# Introduction to Empirical Processes and Semiparametric Inference
# Lecture 08: Stochastic Convergence

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:

- Stochastic processes in metric spaces
- Tightness and separability of stochastic processes
- Gaussian processes
- Weak convergence and portmanteau theorem
- Continuous mapping theorem
- Asymptotic measurability and asymptotic tightness
- Prohorov's theorem
- Slutsky's theorem

# Stochastic Processes in Metric Spaces

Recall that for a stochastic process $\{X(t), t \in T\}$, $X(t)$ is a measurable real random variable for each $t \in T$ on a probability space $(\Omega, \mathcal{A}, P)$.

The sample paths of such a process typically reside in the metric space $\mathbb{D} = \ell^{\infty}(T)$ with the uniform metric.

Often, however, when $X$ is viewed as a map from $\Omega$ to $\mathbb{D}$, it is no longer Borel measurable.

A classic example of this issue comes from Billingsley (1968, Pages 152–153).

The example hinges on the existence of a set $H \subset [0, 1]$ which is not a Borel set (this is true).

Define the stochastic process $X(t) = \mathbf{1}\{U \leq t\}$, where $t \in [0, 1]$ and $U$ is uniformly distributed on $[0, 1]$.

The probability space for $X$ is $(\Omega, \mathcal{B}, P)$, where $\Omega = [0,1]$, $\mathcal{B}$ are the Borel sets on $[0,1]$, and $P$ is the uniform probability measure on $[0,1]$.

A natural metric space for the sample paths of $X$ is $\ell^\infty([0,1])$.

Define the set

$$A = \cup_{s \in H} B_s(1/2),$$

where $B_s(1/2)$ is the uniform open ball of radius $1/2$ around the function

$$t \mapsto f_s(t) \equiv \mathbf{1}\{t \leq s\}.$$

Since $A$ is an open set in $\ell^\infty([0,1])$, and since the uniform distance between $f_{s_1}$ and $f_{s_2}$ is 1 whenever $s_1 \neq s_2$,

$$X(t) \in B_s(1/2)$$

if and only if $U = s$.

Thus the set

$$\{\omega \in \Omega : X(\omega) \in A\}$$

equals $H$.

Since $H$ is not a Borel set, $X$ is not Borel measurable.

This lack of measurability is the usual state for biostatistics.

Many of the associated technical difficulties can be resolved by using of outer measure and outer expectation in the context of weak convergence.

In contrast, most of the limiting processes in biostatistics are Borel measurable.

Hence a brief study of Borel measurable processes is valuable.

Two Borel random maps $X$ and $X'$, with respective laws $L$ and $L'$, are *versions* of each other if $L = L'$.

Recall $BL_1(\mathbb{D})$, the set of all functions $f : \mathbb{D} \mapsto \mathbb{R}$ bounded by 1 and with Lipschitz norm bounded by 1, i.e., $|f(x) - f(y)| \leq d(x, y)$ for all $x, y \in \mathbb{D}$.

When the context is clear, we simply use $BL_1$.

Define a *vector lattice* $\mathcal{F} \subset C_b(\mathbb{D})$ to be a vector space such that if $f \in \mathcal{F}$ then $f \vee 0 \in \mathcal{F}$.

We also say that a set $\mathcal{F}$ of real functions on $\mathbb{D}$ *separates points* of $\mathbb{D}$ if, for any $x, y \in \mathbb{D}$ with $x \neq y$, there exists $f \in \mathcal{F}$ such that $f(x) \neq f(y)$.

LEMMA 1. *Let $L_1$ and $L_2$ be Borel probability measures on a metric space $\mathbb{D}$. TFAE:*

*(i)* $L_1 = L_2$.

*(ii)* $\int f dL_1 = \int f dL_2$ *for all* $f \in C_b(\mathbb{D})$.

*If $L_1$ and $L_2$ are also separable, then (i) and (ii) are both equivalent to*

*(iii)* $\int f dL_1 = \int f dL_2$ *for all* $f \in BL_1$.

*Moreover, if $L_1$ and $L_2$ are also tight, then (i)–(iii) are all equivalent to*

*(iv)* $\int f dL_1 = \int f dL_2$ *for all $f$ in a vector lattice $\mathcal{F} \subset C_b(\mathbb{D})$ that both contains the constant functions and separates points in $\mathbb{D}$.*

# Tightness and Separability of Stochastic Processes

In addition to being Borel measurable, most of the limiting stochastic processes of interest are *tight*.

A Borel probability measure $L$ on a metric space $\mathbb{D}$ is tight if for every $\epsilon > 0$, there exists a compact $K \subset \mathbb{D}$ so that $L(K) \geq 1 - \epsilon$.

A Borel random map $X : \Omega \mapsto \mathbb{D}$ is tight if its law $L$ is tight.

Tightness is equivalent to there being a $\sigma$-compact set that has probability 1 under $L$ or $X$.

$L$ or $X$ is *separable* if there is a measurable and separable set which has probability 1.

$L$ or $X$ is *Polish* if there is a measurable Polish set having probability 1.

Note that tightness, separability and Polishness are all topological properties and do not depend on a metric.

Since both $\sigma$-compact and Polish sets are also separable, separability is the weakest of the three properties.

Whenever we say $X$ has any one of these three properties, we tacitly imply that $X$ is also Borel measurable.

On a complete metric space, tightness, separability and Polishness are equivalent.

For a stochastic process $\{X(t), t \in T\}$, where $(T, \rho)$ is a separable, semimetric space, there is another meaning for separable:

$X$ is *separable* (as a stochastic process) if there exists a countable subset $S \subset T$ and a null set $N$ so that, for each $\omega \notin N$ and $t \in T$, there exists a sequence $\{s_m\} \in S$ with $\rho(s_m, t) \to 0$ and

$$|X(s_m, \omega) - X(t, \omega)| \to 0.$$

It turns out that many empirical processes are separable in this sense, even though they are not Borel measurable and therefore cannot satisfy the other meaning for separable.

The distinction between these two definitions will either be explicitly stated or made clear by the context.

Most limiting processes $X$ of interest will reside in $\ell^\infty(T)$, where the index set $T$ is often a class of real functions $\mathcal{F}$ with domain equal to the sample space.

When such limiting processes are tight, the following lemma demands that $X$ resides on $UC(T, \rho)$, where $\rho$ is some semimetric making $T$ totally bounded, with probability 1.

LEMMA 2. *Let $X$ be a Borel measurable random element in $\ell^\infty(T)$.*
*TFAE:*

*(i) $X$ is tight.*

*(ii) There exists a semimetric $\rho$ making $T$ totally bounded and for which $X \in UC(T, \rho)$ with probability 1.*

*Furthermore, if (ii) holds for any $\rho$, then it also holds for the semimetric $\rho_0(s,t) \equiv \mathsf{E} \arctan |X(s) - X(t)|$.*

A nice feature of tight processes in $\ell^\infty(T)$ is that the laws are completely defined by their finite-dimensional marginal distributions $(X(t_1), \ldots, X(t_k))$, where $t_1, \ldots, t_k \in T$:

LEMMA 3. *Let $X$ and $Y$ be tight, Borel measurable stochastic processes in $\ell^\infty(T)$.*

*Then the Borel laws of $X$ and $Y$ are equal if and only if all corresponding finite-dimensional marginal distributions are equal.*

**Proof.** Consider the collection $\mathcal{F} \subset C_b(\mathbb{D})$ of all functions $f : \ell^\infty(T) \mapsto \mathbb{R}$ of the form $f(x) = g(x(t_1), \ldots, x(t_k))$, where $g \in C_b(\mathbb{R}^k)$ and $k \geq 1$ is an integer.

We leave it as an exercise to show that $\mathcal{F}$ is a vector lattice, an algebra, and separates points of $\ell^\infty(T)$.

The desired result now follows from Lemma 1.$\square$

While the semimetric $\rho_0$ defined in Lemma 2 is always applicable when $X$ is tight, it is frequently not the most convenient choice.

There are also other useful, equivalent choices.

For a process $X$ in $\ell^\infty(T)$ and a semimetric $\rho$ on $T$, we say that $X$ is *uniformly $\rho$-continuous in $p$th mean* if

$$\mathsf{E}|X(s_n) - X(t_n)|^p \to 0$$

whenever $\rho(s_n, t_n) \to 0$.

# Gaussian Processes

Perhaps the most frequently occurring limiting process in $\ell^\infty(T)$ is a *Gaussian* process.

A stochastic process $\{X(t), t \in T\}$ is Gaussian if all finite-dimensional marginals $\{X(t_1), \ldots, X(t_k)\}$ are multivariate normal.

If a Gaussian process $X$ is tight, then by Lemma 2, there is a semimetric $\rho$ making $T$ totally bounded and for which the sample paths $t \mapsto X(t)$ are uniformly $\rho$-continuous.

An interesting feature of Gaussian processes is that this result implies that the map $t \mapsto X(t)$ is uniformly $\rho$-continuous in $p$th mean for all $p \in (0, \infty)$.

For a general Banach space $\mathbb{D}$, a Borel measurable random element $X$ on $\mathbb{D}$ is Gaussian if and only if $f(X)$ is Gaussian for every continuous, linear functional $f : \mathbb{D} \mapsto \mathbb{R}$.

When $\mathbb{D} = \ell^\infty(T)$ for some set $T$, this definition appears to contradict the definition of Gaussianity given in the preceding paragraph, since now we are using all continuous linear functionals instead of just linear combinations of coordinate projections.

These two definitions are not really reconcilable in general, and so some care must be taken in reading the literature.

However, when the process in question is tight, the two definitions are equivalent, as verified in the following proposition.

PROPOSITION 1. *Let $X$ be a tight, Borel measurable map into $\ell^\infty(T)$. TFAE:*

*(i) The vector $(X(t_1), \ldots, X(t_k))$ is multivariate normal for every finite set $\{t_1, \ldots, t_k\} \subset T$.*

*(ii) $\phi(X)$ is Gaussian for every continuous, linear functional $\phi : \ell^\infty(T) \mapsto \mathbb{R}$.*

*(iii) $\phi(X)$ is Gaussian for every continuous, linear map $\phi : \ell^\infty(T) \mapsto \mathbb{D}$ into any Banach space $\mathbb{D}$.*

# Weak Convergence and Portmanteau Theorem

The extremely important concept of weak convergence of sequences arises in many areas of statistics.

To be as flexible as possible, we allow the probability spaces associated with the sequences to change with $n$.

Let $(\Omega_n, \mathcal{A}_n, P_n)$ be a sequence of probability spaces and $X_n : \Omega_n \mapsto \mathbb{D}$ a sequence of maps.

We say that $X_n$ *converges weakly* to a Borel measurable $X : \Omega \mapsto \mathbb{D}$ if

$$\mathsf{E}^* f(X_n) \to \mathsf{E} f(X), \quad \text{for every } f \in C_b(\mathbb{D}). \tag{1}$$

If $L$ is the law of $X$, (1) can be reexpressed as

$$\mathsf{E}^* f(X_n) \to \int_\Omega f(x) dL(x), \quad \text{for every } f \in C_b(\mathbb{D}).$$

Weak convergence is denoted $X_n \rightsquigarrow X$ or, equivalently, $X_n \rightsquigarrow L$.

Weak convergence is equivalent to “convergence in distribution” and “convergence in law.”

By Lemma 1, this definition of weak convergence ensures that the limiting distributions are unique.

Note that the choice of probability spaces $(\Omega_n, \mathcal{A}_n, P_n)$ is important since these dictate the outer expectation.

In most of the settings discussed in this book, $\Omega_n = \Omega$ for all $n \geq 1$.

The above definition of weak convergence does not appear to connect to the standard, statistical notion of convergence of probabilities.

However, this connection does hold for probabilities of sets $B \subset \Omega$ which have boundaries $\delta B$ satisfying $L(\delta B) = 0$.

The boundary $\delta B$ of a set $B$ in a topological space is the closure of $B$ minus the interior of $B$.

THEOREM 1. *(Portmanteau) TFAE:*

*(i)* $X_n \rightsquigarrow L$;

*(ii)* $\liminf \mathsf{P}_*(X_n \in G) \geq L(G)$ *for every open* $G$;

*(iii)* $\limsup \mathsf{P}^*(X_n \in F) \leq L(F)$ *for every closed* $F$;

*(iv)* $\liminf \mathsf{E}_* f(X_n) \geq \int_\Omega f(x) dL(x)$ *for every lower semicontinuous* $f$ *bounded below;*

*(v)* $\limsup \mathsf{E}^* f(X_n) \leq \int_\Omega f(x) dL(x)$ *for every upper semicontinuous* $f$ *bounded above;*

*(vi)* $\lim \mathsf{P}^*(X_n \in B) = \lim \mathsf{P}_*(X_n \in B) = L(B)$ *for every Borel* $B$ *with* $L(\delta B) = 0$;

*(vii)* $\liminf \mathsf{E}_* f(X_n) \geq \int_\Omega f(x) dL(x)$ *for every bounded, Lipschitz continuous, nonnegative* $f$.

*Furthermore, if $L$ is separable, then (i)–(vii) are also equivalent to*

*(viii)* $\sup_{f \in BL_1} \left| \mathsf{E}^* f(X_n) - \mathsf{E} f(X) \right| \to 0.$

# Continuous Mapping Theorem

THEOREM 2. *(Continuous mapping) Let $g : \mathbb{D} \mapsto \mathbb{E}$ be continuous at all points in $\mathbb{D}_0 \subset \mathbb{D}$, where $\mathbb{D}$ and $\mathbb{E}$ are metric spaces.*

*Then if $X_n \rightsquigarrow X$ in $\mathbb{D}$, with $\mathsf{P}_*(X \in \mathbb{D}_0) = 1$, then $g(X_n) \rightsquigarrow g(X)$.*

## Asymptotic Measureability and Asymptotic Tightness

A potential issue is that there may sometimes be more than one choice of metric space $\mathbb{D}$ to work with.

For example, if we are studying weak convergence of the usual empirical process $\sqrt{n}(\hat{F}_n(t) - F(t))$ based on data in $[0, 1]$, we could let $\mathbb{D}$ be either $\ell^\infty([0, 1])$ or $D[0, 1]$.

The following lemma tells us that the choice of metric space is generally not a problem:

LEMMA 4. *Let the metric spaces $\mathbb{D}_0 \subset \mathbb{D}$ have the same metric, and assume $X$ and $X_n$ reside in $\mathbb{D}_0$.*

*Then $X_n \rightsquigarrow X$ in $\mathbb{D}_0$ if and only if $X_n \rightsquigarrow X$ in $\mathbb{D}$.*

A sequence $X_n$ is asymptotically measurable if and only if

$$\mathsf{E}^* f(X_n) - \mathsf{E}_* f(X_n) \quad \rightarrow \quad 0, \tag{2}$$

for all $f \in C_b(\mathbb{D})$.

A sequence $X_n$ is asymptotically tight if for every $\epsilon > 0$, there is a compact $K$ so that $\liminf \mathsf{P}_*(X_n \in K^\delta) \geq 1 - \epsilon$, for every $\delta > 0$, where for a set $A \subset \mathbb{D}$,

$$A^\delta = \{x \in \mathbb{D} : d(x, A) < \delta\}$$

is the "$\delta$-enlargement" around $A$.

Two good properties of asymptotic tightness are that it does not depend on the metric chosen—only on the topology—and that weak convergence often implies asymptotic tightness.

The first of these two properties are verified in the following:

LEMMA 5. $X_n$ *is asymptotically tight if and only if for every* $\epsilon > 0$ *there exists a compact* $K$ *so that* $\liminf \mathsf{P}_*(X_n \in G) \geq 1 - \epsilon$ *for every open* $G \supset K$.

The second good property of asymptotic tightness is part (ii) of the following lemma, part (i) of which gives the necessity of asymptotic measurability for weakly convergent sequences:

LEMMA 6. *Assume* $X_n \rightsquigarrow X$. *Then*

(i) $X_n$ *is asymptotically measurable.*

(ii) $X_n$ *is asymptotically tight if and only if* $X$ *is tight.*

## Prohorov's Theorem

Prohorov's theorem (modernized) tells us that asymptotic measurability and asymptotic tightness together almost gives us weak convergence.

This "almost-weak-convergence" is *relative compactness*.

A sequence $X_n$ is relatively compact if every subsequence $X_{n'}$ has a further subsequence $X_{n''}$ which converges weakly to a tight Borel law.

Weak convergence happens when all of the limiting Borel laws are the same.

THEOREM 3. *(Prohorov's theorem) If the sequence $X_n$ is asymptotically measurable and asymptotically tight, then it has a subsequence $X_{n'}$ that converges weakly to a tight Borel law.*

Note that the conclusion of Prohorov's theorem does not state that $X_n$ is relatively compact, and thus it appears as if we have broken our earlier promise.

However, if $X_n$ is asymptotically measurable and asymptotically tight, then every subsequence $X_{n'}$ is also asymptotically measurable and asymptotically tight.

Thus repeated application of Prohorov's theorem does indeed imply relative compactness of $X_n$.

The following lemma specifies the relationship between marginal and joint processes regarding asymptotic measurability and asymptotic tightness:

LEMMA 7. *Let* $X_n : \Omega_n \mapsto \mathbb{D}$ *and* $Y_n : \Omega_n \mapsto \mathbb{E}$ *be sequences of maps. TFAT:*

*(i)* $X_n$ *and* $Y_n$ *are both asymptotically tight if and only if the same is true for the joint sequence* $(X_n, Y_n) : \Omega_n \mapsto \mathbb{D} \times \mathbb{E}$*.*

*(ii) Asymptotically tight sequences* $X_n$ *and* $Y_n$ *are both asymptotically measurable if and only if* $(X_n, Y_n) : \Omega_n \mapsto \mathbb{D} \times \mathbb{E}$ *is asymptotically measurable.*

# Slutsky's Theorem

A very useful consequence of Lemma 7 is Slutsky's theorem:

THEOREM 4. *(Slutsky's theorem) Suppose $X_n \rightsquigarrow X$ and $Y_n \rightsquigarrow c$, where $X$ is separable and $c$ is a fixed constant. TFAT:*

*(i)* $(X_n, Y_n) \rightsquigarrow (X, c)$.

*(ii) If $X_n$ and $Y_n$ are in the same metric space, then $X_n + Y_n \rightsquigarrow X + c$.*

*(iii) Assume in addition that the $Y_n$ are scalars. Then whenever $c \in \mathbb{R}$, $Y_n X_n \rightsquigarrow cX$. Also, whenever $c \neq 0$, $X_n / Y_n \rightsquigarrow X/c$.*

**Proof.** By completing the metric space for $X$, we can without loss of generality assume that $X$ is tight.

Thus by Lemma 7, $(X_n, Y_n)$ is asymptotically tight and asymptotically measurable.

Thus by Prohorov's theorem, all subsequences of $(X_n, Y_n)$ have further subsequences which converge to tight limits.

Since these limit points have marginals $X$ and $c$, and since the marginals in this case completely determine the joint distribution, we have that all limiting distributions are uniquely determined as $(X, c)$.

This proves Part (i).

Parts (ii) and (iii) now follow from the continuous mapping theorem.$\square$