BIO 244: Unit 3

Stochastic Processes: Introduction

In traditional parametric inference, we are interested in propeties of estimators, say $\hat{\theta}$, of a finite-dimensional parameter θ which, for example, describes the c.d.f. $F(\cdot)$ of a survival time random variable T. We then focus on the distribution of the finite-dimensional random vector $\hat{\theta}$.

Suppose instead we want to nonparametrically estimate $F(\cdot)$. Then the resulting estimator, say $\hat{F}(\cdot)$, is no longer a finite-dimensional random vector but instead a stochastic process. Thus, consideration of the properties of such estimators involves properties of stochastic processes, and use of any large-sample approximations (analogous to normal approximations for MLEs) involves convergence properties of stochastic processes. This and the following unit will give a brief introduction to stochastic processes. The goals are to define a stochastic process, to illustrate how the probabilistic properties of a stochastic process can be more complex than those of a random variable, to define Gaussian processes, and to introduce the concept of convergence of a sequence of stochastic processes.

Stochastic Process: A stochastic process $X(\cdot)$ is a family of random variables indexed by t in some set I; i.e., $X(\cdot) = (X(t) \mid t \in I)$.

In this course, t will denote time and the set I will usually be taken to be [0,∞). Thus, $X(\cdot) = \{X(t) \mid 0 \leq t < \infty]$. For example, if t denotes time since beginning a clinical trial, X(t) might denote a AIDS patient's CD4 cell count at time t or his/her survival status at time t (where X(t)=0 denotes being alive at t and X(t)=1 denotes having died on or before time t).

Commonly, the value, $X(t)$, of the process $X(\cdot)$ at time t is referred to as the "state of the process" at time t. One way of describing a stochastic process $X(\cdot)$ is by the qualitative nature of the set I or of the values of X(t). The following categories are sometimes used:

			Type of Process
X(t)	=	concentration of a drug in the blood t time units after administration	CT, CS
X(t)	=	air pollution level on day t	DT, CS
X(t)	=	# asthma attacks that occur by time t since beginning a treatment	CT, DS
X(t)	=	indicator of a drug-resistant mutation at codon site t	DT, DS

CT	:	continuous time
DT	:	discrete time
CS	:	continuous state
DS	:	discrete state

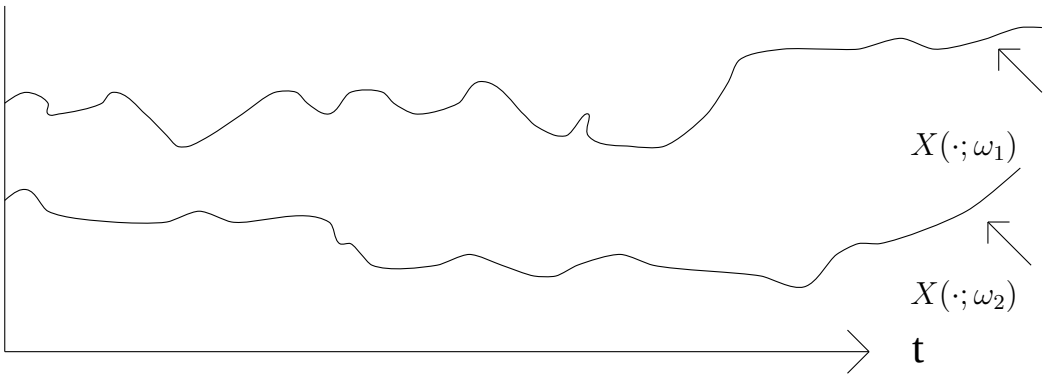
Our focus in this course will be on continuous time processes, usually with discrete states (CT, DS).

Because $X(\cdot)$ is a family of random variables, descriptions of the properties of a stochastic process tend to be more complicated than those of a scalar random variable. To see this, recall that realizations of a scalar random variable Y can be viewed as arising from a mapping from a space Ω to the real line; i.e.,

$$Y : \Omega \to R.$$

For example, we view realizations (observations) from the random variable Y as mappings of different points in Ω; e.g., we might have $Y(\omega_1)=6.5$ and $Y(\omega_2)=14.1$, where ω_1 and ω_2 are two points in Ω.

In contrast, a single realization (also known as 'sample path') of a stochastic process $X(\cdot)$ means the values of $X(t;w)$ for all t (for a fixed value of ω). For example, if we use $X(\cdot;\omega)$ to denote the set $\{X(t;\omega) \mid 0 \leq t < \infty\}$, and if ω_1 and ω_2 are two points in Ω, the corresponding values of $X(\cdot,\omega_1)$ and $X(\cdot,\omega_2)$ might look like the following:



For example, these might denote the systolic blood pressures of 2 patients over a period of time following administration of a treatment.

In most respects, the definition of a stochastic process as a set of random variables defined on an underlying probability space generalizes directly from that used for random variables. That is, suppose that $(\Omega, \mathcal{A})$ is a probability space with probability measure P. That is, Ω is a set, $\mathcal{A}$ is a σ-algebra of subsets of Ω, and P is a probability measure defined on the elements of $\mathcal{A}$. For a review of this set-up see Appendix A.

Then we say that the stochastic process $X(\cdot)$ is defined on $(\Omega, \mathcal{A}, P)$ if the scalar random variable $X(t)$ is defined on $(\Omega, \mathcal{A}, P)$ for every t. That is, for each $t \in [0, \infty)$,

(a) $X(t) : \Omega \to \mathbb{R}$, where the value of $X(t)$ for a particular ω is denoted $X(t;\omega)$, and

(b) $X(t)$ is measurable: for any $x \in \mathbb{R}$, the set $E(x,t) \stackrel{def}{=} \{\omega \in \Omega : X(t,\omega) \leq x\} \in \mathcal{A}$.

Let's consider the analogs of these properties for the entire stochastic process $X(\cdot)$. The analog of (a) is the function

$$X(\cdot;\omega) \stackrel{def}{=} \{X(t;\omega) : 0 \leq t < \infty\} ,$$

a "sample path" of the process $X(\cdot)$. Thus, when we are dealing with a random variable X, we sometimes think of a set of n realizations as the values $X_i(\omega)$ for i=1,2,...,n where $\omega \in \Omega$. In contrast, n realizations or sample paths from the stochastic process $X(\cdot)$ are the values of the n functions $X_i(\cdot;\omega)$, for i=1,2,...,n.

Now consider (b); that is, that the scalar random variable $X(t)$ is $\mathcal{A}$-measurable. It is tempting to conclude from (b) that any event expressible in terms of $X(\cdot)$ is $\mathcal{A}$-measurable, which would be desirable because otherwise the probability of such events would not be defined. However, this need not be the case. For any two times, say t_1 and t_2, consider the random variables $X(t_1)$ and $X(t_2)$. Then for any reals x_1 and x_2, the measurability of $X(t_1)$ and $X(t_2)$ ensures that $E(x_1,t_1)$ and $E(x_2,t_2)$ are elements of $\mathcal{A}$, and hence the event $E = E(x_1,t_1) \cap E(x_2,t_2)$ is also an element of $\mathcal{A}$. Similar arguments hold for the k random variables formed by examining $X(\cdot)$ at k distinct time points, or for that matter for a countably infinite number of random variables since σ-algebras are closed under countably infinite intersections. However, since $X(\cdot)$ contains an uncountably infinite number of random variables, complications can arise because events desribed in terms of values of the process may not correspond to elements of the σ-algebra $\mathcal{A}$, in which case the probability of the event would not be defined. This is illustrated in the following example.

Example 3.1: Suppose that $(\Omega, \mathcal{A}, P)$ is the unit interval probability space (see Appendix A), $I = [0,1]$, and S is some non-measurable subset of $\Omega = [0,1]$. Define the stochastic process $X(\cdot)$ by $X(t;\omega) = 1$ if $\omega \in S$ and $\omega = t$, and $X(t;\omega) = 0$ otherwise. Then it can be verified that $X(t)$ is $\mathcal{A}$-measurable for every t. However, consider the event

$$E \stackrel{def}{=} \{\omega : \sup_t X(t;\omega) \leq 0\}.$$

Since $P\left(X(t) = 0\right) = 1$ for every t, we would want to consider E as an event and to take its probability to be 1. However, $E = \Omega \backslash S$, which is not a measurable set. Therefore, we cannot consider E as an event and take its probability to be 1, because the probability measure P is only defined on elements of $\mathcal{A}$.

Some Definitions and Results

- $X(\cdot) = \{X(t), t\epsilon I\}$ called **continuous** if $P[A] = 1$,
  where $A = \{\omega \ : \ X(\cdot;\ \omega)$ is a continuous function of $t\epsilon I\ \}$

  Similar definitions for **right continuous** and **left continuous**.

  $\longrightarrow$ i.e., the stochastic process is said to have the property if the set of its sample paths which have the property has probability 1.

  We now consider 2 definitions of relationships between processes (taken from Fleming & Harrington, 1991).

- $X(\cdot)$ and $Y(\cdot)$ are **indistinguishable** if $P[B] = 1$,
  where $B = \{\omega\epsilon\Omega \ : \ X(t;\omega) = Y(t;\omega)$ for all $t\epsilon I\}$

- $X(\cdot)$ is a **modification** of $Y(\cdot)$
  if, for every $t\epsilon I$, $P[X(t) = Y(t)] = 1$

  i.e., if for every t, $P[C_t] = 1$, where $C_t = \{\omega\epsilon\Omega \ : \ X(t;\omega) = Y(t;\omega)\}$.

At first glance, one might think that the preceding 2 properties are equivalent. However, this is not necessarily so, as the following example illustrates:

Example 3.2: Let $\Omega = [0, 1]$, and for $A \subset \Omega$, suppose that P(A) denotes Lebesgue measure.

$$\text{Let} \qquad X(t;\omega) \stackrel{\text{def}}{=} \begin{cases} 1 & \text{if } t - [t] = \omega \\ 0 & \text{otherwise,} \end{cases}$$

where [t] denotes the greatest integer less than or equal to t. For example, if $\omega = 0.5$, then $X(t;\omega)$=1 when $t = 0.5, 1.5, 2.5, \ldots$ and $X(t;\omega) = 0$ otherwise.

Also define $Y(t;\omega) = 0$ for all t and ω.

(a) Are $X(\cdot)$ and $Y(\cdot)$ indistinguishable?
No, since $B = \{\omega \ : \ X(\cdot\,;\omega) = Y(\cdot\,;\omega)\}$ = the empty set, so that P(B)=0.

(b) Is $X(\cdot)$ a modification of $Y(\cdot)$?
Yes, since for any specific t,
$C_t = \{\omega : X(t;\omega) = Y(t;\omega)\} = \Omega \setminus \{t - [t]\}$
(i.e., C_t denotes every $\omega \epsilon \Omega$ except the singleton $\omega = t - [t]$)
and hence $P(C_t) = 1$.

The 2 conditions are not equivalent in general. However, it can be shown that if $X(\cdot)$ and $Y(\cdot)$ are each left (or right) continuous, then, $X(\cdot)$ is indistinguishable from $Y(\cdot)$ if and only if $X(\cdot)$ is a modification of $Y(\cdot)$ (3.1)

Note: It's hard to envision any **real** processes for which 'indistinguishability' and 'modification of' are not equivalent. However, we give this example to illustrate how things can be more difficult when dealing with a stochastic process than with a random variable. For our applications in this course, (3.1) will always hold, and so we don't need to make a distinction between indistinguishability and a modification.

- If $X(\cdot)$ is a modification of $Y(\cdot)$, then for any $k > 0$ and $t_1, t_2, \ldots, t_k \in I$, the k-dimensional random vector $(X(t_1), \ldots, X(t_k))$ has the same distribution as $(Y(t_1), \ldots, Y(t_k))$.

Let us next consider how to describe the probabilistic aspects of a stochastic process $X(\cdot)$. This is commonly done by describing all of its **finite-dimensional distributions**; i.e., the distribution of $(X(t_1), X(t_2), \ldots, X(t_k))$ for every k and every $(t_1, t_2, \ldots, t_k)$, where $0 \leq t_1 < t_2 < \cdots < t_k$.

Like with random variables, the concept of moments carries over to stochastic processes. In particular, the mean function $\mu(\cdot)$ and covariance function $C(\cdot, \cdot)$ of the stochastic process $X(\cdot)$ are defined by

$$\mu(t) \stackrel{\text{def}}{=} E(X(t)) \qquad\qquad t \geq 0$$

and

$$C(s,t) = \text{Cov}(X(s), X(t)) \qquad\qquad s, t \geq 0.$$

Gaussian Process: $X(\cdot)$ is a Gaussian process if, for every k and $t_1, \ldots, t_k$,

$$(X(t_1), X(t_2), \ldots, X(t_k))$$

has a multivariate normal distribution. Analogous to the normal distribution for a random variable, the probabilistic properties of a Gaussian process can be characterized by its mean function $\mu(\cdot)$ and covariance function $C(\cdot, \cdot)$.

Within the class of Gaussian distributions, there are several important special cases. In all the Gaussian processes that we consider in this course, we assume that all sample paths are continuous.

Wiener Process: The Gaussian process $X(\cdot)$ is a **Wiener Process** (also known as **Brownian Motion**) if

$$\begin{aligned} \mu(t) &= 0 \qquad \forall t \epsilon [0, \infty) \\ X(0) &= 0, \\ \text{and} \quad C(s,t) &= \min(s,t). \end{aligned}$$

Note that $\text{Var}(X(t)) = C(t,t) = t$.

Example 3.3: (Approximation of Wiener Process/Brownian Motion by a discrete time Random Walk).

Suppose that $Z_1, Z_2, \ldots$ are i.i.d. N(0,1) random variables. For any t and $\epsilon > 0$, define $M = [t/\epsilon]$, where [u] denotes the greatest integer less than or equal to u. Let $X(0) = 0$ and for t=1,2, ..., define the continuous-time process:

$$X(t) = \sqrt{\epsilon}(Z_1 + Z_2 + \ldots + Z_M) \quad \text{for } t > 0.$$

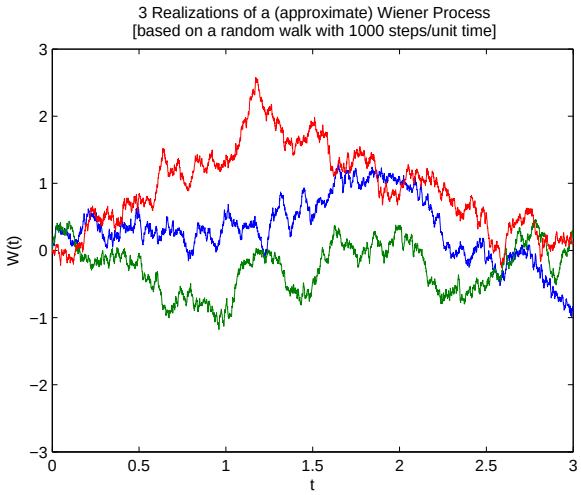
That is, X(t) consists of the sum of M i.i.d. $N(0, \epsilon)$ random variables. Its sample paths are step functions, yet it approximates a continuous-time Wiener process. Then

$$\mu(t) = E(X(t)) = 0$$

and for $t \leq s$ (and hence also for $t > s$),

$$C(s,t) = Cov(X(s), X(t)) = \epsilon M = \epsilon \left[\frac{t}{\epsilon}\right] \approx t = min(s,t).$$

Thus, X(t) is a discrete time version of a Wiener Process. From its construction, we can get a feel for how Wiener processes would tend to behave. This is illustrated below.



Note: If $X(\cdot)$ is a Wiener process, and $s < t$, then

$$Y(s,t) \stackrel{\text{def}}{=} X(t) - X(s)$$

satisfies

$$\begin{aligned} &\text{(a)} \quad Y(s,t) \quad \sim \quad N(0, t-s) \\ &\text{(b)} \quad Y(s_1,t_1) \quad \perp \quad Y(s_2,t_2) \\ &\qquad\qquad\qquad\qquad \text{for } s_1 < t_1 \leq s_2 < t_2. \end{aligned}$$

That is, (a) tells us that displacements in a time interval have a distribution that depends only on width of interval, and (b) tells us that the process $X(\cdot)$ has independent increments (Exercise 1).

Another functional of interest is the supremum Y of a Wiener Process X over a specific time interval; e.g.

$$Y \stackrel{def}{=} sup_{0 \leq t \leq \tau} \mid X(t) \mid \quad . \tag{3.2}$$

One example where this arises is in forming confidence bands for a mean function; we return to this later in the course. It can be shown [see, for example, Hall & Wellner (1980) and Schumacher (1984)] that for $\tau = 1$ and for any $c > 0$

$$P(Y \leq c) = \frac{4}{\pi} \sum_{k=0}^{\infty} \frac{(-1)^k}{(2k+1)} e^{-\pi^2 (2k+1)^2 / 8c^2} \quad .$$

Because of the complexity of this expression, these probabilities have been tabulated for various choices of c and time intervals (see Hall & Wellner and Schumacher).

Another functional of interest is the elapsed time, or passage time, until a process enters a specific state or takes a specific value. For example, when monitoring treatment differences over time in a clinical trial, suppose that $X(\cdot)$ denotes some transformation of a test statistic. Then we might be interested in the first time at which the test statistic attains some critical value x; that is, in the distribution of the random variable

$$T(x) \stackrel{def}{=} inf\{t : X(t) \geq x\} \tag{3.3}$$

for some $x > 0$. In general, T(x) is referred to as the first passage time of $X(\cdot)$ to the point x. Before developing this, we introduce the notion of stopping times.

Definition: Consider a stochastic process $X(\cdot)$ and let $\mathcal{F}$(t) denote the smallest σ-algebra with respect to which $X(s)$ is measurable for each $s \leq t$. Such σ-algebra exists (see Appendix A). Then a random variable T is called a stopping time for $X(\cdot)$ if the event $\{T \leq t\} \in \mathcal{F}$(t) for every t ("at time t, it is known whether $T \leq t$ from the information on the process X").

As we will see, the requirement reflected in this definition is necessary to ensure that $T(x)$ is a measurable random variable.

If the process X is continuous, $T(x)$ is a stopping time:

$$\begin{aligned} \{T(x) \leq t\} &= \cup_{s \leq t} \{X(s) \geq x\} \\ &= \cap_{n \in \mathbb{N}} \cup_{q \leq t, q \in \mathbb{Q}} \left\{ X(q) \geq x - \frac{1}{n} \right\}, \end{aligned}$$

where in both lines we use that X is continuous. The second equality holds since $X(s) \geq x$ for some s if and only if for every $n \in \mathbb{N}$ there is $q \in \mathbb{Q}$, $q \leq s$ such that $X(q) \geq x - \frac{1}{n}$. The last expression is in the σ-algebra because countable unions (defining property) and countable intersections (see Exercises) are. Care is needed here: if X is not continuous, x could be the limit $\lim_{s \uparrow t} X(s)$ without X reaching x by t, or $\lim_{s \downarrow t} X(s)$ could be greater or equal to x while x is not reached yet at t. Assuming continuity avoids these issues. Convince yourself that assuming right continuity is necessary. Among other things, this implies that for continuous processes, the first passage time is a measurable random variable.

When the underlying process is a Wiener process, several useful and interesting results can be derived about first passage times. We discuss just a few here.

Theorem 3.1: Suppose $W(\cdot)$ is a Wiener process and that $T(x)$ denotes the resulting first passage time to the point x. Then the density function of $T(x)$ is given by

$$f_{T(x)}(t) = \frac{\mid x \mid}{\sqrt{2\pi t^3}} exp^{-\frac{x^2}{2t}} \ , \tag{3.4}$$

for $t \geq 0$. Note that because of the symmetry of a Wiener process, T(x) and T(-x) have the same distribution. One can show that this density arises as that of Z^{-2}, where Z has the N(0,x^{-2}) distribution. The proof of Theorem 3.1 is given in Appendix B to this Unit.

Theorem 3.2: Suppose that $W(\cdot)$ is a Wiener process and define the "sup" process $M(\cdot)$ by

$$M(t) = sup\{W(s) : 0 \leq s \leq t\} \ .$$

Then M(t) has the same distribution as $\mid W(t) \mid$, and thus has density function

$$f_{M(t)}(m) = 2\frac{1}{\sqrt{2\pi t}} exp(-\frac{m^2}{2t}) \qquad \text{for } m \geq 0.$$

The proof is given in Appendix B.

These results can be used to prove others, such as the following theorem about the probability that a Wiener process crosses the zero axis during a specified time interval. For the proof, see Grimmett & Stirzaker (2001).

Theorem 3.3: Suppose that $W(\cdot)$ is a Wiener process satisfying W(0)=0, and let $0 \leq t_0 < t_1$. Then

$$P\left(W(t) = 0 \text{ for some } t \in (t_0, t_1)\right) = \frac{2}{\pi} cos^{-1}\left((t_0/t_1)^{1/2}\right).$$

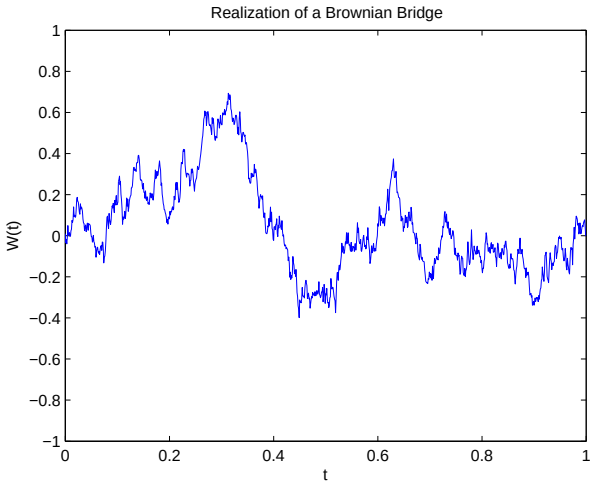
This result illustrates some surprising properties of Wiener processes. For example, when $t_0 = 0$, this says that the probability that a Wiener process equals zero in the interval $(0, t)$ is 1 for every $t > 0$. This is not altogether surprising when one thinks of a Wiener process as a limit of a random walk.

Indeed, there is an analogous theorem about a simple (symmetrically distributed) random walk in discrete time ever crossing the horizontal axis. Continuing with this analogy (and with $t_0 = 0$), it follows that the random variable $T(0) \stackrel{def}{=} inf\{t \neq 0 : W(t) = 0\}$ is zero with probability 1. Further investigation of this phenomenon shows that a Wiener process has infinitely many zeros in any non-empty interval $[0, t]$.

Brownian Bridge: The Gaussian process $X(\cdot)$ is a **Brownian Bridge** on $[0, 1]$ if

- $X(0) = X(1) = 0$
- $E\left(X(t)\right) = 0 \qquad 0 \leq t \leq 1$
- $C(s, t) = s \cdot (1 - t) \quad \text{for} \quad 0 \leq s \leq t \leq 1$

Note: $\text{Var}[X(t)] = t(1 - t)$. A typical sample path for a Brownian Bridge is illustrated below.



$\longrightarrow$ The probabilistic properties of this distribution have been studied and tabulated. Thus, for example, $P\left(\sup_{0 \leq t \leq 1} |X(t)| > 2.5\right)$ cannot be expressed in simple algebraic form, but can be obtained from tables (or via simulation).

We will discuss this further in a later unit.

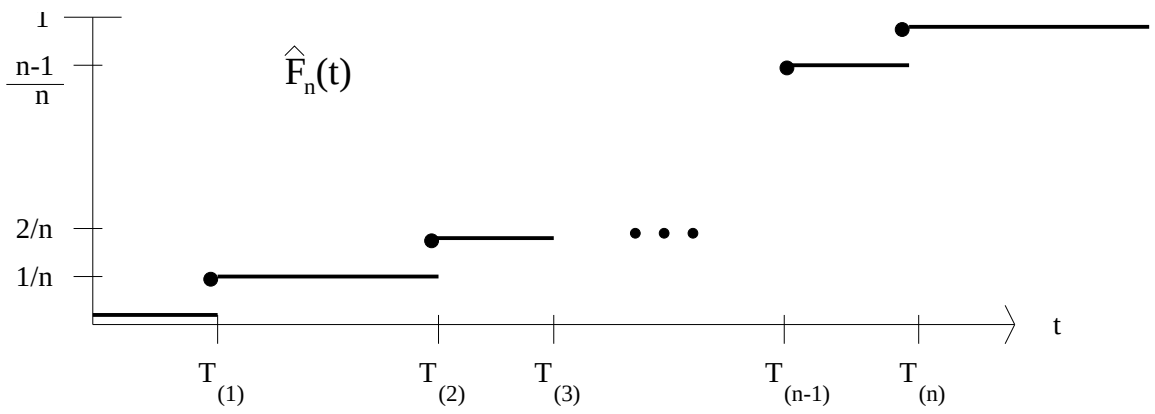
It is easy to show (Exercise 3) that if $W(\cdot)$ is a Wiener process, then $W_0(\cdot)$ defined by $W(t) - t \cdot W(1)$ is a Brownian Bridge.

Convergence of a Sequence of Processes

Let's now consider some concepts of convergence of stochastic processes. We will motivate this by the empirical c.d.f. Suppose that $T_1, T_2, \ldots, T_n$ are i.i.d. random variables with common distribution function $F(\cdot)$. Consider estimating $F(\cdot)$ by the empirical *cdf* ("*ecdf*"), $\hat{F}_n(\cdot)$, defined by:

$$\hat{F}_n(t) = \frac{1}{n}\sum_{i=1}^{n} 1(T_i \leq t) \qquad for\ 0 \leq t < \infty.$$

That is, $\hat{F}_n(t)$ is simply the proportion of observations no greater than t. This is illustrated below.



Let's examine some of its properties. First, fix t and note that if $Z_i = 1(T_i \leq t)$ for i=1,...,n, then $Z_1, Z_2, \cdots, Z_n$ are i.i.d. Bernoulli(p), where $p = F(t)$. Then we can write

$$\hat{F}_n(t) = \frac{1}{n} \sum_{i=1}^{n} Z_i.$$

It follows that

- $E[\hat{F}_n(t)] = E[Z_i] = F(t)$; i.e., $\hat{F}_n$ is unbiased
- $\text{Var}[\hat{F}_n(t)] = \dfrac{V(Z_i)}{n} = \dfrac{F(t) \cdot [1 - F(t)]}{n}$
- $\hat{F}_n(t) \xrightarrow{P} F(t) \quad \text{as } n \to \infty \quad \text{(by the WLLN)}$
- $\sqrt{n} \left\{ \hat{F}_n(t) - F(t) \right\} \xrightarrow{\mathcal{L}} N\left[0, F(t)[1 - F(t)]\right]$
  as $n \to \infty$ (by the ordinary CLT).

Thus, for example, an approximate 95% CI for $p = F(t)$ is given by

$$\hat{F}_n(t) \pm 1.96 \sqrt{\frac{\hat{F}_n(t)[1 - \hat{F}_n(t)]}{n}}.$$

We have been focusing on the distribution of $\hat{F}_n(t)$ for the fixed t; i.e., on the r.v. $Z_n = \hat{F}_n(t)$. Later we will consider the entire stochastic process $\hat{F}_n(\cdot)$. This requires us to define the concepts of 'convergence' for stochastic processes.

Additional Reading and Discussion

The text by Grimmett and Stirzaker (2001) includes some useful results about stochastic processes, including some foundational considerations.

Sample Paths versus Finite Dimensional Distributions: The underlying phenomenon reflected in Example 3.1 leads to more general issues in how to characterize and study the probabilistic properties of a stochastic process $X(\cdot)$. To examine this further, note that for any integer k and any k times, say $0 \leq t_1 < t_2 < \cdots < t_k < \infty$, we can consider the usual distribution function of the k-vector $(X(t_1), X(t_2), \cdots, X(t_k))$. The set of all such distributions, for all integers k and all k-tuples $(t_1, t_2, \cdots, t_k)$, is called the set of **finite dimensional distributions (fdd)** of $X(\cdot)$. Intuitively, one might think that knowledge of the fdd's fully describes all probabilistic aspects of a stochastic process. Alternatively, one could envision studying the probabilistic features of a process by studying the properties of its sample paths. Thus, we could:

(1) Study the properties of the sample paths $X(\cdot;\omega)$ for every $\omega \in \Omega$; or

(2) Study the collection of finite dimensional distributions of $X(\cdot)$.

Because of the uncountably infinite number of random variables $X(t)$ that comprise $X(\cdot)$, it turns out that the collection of fdd's does not necessarily tell us everything of interest about the behavior of $X(\cdot)$. This is illustrated in the following example.

Example 3.4: Suppose that the random variable U is defined on the unit interval probability space and has the Uniform(0,1) distribution. Consider the 2 stochastic process, $X(\cdot)$ and $Y(\cdot)$, defined for $0 \leq t \leq 1$ by $X(t) = 0$ for all t and $Y(t) = 1$ when $U = t$ and zero otherwise. Since $P(U = t) = 0$ for every t, the processes $X(\cdot)$ and $Y(\cdot)$ have the same fdd's; that is, for any k, any $0 \leq t_1 < \cdots < t_k \leq 1$, and any $x_1, \cdots, x_k$, we have that

$$P\left(X(t_1) \leq x_1, \cdots, X(t_k) \leq x_k\right) = P\left(Y(t_1) \leq x_1, \cdots, Y(t_k) \leq x_k\right).$$

However, $X(\cdot)$ and $Y(\cdot)$ are clearly different processes. In particular, $P(X(t) = 0 \text{ for all t }) = 1$ while $P(Y(t) = 0 \text{ for all t }) = 0$.

In general, we say that 2 stochastic processes are **versions** of one another if they have the same set of fdd's. When our interest is in probabilistic features of a process that can be described in terms of its fdd's, such as a patient's blood sugar levels at monthly visits, non-identical processes that are versions of one another will lead to the same probabilities. However, if we were interested in a feature such as the elapsed time between states of a process, two processes that are versions of one another may have different properties. For example, in Example 3.4, consider the time until each process first takes the value 1. The process $X(\cdot)$ never takes such a value while for the process $Y(\cdot)$, the time is just U.

The previous example illustrates that the fdd's of a stochastic process do not necessarily describe all of its probabilistic features. An underlying problem causing this is that σ-algebras need not be closed under an uncountably infinite number of intersections. If all events of interest could be described in terms of a countably infinite number of unions or intersections of events in $\mathcal{A}$, then we could avoid this problem. For example, suppose that the sample paths of a stochastic process are continuous in t. Then since the rationals are dense in $\mathbb{R}$, knowledge of the values of each sample path $X(\cdot;\omega)$ for all rational t would fully determine the sample path. Thus, any events describable by the process would be elements in the σ-algebra $\mathcal{A}$. This suggests that the problems illustrated above can be avoided if we restricted ourselves to processes whose sample paths were continuous. This is further illustrated in the following example.

Example 3.5: Consider the unit interval probability space and suppose that $W(\cdot)$ is a Wiener Process. Define the random variable S to be the elapsed time until $W(\cdot)$ first takes the value 1; that is,

$$S = \inf\{t : W(t) = 1\} \ .$$

It follows that the event $\{S > t\}$ can be expressed as

$$\{S > t\} = \bigcap_{0 \le s \le t} \{W(s) < 1\} \ .$$

As in Example 3.3, we cannot immediately conclude that this event is in $\mathcal{A}$ since σ-algebras need not be closed under uncountably infinite intersections. However, sample paths of $W(\cdot)$ are continuous. Hence, with $\mathbb{Q}$ the rational numbers,

$$\begin{aligned}
\{S > t\} &= \left\{ \max_{s \in [0,t]} W(s) < 1 \right\} \\
&= \bigcup_{n \in \mathbb{N}} \left\{ \max_{s \in [0,t]} W(s) \le 1 - \frac{1}{n} \right\} \\
&= \bigcup_{n \in \mathbb{N}} \bigcap_{s \in [0,t]} \left\{ W(s) \le 1 - \frac{1}{n} \right\} \\
&= \bigcup_{n \in \mathbb{N}} \bigcap_{s \in [0,t] \cap \mathbb{Q}} \left\{ W(s) \le 1 - \frac{1}{n} \right\}
\end{aligned}$$

where in the first line we use that a continuous function on a closed interval attains a maximum, in the second line we use that if the maximum is less than 1, it is less than $1 - 1/n$ for some natural number n, and in the last line we use that if $W(s) \le 1 - 1/n$ for all $s \in [0,t] \cap \mathbb{Q}$, $W(s) \le 1 - 1/n$ for all $s \in [0,t]$. The latter can be proved as follows. Let $s \in [0,t]$ be given. Then $s = \lim_{k \to \infty} s_k$ for some sequence $s_k \in [0,t] \cap \mathbb{Q}$. Because of continuity of W, $W(s) = \lim_{k \to \infty} W(s_k) \le 1 - 1/n$. We conclude that the above equation holds. It follows that the event $\{S > t\}$ is a countably-infinite union of countably-infinite intersections of measurable events, and thus an element in the underlying σ-algebra.

Thus, if a process $X(\cdot)$ were a version of another process $Y(\cdot)$ and each had continuous sample paths, then they would not only have the same fdd's, but any event defined in terms of (possibly uncountably infinite) values of one process would have the same probability as the same event defined in terms of the other process, and the problem illustrated in Example 3.4 could not occur. In practice, however, it is sometimes too restrictive to restrict attention to processes with continuous sample paths. For example, in the stochastic process that represents the number of asthma attacks experienced by a child that we described earlier, the sample paths are right-continuous but not continuous. Fortunately, the nice feature of processes with continuous sample paths illustrated above also applies to processes whose sample paths are right-continuous or processes whose paths are left-continuous. The following theorem (taken from Breiman, 1968, pages 299-300), shows that for many processes there exists a version with right-continuous sample paths with left-hand limits.

Theorem 3.4: Let $X(\cdot)$ be a stochastic process defined for $t \geq 0$ and let D be a subset of $[0, \infty)$ that is dense in $[0, \infty)$. Then if

(a) $X(t+h) \xrightarrow{P} X(t)$ as $h \downarrow 0$ for all t, and

(b) $lim_{s \uparrow t} X(s), s \in D$ and $lim_{s \downarrow t} X(s), s \in D$ exist and are finite for all $t \in [0, \infty)$ where these limits can be defined, possibly except for ω in a set of probability 0 that does not depend on t.

Then there exists a version, $Y(\cdot)$, of $X(\cdot)$ with right-continuous sample paths and left-hand limits in the sense that $lim_{h \downarrow 0} X(t-h)$ exists for all t.

Processes that are right-continuous with left-hand limits are sometimes called "cadlag" processes. This comes from "continue à droite, limite à gauche". This theorem says that if (a) and (b) hold, there exists a probability space and a cadlag process $Y(\cdot)$ defined on this space such that $Y(\cdot)$ has the same fdd's as the original process $X(\cdot)$. It essentially allows us to assume that all processes satisfying (a) and (b) are cadlag. The trivial process $Y(\cdot)$ described

in Example 3.4 is seen to satisfy these conditions. Other processes, such as the empirical c.d.f. $F_n(\cdot)$ are already cadlag. Yet others need not be, yet we will later see that one can find a version that is cadlag or that has continuous sample paths. When we study Gaussian processes, we will usually restrict ourselves to processes with continuous sample paths.

We conclude this discussion with one additional existence consideration. We often begin a discussion with something like "Suppose that $X(\cdot)$ is a Gaussian process ....". But how do we know that such processes exist? With a random variable, say $X \sim F(\cdot)$, this can be shown by simple construction. For example, with the unit interval probability space we can define the Uniform[0,1] random variable U by $U(\omega) = \omega$ and then define $X = F^{-1}(U)$. Is there an analogy for stochastic processes? The answer is yes; for a formal proof see the text by Kingman & Taylor (1973) or the text by Grimmett & Stirzaker (2001). This can be seen heuristically by viewing a Wiener process $W(\cdot)$ as a limiting case of a simple random walk in discrete time. The same can be done with other processes. For example, Brownian Bridge processes, say $W_0(\cdot)$, can be constructed via Wiener processes by $W_0(t) = W(t) - tW(1)$ for $0 \leq t \leq 1$.

Appendix A

Probability Spaces and Random Variables

1. An axiomatic treatment of probability starts with "outcomes" from a random experiment. A random experiment refers to any repeatable mechanism that generates values in some set Ω, called the *sample space*. In case we observe only one survival time, the form of Ω could simply be $\mathbb{R}$. In case we observe n survival times, the form of Ω could be $\mathbb{R}^n$. When researching properties of estimators, we often consider sample spaces on which a countable number of survival times are defined $(n = 1, \ldots, \infty)$, sometimes combined with, possibly even time dependent, covariates. The form of Ω is then more complex. Usually we do not bother about the form of Ω.

2. An event is a subset of the sample space. A probability is a function defined on a collection of events satisfying some axioms. We would like to be able to form new events via operations like taking unions and complements of events and to compute the probability of the new events. It turns out that a theory of probability that is both useful and quite general can be built by allowing *countable* operations. This entails using σ-algebras as the collection of events, and countably additive measures as the probabilities; see below.

3. A *σ-algebra* is a collection $\mathcal{A}$ of subsets of Ω satisfying some properties. Often, especially later in this course, the idea behind a σ-algebra is that after an experiment, some information on $\omega \in \Omega$ is revealed: after the experiment we know for all $A \in \mathcal{A}$ whether $\omega \in A$ or not. The defining properties of a σ-algebra $\mathcal{A}$ are:

   - $\emptyset \in \mathcal{A}$
   - Closed under complementation: $A \in \mathcal{A} \Rightarrow A^c \in \mathcal{A}$.
   - Closed under countable unions: $A_n \in \mathcal{A}, n \in \mathbb{N} \Rightarrow \cup_{n\in\mathbb{N}} A_n \in \mathcal{A}$.

   Sets in the σ-algebra are called *measurable* sets. An example of a σ-algebra on $\Omega = \{1, 2, 3\}$ is the collection of sets $\emptyset$, $\{1\}$, $\{2, 3\}$, and $\{1, 2, 3\}$. The meaning of this is that after the experiment, it is known whether $\omega = 1$. Another example of a σ-algebra on $\Omega = \{1, 2, 3\}$ is the

collection of all subsets of $\{1, 2, 3\}$. The meaning of this is that after the experiment, it is known whether ω is 1, 2, or 3. This way, it is clear that the "richer" σ-algebra, containing more sets, reveals more information on ω.

4. It can be shown (Exercise 4) that *intersections* of elements in the σ-algebra are also elements of the σ-algebra. This is also true for countable intersections.

5. A *probability space* is a sample space Ω along with a σ-algebra $\mathcal{A}$ defined on the sample space: $(\Omega, \mathcal{A})$.

6. A *probability measure* on a probability space $(\Omega, \mathcal{A})$ is a set function $P : \mathcal{A} \to [0, 1]$ satisfying:

   - $P(\Omega) = 1$
   - Countable additivity: if $A_n \in \mathcal{A}, n \in \mathbb{N}$ and A_n pairwise disjoint, then $P\left(\cup_{n\in\mathbb{N}} A_n\right) = \sum_{n\in\mathbb{N}} P(A_n)$.

   The σ-algebra here is important, since it turns out that it is not always possible to meaningfully assign probabilities to all subsets of the sample space; hence, not always all subsets of the sample space are events, or members of $\mathcal{A}$.

7. So, after the experiment, we observe whether $\omega \in A$ for all $A \in \mathcal{A}$, and before the experiment, there is a probability of $\omega \in A$ for all $A \in \mathcal{A}$. $\mathcal{A}$ represents the information available due to the experiment.

8. A scalar *random variable* X on $(\Omega, \mathcal{A}, P)$ is a map X from Ω to $\mathbb{R}$ which is *measurable*: for any $x \in \mathbb{R}$, $\{\omega \in \Omega : X(\omega) \leq x\} \in \mathcal{A}$. Or, equivalently, $X^{-1}\left((-\infty, x]\right) \in \mathcal{A}$.

9. Thus, for a random variable X, after the experiment, it is known whether $X \leq x$, and before the experiment, there is a probability attached to whether or not $X \leq x$.

10. It turns out that if X is a random variable on $(\Omega, \mathcal{A}, P)$, also, for every $x \in \mathbb{R}$:

    - $\{\omega \in \Omega : X(\omega) \geq x\} \in \mathcal{A}$

- $\{\omega \in \Omega : X(\omega) = x\} \in \mathcal{A}$
- $\{\omega \in \Omega : X(\omega) < x\} \in \mathcal{A}$
- $\{\omega \in \Omega : X(\omega) > x\} \in \mathcal{A}$.

Hence, these are all events, and their probabilities are well-defined. These properties follow from the properties of a σ-algebra. See also Exercise 4.

11. Notice that the set of all subsets of Ω is a σ-algebra. For $\Omega = \{1, 2, 3\}$, this would be the collection of sets $\emptyset$, $\{1\}$, $\{2\}$, $\{3\}$, $\{1, 2\}$, $\{1, 3\}$, $\{2, 3\}$, and $\{1, 2, 3\}$. Convince yourself that the intersection of σ-algebras is again a σ-algebra. Hence, for any collection $\mathcal{D}$ of sets of Ω, there exists a smallest σ-algebra containing all sets in $\mathcal{D}$ (it is the intersection of all σ-algebras containing $\mathcal{D}$). This is called the σ-algebra *generated* by $\mathcal{D}$. For example, with $\Omega = \{1, 2, 3\}$, the σ-algebra generated by the set $\{1\}$ consists of the following sets: $\emptyset$, $\{1\}$, $\{2, 3\}$, $\{1, 2, 3\}$. Again, this is the σ-algebra revealing information on whether or not $\omega = 1$.

12. Note: although σ-algebras are defined through countable operations, the σ-algebra generated by a collection of sets is not necessarily obtained by countable operations on the collection. (This is not easy to see).

13. The *Borel σ-algebra* $\mathcal{B}$ on $[0, 1]$ is the σ-algebra generated by the intervals $[0, x]$: $x \in [0, 1]$. Not all subsets of $[0, 1]$ are in the Borel σ-algebra. Convince yourself that points and intervals (open, closed, half open and half closed) are in the Borel σ-algebra.

14. Similarly, the Borel σ-algebra $\mathcal{B}$ on $[0, \infty)$ is the σ-algebra generated by the intervals $[0, x]$: $x \in [0, \infty)$.

15. Some examples in this unit mention the *unit interval probability space* $([0, 1], \mathcal{B}, \mu)$ with $\mathcal{B}$ the Borel σ-algebra and μ the Lebesgue measure. The *Lebesgue measure* assigns probability x to each interval $[0, x]$: $x \in [0, 1]$. One can show that that leads to a unique probability measure on the probability space $([0, 1], \mathcal{B})$. Under the Lebesgue-measure, the probability of ω falling in any interval in $[0, 1]$ equals the length of that interval. This is true for open, closed, or half open and half closed intervals. Thus, all values in $[0, 1]$ are equally likely. And if you consider ω as a random variable, it has the uniform$[0, 1]$-distribution.

For more about random variables and σ-algebras we refer to the book "Probability and Measure" by Billingsly (1995).

Appendix B

Proof of Theorem 3.2: Suppose that $m > 0$ and note that

$$T(m) \leq t \text{ if and only if } M(t) \geq m \ . \tag{3.5}$$

Then

$$P(M(t) \geq m) = P(M(t) \geq m, W(t) - m \geq 0) + P(M(t) \geq m, W(t) - m < 0).$$

However, by (3.5) and since $W(T(m)) = m$,

$$\begin{aligned} P(M(t) \geq m, W(t) - m < 0] &= P[T(m) \leq t, W(t) - W(T(m)) < 0) \\ &= P(W(t) - W(T(m)) < 0 \mid T(m) \leq t) P(T(m) \leq t) \\ &= P(W(t) - W(T(m)) \geq 0 \mid T(m) \leq t) P(T(m) \leq t) \\ &= P(M(t) \geq m, W(t) - m \geq 0), \end{aligned}$$

where we have used the facts that $W(t) - W(T(m))$ is symmetrically distributed about zero whenever $t \geq T(m)$. Thus,

$$P(M(t) \geq m) = 2P(M(t) \geq m, W(t) \geq m) = 2P(W(t) \geq m),$$

since $W(t) \leq M(t)$. Hence

$$P(M(t) \geq m) = P(\mid W(t) \mid \geq m),$$

and so the theorem is proven since $\mid W(t) \mid$ has the $\mid N(0,t) \mid$ distribution.

Proof of Theorem 3.1: Note that for $x \geq 0$ and from Theorem 3.2,

$$P(T(x) \leq t) = P(M(t) \geq x) = P(\mid W(t) \mid \geq x).$$

Thus,

$$P(T(x) \leq t) = (\frac{2}{\pi t})^{1/2} \int_x^\infty exp(-\frac{m^2}{2t}) dm = \int_0^t \frac{\mid x \mid}{\sqrt{2\pi y^3}} exp(-\frac{x^2}{2y}) dy \ ,$$

where we have made the substitution $y = x^2 t / m^2$.

Exercises

1. Suppose $X(\cdot)$ is a Wiener Process and $t \geq s$.

   (a) Prove $Y(s,t) \stackrel{\text{def}}{=} X(t) - X(s)$ is $N(0, t-s)$.

   (b) Prove $Y(s_1,t_1) \perp Y(s_2,t_2)$, where $s_1 < t_1 \leq s_2 < t_2$.

2. Consider the *ecdf*, say $\hat{F}_n(t)$, based on n i.i.d. random variables $T_1, T_2, \ldots, T_n$.

   Find $\text{Cov}\Big(\hat{F}_n(s), \hat{F}_n(t)\Big)$ for $s < t$.

3. Show that if $W(\cdot)$ is a Wiener process, then $W_0(\cdot)$ defined by $W_0(t) = W(t) - t \cdot W(1)$ is a Brownian Bridge.

4. Properties of random variables and σ-algebra's (see Appendix A):

   (a) Show that if X is a random variable, $\{X > x\}$ is an event for all $x \in \mathbb{R}$.

   (b) Show that if X is a random variable, $\{X < x\}$ is an event.

   (c) Show that if A and B are events, also $A \cap B$ is an event.

   (d) Show that if X is a random variable, $\{X = x\}$ is an event.

   (e) Show that if A_n, $n \in \mathbb{N}$, are events, also $\cap_{n\in\mathbb{N}} A_n$ is an event.

   (f) Show that we could as well have defined a random variable starting with events $\{X < x\}$

   (g) Can you give more examples like in f?

   (h) If S is not in a σ-algebra, can S^c (S-complement) be in the σ-algebra?

5. In Example 3.1, show that $X(t)$ is $\mathcal{A}$-measurable for every t.

6. Consider the unit interval probability space. Show that with the definition in Appendix A, under the Lebesgue-measure, the probability of ω falling in any interval in $[0, 1]$ equals the length of the interval. Also, show that the probability that $\omega = x$ is 0 for each x.

7. Show that if $X(\cdot)$ is a modification of $Y(\cdot)$, then for any $k > 0, k \in \mathbb{N}$ and $t_1, t_2, \ldots, t_k \in I$, the k-dimensional random vector $(X(t_1), \ldots, X(t_k))$ has the same distribution as $(Y(t_1), \ldots, Y(t_k))$.

8. Assume Theorem 3.3. Show that the random variable $T(0) \stackrel{def}{=} inf\{t \neq 0 : W(t) = 0\}$ is zero with probability 1.

9. Suppose that $v(\cdot)$ is a nondecreasing and bounded deterministic function for which $v(0) = 0$. If $W(\cdot)$ is a Wiener process, show that

$$X(\cdot) \stackrel{def}{=} W(v(\cdot))$$

is a zero-mean Gaussian process with independent increments and variance funcion $v(\cdot)$.

References

Billingsley P. (1995). Probability and Measure, 3rd Edition, Wiley, New York.

Breiman (1968). Probability.

Cox DR & Miller HD (1965), Theory of Stochastic Processes, Methuen, London.

Fleming TR & Harrington DP (1991). Counting Processes and Survival Analysis, Wiley, NY.

Grimmett G & Stirzaker D (2001). Probability and Random Processes, 3rd Edition, Oxford University Press, Oxford, UK.

Hall & Wellner (1980). Biometrika 67:133-144.

Kingman JFC & Taylor SC (1973). Introduction to Measure and Probability, Cambridge University Press.

Schumacher (1984). Intl. Statis. Review 52:263-281)