_d P$ with P concentrated on a separable Borel subset $\mathcal{X}_0$ of $\mathcal{X}$. Suppose that τ is a map into from $\mathcal{X}$ into another metric space $\mathcal{Y}$ such that: (a) the restriction of τ to $\mathcal{X}_0$ is Borel measurable, and, (b) τ is continuous at P-almost-all points of $\mathcal{X}_0$. Then $\tau(X_n)$ converges in distribution to the probability measure $P\,\tau^{-1}$ (which is defined on the Borel σ-field on the metric space $\mathcal{Y}$).

Proof of the continuous mapping theorem: Invoke the representation theorem to obtain the $\tilde{X}_n$'s, $\tilde{X}$ and the δ_n's. We need to show that for any $f \in \mathcal{U}(\mathcal{Y})$, $\mathbb{P}^\star f(\tau(X_n)) \to (P \circ \tau^{-1})\, f \equiv P\,(f \circ \tau) \equiv P\, h$ where $h = f \circ \tau$. Now $\mathbb{P}^\star f(\tau(X_n)) = \mathbb{P}^\star h(X_n) = \tilde{\mathbb{P}}^\star h(\tilde{X}_n)$ and it suffices to show that this converges to $\tilde{\mathbb{P}}\, h(\tilde{X})$.

Without loss of generality (why?), we can assume that $0 \leq h \leq 1$. Let $\epsilon > 0$ be pre–assigned. For each positive integer k, set G_k to be the set of all $x \in \mathcal{X}$ such that there exist points $y, z \in B(x; 1/k)$ with $|h(y) - h(z)| > \epsilon$. This is an open set: for if x_0 is in G_k and $y_0, z_0 \in B(x_0; 1/k)$ satisfy $|h(y) - h(z)| > \epsilon$, choosing $\delta_0 < \{(1/k - d(x_0, y_0)) \wedge (1/k - d(x_0, z_0))\}$, one ensures that every point in $B(x_0, \delta_0)$ is within distance $1/k$ of each of y_0 and z_0 and by definition is a member of G_k. The sets G_k are nested with G_1 containing G_2 which in turn contains G_3 and so. The sets G_k then shrink to a set, say G_∞, which excludes all continuity points of h and therefore all continuity points of τ. Thus $P(G_\infty) = 0$ and one can therefore find a G_k, for k sufficiently large, such that $P\, G_k < \epsilon$, i.e. $\tilde{\mathbb{P}}(\tilde{X} \in G_k) < \epsilon$. Note that the G_k's are open and therefore measurable sets.

Now, by the definition of G_k, if $\tilde{X}(\tilde{\omega}) \notin G_k$ and $\delta_n(\tilde{\omega}) < 1/k$, so that $d(\tilde{X}_n(\tilde{\omega}), \tilde{X}(\tilde{\omega})) < 1/k$, then $|h(\tilde{X}_n(\tilde{\omega})) - h(\tilde{X}(\tilde{\omega}))| \leq \epsilon$. Conclude that:

$$h(\tilde{X}_n) \leq (\epsilon + \tilde{h}(\tilde{X}))\, 1(\tilde{X} \notin G_k, \delta_n < 1/k) + 1(\tilde{X} \in G_k) + 1(\delta_n \geq 1/k)\,.$$

Taking expectations, we conclude that:

$$\tilde{\mathbb{P}}^\star(h(\tilde{X}_n)) \leq \epsilon + \tilde{\mathbb{P}}\, h(\tilde{X}) + \tilde{\mathbb{P}}(\tilde{X} \in G_k) + \tilde{\mathbb{P}}(\delta_n > 1/k)\,.$$

Since $\tilde{\mathbb{P}}(\tilde{X} \in G_k) < \epsilon$ and $\tilde{\mathbb{P}}(\delta_n > 1/k)$ converges to 0, conclude that:

$$\limsup \tilde{\mathbb{P}}^\star(h(\tilde{X}_n)) \leq 2\,\epsilon + \tilde{\mathbb{P}}\, h(\tilde{X})\,.$$

Since $\epsilon > 0$ is arbitrary, conclude that $\limsup \tilde{\mathbb{P}}^\star(h(X_n)) \leq \tilde{\mathbb{P}}\, h(\tilde{X})$. An analogous argument with h replaced by $1 - h$ leads to the conclusion that $\liminf \tilde{\mathbb{P}}_\star(h(X_n)) \geq \tilde{\mathbb{P}}\, h(\tilde{X})$. These two conclusions, in conjunction, show that $\lim \tilde{\mathbb{P}}^\star\, h(\tilde{X}_n) = \tilde{\mathbb{P}}\, h(\tilde{X})$. $\square$

A full version of the Portmanteau Theorem allowing for potential non-measurability is available in Chapter 1 of van der Vaart and Wellner. See also the more classical portmanteau theorem in Billingsley's weak convergence text. The various (equivalent) characterizations of weak given in the portmanteau theorem prove useful in different settings. A nice necessary and sufficient condition of weak convergence of stochastic processes living in $l^\infty(T)$ for an arbitrary set T will be established in the next section, and will use some of the alternative characterizations of (extended) weak convergence in general metric spaces.

3 Weak Convergence: Key Characterizations

We first state and prove a proposition that provides a useful characterization of sample-bounded stochastic processes indexed by a set T.

Proposition 0: Let $\{X(t), t \in T\}$ be a sample-bounded stochastic process. Then the finite dimensional distributions of X are those of a tight Borel probability measure on $l^\infty(T)$ *if and only if* there exists a pseudometric ρ on T for which (T, ρ) is totally bounded and such that X has a version with almost all its sample paths uniformly continuous for ρ.

Remark: If X and Y are tight Borel measurable maps into $l^\infty(T)$, then X and Y are equal in Borel law if and only if all corresponding marginals of X and Y are equal in law. For a proof see

Section 1.5 of van der Vaart and Wellner. Thus, tight Borel measures are uniquely characterized by their finite dimensional distributions.

Proof: (ONLY IF part) Suppose that the induced probability measure of X on $l^\infty(T)$ is a tight Borel measure P_X. (We tacitly assume here that X is a Borel measurable random element assuming values in $l^\infty(T)$.) Then, we can find an increasing sequence of compact subsets of $l^\infty(T)$, say $\{K_m\}$, such that $P_X(\cup_m K_m) = 1$. Let $K = \cup_m K_m$. Define a pseudo-metric ρ on T by:

$$\rho(s,t) = \sum_{m=1}^{\infty} 2^{-m} \left(1 \wedge \rho_m(s,t)\right),$$

where $\rho_m(s,t) = \sup_{x \in K_m} |x(s) - x(t)|$. It is easy to check that this is indeed a pseudo-metric. We claim that (T, ρ) is totally bounded. Thus, for every $\epsilon > 0$, we seek to produce a finite subset of T, say T_ϵ such that any $t \in T$ is at $\rho-$ distance less than ϵ for some member of T_ϵ.

First, choose k so large that $\sum_{m=k+1}^{\infty} 2^{-m} < \epsilon/4$. Let $x_1, x_2, \ldots, x_r$ be a finite subset of $\cup_{m=1}^{k} K_m = K_k$ such that it is $\epsilon/4$-dense in K_k for the supremum norm. Such a set exists by compactness of K_k in $l^\infty(T)$. Consider the subset A of $\mathbb{R}^r$ defined by: $\{(x_1(t), x_2(t), \ldots, x_r(t)) : t \in T\}$. This is bounded in $\mathbb{R}^r$ (since K_k is compact in $l^\infty(T)$ it is bounded for the sup norm) and therefore totally bounded (as boundedness and total boundedness are equivalent in Euclidean spaces). Thus, there exists a finite subset of A, say $\{(x_1(t_i), x_2(t_i), \ldots, x_r(t_i)) : 1 \le i \le N\}$ where $t_1, t_2, \ldots, t_N$ are points in T, that is $\epsilon/4$ dense in A with respect to the l^∞ norm on $\mathbb{R}^r$. Let $T_\epsilon = \{t_1, t_2, \ldots, t_N\}$. We seek to show that this set is ϵ-dense in T. To this end, for any $t \in T$, find $t_l \in T_\epsilon$ such that $\max_{1 \le i \le r} |x_i(t_l) - x_i(t)| < \epsilon/4$. For $m \le k$, consider $\rho_m(t, t_l) = \sup_{x \in K_m} |x(t) - x(t_l)|$. For any $x \in K_m$, we can find x_j such that $\|x - x_j\|_T < \epsilon/4$. Then $|x(t) - x(t_l)| \le 2\|x - x_j\| + |x_j(t) - x_j(t_l)| < 3\,\epsilon/4$. It follows that $\rho_m(t, t_l) \le 3\,\epsilon/4$ for $m \le k$. Now, by the choice of k and the definition of ρ, it follows readily that $\rho(t, t_l) < \epsilon$. Thus, it has been shown that (T, ρ) is totally bounded.

We next show that every $x \in K$ is uniformly continuous with respect to ρ: given $\epsilon > 0$, we need to produce a $\delta > 0$ such that $\rho(s,t) < \delta$ would imply $|x(s) - x(t)| < \epsilon$. Without loss of generality, let $\epsilon < 1$. Since $x \in K$, $x \in K_m$ for some $m \ge 1$. Now, $\rho(s,t) \ge \sum_{j=m}^{\infty} 2^{-j} (1 \wedge \rho_j(s,t)) \ge \sum_{j=m}^{\infty} 2^{-j} (1 \wedge \rho_m(s,t))$, since $\rho_j(s,t)$ increases with j (owing to the fact that the K_j's are increasing sets). It follows that: $1 \wedge \rho_m(s,t) \le 2^{m-1} \rho(s,t)$. Let $\delta = \epsilon/2^{m-1}$. Then $\rho(s,t) < \delta$ implies that $\rho_m(s,t) < \epsilon$ which, in turn, implies that $|x(s) - x(t)| < \epsilon$, establishing uniform continuity of x.

Since $P_X(K) = 1$, the identity map on $(l^\infty(T), \mathcal{B}, P_X)$ yields a version of X, say $\tilde{X}$, with almost all its sample paths in K and hence in $C_u(T, \rho)$.

(IF part) The total boundedness of (T, ρ) in conjunction with the fact that almost all sample paths of X are ρ-uniformly continuous immediately implies that the process X almost surely has bounded sample paths. We can clearly assume that all sample paths are uniformly continuous by redefining X on the set of ω's where uniform ρ–continuity is violated to be the identically zero function. This does not change the finite dimensional distributions of X. We continue to use X for this tweaked version. Now, consider X as a map from its domain to the space $C_u(T, \rho)$ equipped with its Borel σ-field with respect to the uniform metric. For $k \in \mathcal{N}$ and $(t_1, t_2, \ldots, t_k) \in T^k$, let the projection operator $\pi_{t_1,t_2,\ldots,t_k}(x) \equiv (x(t_1), x(t_2), \ldots, x(t_k))$. Consider the class of subsets of the Borel σ-field on $C_u(T, \rho)$ given by $\Pi \equiv \{\pi^{-1}_{t_1,t_2,\ldots,t_k}(B) : B \in \mathcal{B}_{\mathbb{R}^k}, k \in \mathcal{N}\}$. The inverse images of these sets under the map X are measurable in the probability space on which X is defined since, by assumption, each $X(t)$ is a (measuable) random variable. Since Π generates the Borel σ-field on $C_u(T, \rho)$, this implies that X is a Borel measurable map into $C_u(T, \rho)$. The induced probability measure P_X on the Borel subsets of $C_u(T, \rho)$ is tight by Ulam's theorem, since $C_u(T, \rho)$ is complete and separable. If we now consider X as a map into the larger space $l^\infty(T)$ equipped with the Borel σ-field with respect to the sup metric, it continues to induce a tight measure in this space, since compactness of a subset of $C_u(T, \rho)$ implies its compactness as a subset of $l^\infty(T)$. □

Remark: Note that though Π generates the Borel σ-field on $C_u(T, \rho)$, it does not generate the Borel sigma-field in $l^\infty(T)$, this being much larger. Measurability of all finite dimensional marginals of an arbitrary random map (from some probability space) assuming values in $l^\infty(T)$ therefore does not necessarily imply measurability of the map itself.

Our next theorem characterizes weak convergence of a sequence of stochastic processes to a tight Borel measurable random element assuming values in $l^\infty(T)$.

Theorem 3.1 *Let $\{X_n\}$ be a sequence of sample-bounded stochastic processes assuming values in $l^\infty(T)$. The following are equivalent:*

(1): *All finite dimensional distributions of the sample-bounded stochastic processes converge in law and there exists a pseudo-metric ρ on T such that: (1) (T, ρ) is totally bounded, and, (2) The processes X_n are asymptotically ρ–equicontinuous in probability, i.e: for every $\epsilon > 0$,*

$$\lim_{\delta \to 0} \limsup_{n \to \infty} P^\star\{\sup_{\rho(s,t)<\delta} |X_n(s) - X_n(t)| > \epsilon\} = 0\,.$$

(2): *There exists a process X with tight Borel probability distribution in $l^\infty(T)$ such that $X_n \Rightarrow X$ in the space $l^\infty(T)$.*

Furthermore, if (1) holds, then the process X in (2) – this being completely determined by the limiting finite dimensional distributions of $\{X_n\}$ – has a version with sample paths in $C_u(T,\rho)$: the subspace of $l^\infty(T)$ that comprises all functions that are uniformly continuous with respect to the semimetric ρ. On the other hand, if X in (2) has sample functions in $C_u(T,\gamma)$ for some pseudo-metric γ for which (T,γ) is totally bounded, then (1) holds with γ taking the role of the pseudo-metric ρ.

Proof: Assume that (1) holds. Since (T,ρ) is totally bounded, we can find a countably ρ-dense subset of T. Call this T_∞, and let T_k be a sequence of subsets of T_∞ that increase to T_∞. If $\{t_1,t_2,\ldots\}$ is a numbering of T_∞, we can take T_k to be $\{t_1,t_2,\ldots,t_k\}$. The limit distributions of the processes X_n are consistent and by the Kolmogorov consistency theorem, we can construct a stochastic process (a collection of random variables) $\{X(t) : t \in T\}$ defined on some $(\Omega,\mathcal{F},\mathbb{P})$ such that, for each $(s_1,s_2,\ldots,s_m) \in T^m$, $(X(s_1),X(s_2),\ldots,X(s_m))$ is distributed like $\nu_{s_1,s_2,\ldots,s_m}$, this being the limit distribution of $(X_n(s_1),X_n(s_2),\ldots,X_n(s_m))$, as $n \to \infty$. Now, note that:

$$\begin{aligned}\mathbb{P}\,(\max_{\rho(s,t)\le\delta;s,t\in T_k}|X(s)-X(t)|>\epsilon) &\le \liminf_{n\to\infty} P\,(\max_{\rho(s,t)\le\delta;s,t\in T_k}|X_n(s)-X_n(t)|>\epsilon)\\ &\le \liminf_{n\to\infty} P\,(\sup_{\rho(s,t)\le\delta;s,t\in T_\infty}|X_n(s)-X_n(t)|>\epsilon)\,,\end{aligned}$$

where the first inequality is a direct consequence of the portmanteau theorem for finite dimensional random variables. Next, letting $V_k = \max_{\rho(s,t)\le\delta;s,t\in T_k}|X(s)-X(t)|$ and $V_\infty = \sup_{\rho(s,t)\le\delta;s,t\in T_\infty}|X(s)-X(t)|$, we see that V_k increases almost surely to V_∞. Thus, $\mathbb{P}(V_\infty > \epsilon) \le \mathbb{P}(\cup_k\{V_k>\epsilon\})$ and this latter probability is given by $\lim_{k\to\infty}\mathbb{P}(V_k>\epsilon)$. It follows that:

$$\begin{aligned}\mathbb{P}\,(\sup_{\rho(s,t)\le\delta;s,t\in T_\infty}|X(s)-X(t)|>\epsilon) &\le \lim_{k\to\infty} P\,(\max_{\rho(s,t)\le\delta;s,t\in T_k}|X(s)-X(t)|>\epsilon)\\ &\le \liminf_{n\to\infty} P\,(\sup_{\rho(s,t)\le\delta;s,t\in T_\infty}|X_n(s)-X_n(t)|>\epsilon)\,,\end{aligned}$$

using the first display. Now, using the asymptotic equicontinuity condition, one readily obtains a sequence δ_m decreasing to 0, such that,

$$\mathbb{P}\,(\sup_{\rho(s,t)\le\delta_m;s,t\in T_\infty}|X(s)-X(t)|>\epsilon)\le 2^{-m}\,.$$

It follows from the first Borel-Cantelli lemma that: $\{\omega : \sup_{\rho(s,t)\le\delta_m,s,t\in T_\infty}|X(s)-X(t)| \le \epsilon, \forall \text{ sufficiently large } m\}$ has $\mathbb{P}$–probability 1. Call this set C_ϵ. Then $\mathcal{C} \equiv \bigcap_{m=1}^\infty C_{1/m}$ also has

$\mathbb{P}$-probability one and the restriction of X to T_∞ is ρ uniformly continuous on this set. Since T_∞ is ρ–dense in T, we can extend X, by continuity, to the entire set T, on the event $\mathcal{C}$. This gives a stochastic process, say $\tilde{X}$, defined on T (take $\tilde{X}$ to be the extension of X on $\mathcal{C}$ and to be the identically 0 function on its complement) that has ρ uniformly continuous sample paths. (Check that the process $\tilde{X}$ is well–defined and that uniform continuity is preserved.) At this point, we cannot say that $\tilde{X}$ is a version of X, i.e. their finite dimensional marginals coincide. However, by Proposition 0 at the beginning of this section, $\tilde{X}$ induces a tight Borel measure on $l^\infty(T)$ (and the measure concentrates on a separable subspace of $l^\infty(T)$: namely the subspace of ρ uniformly continuous functions on T).

We next show that X_n converges in distribution to $\tilde{X}$, subsequently denoted by X in an abuse of notation, by proving that for any bounded continuous function $H : l^\infty(T) \to \mathbb{R}$, $E^\star(H(X_n)) \to E(H(X))$. Once this is accomplished, it follows that the new X, i.e. $\tilde{X}$ that we constructed above, is a version of the old X, since convergence in $l^\infty(T)$ imples convergence of all finite–dimensional marginals. Our approach follows Wellner's Torgnon notes and should be contrasted with the approach in Theorem 10.2 of Pollard's monograph which uses techniques involving almost sure representations. The following fact plays a central role.

Fact: if $H : l^\infty(T) \to \mathbb{R}$ is bounded and continuous, and $K \subset l^\infty(T)$ is compact, then for every $\epsilon > 0$, there exists $\delta > 0$ such that: if $x \in K$ and $\|x - y\|_T < \delta$, then $|H(x) - H(y)| < \epsilon$.

The main idea rests on approximating the processes X_n by what we call δ-discrete approximations: discrete processes assuming finitely many values for each ω but on a sufficiently fine partition indexed by a positive number δ. Let $X_{n,\delta}$ denote the δ-discrete approximation to X_n and X_δ, the corresponding approximation to X. Then,

$$
\begin{aligned}
&|E^\star H(X_n) - E\,H(X)| \\
&\qquad \leq \; |E^\star H(X_n) - E\,H(X_{n,\delta})| + |E\,H(X_{n,\delta}) - E\,H(X_\delta)| + |E\,H(X_\delta) - E\,H(X)| \\
&\qquad \equiv \; I_{n,\delta} + II_{n,\delta} + III_{n,\delta}\,.
\end{aligned}
$$

In what follows, we let δ go to 0 along a countable set. So, consider a sequence δ_m that goes to 0. We next show that $\lim_{\delta_m \to 0} \limsup_{n\to\infty} I_{n,\delta_m} = 0$ and that the same holds true for II_{n,δ_m} and III_{n,δ_m}. It then follows that $|E^\star H(X_n) - E\,H(X)|$ goes to 0. The middle term II_{n,δ_m} is the easiest to deal with while the remaining terms need more work. Before we proceed further, we need to define the discrete approximating processes. By total-boundedness of (T,ρ), for every $\delta > 0$, there exists a δ-net, $t_1, t_2, \ldots, t_{n(\delta)}$, such that every $t \in T$ satisfies $\rho(t, t_j) < \delta$ for some t_j (depending on

t). Choose and fix such a t_j for each t and denote it as $\pi_\delta(t)$. Then the δ-discrete approximation to X_n is defined as: $X_{n,\delta}(t) = X_n(\pi_\delta(t))$. Similarly define $X_\delta(t)$. Note that these are measurable processes. **Now, by the finite dimensional convergence of X_n to X, it follows that $X_{n,\delta} \to X_\delta$ in the space $l^\infty(T)$, for every fixed $\delta > 0$.** This can be proved, for example, by invoking the Skorohod representation for random vectors and is left as an exercise. Thus, II_{n,δ_m} goes to 0 for every δ_m, as $n \to \infty$, and it follows trivially that $\lim_{\delta_m \to 0} \limsup_{n\to\infty} I_{n,\delta_m} = 0$.

The uniform continuity of the sample paths of X implies that $\lim_{\delta_m \to 0} \|X - X_{\delta_m}\|_T = 0$, almost surely. We next show that $\lim_{\delta_m\to 0} III_{n,\delta_m} = 0$ (note that $III_{n,\delta}$ does not depend on n, so the n in the subscript is really redundant). Since $X_{\delta,m}$ converges almost surely to X in the metric on $l^\infty(T)$ and H is a bounded continuous function, the real-valued random variables $H(X_{\delta,m})$ converge a.s. to $H(X)$. The boundedness of H then allows the DCT to be applied yielding that $E\,H(X_{\delta,m})$ converges to $E\,H(X)$ as $m \to \infty$.

Finally, to show that $\lim_{\delta_m\to 0} \limsup_{n\to\infty} I_{n,\delta_m} = 0$, we choose ϵ, τ and K as in the above paragraph. Let $K_{\tau/2}$ be the $\tau/2$ open neighborhood of the set K for the uniform metric. Then we have:

$$\begin{aligned} |E^\star H(X_n) - EH(X_{n,\delta_m})| \;\leq\; & 2\,\|H\|_\infty \{P^\star(\|X_n - X_{n,\delta_m}\|_T \geq \tau/2) + P(X_{n,\delta_m} \in K^c_{\tau/2})\} \\ & +2\,\sup\{|H(x) - H(y)| : x \in K, \|x-y\|_T < \tau\}\,. \end{aligned}$$

To see this, note that if $\|X_n - X_{n,\delta_m}| < \tau/2$ and $X_{n,\delta_m} \in K_{\tau/2}$, then there exists $x_0 \in K$ with $\|x_0 - X_{n,\delta_m}\|_T < \tau/2 < \tau$ and therefore $|H(x_0) - H(X_{n,\delta_m})| \leq \sup\{|H(x) - H(y)| : x \in K, \|x-y\|_T < \tau\}$. ALso, $\|x_0 - X_n\|_T < \tau$ by the triangle inequality, whence $|H(x_0) - H(X_n)| \leq \sup\{|H(x) - H(y)| : x \in K, \|x - y\|_T < \tau\}$. By our choice of τ, the second term on the right side of the above display is no larger than 2ϵ. We next take care of the first term. By the asymptotic equicontinuity condition, for all sufficiently large m, and therefore sufficiently small δ_m, $\limsup_{n\to 0\infty} P^\star(\|X_n - X_{n,\delta_m}\|_T \geq \tau/2) \leq \epsilon$ for all sufficiently large m. Finally, by the Portmanteau Theorem $\limsup_{n\to\infty} P(X_{n,\delta_m} \in K^c_{\tau/2}) \leq P(X_{\delta_m} \in K^c_{\tau/2})$. Now, $\limsup_{\delta_m \to 0} P(X_{\delta_m} \in K^c_{\tau/2}) \leq P(X \in K^c_{\tau/2}) \leq P(X \in K^c) < \epsilon$, showing that $\limsup_{\delta_m\to 0} \limsup_{n\to\infty} P(X_{n,\delta_m} \in K^c_{\tau/2}) < \epsilon$. Since $\epsilon > 0$ is arbitrary, it follows that $\lim_{\delta_m\to 0} \limsup_{n\to\infty} I_{n,\delta_m} = 0$.

The converse part of the proposition can be deduced, for example, by using the representation theorem. Since the law of X induces a tight Borel probability measure on $l^\infty(T)$, by Proposition 0, there exists a pseudo-metric ρ such that (T,ρ) is totally bounded and X assumes values (with probability 1) in the space of all uniformly ρ-continuous functions on T equipped with the uniform metric which is a separable subspace of $l^\infty(T)$. By the representation theorem above, we can construct random elements $\tilde{X}_n$ and a tight Borel measurable random element $\tilde{X}$ with the same distribution as X on a common probability space $(\tilde{\Omega}, \tilde{A}, \tilde{\mathbb{P}})$. Since $\|\tilde{X}_n(\tilde{\omega}) - \tilde{X}(\tilde{\omega})\| \leq \tilde{\delta}_n(\tilde{\omega}) \to 0$, it follows trivially that $\{\tilde{X}_n(t_i)\}_{i=1}^k \to_{a.s.} \{\tilde{X}(t_i)\}_{i=1}^k$. But $\{\tilde{X}_n(t_i)\}_{i=1}^k$ and $\{\tilde{X}(t_i)\}_{i=1}^k$ have the same distributions as $\{X_n(t_i)\}_{i=1}^k$ and $\{X(t_i)\}_{i=1}^k$ respectively, showing the convergence of finite-dimensional distributions. Next, consider

$$P^\star(\sup_{\rho(s,t)\leq\delta} |X_n(s) - X_n(t)| > \epsilon) = \tilde{\mathbb{P}}^\star(\sup_{\rho(s,t)\leq\delta} |\tilde{X}_n(s) - \tilde{X}_n(t)| > \epsilon)\,. \quad (\star\star)$$

Now,

$$\sup_{\rho(s,t)\leq\delta} |\tilde{X}_n(s) - \tilde{X}_n(t)| \leq 2\,\|\tilde{X}_n - \tilde{X}\|_T + \sup_{\rho(s,t)\leq\delta} |\tilde{X}(s) - \tilde{X}(t)|\,.$$

Thus, $\tilde{P}^\star(\sup_{\rho(s,t)\leq\delta} |\tilde{X}_n(s) - \tilde{X}_n(t)| > \epsilon)$ is no larger than:

$$\tilde{\mathbb{P}}^\star(\|\tilde{X}_n - \tilde{X}\|_T > \epsilon/4) + \tilde{\mathbb{P}}(\sup_{\rho(s,t)\leq\delta} |\tilde{X}(s) - \tilde{X}(t)| > \epsilon/2)\,.$$

Since $\tilde{\mathbb{P}}(\delta_n(\tilde{\omega}) > \epsilon/4) \to 0$, it follows that $\lim_{\delta\to 0} \limsup_{n\to\infty} \tilde{\mathbb{P}}^\star(\|\tilde{X}_n - \tilde{X}\|_T > \epsilon/4) = 0$. Also, by the almost sure ρ-uniform-continuity of the sample paths of $\tilde{X}$, $\lim_{\delta\to 0} \tilde{\mathbb{P}}\,(\sup_{\rho(s,t)\leq\delta} |\tilde{X}(s) - \tilde{X}(t)| > \epsilon/2) = 0$. The asymptotic equicontinuity condition now follows from $(\star\star)$ above. □

Remark: The case when X is a Gaussian process in $l^\infty(T)$ is of special interest as this is the case that arises repeatedly in empirical process applications. Formally, a stochastic process X in $l^\infty(T)$ is called *Gaussian* if for each $(t_1, t_2, \ldots, t_k), \tau_i \in T, k \in \mathcal{N})$, the random vector $(X(t_1), X(t_2), \ldots, X(t_k))$ is a multivariate normal random variable. Consider, a family of semi-metrics on T given by $\rho_p(s,t) = E(|X(s) - X(t)|^p)^{1/p\vee 1}$. The following result is from Page 41 of van der Vaart and Wellner:

A (Borel measurable) Gaussian process X in $l^\infty(T)$ is tight if and only if (T, ρ_p) is totally bounded and almost all paths $t \mapsto X(t,\omega)$ are uniformly ρ_p-continuous for some p and then for all p.

In view of the above theorem, we can therefore restrict ourselves to the ρ_p semimetrics for processes that have tight Gaussian limits. Typically, the ρ_2 semimetric is used in the setting of empirical process theory: this is just the L_2 distance between $X(s)$ and $X(t)$. For an illuminating discussion of the role of the ρ_p semimetrics, we refer the reader to the discussion on pages 39–41 of van der Vaart and Wellner. The following corollary is immediate.

Corollary: Let $\mathcal{F}$ be a class of real-valued measurable functions on $(\mathcal{X}, \mathcal{A})$. Then the following are equivalent:

(a) $\mathcal{F}$ is P–Donsker: $\mathbb{G}_n \Rightarrow \mathbb{G}$ in $l^\infty(\mathcal{F})$ where $\mathbb{G}$ is a Borel-measurable tight random element taking values in $l^\infty(\mathcal{F})$ with finite-dimensional Gaussian marginals.

(b)$(\mathcal{F}, \rho_2)$ (where $\rho_2(f,g) = \{E(\mathbb{G}(f) - \mathbb{G}(g))^2\}^{1/2} \equiv \{\text{Var}(f(X_1) - g(X_1))\}^{1/2}$) is totally bounded and $\mathbb{G}_n$ is asymptotically equicontinuous with respect to ρ_2 in probability; i.e:

$$\lim_{\delta \to 0} \lim \sup_{n \to \infty} P^\star\{ \sup_{f,g \in \mathcal{F}: \rho_2(f,g) < \delta} |\mathbb{G}_n(f) - \mathbb{G}_n(g)| > \epsilon\} = 0\,,$$

for every given $\epsilon > 0$.

4 Problems

- 1. Prove the **Fact** that was crucially used to establish that (A) implies (B) in the proof of Theorem 3.1.
- 2. Show that $C_u(T, \rho)$, the class of ρ-uniformly-continuous functions from T to $\mathbb{R}$, where ρ is a pseudometric on T with respect to which it is completely bounded is a complete and separable subspace of $l^\infty(T)$.
- 3. Prove the claim in bold letters on the convergence of $X_{n,\delta}$ to X_δ (in the proof of Theorem 3.1).
- 4. A random variable X is said to be weak L_2 if $x^2\,P(|X| > x) \to 0$ as $x \to \infty$. Show that for such an X, $E(|X|^r) < \infty$ for all $r < 2$.