# Notes on Weak Convergence and the Skorokhod Topology

SooJean Han*

January 29, 2020

**Abstract**

Notes on (stochastic/probabilistic) metric spaces, topologies on $C[0,1]$, the space of continuous functions on the interval $[0,1]$, and a new definition $D[0,1]$ which includes discontinuous functions on $[0,1]$.

# 1 Background

## 1.1 Four Notions of Convergence for Random Variable Sequences

Define a sequence of random variables $\{X_n(\omega), n \geq 1\}$ on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$. There are four main notions of convergence for random variables. We will briefly review them for the specific case where the $X_n$ are scalar and real-valued.

1. The strongest, most restrictive notion is to require $X_n(\omega)$ to converge for each fixed sample path $\omega$, except for the events that occur with probability zero.

   **Definition 1** (Almost-Sure Convergence). *$\{X_n(\omega), n \geq 1\}$ is said to converge* ***almost surely*** *to some random variable $X$, defined on the same probability space as every $X_n$, if*

   $$\mathbb{P}\left(\omega \,:\, \lim_{n\to\infty} X_n(\omega) = X(\omega)\right) = 1$$

   *Almost-sure convergence is denoted as $\lim_{n\to\infty} X_n = X$ a.s. or $X_n \xrightarrow{a.s.} X$.*

   **Example 1** (Almost-Sure Convergence). *Let $\{X_n, n \geq 1\}$ be the sequence $X_n(\omega) = \omega^n$, defined on the uniform unit-interval probability space $\Omega = [0,1]$. The sequence is shown in Figure 1. Note that it converges for all $\omega \in \Omega$, with the limit*

   $$\lim_{n\to\infty} X_n(\omega) = \begin{cases} 0 & \textit{if } 0 \leq \omega < 1 \\ 1 & \textit{if } \omega = 1 \end{cases}$$

*SooJean Han is with the Department of Control and Dynamical Systems, California Institute of Technology, Pasadena, CA 91125, USA. Email: soojean@caltech.edu

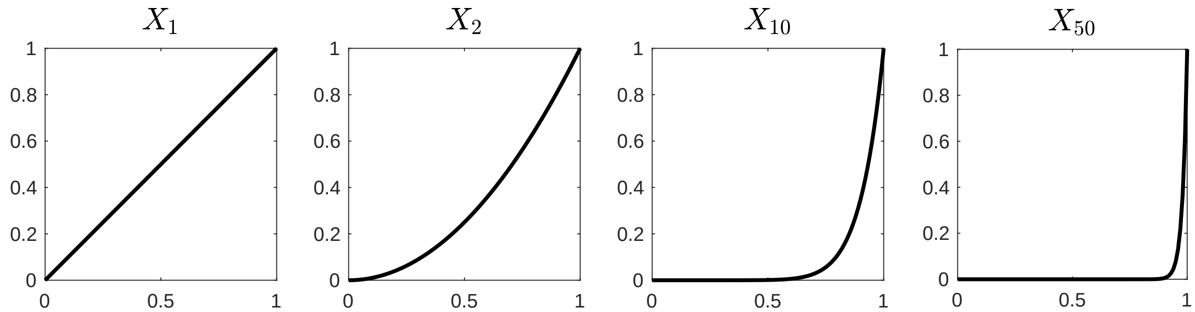


**Figure 1:** The sequence of random variables on $[0,1]$, defined as $X_n(\omega) = \omega^n$.

*One can compute the corresponding probability density functions (PDFs) $f_n$ for each random variable $X_n$. The uniform unit-interval probability space $\Omega$ is transformed to a nonuniform one via the mapping given by $X_n$. Hence:*

$$F_n(x) = \mathbb{P}(X_n \leq x) = \sqrt[n]{x} \quad \Longrightarrow \quad f_n(x) = \frac{1}{n} x^{\frac{1}{n}-1}$$

*The set $\{1\}$ has probability of occurrence zero. Thus, we can see that $\{X_n\}$ converges a.s. to $X$, where $X(\omega) = 0$ for all $\omega \in [0,1]$.*

*Alternatively, we can define a sequence of PDFs $f_n$ by rescaling each $X_n$ so that the area under the curve is 1:*

$$f_1(x) = 2x, f_2(x) = 3x^2, f_3(x) = 11x^{10}, \quad \Longrightarrow \quad f_n(x) = (n+1)x^n$$

*This sequence of PDFs converge almost-surely to the point density function with all its probability mass at $x = 1$: $f(1) = 1, f(x) = 0$ for all $x \in [0,1)$.*

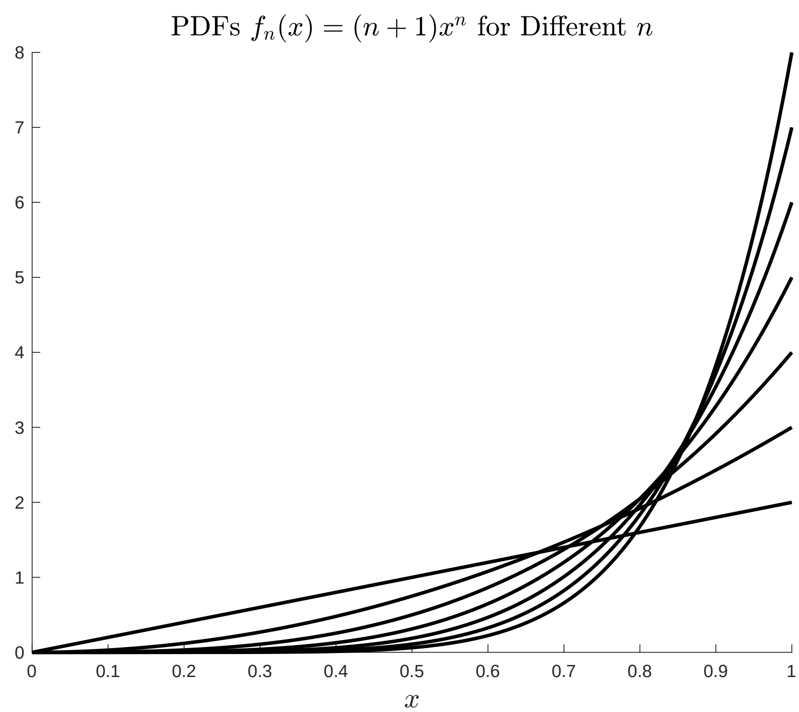


**Figure 2:** The sequence of PDFs on $[0,1]$, defined as $f_n(x) = (n+1)x^n$ for $n = 1, \cdots, 7$.

2. We now define a weaker notion of convergence than almost-sure.

**Definition 2** (Convergence in Probability). *$\{X_n(\omega), n \geq 1\}$ is said to converge **in probability** to some random variable $X$, defined on the same probability space as every $X_n$, if for any $\varepsilon > 0$*

$$\lim_{n\to\infty} \mathbb{P}(\omega \,:\, |X(\omega) - X_n(\omega)| \geq \varepsilon) = 0$$

*Convergence in probability is denoted as $\lim_{n\to\infty} X_n = X$ p. or $X_n \xrightarrow{p.} X$.*

Essentially, we must have $|X - X_n|$ be small with high probability in order for the sequence $\{X_n\}$ to be convergent in probability.

**Example 2** (Moving Rectangles [1]). *Let $\{f_n, n \geq 1\}$ be the sequence of probability density functions defined below:*

$$f_n(x) = \begin{cases} 2^{\lfloor \log_2 n \rfloor} & \text{on } \omega \in \left[\frac{k}{2^m}, \frac{(k+1)}{2^m}\right] \\ 0 & \text{everywhere else} \end{cases}$$

*where $m = \lfloor \log_2 n \rfloor$ and $k = mod(n, 2^m)$.*

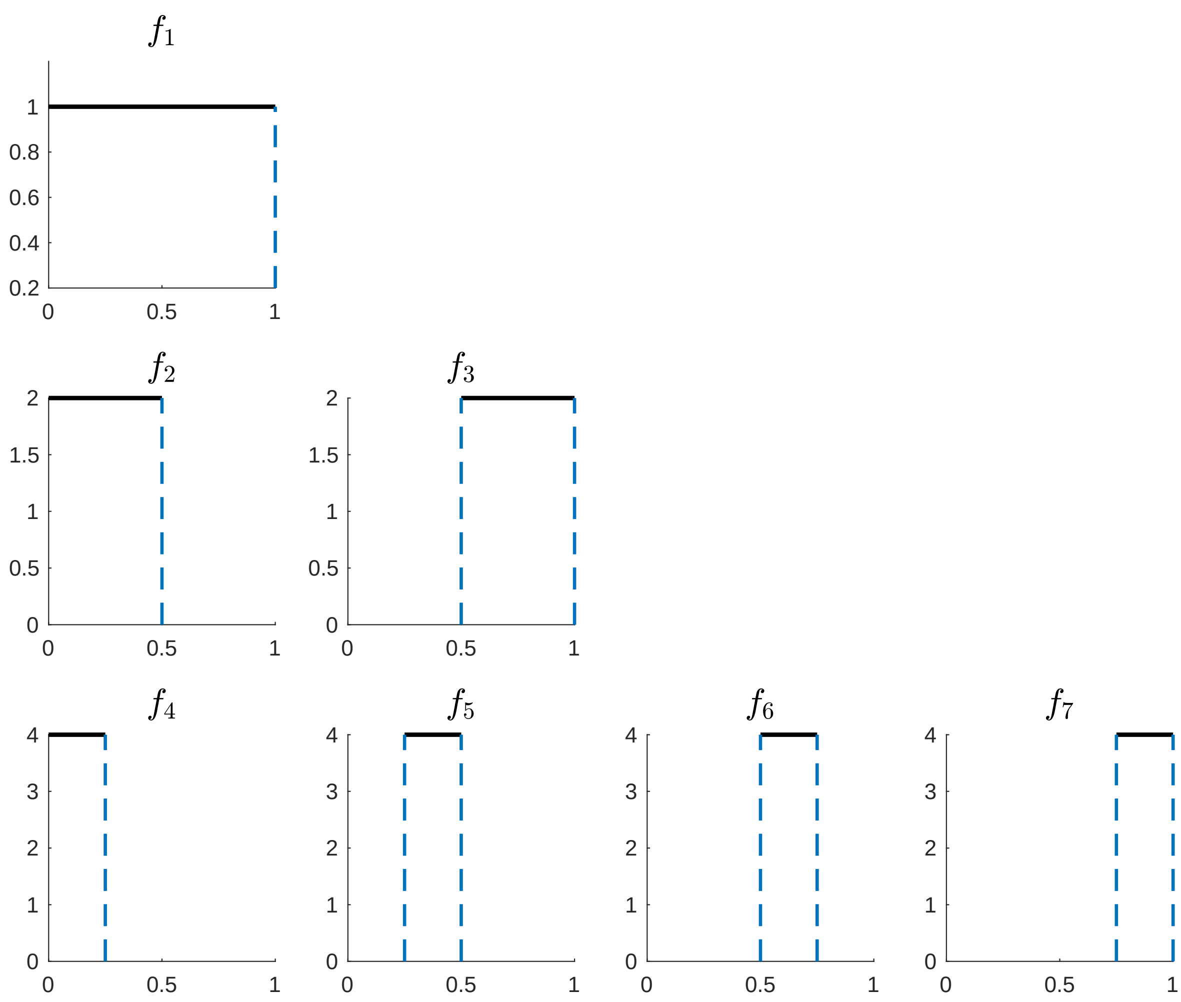


**Figure 3:** The moving rectangles sequence of PDFs on $[0, 1]$.

*This sequence does not converge a.s. because at each point $\omega \in [0,1]$, the value of $X_n$ alternates aperiodically with $n$ between 0 and a nonzero value.*

*However, note that the probability $\lim_{n\to\infty} \mathbb{P}(f_n = 0)$ approaches one because the length of the subinterval in which $f_n \neq 0$ approaches 0. Hence, $f_n \xrightarrow{p.} 0$.*

3. For convergence in probability, there may be a small probability event where $|X - X_n|$ is not small. For some applications, the value within this small event may be unacceptably large; we've seen this in the previous example, where the height of the rectangle increases with increasing $n$ in order to preserve the area to be 1.

   This motivates the next notion of convergence, which ensures that even if the individual value of $|X - X_n|$ is not small for this event, then at least the $p$th moment should be, where $p \in \mathbb{N}$.

   **Definition 3** (Convergence in $L^p$)**.** *Given a fixed $p \geq 1$, the sequence $\{X_n(\omega), n \geq 1\}$, where $\mathbb{E}[|X_n|^p] < \infty$ for all $n$, is said to converge **in** $L^p$ to some random variable $X$, defined on the same probability space as every $X_n$, if*

$$\lim_{n\to\infty} \mathbb{E}\left[|X - X_n|_p\right] = 0$$

   *where the $L^p$ norm is defined as*

$$|X - X_n|_p = \left( \int_\Omega |X_n(\omega) - X(\omega)|^p dP(\omega) \right)^{\frac{1}{p}}$$

   *$L^p$ convergence is denoted as $X_n \xrightarrow{L^p} X$.*

   The most commonly used case is when $p = 2$, where it is referred to as **mean-square convergence**. Mean-square convergence is denoted as $\lim_{n\to\infty} X_n = X$ m.s. or $X_n \xrightarrow{\text{m.s.}} X$.

   **Example 3** (Convergence in Mean-Square)**.** *Let $\{X_n, n \geq 1\}$ be the sequence of moving rectangles on $[0,1]$ which are decreasing in size, similar in concept to Example 2. In this case, however, we take the width of the rectangle to be $\frac{1}{n}$, and the height is a deterministic sequence $h_n$. The location of the interval is not specified.*

   *If $h_n > 0$ for all $n$, then $X_n$ does not converge almost-surely to 0 for the reason specified in Example 2: for any $\omega \in [0,1]$, the value of $X_n$ alternates between 0 and a nonzero value as $n$ increases.*

   *On the other hand, the sequence converges to zero in probability for any sequence $\{h_n\}$ because for any $\varepsilon > 0$,*

$$\mathbb{P}(|X_n - 0| \geq \varepsilon) \leq \mathbb{P}(X_n \neq 0) = \frac{1}{n} \to 0$$

   *Now we consider mean-square convergence. We have*

$$\mathbb{E}[|X_n - 0|^2] = \frac{1}{n} h_n^2 \to 0 \textit{ only if } \lim_{n\to\infty} \frac{h_n^2}{n} = 0$$

   *So for heights which grow on an order strictly less than $\sqrt{n}$ (e.g. $h_n = \ln(n)$ or $h_n = n^p$ for $p < 1/2$), the sequence converges in mean-square.*

4. The final, weakest notion of convergence is defined as follows. We will be using this notion of convergence most often throughout this manuscript, and we study it in much more detail within the context of metric spaces in the following subsection 1.2.

**Definition 4** (Weak Convergence). *$\{X_n(\omega), n \geq 1\}$ is said to converge **in distribution** (or **weakly**) to some random variable $X$ if*

$$\lim_{n\to\infty} F_n(x) = F(x) \text{ at all continuity points } x \text{ of } F$$

*where $F_n$ and $F$ are the CDFs of $X_n$ and $X$, respectively. Convergence in distribution is denoted as $\lim_{n\to\infty} X_n = X$ d. or $X_n \xrightarrow{d.} X$.*

**Example 4.** *Let $\{X_n(t), n \geq 2 \text{ even}\}$ be a sequence defined by*

$$X_n(t) := \begin{cases} 0 & \text{for } t \in \left[\frac{2k}{n}, \frac{2k+1}{n}\right) \\ n^{-\frac{1}{3}} & \text{for } t \in \left[\frac{2k+1}{n}, \frac{2k+2}{n}\right) \end{cases}$$

*for $0 \leq k \leq \frac{n}{2} - 1$ and on $t \in [0, 1)$.*

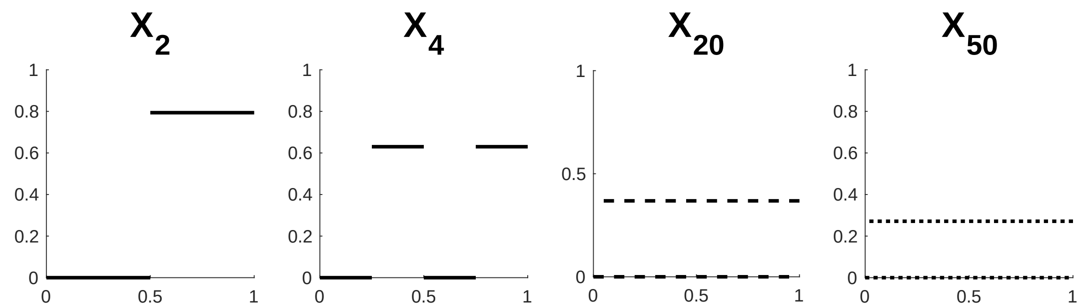


**Figure 4:** Sequence of random variable rectangles on $[0, 1]$ for Example 4.

*Note that it converges in distribution to $X(t) \equiv 0$.*

**Example 5** (Weak Convergence: Cosine). *Consider the sequence of random variables $X_n(\omega)$ with probability density functions given by $f_n(x) = 1 - \cos(2\pi n x)$ on the interval $x \in [0, 1]$. See Figure 5 for visualizations of three chosen values of $n$.*

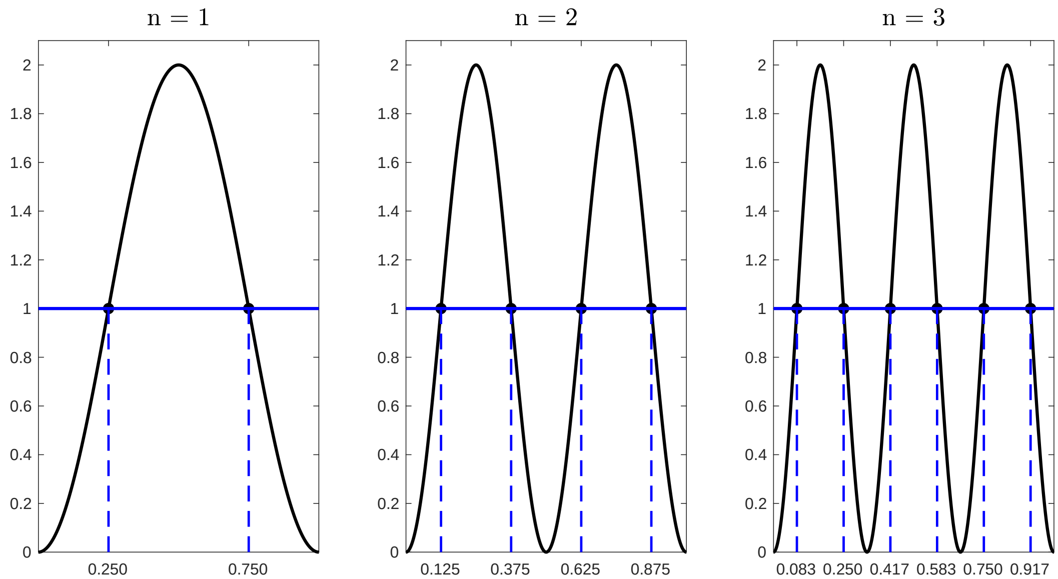


**Figure 5:** Sequence of functions $f_n(x) = 1 - \cos(2\pi n x)$ for $n = 1, 2, 3$.

*Let $f(x) = 1$ be the uniform probability density function on $[0, 1]$. We will show that $f_n$ converges weakly to $f$.*

*Using Definition 4 above, we compare the CDFs $F_n(x)$ of the sequence and the CDF $F(x)$ of the limiting function:*

$$F_n(x) = \int_0^x f_n(t)dt = \int_0^x (1 - \cos(2\pi n t))\, dt = x - \frac{1}{2\pi n}\sin(2\pi n x)$$
$$F(x) = \int_0^x f(t)dt = \int_0^x 1dt = x$$

*Clearly, the sequence of CDFs $F_n$ converge pointwise to $F$ as $n \to \infty$. See Figure 6.*

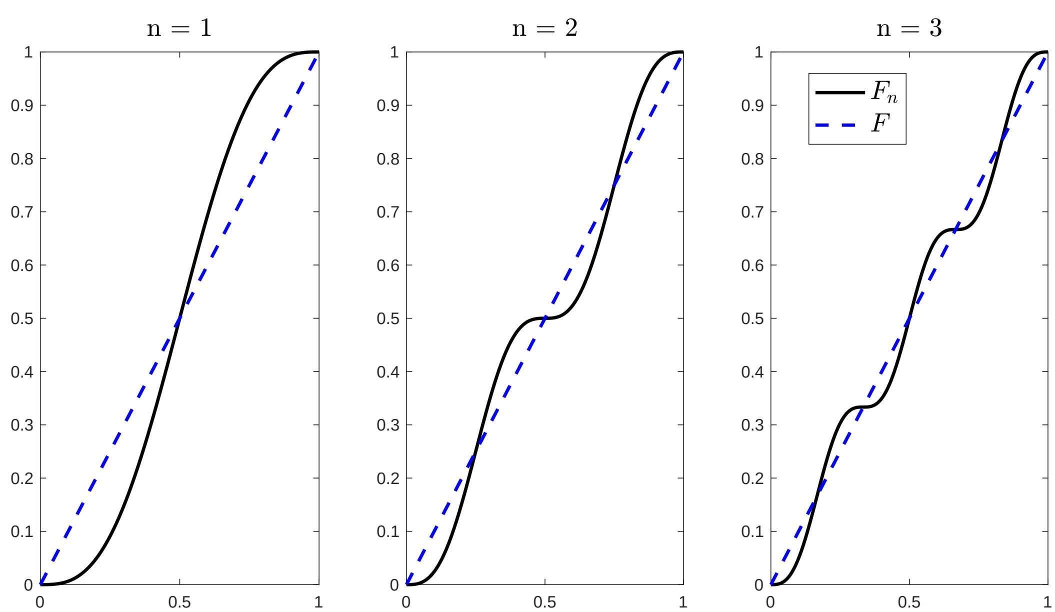


**Figure 6:** The CDFs of the sequence of functions $f_n(x) = 1 - \cos(2\pi n x)$ [in black] along with the CDF of the uniform density $F(x) = x$ [in blue dashed].

*On the other hand, does $f_n(x)$ converge to $f(x)$ almost-surely? As we can see in Figure 5, only a countably many number of points $x \in [0,1]$ satisfy $f_n(x) = 1$ for each $n$. For instance, when $n = 1$, $f_1(x) = 1$ only at the points $x = \frac{1}{4}, \frac{3}{4}$ and when $n = 2$, $f_2(x) = 1$ only at the points $x = \frac{1}{8}, \frac{3}{8}, \frac{5}{8}, \frac{7}{8}$. For general $n$, $f_n(x)$ has $2n$ points at which it takes value 1, and they are given by $x = \frac{2k-1}{4n}$ for $k = 1, \cdots, 2n$. But an uncountably-many number of points in $[0,1]$ do not satisfy $f_n(x)$, no matter how large $n$ grows. Thus, $f_n \not\to f$ a.s.*

**Example 6** (Weak Convergence on a Hilbert Space)**.** *We can also show that the sequence $\{f_n\}$ of cosine functions defined in the previous example converges weakly to $f$ when they are considered elements of the space $L^2([0,1])$ equipped with the usual integral inner product.*

*Partition the interval into $N$ subintervals as $0 = x_0 < x_1 < \cdots < x_N = 1$. Let any $g$ be a piecewise-linear function in $L^2([0,1])$: $g(x) = \sum_{j=1}^{N} c_j \mathbb{1}(x_{j-1} \leq x < x_j)$ where $c_j$ are constants.*

*Compute:*

$$\lim_{n\to\infty} \int_0^1 g(x) f_n(x) dx = \lim_{n\to\infty} \sum_{j=1}^{N} \int_{x_{j-1}}^{x_j} g(x) f_n(x) dx = \lim_{n\to\infty} \sum_{j=1}^{N} c_j \int_{x_{j-1}}^{x_j} f_n(x) dx$$

*The integral itself is computed to be:*

$$\begin{aligned}\int_{x_{j-1}}^{x_j} 1 - \cos(2\pi n x) dx &= \left( x - \frac{1}{2\pi n} \sin(2\pi n x) \right) \Big|_{x_{j-1}}^{x_j} \\ &= (x_j - x_{j-1}) - \frac{1}{2\pi n} \left( \sin(2\pi n x_j) - \sin(2\pi n x_{j-1}) \right)\end{aligned}$$

*and note that the second term tends to 0 as $n \to \infty$. Thus:*

$$\lim_{n\to\infty} \int_0^1 g(x) f_n(x) dx = \lim_{n\to\infty} \sum_{j=1}^{N} c_j (x_j - x_{j-1}) \tag{1}$$

*which is exactly the Riemann sum area of $g(x)$ by itself, i.e.*

$$(1) = \lim_{n\to\infty} \int_0^1 g(x) f(x) dx$$

*By the Monotone Convergence Theorem, this can be shown true for all general $g$, not necessarily a step function. Therefore, we can conclude that $f_n(x) \Rightarrow f(x)$.*

**Example 7** (Weak Convergence: Sine)**.** *Analogously, we can consider the sequence of PDFs $f_n(x) = 1 - \sin(2\pi n x)$ on the interval $x \in [0,1]$. See Figure 7 for three chosen values of $n$.*

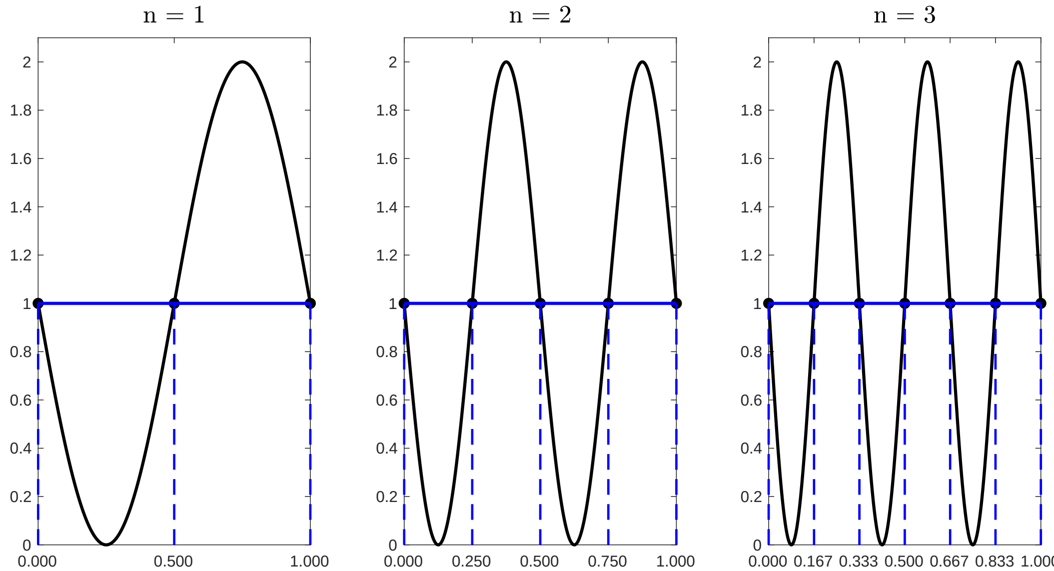


**Figure 7:** Sequence of functions $f_n(x) = 1 - \sin(2\pi nx)$ for $n = 1, 2, 3$.

*A quick integral calculation shows that each sequence is indeed a probability density function:*

$$F_n(x) = \int_0^x f_n(t)dt = \int_0^x (1 - \sin(2\pi nt))\, dt = x + \frac{1}{2\pi n}(\cos(2\pi nx) - 1)$$
$$\implies F_n(0) = 0,\ \ F_n(1) = 1$$

*Again, let $f(x) = 1$ be the uniform probability density function on $[0, 1]$. Using Definition 4 above, compare the CDFs $F_n(x)$ of the sequence and the CDF $F(x)$ of the limiting function:*

$$F_n(x) = \int_0^x f_n(t)dt = \int_0^x (1 - \sin(2\pi nt))\, dt = x + \frac{1}{2\pi n}(\cos(2\pi nx) - 1)$$

*The sequence of CDFs $F_n(x)$ again converge to $F(x)$ as $n \to \infty$. See Figure 8 below.*

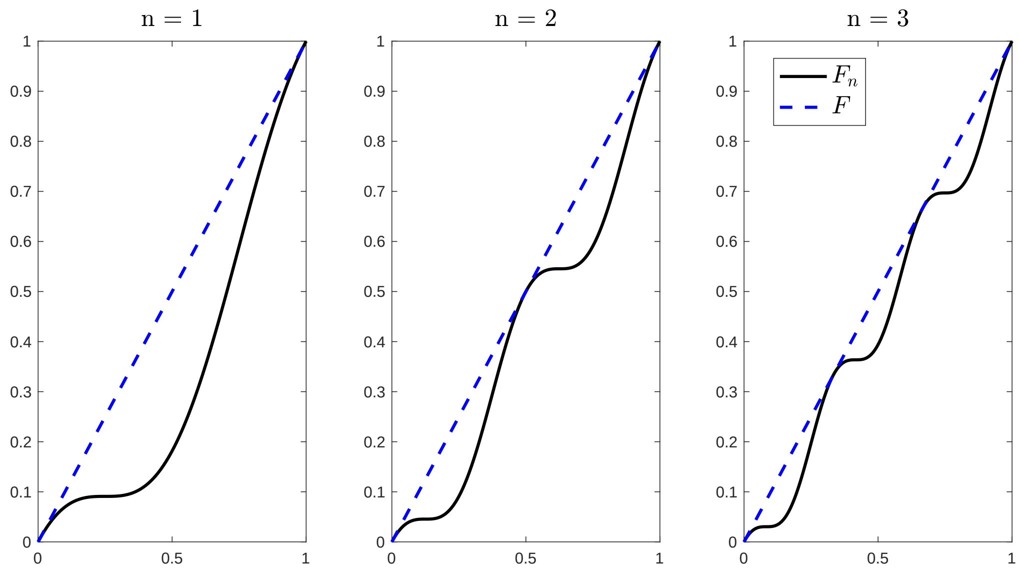


**Figure 8:** The CDFs of the sequence of functions $f_n(x) = 1 - \sin(2\pi n x)$ [in black] along with the CDF of the uniform density $F(x) = x$ [in blue dashed].

*From Figure 7, we note that for each $n$, only a countably many number of points $x \in [0, 1]$ satisfy $f_n(x) = 1$. For instance, when $n = 1$, $f_1(x) = 1$ only at the points $x = 0, \frac{1}{2}, 1$ and when $n = 2$, $f_2(x) = 1$ only at the points $x = 0, \frac{1}{4}, \frac{1}{2}, \frac{3}{4}, 1$. For general $n$, $f_n(x)$ has $2n + 1$ points at which it takes value 1, and they are given by $x = \frac{k}{2n}$ for $k = 0, \cdots, 2n$. But an uncountably-many number of points in $[0, 1]$ do not satisfy $f_n(x)$, no matter how large $n$ grows. Thus, $f_n \not\to f$ a.s.*

**Example 8** (Alternating CDFs)**.** *Now consider the sequence of functions $f_n$ defined as follows:*

$$f_n(t) = \begin{cases} 2 \cdot \mathbb{1}\left\{t \in \left[0, \frac{1}{2}\right]\right\} & \text{if } n \text{ odd} \\ 2 \cdot \mathbb{1}\left\{t \in \left[\frac{1}{2}, 1\right]\right\} & \text{if } n \text{ even} \end{cases}$$

*Clearly they are probability density functions because $f_n(x) \geq 0$ for all $x \in [0, 1]$ and the integral of $f_n$ over $[0, 1]$ is 1.*

*Let us check if convergence in distribution holds. We have two separate CDFs depending on whether $n$ is even or odd. See Figure 9 for illustration.*

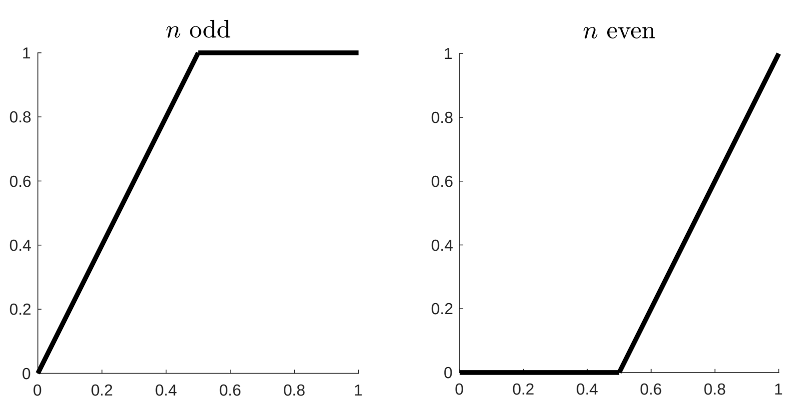


**Figure 9:** The two different CDFs for $n$ odd and $n$ even.

*As $n \to \infty$, there is no pointwise convergence of $F_n(t)$ anywhere for $t \in (0,1)$. For example, it is unclear what happens at $t = \frac{1}{2}$: $F_n(1/2) = 0$ when $n$ is even but becomes $1$ when $n$ is odd. Thus, it this sequence does not converge in distribution to any limiting function.*

**Example 9.** *Consider the sequence $\{X_n, n \geq 1\}$ of binomially-distributed random variables:*

$$\mathbb{P}\left[X_n = k\right] = \binom{n}{k} p^k (1-p)^{n-k}$$

*We will show that taking the limit of this distribution with respect to $n$ yields a Poisson distribution, so that $X_n$ converges in distribution to a Poisson random variable $X$ with intensity $\lambda$.*

*Recall Stirling's approximation*

$$n! \approx \sqrt{2\pi n}\left(\frac{n}{e}\right)^n$$

*Then we get:*

$$\begin{aligned}
\lim_{n\to\infty} \mathbb{P}\left[X_n = k\right] &= \lim_{n\to\infty} \frac{n!}{k!(n-k)!} p^k (1-p)^{n-k} \\
&\approx \lim_{n\to\infty} \frac{\sqrt{2\pi n}\left(\frac{n}{e}\right)^n}{\sqrt{2\pi (n-k)}\left(\frac{n-k}{e}\right)^{n-k}} \cdot \frac{p^k}{k!}(1-p)^{n-k} \\
&= \lim_{n\to\infty} \sqrt{\frac{n}{n-k}} \cdot \frac{n^n}{(n-k)^{n-k} e^k} \cdot \frac{p^k}{k!}(1-p)^{n-k} \qquad (2)
\end{aligned}$$

*Note that as $n \to \infty$, $\sqrt{n/(n-k)} \to 1$. We will additionally define $\lambda$ to be such that $np \to \lambda$ as $n \to \infty$. Continuing:*

$$\begin{aligned}
(2) &= \lim_{n\to\infty} \frac{n^n}{(n-k)^{n-k}} \cdot \frac{\left(\frac{\lambda}{n}\right)^k}{k!}\left(1 - \frac{\lambda}{n}\right)^{n-k} e^{-k} \\
&= \lim_{n\to\infty} \frac{n^n}{n^{n-k}\left(1-\frac{k}{n}\right)^{n-k}} \cdot \frac{\left(\frac{\lambda}{n}\right)^k}{k!}\left(1 - \frac{\lambda}{n}\right)^{n-k} e^{-k} \\
&= \lim_{n\to\infty} \frac{1}{\left(1-\frac{k}{n}\right)^{n-k}} \cdot \frac{\lambda^k}{k!}\left(1 - \frac{\lambda}{n}\right)^{n-k} e^{-k} = \frac{1}{e^{-k}} \cdot e^{-k}\frac{\lambda^k}{k!}e^{-\lambda} = e^{-\lambda}\frac{\lambda^k}{k!}
\end{aligned}$$

*This is indeed the probability density of a Poisson random variable with intensity $\lambda$. Thus, $X_n \xrightarrow{d.} X$.*

**Example 10** (Poisson Distribution)**.** *We can plot the Poisson PDF and CDF for when $\lambda = 5$.*

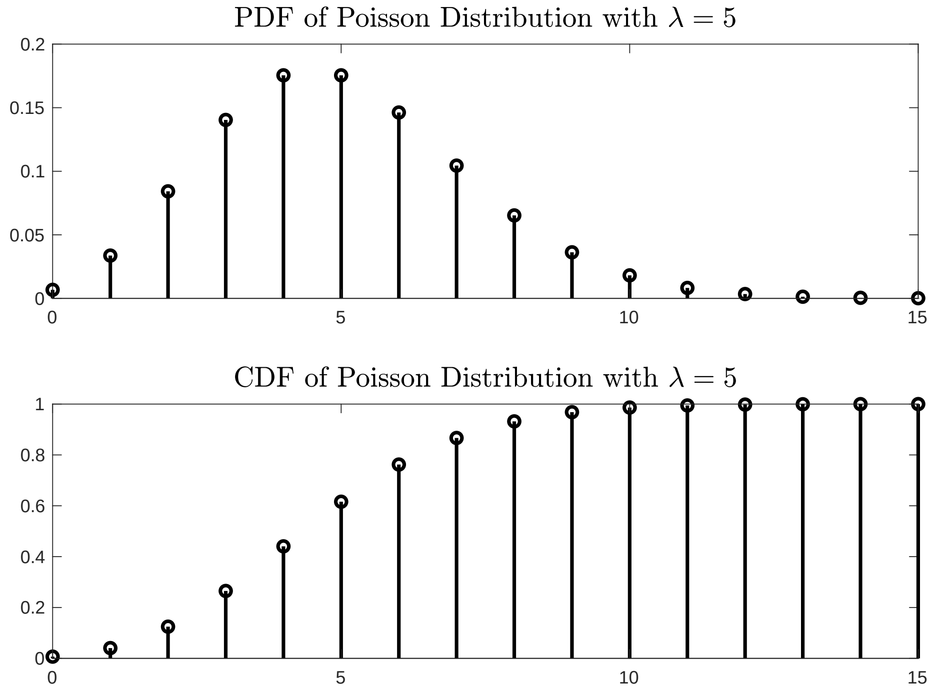


**Figure 10**

**Proposition 1.** *We have the following relationships between the four notions of convergence.*

1. *If* $X_n \xrightarrow{a.s.} X$, *then* $X_n \xrightarrow{p.} X$.
2. *If* $X_n \xrightarrow{m.s.} X$, *then* $X_n \xrightarrow{p.} X$.
3. *If* $\mathbb{P}(|X_n| \leq Y) = 1$ *for all* $n \geq 1$ *for some fixed random variable* $Y$ *with* $\mathbb{E}[Y^2] < \infty$ *and if* $X_n \xrightarrow{p.} X$, *then* $X_n \xrightarrow{m.s.} X$.
4. *If* $X_n \xrightarrow{p.} X$, *then* $X_n \xrightarrow{d.} X$.
5. *Suppose* $X_n \to X$ *in the p., m.s., or a.s. sense and* $X_n \to Y$ *in the p., m.s., or a.s. sense. Then* $\mathbb{P}(X = Y) = 1$. *That is, if the differences between sets of probability zero are ignored, then limits are unique.*
6. *Suppose* $X_n \xrightarrow{d.} X$ *and* $X_n \xrightarrow{d.} Y$. *Then* $X$ *and* $Y$ *have the same distribution.*

## 1.2 Metric Spaces

We restate several fundamental mathematical concepts as background for our upcoming discussion of the Skorokhod topologies. Many basic definitions were obtained from [2], but can be found in any standard reference on real analysis.

**Definition 5** (Metric Space)**.** *If, for a given set* $\mathcal{X}$ *and a given function* $d : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$, *the following properties hold:*

1. $d(x, y) \geq 0$ *for all* $x, y \in \mathcal{X}$ *and* $d(x, y) = 0$ *iff* $x = y$

2. $d(x, y) = d(y, x)$ *for all* $x, y \in \mathcal{X}$ *(symmetry)*

3. $d(x, z) \leq d(x, y) + d(y, x)$ *for all* $x, y, z \in \mathcal{X}$ *(Triangle Inequality)*

*then the pair* $(\mathcal{X}, d)$ *(or, for simplicity, just* $\mathcal{X}$*) is called a* ***metric space*** *and d is called a* ***metric****.*

**Definition 6** (Completeness)**.** *A sequence* $\{x_n\} \subset \mathcal{X}$ *is said to be a* ***Cauchy sequence*** *if for every* $\epsilon > 0$*, there exists some* $N \in \mathbb{N}$ *such that for all* $n, m \geq N$*, we have* $d(x_n, x_m) < \epsilon$*. A metric space is said to be* ***complete*** *if every Cauchy sequence converges to a limit* $x$ *which belongs in* $\mathcal{X}$*.*

*An example of a metric space which is not complete is the open interval* $(0, 1)$ *equipped with the usual absolute value metric. The Cauchy sequence* $\{\frac{1}{n}\}$ *converges to* $0$ *in the limit, but* $0 \notin (0, 1)$*.*

**Definition 7** (Separablility)**.** *A subset* $S \subset \mathcal{X}$ *is said to be* ***dense*** *in* $\mathcal{X}$ *if for every* $x \in \mathcal{X}$*, there exists a sequence of points* $\{x_n\} \subset S$ *such that* $x_n \to x$*. A metric space is said to be* ***separable*** *if it contains a dense, countable subset* $S$*.*

*A well-known example of a separable space is* $\mathbb{R}$*, with the rational numbers* $\mathbb{Q}$ *as a dense subset. It can be proven that every irrational number can be expressed as the limit of a sequence of rational numbers. For instance,* $\pi = 3 + 0.1 + 0.04 + \cdots$

**Definition 8** (Compactness)**.** *A metric space* $\mathcal{X}$ *is* ***compact*** *if for every collection* $\mathscr{A}$ *of open subsets* $A$ *of* $\mathcal{X}$ *such that* $\mathcal{X} = \bigcup_{A \in \mathscr{A}} A$*, there exists a finite subset* $\mathscr{B} \subset \mathscr{A}$ *such that* $\mathcal{X} = \bigcup_{B \in \mathscr{B}} B$*. This can be restated as follows: every open cover of* $\mathcal{X}$ *has a finite subcover.*

*A set* $S$ *is* ***relatively compact*** *if its closure* $\overline{S}$ *is compact.*

*Every closed interval* $[a, b] \subset \mathbb{R}$ *is compact. Every open interval* $(a, b) \subset \mathbb{R}$ *is not compact but is relatively compact.*

We are particularly interested in the space $C := C[0, T]$ of continuous functions on the interval $[0, T]$ (or $C_\infty := C[0, \infty)$) with the distance metric between two elements $f, g \in C$ defined to be $d(f, g) := \|f - g\| := \sup_{t \in [0,T]} |f(t) - g(t)|$.

**Definition 9** (Uniform Continuity)**.** *A function* $f \in C_\infty$ *(i.e.* $f : [0, \infty) \to \mathbb{R}$*), is said to be* ***uniformly continuous*** *on* $[0, \infty)$ *if for every* $\epsilon > 0$*, there exists a* $\delta > 0$ *such that*

$$|f(s) - f(t)| < \epsilon \ \forall \ s, t \in [0, \infty) \ s.t. \ |s - t| < \delta$$

*Intuitively, this differs from the standard notion of continuity in that the property described above should hold for a single value of* $\delta$ *in the entire domain of the function. Furthermore, note that every function in* $C[0, T]$ *is uniformly continuous because* $[0, T]$ *is a compact set and any continuous function defined on a compact set is bounded.*

**Definition 10** (Uniform Convergence)**.** *A sequence of functions* $\{f_n\} \subset C[0, \infty)$ *is said to be* ***uniformly convergent*** *on* $[0, \infty)$ *with limit* $f : [0, \infty) \to \mathbb{R}$ *if for every* $\epsilon > 0$*, there exists* $N \in \mathbb{N}$ *such that*

$$|f_n(t) - f(t)| < \epsilon \ \forall \ t \in [0, \infty), \ \text{and} \ \forall \ n \geq N$$

Roughly speaking, a **uniform topology (space)** generalizes the metric space with an additional structure that allows for characterizations of uniformity such as uniform continuity, uniform convergence, and completeness. This is an important distinction to make from the Skorokhod topologies that we will define later.

**Definition 11** (Equicontinuous)**.** *A family $\mathcal{F}$ of functions on $C[0,T]$ is said to be **equicontinuous** if for every $\varepsilon > 0$, there exists a $\delta > 0$ such that*

$$|f(s) - f(t)| < \varepsilon \ \forall s, t \in [0,T] \ s.t. \ |s-t| < \delta \ \forall f \in \mathcal{F}$$

**Example 11** (Uniform Continuity Does Not Imply Equicontinuity)**.** *Definition 11 suggests that a single value of $\delta$ should be used for every function in $\mathcal{F}$ and every point in the domain the functions in $\mathcal{F}$. Hence, every member $f$ of an equicontinuous family is uniformly continuous.*

*However, the reverse implication does not hold. One example of this is given by $\mathcal{F} = \{f_n(t)\} = \{\sin(nt)\}$. Each $f_n$ is uniformly continuous on the interval $[0, 2\pi]$ because $|f_n'(t)| = |n\cos(nt)| \leq n$ which implies that $|f_n(t) - f_n(s)| \leq n|t-s|$, which is always strictly less than $\varepsilon > 0$ provided that we choose $\delta < \varepsilon/n$. However, note how the value of $\delta$ depends on $n$. So if we attempt to fix a constant $\delta$ for the entire family $\mathcal{F}$, there always exists an $n$ sufficiently large such that the condition $|f_n(t) - f_n(s)| < \varepsilon$ does not hold.*

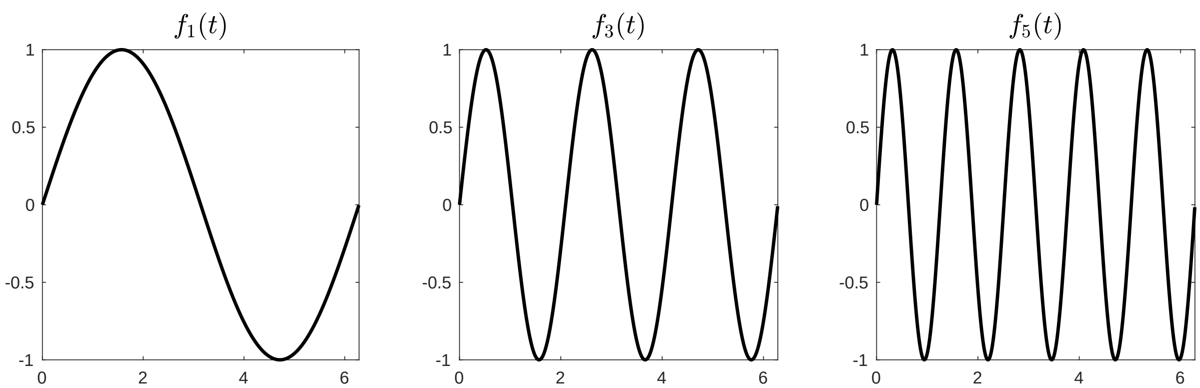


**Figure 11:** A family of uniformly-continuous functions which is not equicontinuous.

*Another example is the family of functions $f_n(t) = t^n$ defined on the closed interval $[0,1]$. Illustrations of some members of the sequence have been shown previously in Figure 1. Again, because each $f_n$ is a continuous function defined on a compact space, each $f_n$ is uniformly continuous on $[0,1]$. We have $|f_n'(t)| = |nt^{n-1}| \leq n$ since $t \in [0,1]$. Now, this family is not equicontinuous for the same reason as in the family of sines: the value of $\delta$ depends on $n$.*

The concept of open covers and finite subcovers is difficult to apply to many metric spaces, and especially for $C[0,T]$. What follows is the Arzelá - Ascoli Theorem, in which equivalent conditions for compactness of $\mathcal{X}$ is instead expressed in terms sequences of elements in $\mathcal{X}$.

**Theorem 1** (**Arzelá - Ascoli**)**.** *Let $\mathcal{F} := \{f_n\} \subset C$ be a sequence of functions such that*

1. *$\mathcal{F}$ is equicontinuous*

2. $\mathcal{F}$ *is* ***pointwise-bounded****: for each* $t \in [0, T]$*, there exists a function* $\varphi(t) > 0$ *such that* $|f_n(t)| \leq \varphi(t)$ *for all* $n \in \mathbb{N}$

*Then the following conditions hold:*

1. $\mathcal{F}$ *is* ***uniformly bounded****: there exists* $M > 0$ *such that* $|f_n(t)| \leq M$ *for all* $t \in [0, T]$ *and all* $n \in \mathbb{N}$

2. $\mathcal{F}$ *contains a uniformly convergent subsequence.*

*If we can additionally show that* $\mathcal{F}$ *is closed, then by definition, we can further conclude* $\mathcal{F}$ *to be compact.*

**Example 12** (Application of Arzelá - Ascoli)**.** *Consider the family of functions*

$$\mathcal{F} := \left\{ f : [0,1] \to \mathbb{R} \middle| |f(x) - f(y)| \leq |x - y| \text{ and } \int_0^1 f(x)dx = 1 \right\}$$

*We can show that* $\mathcal{F}$ *is a compact subset of* $\mathcal{C}[0,1]$ *by invoking the Arzelá - Ascoli Theorem.*

1. ***equicontinuous****: This follows by choosing* $\delta < \varepsilon/2$*. Then*

$$|f(t) - f(s)| \leq |t - s| < \delta = \frac{\varepsilon}{2} < \varepsilon$$

2. ***pointwise-bounded****: Notice* $|f(t) - f(0)| \leq |t| = t$ *since* $t \in [0, 1]$*. This implies* $f(0) \leq t + f(t)$ *and* $f(t) \leq t + f(0)$*. From the first inequality, we get*

$$\int_0^1 f(0) \leq \int_0^1 tdt + \int_0^1 f(t)dt \leq \frac{1}{2} + 1 = \frac{3}{2}$$

*and from the second inequality, this further yields*

$$f(t) \leq f(0) + t \leq \frac{3}{2} + 1 = \frac{5}{2}$$

3. ***closed****: Consider a sequence* $\{f_n\} \subset \mathcal{F}$ *which converges uniformly to* $f \in \mathcal{C}[0,1]$*. We want to show that* $f \in \mathcal{F}$*.*

*By uniform convergence of* $f_n$ *to* $f$*, we can interchange the limit and the integral:*

$$\int_0^1 f(t)dt = \int_0^1 \lim_{n\to\infty} f_n(t)dt = \lim_{n\to\infty} \int_0^1 f_n(t)dt = \lim_{n\to\infty} 1 = 1$$

*Furthermore, we have for* $|t - s| < \delta$

$$\begin{aligned} |f(s) - f(t)| &\leq |f(s) - f_n(s)| + |f_n(s) - f_n(t)| + |f_n(t) - f(t)| \text{ by Triangle Inequality} \\ &< \frac{\varepsilon}{3} + \frac{\varepsilon}{3} + \frac{\varepsilon}{3} \text{ by uniform convergence and Lipschitz property of each } f_n \end{aligned}$$

*With these two properties satisfied, we have* $f \in \mathcal{F}$*.*

*Overall, the family* $\mathcal{F}$ *is compact.*

## 1.3 Weak Convergence in Metric Spaces

In Section 1.1, we studied the notion of weak convergence for a sequence of random variables. Now, we study weak convergence for probability measures on a metric space.

Denote $(\mathcal{S}, d)$ to be a metric space and denote $\mathbb{B}(\mathcal{S})$ to be the $\sigma$-field generated by open subsets of $\mathcal{S}$. Further denote $\mathcal{P}(\mathcal{S})$ to be the family of probability measures defined on $\mathcal{S}$.

**Definition 12** (Tightness). *A subset $\mathcal{M} \subset \mathcal{P}(\mathcal{S})$ is said to be **tight** if for every $\varepsilon > 0$, there exists a compact set $K \subset \mathcal{S}$ such that for all $P \in \mathcal{M}$, $P(K) \geq 1 - \varepsilon$.*

An equivalent condition is that $P$ is tight iff for each $A \subset \mathcal{S}$, $PA$ is the supremum of $PK$ over all compact subsets $K \subset A$. Furthermore, it turns out that if $\mathcal{S}$ is separable and complete, then every probability measure $P \in \mathcal{P}(\mathcal{S})$ is tight. These are proven in [3].

**Example 13** ([3], Example 1.3). *Consider the space $C[0,1]$ metrized by the uniform metric described previously:*

$$d(f,g) := \|f - g\| := \sup_{t \in [0,1]} |f(t) - g(t)|$$

*Denote the set $\mathcal{D}_m$ to be the set of piecewise-linear functions over subinterval partitions $I_{k,m} := [\frac{k-1}{m}, \frac{k}{m}]$ for $k = 1, \cdots, m$ with rational values taken at the endpoints. Since $\mathbb{Q}$ is countable, it follows that $\bigcup_m \mathcal{D}_m$ is a countable set.*

*We will now show that it is dense. For a given $f \in C[0,1]$ and $\epsilon > 0$, choose $m$ such that $|f(t) - f(k/m)| < \epsilon/2$ for $t \in I_{k,m}$. This is possible to do by uniform continuity: there exists a $\delta$ such that $|t - (k/m)| < \delta$ implies $|f(t) - f(k/m)| < \epsilon/2$.*

*Now we can choose $g \in \mathcal{D}_m$ such that $|g(k/m) - f(k/m)| < \epsilon/2$ for each $k \leq m$. Overall, we have*

$$\left| f(t) - g\left(\frac{k}{m}\right) \right| \leq \left| f(t) - f\left(\frac{k}{m}\right) \right| + \left| f\left(\frac{k}{m}\right) - g\left(\frac{k}{m}\right) \right| < \epsilon$$

*We can repeat similarly for $g((k-1)/m)$. Note that $g(t)$ is an affine combination of $g((k-1)/m)$ and $g(k/m)$. Hence, we can construct a sequence of $g_m \in \mathcal{D}_m$ such that it converge to $f \in C[0,1]$. The space $C[0,1]$ is then separable.*

*Suppose $f_n$ was a Cauchy sequence in $C[0,1]$ with a limit $f$ on the interval $[0,1]$. To show that $f \in C[0,1]$, we need to show that it is continuous. This is straightforward: since the Cauchy sequence converges, we have for all $t \in [0,1]$ $|f_n(t) - f_m(t)| \leq \epsilon_n$ with $\epsilon_n \to 0$. Taking $m \to \infty$ yields $|f_n(t) - f(t)| \leq \epsilon_n \to 0$, which implies uniform convergence of $f_n$ to $f$. By the Uniform Limit Theorem, because $f_n$ are each continuous, $f$ must be continuous as well. Hence, $f \in C[0,1]$, and we have that $C$ is complete.*

*Since $C[0,1]$ is a separable and complete metric space, it follows that any probability measure defined on it is tight.*

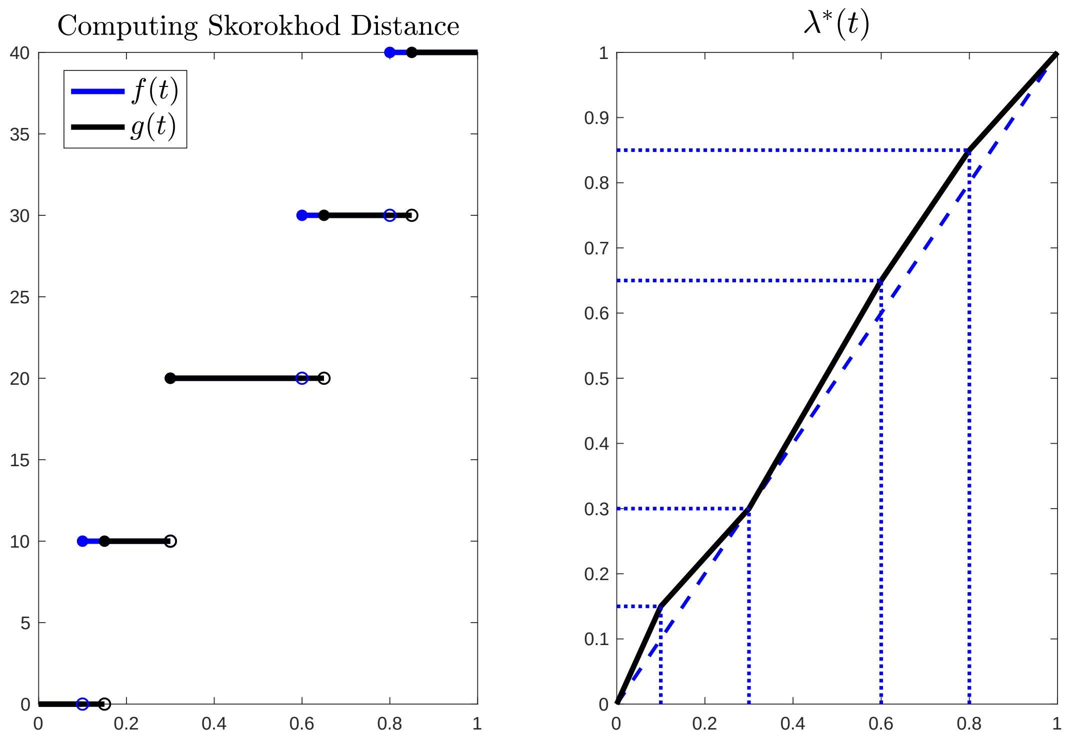


**Figure 12:** [Left] Two functions $f(t)$ and $g(t)$ of Example 14 and [right] the corresponding $\lambda^*(t)$ used to compute a specific distance.

**Definition 13** (Weak Convergence of Probability Measures)**.** *For $P_n \in \mathcal{P}(\mathcal{S})$ and $P \in \mathcal{P}(\mathcal{S})$, define the notation $P_n \Rightarrow P$ to mean **weak convergence** in the sense that $P_n f \to Pf$ for each bounded, uniformly continuous real $f$ on $\mathcal{S}$. In other words, $P_n \Rightarrow P$ if*

$$\lim_{n\to\infty} \int f dP_n = \int f dP$$

*for every bounded, uniformly continuous real $f$ on $\mathcal{S}$.*

In the definition above, we take $\mathcal{S} := C[0,1]$, so that the notions of boundedness and uniform continuity make sense.

## 2 Skorokhod Topologies

**Example 14.** *Using the traditional continuous space $C[0,1]$ and the corresponding uniform metric for estimating "distances" between functions which describe discrete processes is incorrect. Instead, the Skorokhod topology allows for a more realistic comparison between such functions. We illustrate this concept in this example, which was adapted from [4].*

*Let $f(t)$ and $g(t)$ be piecewise-constant functions which describe a product manufacturing process for the independent plants $A$ and $B$, respectively. For each plant, a "jump" in the process occurs whenever 10 products are completed. A specific sample path throughout interval of time $0 \leq t \leq 1$ for each plant is shown in the left subfigure of Figure 12.*

*We seek a metric that can be used to quantify the difference in productivity between the two plants. According to the uniform distance metric, $d_u(f,g) = \sup_{t\in[0,1]}|f(t) - g(t)| = 10$, and so given a certain interval of time, one plant manufactures 10 parts more than the other plant. This is clearly not true; the two processes are unsynchronized in time but otherwise, they manage to produce the same number of products.*

*We will now compute the Skorokhod distance metric between the two processes. Consider the specific $\lambda^*$:*

$$\lambda^*(t) = \begin{cases} 1.5t & \text{if } t \in [0, 0.1) \\ 0.75t + 0.075 & \text{if } t \in [0.1, 0.3) \\ \frac{7}{6}t - 0.05 & \text{if } t \in [0.3, 0.6) \\ t + 0.05 & \text{if } t \in [0.6, 0.8) \\ 0.75t + 0.25 & \text{if } t \in [0.8, 1] \end{cases}$$

*We have $\|\lambda^*(t) - I\| = |\lambda^*(0.1) - 0.1| = 0.05$. Note that under this $\lambda^*$, $g(\lambda^*(t))$ aligns exactly with $f(t)$. Hence, $\|f - g \circ \lambda^*\| = 0$. Therefore, $d_S(f,g) \leq 0.05$.*

*Under the Skorokhod metric, the similarity between the two processes can be described as follows: the functions $f(t)$ and $g(t)$ are near one another in $D[0,1]$ in the sense that one graph can be transformed into the other by a small deformation in the time scale.*

## 2.1 Four Main Topologies

As the previous example showed, the space $C$ is insufficient to use in describing functions or processes which contain jumps (e.g., the Poisson process). Instead we define $D := D[0,1]$ to be the space of real functions $f$ on the unit interval that are *cádlág*:

1. for $t \in [0,1)$, $f(t+) = \lim_{s\downarrow t} f(s)$ exists and $f(t+) = f(t)$.
2. for $t \in (0,1]$, $f(t-) = \lim_{s\uparrow t} f(s)$ exists

Clearly, $C \subset D$.

Two functions $f$ and $g$ are considered "*near*" each other in the uniform topology on $C[0,1]$ if the graph of $f(t)$ can be translated along the vertical axes onto the graph of $g(t)$ by a uniformly small amount, with the time axes kept fixed.

In contrast, two functions $f$ and $g$ are "*near*" each other in the uniform topology on $D[0,1]$ if (again) $f(t)$ can be shifted along the vertical axes onto $g(t)$ by a uniformly small amount, and also a uniformly small deformation of the time axes. See Figure 13 for an example illustration depicting the difference between the two notions.

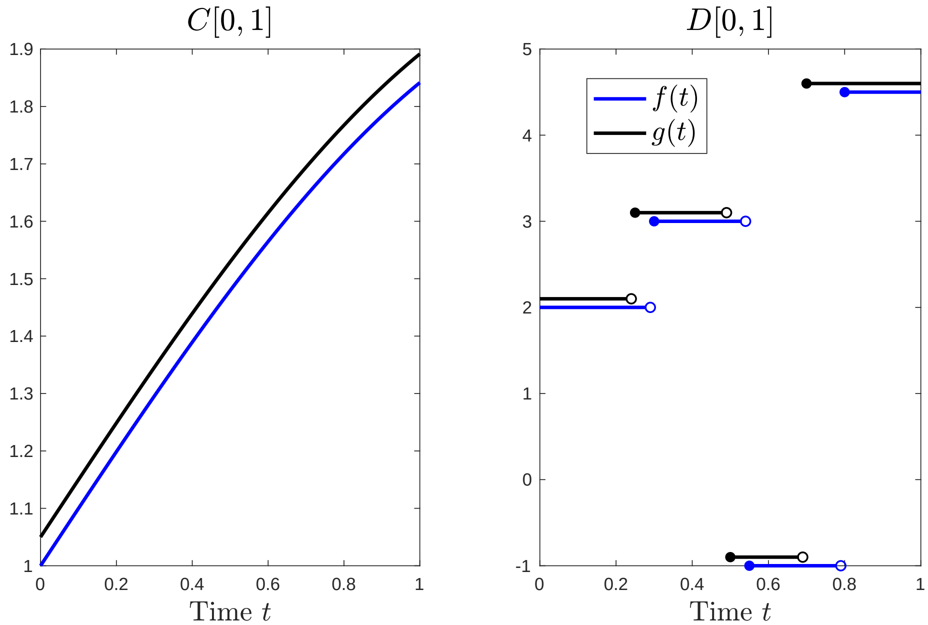


**Figure 13:** Two functions $f(t)$ [blue] and its deformation $g(t)$ [black] for two cases: $f, g \in C[0,1]$ and $f, g \in D[0,1]$. The deformation in $C[0,1]$ is simply a vertical shift upwards whereas the deformation in $D[0,1]$ also includes a time rescaling.

Let $\Lambda$ denote the class of strictly increasing (and one-to-one) continuous mappings $\lambda : [0,1] \to [0,1]$ which satisfy $\lambda(0) = 0$ and $\lambda(1) = 1$. Use the norm $\|x\| = \sup_t |x(t)|$ on $[0,1]$. Further define

$$\|\lambda - I\| = \sup_t |\lambda(t) - t|, \quad \|x - y \circ \lambda\| = \sup_t |x(t) - y(\lambda(t))| \tag{3}$$

where $x, y \in D[0,1]$ and $I \in \Lambda$ denotes the identity map.

**Definition 14.** *In [5], Skorokhod introduced four topologies over $D[0,1]$, known as the $J_1, J_2, M_1$, and $M_2$ topologies. We define each individual topology below by explaining what it means for a sequence of functions to converge in each space.*

1. *The most common $J_1$ **topology** is the space $D$ equipped with the metric*

$$d_{J_1}(f,g) = \inf_{\lambda \in \Lambda} \left( \max\{\|\lambda - I\|, \|f - g \circ \lambda\|\} \right), \ \forall f, g \in D \tag{4}$$

   *Under this metric, however, the metric space $D$ is not complete. We will define an alternative distance metric later.*

   *A sequence $\{f_n(t)\}$ converges to the limit function $f(t)$ in the $J_1$ sense iff there exist a sequence of functions $\{\lambda_n(t)\} \subset \Lambda$ such that the following two conditions hold:*

$$\lim_{n\to\infty} f_n(\lambda_n(t)) = f(t) \tag{5a}$$

$$\lim_{n\to\infty} \lambda_n(t) = t \tag{5b}$$

2. *In the $J_2$ **topology**, we have the same metric and notion of convergence as for the $J_1$ case, but the class $\Lambda$ is extended to include discontinuous functions mapping from $[0,1]$ to $[0,1]$.*

3. *Define* $R[(x_1,t_1);(x_2,t_2)] = |x_1 - x_2| + |t_1 - t_2|$. *Suppose there exists a pair of functions* $(x(s), t(s))$ *for* $s \in [0,1]$ *such that* $x$ *is continuous, and* $t$ *is continuous and monotonically increasing. Then we can define the parametric representation* $\Gamma_{f(t)}$ *of a function* $f \in D$.

   *If there exist parametric representations* $(x(s), t(s))$ *of* $\Gamma_{f(t)}$ *and* $(x_n(s), t_n(s))$ *of* $\Gamma_{f_n(t)}$ *such that*

$$\lim_{n\to\infty} \sup_{s\in[0,1]} R[(x_n(s), t_n(s)); (x(s), t(s))] = 0$$

   *then we say that* $f_n$ *converges to* $f$ *in the* $M_1$ ***topology**. Accordingly, we can define the* $M_1$ *metric:*

$$d_{M_1}(f,g) = \sup_{s\in[0,1]} R[(x_f(s), t_f(s)); (x_g(s), t_g(s))]$$

4. *The* $M_2$***-topology** *is similar to the* $M_1$ *topology except the parameterizations do not necessarily need to be with respect to the same* $s$. *If there exist parametric representations* $(x(s), t(s))$ *of* $\Gamma_{f(t)}$ *and* $(x_n(s), t_n(s))$ *of* $\Gamma_{f_n(t)}$ *such that*

$$\lim_{n\to\infty} \sup_{(x_1,t_1)\in\Gamma_{f(t)}} \sup_{(x_2,t_2)\in\Gamma_{f_n(t)}} R[(x_1(s), t_1(s)); (x_2(s), t_2(s))] = 0$$

   *then we say that* $f_n$ *converges to* $f$ *in the* $M_2$ *sense. We can then define the* $M_2$ *metric:*

$$d_{M_2}(f,g) = \sup_{(x_1,t_1)\in\Gamma_{f(t)}} \sup_{(x_2,t_2)\in\Gamma_{g(t)}} R[(x_1(s), t_1(s)); (x_2(s), t_2(s))]$$

**Example 15** ($J_1$-Convergence)**.** *Consider the sequence of functions* $f_n(t) = \mathbb{1}\{t \in [1-(1/n), 2]\}$ *and the function* $f(t) = \mathbb{1}\{t \in [1,2]\}$.

*Note that* $f_n$ *does not converge uniformly to* $f$ *because the sequence* $V_n := \sup_{t\in[0,2]}|f_n(t) - f(t)|$ *is equal to 1 for all* $n \in \mathbb{N}$, *and so* $V_n \not\to 0$ *as* $n \to \infty$.

*Does* $f_n \to f$ *in the Skorokhod metric? Consider the specific* $\lambda^*$ *as follows:*

$$\lambda_n^*(t) = \begin{cases} \left(1 - \frac{1}{n}\right) t & \text{if } t \in [0,1) \\ \left(1 + \frac{1}{n}\right) t - \frac{2}{n} & \text{if } t \in [1,2] \end{cases}$$

*Note that* $\|\lambda_n^* - I\| = \sup_{t\in[0,2]}|\lambda_n^*(t) - t| = |\lambda_n^*(1) - 1| = \frac{1}{n}$. *Furthermore, under this* $\lambda_n^*$, $f_n(\lambda_n^*(t))$ *aligns exactly with* $f(t)$; *see Figure 14. Hence,* $\|f - f_n \circ \lambda_n^*\| = 0$ *for all* $n$. *Therefore,* $d_{J_1}(f, f_n) = \frac{1}{n}$, *which tends to 0 as* $n \to \infty$. *Thus, under the Skorokhod metric,* $f_n$ *converges to* $f$.

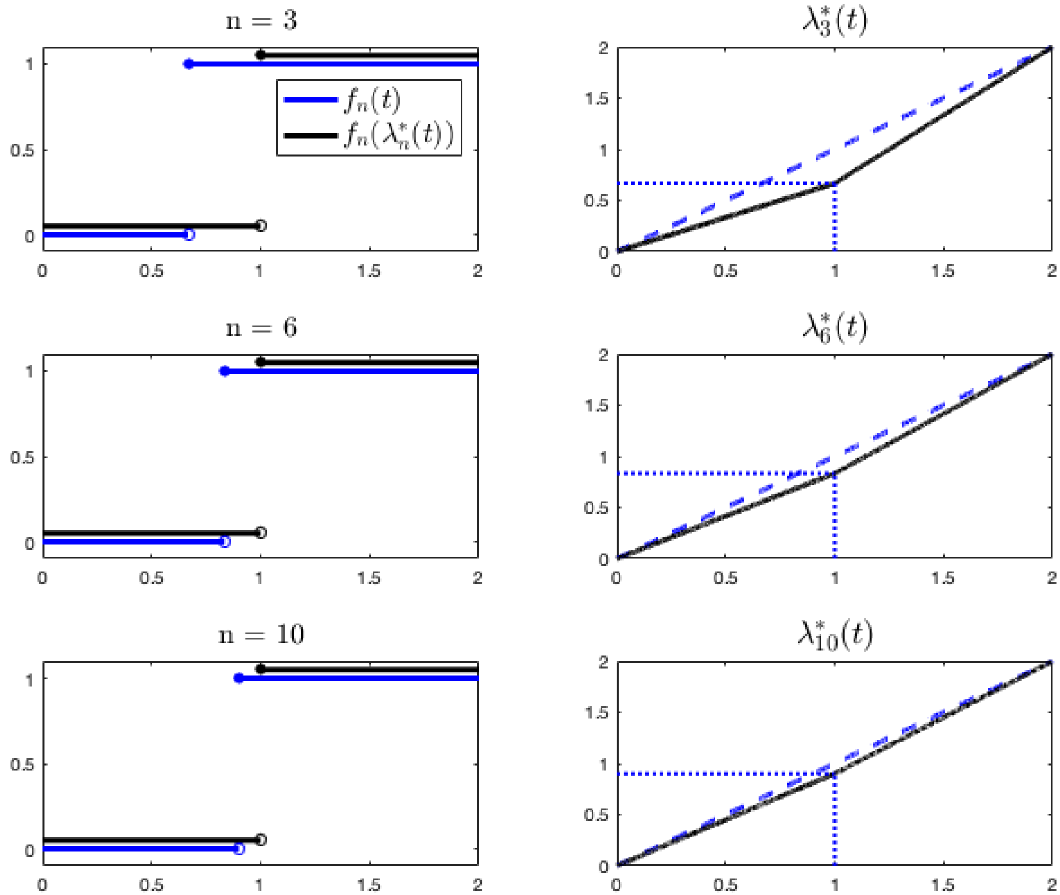


**Figure 14:** A comparison of $f_n(t)$ and $f_n(\lambda_n^*(t))$ for three values of $n$, and the corresponding $\lambda_n^*(t)$.

*Alternatively, because the Skorokhod metric is a metric and thus symmetric, one can also find a sequence of $\tilde{\lambda}_n(t)$ such that $\left\| f_n - f(\tilde{\lambda}) \right\| = 0$. Consider the following:*

$$\tilde{\lambda}_n(t) = \begin{cases} \left(\frac{n}{n-1}\right) t & \text{if } t \in [0, \frac{n-1}{n}) \\ \left(\frac{n}{n+1}\right) t + \frac{2}{n+1} & \text{if } t \in [\frac{n-1}{n}, 2] \end{cases}$$

*Again, we have that $\left\| \tilde{\lambda}_n - I \right\| = \sup_{t\in[0,2]} \left| \tilde{\lambda}_n(t) - t \right| = \left| \tilde{\lambda}_n(1 - (1/n)) - (1 - (1/n)) \right| = \frac{1}{n}$. Under this $\tilde{\lambda}_n$, $f(\tilde{\lambda}_n(t))$ aligns exactly with $f(t)$; see Figure 15. Hence, $\left\| f_n - f \circ \tilde{\lambda}_n \right\| = 0$ for all $n$. Therefore, $d_{J_1}(f_n, f) = \frac{1}{n}$, which tends to 0 as $n \to \infty$. We have that under the Skorokhod metric, $f_n$ converges to $f$.*

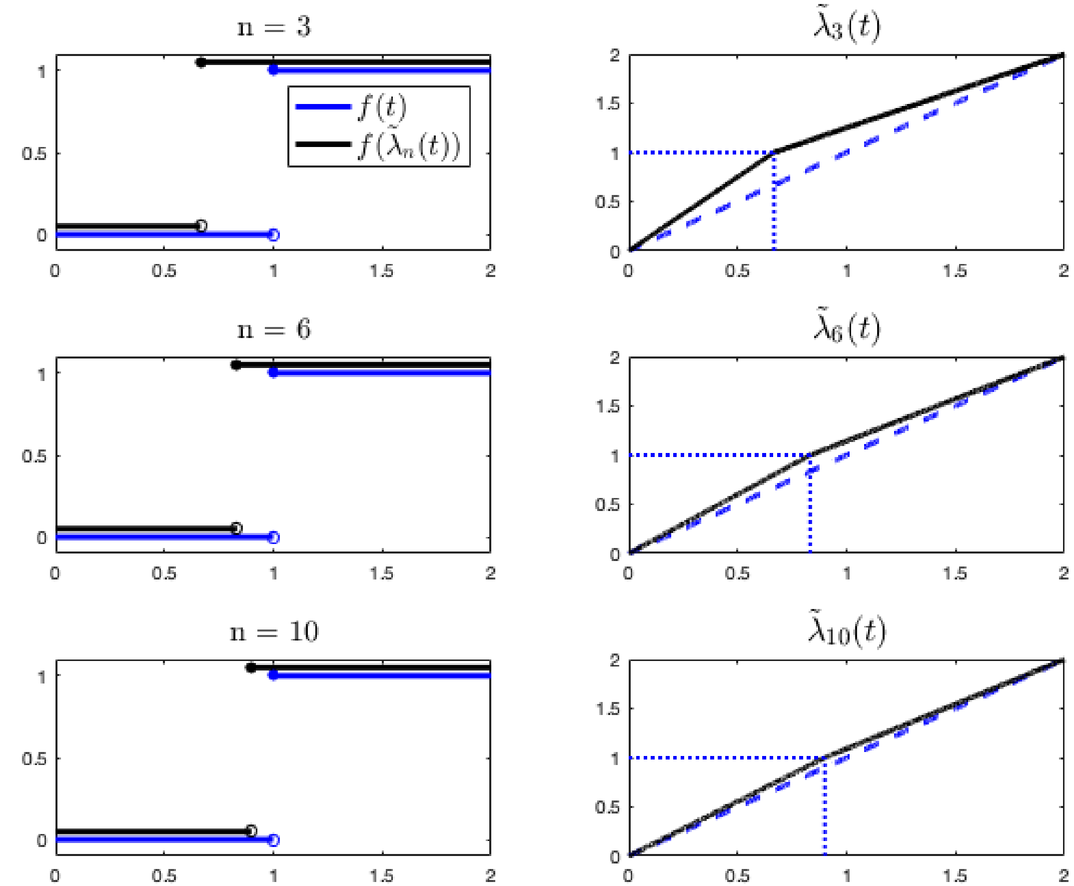


**Figure 15:** An alternative comparison of $f(t)$ and $f(\tilde{\lambda}_n(t))$ for three values of $n$, and the corresponding $\tilde{\lambda}_n(t)$.

We mentioned previously in Definition 14 that the Skorokhod topology is not complete with metric $d_{J_1}$. The following example illustrates this.

**Example 16.** *Suppose we have a sequence of functions* $f_n(t) = \mathbb{1}\{t \in [0, \frac{1}{4} + 1/(2^n)\}$.

*Consider the specific* $\lambda_n(t)$ *defined as follows:*

$$\lambda_n(t) = \begin{cases} \frac{2^{n-1}+1}{2^{n-1}+2} & \textit{if } t \in \left[0, \frac{1}{4} + \frac{1}{2^n}\right) \\ \frac{3(2^n)-2}{3(2^n)-4} + \frac{1}{2^{n-1}(3(2^{n-1})-4)} & \textit{if } t \in \left[\frac{1}{4} + \frac{1}{2^n}, 1\right] \end{cases}$$

*Then* $\|\lambda_n - I\| = |\lambda_n(1/2^n) - 1/2^n| = 1/2^{n+1}$. *Furthermore, with this* $\lambda_n$, *we have that* $f_{n+1}(\lambda_n(t))$ *maps exactly to* $f_n(t)$. *Hence* $\|f_n - f_{n+1} \circ \lambda\| = 0$. *Therefore, we have a Cauchy sequence* $\{f_n\}$ *in* $D[0, 1]$.

*Note that* $f_n$ *converges in the Skorokhod metric to the limit function*

$$f(t) = \begin{cases} 1 & \textit{if } t \in \left[0, \frac{1}{4}\right] \\ 0 & \textit{if } t \in \left(\frac{1}{4}, 1\right] \end{cases}$$

*but this function is not cádlág. Thus* $f \notin D[0, 1]$, *and not every Cauchy sequence in* $D[0, 1]$ *converges in* $D[0, 1]$.

In the discussion that follows, we consider a new metric $d^\circ$, defined in [3].

**Definition 15.** *Define the norm:*

$$\|\lambda\|^\circ := \sup_{s<t}\left|\log\left(\frac{(\lambda(t)-\lambda(s))}{(t-s)}\right)\right|$$

*and define the metric*

$$d^\circ(f,g) = \inf_{\lambda\in\Lambda}\left\{\max\left(\|\lambda\|^\circ, \|f-g\circ\lambda\|\right)\right\}$$

*Intuitively, we now take a time-deformation $\lambda$ such that the slope $(\lambda(t)-\lambda(s))/(t-s)$ of each chord is close to* 1*.*

In [3], it is shown with proof that the space $D[0,1]$ is separable under $d$ and $d^\circ$ and complete only in $d^\circ$.

## 2.2 Semimartingale Topology

The following example shows that the Skorokhod topologies are insufficient for describing the convergence of Stieltjes, and also stochastic, integrals.

**Example 17** (Convergence of Stieltjes Integrals)**.** *Let $f_n(t) = \mathbb{1}\{t \in [1-1/n, 2]\}$ and $f(t) = \mathbb{1}\{t \in [1,2]\}$. Additionally, let $A_n(t) = \mathbb{1}\{t \in [1,2]\}$ for $n = 0, 1, 2, \cdots$. We've seen in Example 15 that $f_n \to f$ in the $J_1$ Skorokhod sense, and clearly $A_n(t) \to A_0(t)$ because it is constant for all $n$. However, the Stieltjes integrals do not converge. given a partition $0 = t_0 < t_1 < \cdots < t_m = 2$*

$$\int_0^t f_n(s-)dA_n(s) := \sum_{i=1}^m f_n(t_{i-1}-)\left(A_n(t_i) - A_n(t_{i-1})\right) = f(t) \not\to \int_0^t f(s-)dA_0(s) = 0$$

*because the value of the integral is essentially determined by the longest range of $t$ which yields value zero. On the left-hand-side, the integral is $x(t)$ because $1-1/n$ is always less than 1, and so there is a discrepancy. This demonstrates that the convergence of the integrands and the integrators in $J_1$ does not necessarily imply convergence of the integrals themselves.*

*Instead, consider an alternative sequence $\tilde{f}_n(t) = \mathbb{1}\{t \in [1+1/n, 2]\}$ with the same $f(t)$ and $A_n(t)$. Now, $1+1/n > 1$, so we have convergence between the integrals:*

$$\int_0^t \tilde{f}_n(s-)dA_n(s) := 0 \to \int_0^t f(s-)dA_0(s) = 0$$

*However, the pairs $(f_n, A_n)$ do not jointly converge to $(f, A_0)$ since $A_n \equiv A_0$ is already converged to $A_0$ whereas $f_n \not\equiv f$ no matter how large $n$ grows. This demonstrates that the convergence of the integral in $J_1$ does not necessarily imply joint convergence of the integrands and the integrators.*

Motivated by this lack of characterization of integrals in the Skorokhod topology, we define a new topology as follows.

**Definition 16** (Semimartingale Topology, [6, 7]). *Let $f_n, f \in D[0,1]$. If the following three properties hold:*

1. *for all $\epsilon > 0$, there exist bounded-variation functions $A_{n,\epsilon}$ and $A_\epsilon$ such that*

$$\sup_{t\in[0,1]} |f_n(t) - A_{n,\epsilon}| \le \epsilon, \quad \textit{and} \quad \sup_{t\in[0,1]} |f(t) - A_\epsilon| \le \epsilon$$

2. *for every continuous function $f : [0,1] \to \mathbb{R}$:*

$$\int_0^1 f(t) dA_{n,\epsilon}(t) \to \int_0^1 f(t) dA_\epsilon(t)$$

*then we say that $f_n$ is $S$-**convergent** to $f$.*

# References


[1] B. Hajek, "Notes for ECE534: An exploration of random processes for engineers," 2006. [Online]. Available: http://www.ifp.uiuc.edu/~hajek/Papers/randomprocJuly06.pdf

[2] W. Rudin, *Principles of Mathematical Analysis, 3rd Ed.* USA: McGraw-Hill, Inc., 1976.

[3] P. Billingsley, *Convergence of probability measures*, 2nd ed., ser. Wiley Series in Probability and Statistics: Probability and Statistics. New York: John Wiley & Sons Inc., 1999.

[4] L. V. Contreras, "Skorokhod Metric," Master's thesis, California State University, Channel Islands, 2009.

[5] A. V. Skorokhod, "Limit theorems for stochastic processes," *Theory of Probability and its Applications*, vol. 1, no. 3, pp. 261–290, Oct 1956.

[6] A. Jakubowski, "A non-Skorohod topology on the Skorohod space," *Electronic Journal of Probability, 2 (1997), s. 1-21*, vol. 2, Jul 1997.

[7] ——, "New characterizations of the S topology on the Skorokhod space," *Electronic Communications in Probability*, vol. 23, Sep 2016.