

Lecture 2: The Space of Continuous Functions

October 24, 2019

Lecturer: Burak Büke

1 Sample Path View of Stochastic Processes

A stochastic process is usually defined to be *an indexed collection of random variables with index set $\mathcal{T}$* . In general the index is referred as the time. Here are some examples with a continuous index set justifying this terminology, :

- (i) Consider the modelling of a price of certain stock in the stock market. Fix a reference time $t = 0$ and measure the time in days. Then $\{X(t)\}_{t \in \mathcal{T}}$ can be treated a continuous-time stochastic process, where X_t is the price of the stock t days after your reference point.
- (ii) Again fix a reference time $t = 0$ and measure the time in days. Then if Y_t is the temperature in our classroom in Celsius t days after our reference point, then $\{Y(t)\}_{t \in \mathcal{T}}$ can be treated as a continuous-time stochastic process.
- (iii) Suppose you are observing a queue at a supermarket till and $t = 0$ is the time when you start your observation. Setting Q_t as the number of people in the queue t days after your observation, $\{Q(t)\}_{t \in \mathcal{T}}$ can be treated as a stochastic process.

Instead of thinking of a stochastic process as a “collection” of random variables, we can think of it as a single random element whose values are functions from $\mathcal{T}$ to $\mathbb{R}$ and consider spaces of functions as metric spaces. In this chapter and the next, we consider continuous and right-continuous functions, respectively.

2 The Space of Continuous Functions on Bounded Closed Intervals

In this section, we consider continuous functions from the closed interval $[0,1]$ to $\mathbb{R}$, which we denote as $C_{\mathbb{R}}[0,1]$. The results, we present here trivially follows when the unit interval is replaced with any bounded closed interval. However, extension to infinite interval $[0, \infty)$ requires some additional work and will be presented separately in Section 4. We also briefly discuss how these results can be generalized to functions taking values in more general metric spaces.

The space $C_{\mathbb{R}}[0,1]$ is equipped with the uniform topology, i.e., the topology of uniform convergence, which can be metricized by the supremum norm

$$\rho(x, y) = \sup_{t \in [0,1]} |x(t) - y(t)|.$$

$C_{\mathbb{R}}[0,1]$ with this topology has the following basic properties:

1. $C_{\mathbb{R}}[0, 1]$ is separable. This can be proven in two ways. First way is to use Stone-Weierstrass Theorem which states any function in $C_{\mathbb{R}}[0, 1]$ can be approximated with polynomials (with rational coefficients) which are countable.

A second way to prove this is to use polygonal functions. For each polygonal function f there exists a k where $f(i/k) = a_i$ rational for each $i = 0, \dots, k$ and f is linear over each segment $[(i-1)/k, i/k]$ for $i = 1, \dots, k$. The set of polygonal functions is countable (why?). Now, any continuous function $x(\cdot) \in C_{\mathbb{R}}[0, 1]$ is uniformly continuous, i.e., for each $\epsilon > 0$ we can find k such that such that

$$\sup_{t,s \in [0,1]: |t-s| < 1/k} |x(t) - x(s)| \leq \epsilon/2.$$

Choosing k as such and polygonal function $|f(i/k) - x(i/k)| < \epsilon/2$, we can see that

$$\sup_{t \in [0,1]} |f(t) - x(t)| < \epsilon.$$

2. $C_{\mathbb{R}}[0, 1]$ is complete. Suppose that $\{x_n\}_{n=1}^{\infty}$ is a Cauchy sequence, i.e., for each $\epsilon > 0$, there exists N_{ϵ} such that $n, m > N_{\epsilon}$ implies $\sup_{t \in [0,1]} |x_n(t) - x_m(t)| < \epsilon$. For any fixed t , then $\{x_n(t)\}_{n=1}^{\infty}$ is also Cauchy and using the completeness of $\mathbb{R}$ we can define $x(t) = \lim_{n \rightarrow \infty} x_n(t)$. Now, we can prove that this function is continuous.
3. Any probability measure $\mathbb{P}$ on $C_{\mathbb{R}}[0, 1]$ is tight. This follows from Theorem 5.6 from Lecture 1 and completeness and separability of $C_{\mathbb{R}}[0, 1]$.
4. Finite dimensional cylindrical sets is a separating class. We define finite dimensional cylindrical sets as A , such that there exists a finite collection $t_1, \dots, t_n \in [0, 1]$ and $B_1, \dots, B_n \in \mathcal{B}(\mathbb{R})$ such that

$$A = \{x(t_1) \in B_1, \dots, x(t_n) \in B_n\},$$

and denote the class of finite dimensional cylinders as $\mathcal{A}_F$ denote the Borel σ -algebra on $C_{\mathbb{R}}[0, 1]$ shortly as $\mathcal{B}(C)$. We can show $\mathcal{A}_F$ is a separating class if we can show that it generates $\mathcal{B}(C)$.

First, we need to show that $\mathcal{A}_F \subset \mathcal{B}(C)$. To do so, we only need to concentrate on one dimensional cylindrical sets where $A = \{x(t_1) \in B_1\}$ for some $t_1 \in [0, 1]$ and $B_1 \in \mathcal{B}(\mathbb{R})$ as more general sets in $\mathcal{A}_F$ are finite intersections of sets of this type. First, realize that sets of the form $A = \{x(t_1) \in (a, b)\}$ is an open set in $\mathcal{B}(C)$. To see this take any function with $a + \epsilon \leq x(t_1) = c \leq b - \epsilon$ for some $\epsilon > 0$. If $\rho(x, y) < \epsilon/2$ then $y \in A$ which implies that A is open. Hence, the class of sets $\{B : \{x(t_1) \in B\} \in \mathcal{B}(C)\}$ is a σ -algebra containing all the open sets, which implies that $\{x(t_1) \in B\} \in \mathcal{A}_F \subset \mathcal{B}(C)$.

Now, we show that $\mathcal{B}(C)$ is contained in the σ -algebra generated by $\mathcal{A}_F$ which we denote as $\sigma(\mathcal{A}_F)$. Using separability of $C_{\mathbb{R}}[0, 1]$ it is enough to show that any closed ball is in $\sigma(\mathcal{A}_F)$. Considering an enumeration of rationals $q_1, q_2, \dots$ and a sequence of Borel sets of $\mathbb{R}$, $\{B_i\}_{i=1}^n$, it is trivial from the definition that $\{x(q_1) \in B_1, x(q_2) \in B_2, \dots\} \in \sigma(\mathcal{A}_F)$. Notice that

$$\overline{B_x(\epsilon)} = \{y : \rho(x, y) \leq \epsilon\} = \{y : |y(q_i) - x(q_i)| \leq \epsilon, i = 1, 2, \dots\},$$

which proves our claim.

This fact shows that finite dimensional distributions of any stochastic process uniquely determines the law of a stochastic process.

Notice that the properties above heavily rely on the separability of completeness of $\mathbb{R}$. If we consider $C_{\mathbb{X}}[0, 1]$, the space of continuous functions taking values in a complete and separable space, then the same results can be proven.

We have seen that the concept of tightness plays central role for weak convergence. To be able to define the concept of tightness for random elements in $C_{\mathbb{R}}[0, 1]$, we need to characterize the relatively compact sets (sets whose closure is compact) in this space. The famous Arzelà-Ascoli Theorem completely characterizes these sets.

Theorem 2.1 (Arzelà-Ascoli Theorem.). *A collection of functions $\{x_{\alpha}(\cdot)\}_{\alpha \in \mathcal{A}}$ in $C_{\mathbb{R}}[0, 1]$ is relatively continuous if and only if they are equibounded and equicontinuous, i.e.,*

- (i) *There exists an $M < \infty$ such that $\sup_{t \in [0, 1]} |x_{\alpha}(t)| < M$ for all $\alpha \in \mathcal{A}$.*
- (ii) *For any $\epsilon > 0$, there exists a δ_{ϵ} such that $|t_1 - t_2| < \delta_{\epsilon}$ implies $|x_{\alpha}(t_1) - x_{\alpha}(t_2)| < \epsilon$ for all $\alpha \in \mathcal{A}$.*

We omit the proof of this famous theorem here and refer the interested reader to Billingsley [1] and Rudin [2]. We can re-write (i) just based on $\{x_{\alpha}(0)\}_{\alpha \in \mathcal{A}}$ using (ii). To do so, take $\epsilon > 0$ find an N such that $1/N < \delta_{\epsilon}$ and write

$$\sup_{t \in [0, 1]} |x_{\alpha}(t)| \leq |x_{\alpha}(0)| + \sum_{i=1}^N \sup_{(i-1)/N \leq t, s < i/N} |x_{\alpha}(t) - x_{\alpha}(s)| \leq |x(0)| + N\epsilon.$$

to replace (i) with the condition that there exists an $M < \infty$ such that $\sup_{\alpha \in \mathcal{A}} |x_{\alpha}(0)| < M$.

We know that all continuous functions on a bounded closed interval are uniformly continuous, i.e., for any $x \in C_{\mathbb{R}}[0, 1]$ and $\epsilon > 0$, there exists a unique δ_{ϵ} such that $|t - s| < \delta$ implies $|x(t) - x(s)| < \epsilon$. The term “uniformity” refers to the fact that the δ is not dependent on t and s under consideration. Equicontinuity of a set of functions means that this δ_{ϵ} can be chosen to apply to all functions in this set. Motivated by this, we define the modulus of continuity

$$w(x, \delta) = \sup_{|t-s| < \delta} |x(t) - x(s)|,$$

It is easy to see that the function x is uniformly continuous if and only if $w_x(\delta) \rightarrow 0$ as $\delta \rightarrow 0$. Similarly equicontinuity of a set of functions $\{x_{\alpha}\}_{\alpha \in \mathcal{A}}$ can be written as

$$\lim_{\delta \rightarrow 0} \sup_{\alpha \in \mathcal{A}} w(x_{\alpha}, \delta) = 0.$$

Rewriting Arzelà-Ascoli Theorem in the new mathematical notation, $\{x_{\alpha}\}_{\alpha \in \mathcal{A}}$ is relatively compact if and only if

- (i) $\sup_{\alpha \in \mathcal{A}} |x_{\alpha}(0)| < M$ for some finite M and
- (ii) $\lim_{\delta \rightarrow 0} \sup_{\alpha \in \mathcal{A}} w(x_{\alpha}, \delta) = 0$.

Now, we are ready to state the tightness conditions in $C_{\mathbb{R}}[0, 1]$.

Theorem 2.2. *A set of probability measures $\{\mathbb{P}_{\alpha}\}_{\alpha \in \mathcal{A}}$ is tight if and only if*

- (i) [(**Modified**) **Compact Containment Condition.**] For each $\eta > 0$, there exists an M_η such that

$$\sup_{\alpha \in \mathcal{A}} \mathbb{P}_\alpha(\{x : |x(0)| > M_\eta\}) < \eta, \text{ and}$$

- (ii) [**Modulus of Continuity Condition.**] For each ϵ and η , there exists a $\delta_{\epsilon, \eta}$ such that

$$\sup_{\alpha \in \mathcal{A}} \mathbb{P}_\alpha(\{x : w(x, \delta_{\epsilon, \eta}) > \epsilon\}) < \eta.$$

The term “modified” in the paranthesis in the first condition refers to the fact that when we are dealing with continuous functions, we need only to check whether $x(0)$ lies in a compact region. Now, we can use tightness and the fact that finite dimensional distributions characterize the law of a process to prove weak convergence on $C_{\mathbb{R}}[0, 1]$. This main idea is expressed in the following theorem.

Theorem 2.3. Suppose $\{X^n\}_{n=1}^\infty$ is a sequence of stochastic processes whose sample paths lies in $C_{\mathbb{R}}[0, 1]$ such that for any finite collection of $t_1, \dots, t_k$

$$(X^n(t_1), \dots, X^n(t_k)) \Rightarrow (X(t_1), \dots, X(t_k)),$$

when thought as vectors in $\mathbb{R}^k$, and for any $\epsilon > 0$

$$\lim_{\delta \downarrow 0} \limsup \mathbb{P}(w(X^n, \delta) \geq \epsilon) = 0.$$

Then, $X^n \Rightarrow X$.

Proof. The first condition imply that $X^n(0) \Rightarrow X(0)$ and hence $\{X^n(0)\}_{n=1}^\infty$ is tight, i.e., for any η there exists

$$\sup_{n \in \mathbb{Z}_+} \mathbb{P}(|X^n(0)| > M_\eta) < \eta.$$

The second condition in the statement of the theorem is equivalent to the modulus of continuity condition and hence, the set of laws of $\{X_n\}_{n=1}^\infty$ is tight and hence for any sequence, we can find a further subsequence that converges to some law $\tilde{\mathbb{P}}$. Using the fact that $C_{\mathbb{R}}[0, 1]$ is separable and complete, we can find $\tilde{X}$ which has law $\tilde{\mathbb{P}}$. The first condition implies that all finite dimensional distributions agree and hence $\tilde{X}$ should have the same law as X . This means that all the subsequences converge to the same limit and hence the original sequence should converge to the same limit and the resul follows. $\square$

3 Example: Existence of Brownian Motion

We define Brownian Motion $\{B(t)\}_{t \in [0, 1]}$ as a stochastic process having the following properties:

- (i) $\mathbb{P}(B(0) = 0) = 1$.
- (ii) The paths of Brownian motion is continuous.
- (iii) For $s \leq t$, $B(t) - B(s)$ is independent of $\{B(u) : 0 \leq u \leq s\}$ and distributed normally with mean 0 and variance $t - s$.

Conditions (ii) and (iii) is fundamental in the definition of Brownian motion, but (i) can be relaxed in different forms. Now, the question we face is whether such a process really exists. Weak convergence techniques can be used to prove this existence result. To do so, consider a sequence of stochastic processes defined by polygonal paths as follows. Let $\{X_i^n\}_{i=1}^\infty$ be a collection of i.i.d. random variables with mean 0 and variance 1, $S_k^n = \sum_{i=1}^k X_i^n$, and

$$B^n(k/n) = \frac{S_k^n}{\sqrt{n}} \text{ for all } k = 0, \dots, n \text{ and} \quad (1)$$

$$B^n(t) = \frac{S_k^n}{\sqrt{n}} + \frac{(nt - k)X_{k+1}}{\sqrt{n}} \text{ when } t \in [k/n, (k+1)/n). \quad (2)$$

We will also need the following two results:

Lemma 3.1 (Etemadi's Inequality). *Let S_k^n be defined as above. Then,*

$$\mathbb{P}(\max_{1 \leq k \leq n} |S_k^n| > 3\epsilon) \leq 3 \max_{1 \leq k \leq n} \mathbb{P}(|S_k^n| > \epsilon).$$

The above result holds more generally when each X_i has different variance by exchanging $n\sigma^2$ with the sum of the variances.

Lemma 3.2. *Fix $0 = t_0 < t_1 < \dots < t_n$ so that $\min_{1 \leq k < n} t_k - t_{k-1} \geq \delta$. Then, for any function $x \in C_{\mathbb{R}}[0, 1]$,*

$$w(x, \delta) \leq 3 \max_{1 \leq k \leq n} \sup_{t_{k-1} \leq s \leq t_k} |x(s) - x(t_{k-1})|,$$

and hence for any probability measure $\mathbb{P}$,

$$\mathbb{P}(\{x : w(x, \delta) > 3\epsilon\}) \leq \sum_{k=1}^n \mathbb{P}\left(\left\{x : \sup_{t_{k-1} \leq s \leq t_k} |x(s) - x(t_{k-1})| > \epsilon\right\}\right).$$

Proof. Consider s, t such that $|t - s| < \delta$. The condition on the distance between t_k s ensures that either s and t are in the same interval $[t_{k-1}, t_k]$ or they are in adjacent intervals. If both $s, t \in [t_{k-1}, t_k]$, we have

$$|x(t) - x(s)| \leq |x(t) - x(t_{k-1})| + |x(s) - x(t_{k-1})|, \text{ and}$$

if they are in adjacent intervals, $s \in [t_{k-1}, t_k]$ and $t \in [t_k, t_{k+1}]$, we have

$$|x(t) - x(s)| \leq |x(t) - x(t_k)| + |x(t_k) - x(t_{k-1})| + |x(s) - x(t_{k-1})|.$$

The result follows by bounding the terms on the righthand side with the maximum in the theorem. The conclusion regarding the probability law follows from the subadditivity. $\square$

For simplicity of the proof we assume that we have finite fourth moment for X_k

Theorem 3.3 (Donsker's Theorem). *Let B^n be as defined in (1) and (2) with independent and i.i.d. random variables X_i^n . Then $B^n \Rightarrow B$, where B is a Brownian motion.*

Proof. Using Theorem 2.3, we first show that the finite dimensional distributions converge and then show that the sequence $\{B^n\}$ satisfy the modulus of continuity condition.

First, realize that $\mathbb{E} \left[\left(\frac{(nt-k)X_{k+1}}{\sqrt{n}} \right) \right] \leq 1/n$ and using Chebyshev's inequality $\frac{(nt-k)X_{k+1}}{\sqrt{n}} \Rightarrow 0$. Then, using the central limit theorem

$$B^n(t) - \frac{S^n_{\lfloor nt \rfloor}}{\sqrt{n}} + \frac{S^n_{\lfloor nt \rfloor}}{\sqrt{n}} \Rightarrow \sqrt{t}Z,$$

where Z is a standard normal random variable. Proceeding in the same fashion, we can prove

$$(B^n(s), B^n(t) - B^n(s)) \Rightarrow (\sqrt{t}Z_1, \sqrt{t-s}Z_2),$$

where Z_1 and Z_2 are independent standard normal random variables. Repeating the same argument proves the convergence of finite dimensional distributions.

To prove the modulus of continuity condition, fix ϵ and η and choose an $m = \frac{3^{11}}{\epsilon^4 \eta}$ (the reason for this specific choice will be clear later). Choose $t_k = k/m$ and realize that the $B^n(t_k) - B^n(t_{k-1})$ is the sum of n/m independent random variables. Using Lemma 3.2 for any $\delta \leq 1/m$, we have

$$\begin{aligned} \mathbb{P}(w(B^n, \delta) > \epsilon) &\leq \sum_{k=1}^m \mathbb{P} \left(\sup_{t_{k-1} \leq s \leq t_k} |B^n(s) - B^n(t_{k-1})| > \epsilon/3 \right) \\ &= m \mathbb{P} \left(\sup_{0 \leq s \leq 1/m} |B^n(s)| > \epsilon/3 \right) \\ &= m \mathbb{P} \left(\max_{1 \leq k \leq n/m} |S_k^n| > \sqrt{n}\epsilon/3 \right) \\ &\leq 3m \max_{1 \leq k \leq n/m} \mathbb{P}(|S_k^n| > \sqrt{n}\epsilon/9) \\ &\leq 3m \mathbb{P}(|S_{n/m}^n| > \sqrt{n}\epsilon/9). \end{aligned}$$

Now, we know from the central limit theorem that $S_k^n/\sqrt{k} \Rightarrow Z$, where Z is a standard normal random variable. Hence, using Markov's inequality with $f(z) = z^4$ and $\mathbb{E}(Z^4) < 3$, we have a n_m such that $n > n_m$ imply

$$3m \mathbb{P}(|S_{n/m}^n| > \sqrt{n}\epsilon/9) < \frac{3n^2}{m^2 n^2 \epsilon^4 / 9^4} < \eta.$$

For the remaining finitely many $B^n(t)$, $1 \leq n \leq n_m$, we use tightness of each probability measure and find a δ_1 such that

$$\sup_{1 \leq n \leq n_m} \mathbb{P}(w(B^n, \delta_1) > \epsilon) < \eta.$$

Choosing $\delta = \min\{1/m, \delta_1\}$ the result follows. $\square$

3.1 An Application of Continuous Mapping Theorem

Suppose that we define S_k as the sum of i.i.d. X_i s with mean 0 and variance σ^2 and we are interested in analyzing how $\max_{1 \leq k \leq n} S_k$ behaves as $n \rightarrow \infty$, i.e., what is the rate of growth and

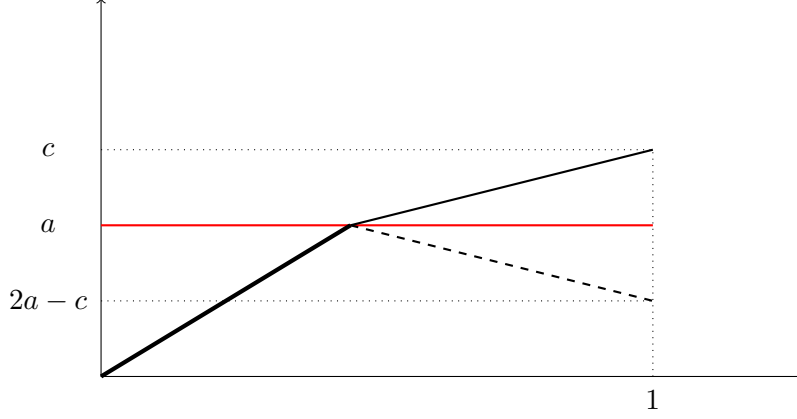


Figure 1: Reflection Principle

for large n how can we approximate $\mathbb{P}(\max_{1 \leq k \leq n} S_k > a)$, etc. Donsker's theorem combined with continuous mapping theorem enables us to answer such questions.

Again using our interpolation idea, we can define $B^n(\cdot)$ as in (1) and (2). It is easy to see that

$$\frac{\max_{1 \leq k \leq n} S_k}{\sqrt{n}} = \max_{0 \leq t \leq 1} B^n(t)$$

and thinking of $\max(\cdot) : C_{\mathbb{R}}[0, 1] \rightarrow \mathbb{R}$ as a function it is easy to see that $\max(\cdot)$ is continuous. Hence, Donsker's theorem implies

$$\frac{\max_{1 \leq k \leq n} S_k}{\sqrt{n}} \Rightarrow \max_{0 \leq t \leq 1} B(t).$$

This means that the maximum $\max_{1 \leq k \leq n} S_k$ is $O(\sqrt{n})$, moreover for large n

$$\mathbb{P}\left(\max_{1 \leq k \leq n} S_k > a\right) \approx \mathbb{P}\left(\max_{0 \leq t \leq 1} B(t) > \frac{a}{\sqrt{n}}\right).$$

This result is not helpful if calculating the probability on the righthand side is as hard as the one on the left. Fortunately, Brownian motion has many nice properties and the probability on the righthand side can be easily calculated using the so-called reflection principle.

Suppose, $\tau_a = \inf\{t : B(t) > a\}$. Then, the reflection principle argues that the number of paths that has $0 < \tau_a < 1$ and $B(1) > c \geq a$ is equal to the number of paths that has $0 < \tau_a < 1$ and $B(1) < a - (c - a) = 2a - c$.

Using reflection principle, we can get

$$\mathbb{P}(\tau_a < 1, B(1) < 2a - c) = \mathbb{P}(\tau_a < 1, B(1) > c) = \mathbb{P}(B(1) > c),$$

and taking $c = a$ this implies

$$\mathbb{P}\left(\max_{0 \leq t \leq 1} B(t) > a\right) = \mathbb{P}(\tau_a < 1) = \mathbb{P}(\tau_a < 1, B(1) < a) + \mathbb{P}(\tau_a < 1, B(1) > a) = 2\mathbb{P}(B(1) > a).$$

4 Continuous Functions on $[0, \infty)$

Our results so far extends to any bounded closed interval by a simple normalization. However, one needs to be careful when dealing with the unbounded interval $[0, \infty)$. This case was studied by Whitt [3] and this space is metricized by

$$\rho(x, y) = \sum_{j=1}^{\infty} \frac{\rho_j(x, y)}{2^{-j}(1 + \rho_j(x, y))},$$

where $\rho_j(x, y) = \sup_{0 \leq t \leq j} |x(t) - y(t)|$. This metric is the metric of *uniform convergence on compact sets*. This implies that if you wish to prove that a sequence of functions $\{x_n(\cdot)\}_{n=1}^{\infty} \subset C_{\mathbb{R}}[0, \infty)$ converges to x , then we need to prove uniform convergence on compact sets using the techniques above.

Remark 4.1. *In most cases, we need to associate weak convergence of processes and the convergence of stationary distributions. As we are using uniform convergence on compact sets, one should realize that this relationship is not trivial and additional arguments are needed to conclude convergence of stationary distributions from weak convergence.*

References

- [1] P. Billingsley. *Convergence of Probability Measures*. John Wiley and Sons, New York, 1997.
- [2] W. Rudin. *Functional Analysis*. McGraw-Hill, Boston, 1991.
- [3] W. Whitt. Weak convergence of probability measures on the function space $C[0, \infty)$. *The Annals of Mathematical Statistics*, 41(3):939–944, 1970.