Stochastic convergence

Xinjie Qian

June 16, 2021

Outline

Stochastic Processes in Metric Spaces

- A stochastic process $\{X(t), t \in T\}$is a measurable real random variable for each $t \in T$ on a probability space $(\Omega, \mathscr{A}, P)$.
- The sample paths of such a process typically reside in the metric space $\mathbb{D} = l^{\infty}(T)$ with the uniform metric.
- **However**, in many cases, when X is viewed as a map from Ω to $\mathbb{D}$, it is no longer Borel measurable.
- An example is given from Billingsley.

Example

- Fact: there exists a set $H \subset [0,1]$ which is not a Borel set.
- Define $X(t) = 1\{U \leq t\}$, where $t \in [0,1]$ and U is uniformly distributed on $[0,1]$.
- Let $\Omega = [0,1]$, $\mathscr{B}$ are the Borel sets on $[0,1]$, P is the uniform probability measure on $[0,1]$ and $\mathbb{D}$ is $l^\infty([0,1])$.
- 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

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

Example

- A is an open set in $l^{\infty}([0,1])$
  The uniform distance between f_{s_1} and f_{s_2} is 1 whenever $s_1 \neq s_2$
  $\Longrightarrow X(t) \in B_s(1/2)$ if and only if $U = s$.
- So the set $\{\omega \in \Omega : X(\omega \in A)\}$ equals H.
- Thus, X is not Borel measurable.

- The example shown above is the usual state for most of the empirical processes we are interested in.
- Many of the associated technical difficulties can be resolved by using outer measure and outer expectation in the context of weak convergence.
- In contrast, most of the limiting processes we will be studying are Borel measurable.

- Define a *vector lattice* $\mathscr{F} \subset C_b(\mathbb{D})$ to be a vector space for which if $f \in \mathscr{F}$ then $f \vee 0 \in \mathscr{F}$.
- Define a set $\mathscr{F}$ 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 $f \in \mathscr{F}$ such that $f(x) \neq f(y)$.

Lemma 1

Let L_1 and L_2 be Borel probability measures on a metric space $\mathbb{D}$. The following are equivalent:
(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 $\mathscr{F} \subset C_b(\mathbb{D})$ that both contains the constant functions and separates points in $\mathbb{D}$.

Note: This lemma, provides two ways of establishing equivalence between Borel probability measures.

Tight and separable

Definition

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.

Definition

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

Definition

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

Note: separability is the weakest of the three properties.

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

Definition

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

Note: Many of the empirical processes we will study are separable in this sense, even though they are not Borel measurable and therefore cannot satisfy the other meaning for separable.

Lemma 2

Let X be a Borel measurable random element in $l^{\infty}(T)$. Then the following are equivalent:

(i) X is tight.

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

Furthermore, if (ii) holds for any ρ, then it also holds for the semimetric $\rho_0(s,t) \equiv Earctan|X(s) - X(t)|$.

Lemma 3

Let X and Y be tight, Borel measurable stochastic processes in $l^\infty(T)$. Then the Borel laws of X and Y are equal if and only if all corresponding finite-dimensional marginal distributions are equal.

Tight processes are completely defined by their finite-dimensional marginal distributions $(X(t_1), ..., X(t_k))$, where $t_1, ..., t_k \in T$ and $k \geq 1$ is an integer.

Definition

For a process X in $l^\infty(T)$ and a semimetric ρ on T, we say that X is *uniformly ρ-continuous in pth mean* if $E\|X(s_n) - X(t_n)\|^p \to 0$, whenever $\rho(s_n, t_n) \to 0$

Gaussian process

Definition

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

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

Definition

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 map $f : \mathbb{D} \to \mathbb{R}$.

Proposition 4

Let X be a tight, Borel measurable map into $l^{\infty}(T)$. Then the following are equivalent:

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

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

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

Note: when the process in question is tight, the two definitions are equivalent.

Proof.

(ii)$\Rightarrow$ (iii):

1.Fix any Banach space $\mathbb{D}$ and any continuous, linear map $\phi : l^\infty(T) \mapsto \mathbb{D}$.

2.For any continuous, linear map $\psi : \mathbb{D} \mapsto \mathbb{R}$, the composition map $\psi \circ \phi : l^\infty(T) \mapsto \mathbb{R}$ is continuous and linear.

3.Thus, by (ii), we have that $\psi(\phi(X))$ is Gaussian.

4.Since ψ is arbitrary, and by the definition of a Gaussian process on a Banach space, we can get that $\phi(X)$ is Gaussian.

Since both $\mathbb{D}$ and ϕ were also arbitrary, conclusion (iii) follows.

(iii)$\Rightarrow$(i):

Multivariate coordinate projections are special examples of continuous, linear maps into Banach spaces. □

Weak Convergence

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

Definition

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

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

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

$$E^* f(X_n) \to \int_\Omega f(x) dL(x), for\ every\ f \in C_b(\mathbb{D}). \tag{2}$$

This weak convergence is denoted $X_n \rightsquigarrow x$ or, equivalently, $X_n \rightsquigarrow L$.

Theorem 5 (Portmanteau)

The following are equivalent:
(i) $X_n \rightsquigarrow L$;
(ii) $lim\ inf P_*(X_n \in G) \geq L(G)$ for every open G;
(iii) $lim\ sup P^*(X_n \in F) \leq L(F)$ for every closed F;
(iv) $lim\ inf E_* f(X_n) \geq \int_\Omega f(x)dL(x)$ for every lower semicontinuous f bounded below;
(v) $lim\ sup E^* f(X_n) \leq \int_\Omega f(x)dL(x)$ for every upper semicontinuous f bounded above;
(vi) $lim P^*(X_n \in B) = lim P_*(X_n \in B) = L(B)$ for every Borel B with $L(\delta B) = 0$;
(vii) $lim\ inf E_* f(X_n) \geq \int_\Omega f(x)dL(x)$ for every bounded, Lipschitz continuous, nonnegative f.
Furthermore, if L is separable, then (i)–(vii) are also equivalent to
(viii) $sup_{f \in BL_1} \|E^* f(X_n) - Ef(X)\| \to 0$.

Note: depending on the setting, one or more of these alternative definitions will prove more useful than the others.

Theorem 6 (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 $P_*(X \in \mathbb{D}_0) = 1$, then $g(X_n) \rightsquigarrow g(X)$.

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

Example: when considering 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 $l^\infty([0, 1])$ or $D[0, 1]$.

Lemma 7

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 $\mathbb{D}$.

Proof. Since any set $B_0 \in \mathbb{D}_0$ is open if and only if it is of the form $B \cap \mathbb{D}_0$ for some open B in $\mathbb{D}$, the result follows from Part (ii) of the portmanteau theorem.

Note: this lemma tells us that the choice of metric space is generally not a problem.

Asymptotic Measureability and Asymptotic Tightness

Recall: a sequence X_n is asymptotically measurable if and only if $E^* f(X_n) - E_* f(X_n) \to 0$, for all $f \in C_b(\mathbb{D})$.

Definition

A sequence X_n is asymptotically tight if for every $\epsilon > 0$, there is a compact K so that $lim\ inf P_*(X_n \in K^\delta) \geq 1 - \epsilon$, for every $\delta > 0$, where for a set $A \subset \mathbb{D}$, $A^\delta = \{x \in \mathbb{D} : d(x, A) < \delta\}$ is the "δ-enlargement" around A.

Properties of asymptotic tightness :

- It does not depend on the metric chosen.
- Weak convergence often implies asymptotic tightness.

These two lemmas show the properties of asymptotic tightness:

Lemma 8

X_n is asymptotically tight if and only if for every $\epsilon > 0$ there exists a compact K so that $lim\ inf P_*(X_n \in G) \geq 1 - \epsilon$ for every open $G \supset K$.

Lemma 9

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.

Proof of Lemma 8

Proof.

First assume that X_n is asymptotically tight.($\Rightarrow$)
1. Fix $\epsilon > 0$, and let the compact set K satisfy $lim\ inf P_*(X_n \in K^\delta) \geq 1 - \epsilon$, for every $\delta > 0$. 2. If $G \supset K$ is open, then there exists a $\delta_0 > 0$ so that $G \supset K^{\delta_0}$.
(If this were not true, then there would exist a sequence $\{x_n\} \notin G$ so that $d(x_n, K) \to 0$. This implies the existence of a sequence $\{y_n\} \in K$ so that $d(x_n, y_n) \to 0$. Thus, since K is compact and the complement of G is closed, there is a subsequence $n^{'}$ and a point $y \notin G$ so that $d(y_{n'}, y) \to 0$, but this is impossible.)
3. Hence $lim\ inf P_*(X_n \in G) \geq 1 - \epsilon$.
Now assume that X_n satisfies the alternative definition.($\Leftarrow$)
1. Fix $\epsilon > 0$, and let the compact set K satisfy $lim\ inf P_*(X_n \in G) \geq 1 - \epsilon$, for every open $G \supset K$.
2. For every $\delta > 0$, K^δ is an open set. 3. Thus $lim\ inf P_*(X_n \in K^\delta) \geq 1 - \epsilon$, for every $\delta > 0$. □

Proof of Lemma 9

Proof.

(i):

Fix $f \in C_b(\mathbb{D})$. Note that weak convergence implies both
$E^*f(X_n) \to Ef(X)$ and
$E_*f(X_n) = -E^*[-f(X_n)] \to -E[-f(X)] = Ef(X)$.

(ii):

Assume X is tight.($\Rightarrow$)

1. Fix $\epsilon > 0$. Assume X is tight, and choose a compact K so that $P(X \in K) \geq 1 - \epsilon$.
2. By Part(ii) of the portmanteau theorem, $lim\ inf P_*(X_n \in K^{\delta}) \geq P(X \in K^{\delta}) \geq 1 - \epsilon$ for every $\delta > 0$.
3. Hence X_n is asymptotically tight.

Now assume that X_n is asymptotically tight. ($\Leftarrow$)

1. Fix $\epsilon > 0$, and choose a compact K so that $lim\ inf P_*(X \in K^{\delta}) \geq 1 - \epsilon$ for every $\delta > 0$.
2. By Part (iii) of the portmanteau theorem, $P(X \in \bar{K}^{\delta}) \geq lim\ sup P^*(X_n \in \bar{K}^{\delta}) \geq lim\ inf P_*(X_n \in \bar{K}^{\delta}) \geq 1 - \epsilon$.

Theorem 10 (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: Prohorov's theorem shows that asymptotic measurability and asymptotic tightness together almost gives us weak convergence, which is *relative compactness*.

Definition

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.

Question: under what circumstances does asymptotic measurability or tightness of the marginal sequences X_n and Y_n imply asymptotic measurability or tightness of the joint sequence (X_n, Y_n)?

Lemma 11

Let $X_n : \Omega_n \mapsto \mathbb{D}$ and $Y_n : \Omega_n \mapsto \mathbb{E}$ be sequences of maps. Then the following are true:
(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.

Theorem 12 (Slutsky's theorem)

Suppose $X_n \rightsquigarrow X$ and $Y_n \rightsquigarrow c$, where X is separable and c is a fixed constant. Then the following are true:
(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 $c \neq 0$, $X_n/Y_n \rightsquigarrow X/c$.

Note: We can use Prohorov's theorem and the continuous mapping theorem to prove Slutsky's theorem.

Theorem 12 (Slutsky's theorem)

Suppose $X_n \rightsquigarrow X$ and $Y_n \rightsquigarrow c$, where X is separable and c is a fixed constant. Then the following are true:
(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.

(i):
1.By completing the metric space for X, we can without loss of generality assume that X is tight.
2. By Lemma 11, (X_n, Y_n) is asymptotically tight and asymptotically measurable.
3. By Prohorov's theorem, all subsequences of (X_n, Y_n) have further subsequences which converge to tight limits.
(ii) and (iii): Continuous mapping theorem. □