BROWNIAN MOTION AND RELATED PROCESSES

Karlin & Taylor, *A First Course in Stochastic Processes, ch 7, 15*

Brownian Motion: Definitions

Brownian motion can be defined and constructed in many ways. Some of these include:

1. A stochastic process X_t with independent, normally distributed increments $X_t - X_s \sim N(0, t-s)$, continuous paths, and initial value $X_0 = 0$;
2. A Gaussian stochastic process with mean $\mathsf{E}X_t = 0$, covariance $\mathsf{E}X_s X_t = \min(s,t)$, and continuous paths;
3. A Markov process X_t with initial value $X_0 = 0$, transition probability

$$\mathsf{P}[X_t \le r \mid X_s = x] = \int_{-\infty}^{r} e^{-(y-x)^2/2(t-s)} \frac{dy}{\sqrt{2\pi(t-s)}}$$

and continuous paths;

4.* A process X_t with (a.) independent and (b.) stationary increments (*i.e.*, the random variables $[X_{t_i} - X_{t_{i-1}}]$ are independent and have distributions depending only on $(t_i - t_{i-1})$), with continuous paths.
5. A martingale X_t such that $X_t^2 - t$ is also a martingale, with initial value $X_0 = 0$ and continuous paths.

The only tricky part of constructing X_t is getting continuous paths; it's pretty easy to get a process with the right joint distribution for all times t. Note that definitions (4.) and (5.) don't even mention the normal distribution; that follows from the other requirements as a consequence of the Central Limit Theorem.

Here's one construction, for $0 \le t \le 1$. The idea is to construct a sequence of piecewise-linear approximations $X_t^{(n)}$ to X_t, with exactly the correct distribution on all dyadic rationals of degree n (those of the form $\frac{i}{2^n}$). The key computation about Brownian motion is that for any $0 \le a \le b \le c < \infty$, the conditional distribution of X_b given X_a and X_c is normal with mean $\mu_b = \frac{(c-b)X_a + (b-a)X_c}{c-a}$ (the linear interpolate) and variance $\sigma_b^2 = \frac{(c-b)(b-a)}{c-a}$; for $a = \frac{i}{2^n}$, $c = \frac{i+1}{2^n}$, and $b = \frac{a+c}{2} = \frac{2i+1}{2^{n+1}}$, we have $\mu_b = 1/2[X_a + X_c]$ and $\sigma_b^2 = (1/2)^{2+n}$.

Let z_i be an *iid* sequence of $N(0,1)$ random variables and for each n define random variables x_i^n, $1 \le i \le 2^n$, by:

Even i:	$x_0^0 = 0$	$x_{2i}^{n+1} = x_i^n$	$0 \le i \le 2^n$
Odd i:	$x_1^0 = z_1$	$x_{2i-1}^{n+1} = 1/2[x_{i-1}^n + x_i^n] + (1/2)^{1+n/2} z_{2^n+i}$	$1 \le i \le 2^n$

Now define a sequence of processes $X_t^{(n)}$ by linearly interpolating the $X_{i/2^n}^{(n)} = x_i^n$'s:

$$X_t^{(n)} = (i - t2^n)x_{i-1}^n + (1 - i + t2^n)x_i^n. \qquad \frac{i-1}{2^n} < t \le \frac{i}{2^n}$$

By construction $X_t^{(n)}$ is a Gaussian process with continuous paths, initial value zero, and the right probability distribution at each n^{th}-order dyadic rational; it remains to show that the $X_t^{(n)}$ converge uniformly *a.s.* and that the limit is Brownian motion. We'll turn to that next lecture.

$*$ With this definition the process will have mean $\mathsf{E}X_t = t\mu$ and variance $\mathsf{E}(X_t - t\mu)^2 = t\sigma^2$ for some constants μ, σ^2; the rescaled process $[X_t - t\mu]/\sigma$ has the usual normalization, $\mu = 0$ and $\sigma^2 = 1$.

Actually, constructing Brownian motion is in some sense very easy— if Y_t is *any* square-integrable mean-zero stochastic process starting at zero with independent increments, the function $\sigma_s^2 = \mathsf{E}[Y_s^2]$ must be increasing since, for $0 \le s \le t$, $\sigma_t^2 = \mathsf{E}[\Big(X_s + (X_t - X_s)\Big)^2] = \sigma_s^2 + \mathsf{E}(X_t - X_s)^2 \ge \sigma_s^2$. If $\sigma_s^2 \to \infty$ as $s \to \infty$, then for each $n \ge 1$ and $t \ge 0$ we can set

$$s_n(t) = \inf[s : \sigma_s^2 \ge nt]$$

and define

$$X_t^{(n)} = \frac{1}{\sqrt{n}} Y_{s_n(t)}; \tag{$*$}$$

for every n, $X_t^{(n)}$ has independent increments with mean zero and approximately the right covariance (*exactly* the right covariance if σ_s^2 is strictly increasing). In the limit as $n \to \infty$, the covariance becomes exactly correct and moreover the Central Limit Theorem applies: for large n each $s_n(t)$ becomes large, and for $s < t$ the increment $[Y_{s_n(t)} - Y_{s_n(s)}]$ can be thought of as the sum of very many small and independent increments. Thus the time-change $(*)$ makes almost any independent-increment process converge to Brownian motion, and in particular we can construct Brownian Motion as a limit of random walks, Markov chains, or Poisson processes. For the simple symmetric random walk starting at zero, $s_n = \lceil nt \rceil$ and

$$X_t = n^{-1/2} Y_{\lceil nt \rceil}$$

converges to Brownian Motion.

Continuous Paths

Most things we might want to compute about any random variable X defined on some probability space $(\Omega, \mathcal{F}, \mathsf{P})$ don't really depend on $(\Omega, \mathcal{F}, \mathsf{P})$ at all, but only on the *probability distribution*, the induced measure $\mu_X = \mathsf{P} \circ X^{-1}$ on the real line $(\mathbb{R}, \mathcal{B})$. The random variable $\xi(\omega) = \omega$ on the probability space $(\mathbb{R}, \mathcal{B}, \mu_X)$ has the same probability distribution as X, and so we can usually study features of X without worrying about $(\Omega, \mathcal{F}, \mathsf{P})$ by using this "canonical probability space" $(\mathbb{R}, \mathcal{B}, \mu_X)$. If we have not one but several random variables X_1, X_2,..., X_n, the same idea works in n-dimensional space: if $\mu_{\mathcal{X}}$ denotes the joint probability distribution, the canonical space is $(\mathbb{R}^n, \mathcal{B}^n, \mu_{\mathcal{X}})$ on which the random variables $\xi_i(\omega) = \omega_i$ have the same joint distribution as the X_i.

What about *infinitely many* random variables, especially the uncountable infinity of random variables X_t for Brownian Motion?

Consider the set Ω of continuous real-valued functions on the unit interval starting at zero: $\Omega = \{\text{Continuous } \omega_t : [0,1] \to \mathbb{R},\ \omega_0 = 0\}$. The supremum gives a natural notion of distance from one ω to another in Ω, leading to the topological notion of open sets; let $\mathcal{F}$ be the smallest σ-algebra (or Borel Field, BF) containing these open sets, *i.e.*, containing for each $\omega_0 \in \Omega$ and $\epsilon > 0$ the set

$$[\omega \in \Omega : \sup_{0 \le s \le 1} |\omega(s) - \omega_0(s)| \le \epsilon].$$

(Don't worry if this seems obscure). The *distribution* of Brownian Motion is just the probability measure P on $(\Omega, \mathcal{F})$ such that $\xi_t(\omega) = \omega(t)$ is a Brownian Motion on $(\Omega, \mathcal{F}, \mathsf{P})$.

One way to construct P, and with it Brownian Motion, is to look at the distribution P_n induced by the process $X^{(n)}$ defined earlier; if we can show that these measures converge, we can define P to be their limit and verify that it has the right properties.

PATH CONTINUITY AND NON-DIFFERENTIABILITY

Richard Durret, *Brownian Motion and Martingales in Analysis, pp 1–7*

Introduction

Last time we defined Brownian Motion in five ways, including

I. A stochastic process X_t with initial value $X_0 = 0$ with
  A. Stationary independent increments $[X_t - X_s]$;
  B. Normally distributed increments $[X_t - X_s] \sim N(0, t - s)$;
  C. Continuous paths, almost surely.

Are these consistent? If a process X_t has independent increments, can they also have the specified Gaussian distributions? If they do, can the process also have continuous paths? If so, is path continuity a consequence of a. and b.? As we will see, the answers are *Yes*, *Yes*, and *No*, respectively. To clarify the issues let's consider other SII processes satisfying a. above; three possibilities are:

$$\text{Brownian Motion} \quad X_t: \quad \mathsf{P}[X_t - X_s \in A] = \int_A e^{-x^2/2(t-s)} \frac{dx}{\sqrt{2\pi(t-s)}}$$

$$\text{Cauchy Process} \quad C_t: \quad \mathsf{P}[C_t - C_s \in A] = \int_A \frac{(t-s)\,dx}{\pi[(t-s)^2 + x^2]}$$

$$\text{Poisson Process} \quad N_t: \quad \mathsf{P}[N_t - N_s \in A] = \sum_{x \in A} e^{-(t-s)} \frac{(t-s)^x}{x!}$$

All three distributions are possible for SII processes; to see this it is only necessary to check that for $t_1 < t_2 < t_3$, the indicated distribution for $[X_{t_3} - X_{t_1}]$ is the same as that of the sum of independent random variables with the distributions indicated for $[X_{t_2} - X_{t_1}]$ and $[X_{t_3} - X_{t_2}]$. It turns out that this is equivalent to requiring that the characteristic function $\mathsf{E}[e^{iaX_t}]$ be of the form $e^{-t\phi(a)}$; for these three distributions the characteristic functions are indeed of that form with $\phi(a) = a^2/2$, $|a|$, and $[1 - e^{ia}]$, respectively.

All three distributions are also almost-surely continuous at every point, in the sense that for every fixed t, $\mathsf{P}[X_t = \lim_{s \to t} X_s] = \mathsf{P}[C_t = \lim_{s \to t} C_s] = \mathsf{P}[N_t = \lim_{s \to t} N_s] = 1$. Note that the Poisson process is constant except for jumps of size one, and so its paths are *not* continuous—they are continuous *at each fixed* t, because the jump times have continuous distributions, but any interval of length L will contain at least one jump with probability $1 - e^{-L}$ and so the path will not be *a.s.* continuous on that interval. There is no way to construct a Poisson process with continuous paths; it turns out that there is no way to construct a Cauchy process with continuous paths, either. What about Brownian Motion?

We have constructed Brownian Motion already on the dyadic rationals $\mathbb{Q}_2$ from an IID sequence z_k of $N(0,1)$ random variables by setting $X_0 = 0$ and $X_1 = z_1$ and, recursively, defining X_t for $t = \frac{2i-1}{2^{n+1}}$ by $X_t = {}^1\!/\!_2[X_{(i-1)/2^n} + X_{i/2^n} + z_{i+2^n}/\sqrt{2^n}]$. Can we extend the definition to all $0 \le t \le 1$ by continuity, *i.e.*, set $X_t \equiv \lim_{\mathbb{Q}_2 \ni s \to t} X_s$?

Any continuous function $f(x)$ is uniformly continuous when restricted to a compact set like $[0,1]$, and any *uniformly* continuous function $g(x)$ defined on a set D can be extended to a uniformly continuous function on the closure $\bar{D}$, but in general a function that is merely continuous on a set D cannot be extended to be continuous on $\bar{D}$. Pick an irrational $\omega \in (0,1)$ (perhaps $\frac{1}{\pi}$) and think about the function $g(x) = 1_{[\omega,1]}(x)$ defined on the dyadic rationals $x \in \mathbb{Q}_2$; $g(x)$ is continuous at every rational x, since $|g(y) - g(x)| < \epsilon$ whenever $|y - x| < \delta_x = |x - \omega|$, but is not *uniformly* continuous since no single δ will work for all x.

To extend X_s continuously to $\overline{\mathbb{Q}_2} = [0,1]$ we must show that X_s is almost surely *uniformly continuous* on $\mathbb{Q}_2$, *i.e.*, that for *a.e.* ω, $\forall \epsilon \exists \delta_\omega$ such that $0 \le s < t \le 1, (t-s) < \delta_\omega \Rightarrow |X_t(\omega) - X_s(\omega)| < \epsilon$. Note that this is *not* true for the Poisson process, despite the almost-sure continuity at each point. The argument for Brownian Motion hinges on the Borel-Cantelli lemma and the routine calculation for normally-distributed random variables $X \sim N(0, \sigma^2)$ and real numbers $p > -1$,

$$\begin{aligned}
\mathsf{E}[|X|^p] &= \int_{-\infty}^{\infty} |x|^p \, e^{-x^2/2\sigma^2} \, \frac{dx}{\sqrt{2\pi\sigma^2}} \\
&= \frac{2}{\sqrt{\pi}} \big(2\sigma^2\big)^{p/2} \int_0^\infty \Big(\frac{x^2}{2\sigma^2}\Big)^{p/2} e^{-x^2/2\sigma^2} \, \frac{dx}{\sqrt{2\sigma^2}} \\
&= \frac{2}{\sqrt{\pi}} \big(2\sigma^2\big)^{p/2} \int_0^\infty \Big(\frac{x^2}{2\sigma^2}\Big)^{(p-1)/2} e^{-x^2/2\sigma^2} \, \frac{x\,dx}{2\sigma^2} \\
&= \frac{\Gamma\big(\frac{p+1}{2}\big)\big(2\sigma^2\big)^{p/2}}{\sqrt{\pi}} = c_p (\sigma^2)^{p/2} \qquad (\star)
\end{aligned}$$

and so, for any $\gamma > 0$ and $\delta > 0$,

$$\begin{aligned}
&\mathsf{P}\Big[|X_{j/2^n} - X_{i/2^n}| > \Big(\frac{j-i}{2^n}\Big)^\gamma \text{ for some } 0 \le i < j \le 2^n,\ (j-i) \le 2^{\delta n}\Big] \\
&\quad \le \sum_{i=0}^{2^n-1} \sum_{j=i+1}^{i+2^{n\delta}} \mathsf{P}\Big[|X_{j/2^n} - X_{i/2^n}| > \Big(\frac{j-i}{2^n}\Big)^\gamma\Big] \qquad \text{(by subadditivity)} \\
&\quad \le \sum_{i=0}^{2^n-1} \sum_{j=i+1}^{i+2^{n\delta}} \frac{\mathsf{E}[|X_{j/2^n} - X_{i/2^n}|^p]}{\Big(\frac{j-i}{2^n}\Big)^{\gamma p}} \qquad \text{(by Chebychev)} \\
&\quad = \sum_{i=0}^{2^n-1} \sum_{j=i+1}^{i+2^{n\delta}} \frac{c_p\Big(\frac{j}{2^n} - \frac{i}{2^n}\Big)^{p/2}}{\Big(\frac{j-i}{2^n}\Big)^{\gamma p}} \qquad \text{(by } (\star)\text{)} \\
&\quad = c_p \sum_{i=0}^{2^n-1} \sum_{j=i+1}^{i+2^{n\delta}} \Big(\frac{j-i}{2^n}\Big)^{p(1/2-\gamma)} \\
&\quad \le c_p 2^n 2^{n\delta} \Big(\frac{2^{n\delta}}{2^n}\Big)^{p(1/2-\gamma)} = c_p 2^{-n\epsilon}
\end{aligned}$$

where $\epsilon = -(1+\delta) + (1-\delta)p(1/2 - \gamma)$. For $\gamma < 1/2$ and $\delta < 1$ we can insure $\epsilon > 0$ by taking $p > \frac{1+\delta}{(1/2-\gamma)(1-\delta)}$. By the Borel-Cantelli lemma, for *a.e.* ω $\exists N_\omega$ $\forall n \ge N_\omega$ $\forall q = i/2^n$, $r = j/2^n$ s.t. $|q - r| < 2^{-n(1-\delta)}$, $|X_q - X_r| \le (q-r)^\gamma$. It follows (see Durrett) that there exists a number c_ω such that $\forall q, r \in \mathbb{Q}_2 \cap [0,1]$,

$$|X_q - X_r| \le c_\omega (q-r)^\gamma$$

i.e., that the restriction of X_t to $\mathbb{Q}_2$ is *a.s.* uniformly Hölder continuous of index γ for every $\gamma < 1/2$. In fact this is about the best we can do: X_t is *a.s. not* Hölder continuous of index $1/2$ at any point t, and in particular is not differentiable at any point t. In fact, one of the Laws of the Iterated Logarithm gives

$$1 = \limsup_{s \to t} \frac{\pm(X_t - X_s)}{\sqrt{2(t-s)\log\log 1/(t-s)}}, \text{ so } \limsup_{s \to t} \frac{|X_t - X_s|}{\sqrt{t-s}} = +\infty \ a.s.$$

BROWNIAN SCALING AND REFLECTION

Karlin & Taylor, *A First Course in Stochastic Processes, pp 345–351*

Introduction

We have just constructed a Brownian Motion process, *i.e.*,

II. A stochastic process X_t with initial value $X_0 = 0$ with
   A. Stationary independent increments $[X_t - X_s]$;
   B. Normally distributed increments $[X_t - X_s] \sim N(0, t-s)$;
   C. Continuous paths, almost surely.

Now pick any $c \in \mathbb{R}$, $c \neq 0$, and $h > 0$ and define four processes $X_k(t)$ from $X_t = X(t)$ as follows:

III. $X_1(t) = cX(t/c^2)$;
IV. $X_2(t) = tX(1/t)$ for $t > 0$, $X_2(0) = 0$;
V. $X_3(t) = X(t+h) - X(h)$;
VI. $X_4(t) = (t+1)X\big(\frac{1}{t+1}\big) - X(1)$.

It is straightforward to verify that each of these is a Brownian motion satisfying *a.*, *b.*, *c.* above; by the way, $X_4(t)$ only depends on $X(s)$ for $0 < s \le 1$, and yields a Brownian motion for $0 \le t < \infty$ from our earlier construction of Brownian motion only for $0 \le s \le 1$.

It turns out that Property 3. above is true, not only for fixed $h > 0$, but also for random $\tau = \tau(\omega) > 0$ provided τ is a Markov time (a.k.a. stopping time); in particular, it holds for first hitting times $\tau_a = \inf[s > 0 : X_s = a]$. Since a Brownian Motion has probability ½ of being positive at any time $t > 0$, it follows that for any time $t > 0$ and level $a \ge 0$,

$$\begin{aligned} \mathsf{P}\big([X_t > a]\big) &= \mathsf{P}\big([X_t > a] \cap [\tau_a \le t]\big) \\ &= \mathsf{P}\big([\tau_a \le t]\big)\mathsf{P}\big([X_t > a]\big|[\tau_a \le t]\big) \\ &= \mathsf{P}\big([\tau_a \le t]\big)\big(\frac{1}{2}\big) \end{aligned}$$

so, turning things around,

$$\begin{aligned} \mathsf{P}\big([\tau_a \le t]\big) &= 2\mathsf{P}\big([X_t > a]\big) \\ &= 2\Phi\big(\frac{-a}{\sqrt{t}}\big) \end{aligned}$$

Starting at $X_0 = 0$, let's find the probability that $X_t = 0$ for any $t \in [t_0, t_1]$. One way to make this precise is to think about the Markov time $\tau = \inf[t \ge t_0 : X_t = 0]$ and calculate $\mathsf{P}[\tau \le t_1]$. If we condition on the value a of X_{t_0}, this is just the probability that, in time $t_1 - t_0$, the Brownian motion $X_{t+t_0} - X_{t_0}$ ever reaches the value $|a|$ in time $[t_1 - t_0]$: we just calculated that this is $\mathsf{P}[\tau \le t_1 \mid X_{t_0} = a] = 2\Phi\big(\frac{-|a|}{\sqrt{t_1 - t_0}}\big)$. Thus the desired probability is

$$\begin{aligned} \mathsf{P}[\tau \le t_1] &= \mathsf{E}\big[2\Phi\big(\frac{-|X_{t_0}|}{\sqrt{t_1 - t_0}}\big)\big] \\ &= \int 2\Phi\big(\frac{-|z|}{\sqrt{t_1 - t_0}}\big)e^{-z^2/2t_0}\,\frac{dz}{\sqrt{2\pi t_0}} \\ &= \frac{2}{\pi}\arccos\big(\frac{t_0}{t_1}\big) && \text{(see text, p.348).} \end{aligned}$$

These are intended to illustrate that many features of Brownian motion are amenable to analytic treatment and exact calculation: this isn't true for most other processes, but we can often use calculations for Brownian motion as approximations for other processes. For example, Lèvy showed that for any $t > 0$, the Lebesgue measure of the set of times $s \le t$ at which X_s is positive exactly satisfies, for $0 \le \theta \le 1$, the relation

$$\mathsf{P}\Big[\frac{\lambda[s \le t \mid X_s > 0]}{t} \le \theta\Big] = \frac{2}{\pi}\arcsin(\theta);$$

Kakutani showed that the fraction of $k \le n$ for which $S_k > 0$ has approximately that same distribution, for any sum of *i.i.d.* rv's with zero mean and finite variance.

Processes Related to Brownian Motion

1. Brownian Motion with Drift.

Let X_t be a Brownian motion and let $x_0 \in \mathbb{R}$, $\mu \in \mathbb{R}$, and $\sigma^2 > 0$ be arbitrary; the process

$$X_1(t) = x_0 + \mu t + \sigma X_t$$

is called *Brownian motion with drift.* It has stationary independent increments (with the normal $N(\mu(t-s), \sigma^2(t-s))$ distribution) and continuous paths starting at $X_1(0) = x_0$.

2. Geometric Brownian Motion.

Let $X_1(t)$ be a Brownian motion with drift and set

$$X_2(t) = e^{X_1(t)} = e^{x_0 + \mu t + \sigma X_t}.$$

This is called *Geometric Brownian motion*, and is useful in modeling positive quantities whose *fractional* change is independent over different periods; it is often used in the mathematical theory of finance, and in modeling reservoir levels and related phenomena.

3. Reflected Brownian Motion.

Let $X(t)$ be a Brownian motion and set

$$X_3(t) = |X(t)|$$

This is called *Reflected Brownian motion.* The process is positive and Markovian: in fact, for any $x > 0$ and $y > 0$ and $t > s > 0$,

$$\begin{aligned}
\mathsf{P}[X_3(t) \le y | X_3(s) = x] &= \mathsf{P}\big[[|X(t)| \le y] \cap [X(s) = x] | X_3(s) = x\big] \\
&+ \mathsf{P}\big[[|X(t)| \le y] \cap [X(s) = -x] | X_3(s) = x\big] \\
&= \tfrac{1}{2}\mathsf{P}\big[[|X(t)| \le y] | X(s) = x\big] + \tfrac{1}{2}\mathsf{P}\big[[|X(t)| \le y] | X(s) = -x\big] \\
&= \Phi\big(\frac{y-x}{\sqrt{t-s}}\big) - \Phi\big(\frac{-y-x}{\sqrt{t-s}}\big) \\
&= \int_0^y p_{t-s}(z|x)\, dz
\end{aligned}$$

where the conditional *pdf* is given by differentiation as

$$p_u(y|x) = \frac{1}{\sqrt{2\pi u}}\big[e^{-(x-y)^2/2u} + e^{-(x+y)^2/2u}\big].$$

THE BROWNIAN BRIDGE

Introduction

Let X be a Brownian Motion process and consider two processes defined as follows for $0 \le t \le 1$:

$$X_1(t) = X(t) - tX(1) \qquad\qquad X_2(t) = (1-t)X\big(\frac{t}{1-t}\big).$$

Obviously each of these is a mean-zero Gaussian process, since X is; the finite-dimensional distributions will be determined completely once we identify the covariance functions

$$\begin{aligned}
\mathsf{E}[X_1(s)X_1(t)] &= \mathsf{E}\big[\big(X(s) - sX(1)\big)\big(X(t) - tX(1)\big)\big] \\
&= \mathsf{E}\big[X(s)X(t) - sX(1)X(t) - X(s)tX(1) + stX(1)X(1)\big] \\
&= \big[(s \wedge t) - s(1 \wedge t) - t(s \wedge 1) + st(1 \wedge 1)\big] \\
&= s \wedge t - st \\
\mathsf{E}[X_2(s)X_2(t)] &= (1-s)(1-t)\mathsf{E}\big[X\big(\frac{s}{1-s}\big)X\big(\frac{t}{1-t}\big)\big] \\
&= (1-s)(1-t)\big[\big(\frac{s}{1-s}\big) \wedge \big(\frac{t}{1-t}\big)\big] \\
&= (1-s)(1-t)\big(\frac{s}{1-s}\big) && \text{if, say, } s \le t \\
&= s - st \\
&= s \wedge t - st && \text{for any } s,\ t.
\end{aligned}$$

Thus *both* processes have continuous sample paths and mean-zero Normal finite-dimensional distributions with covariance $s \wedge t - st$; such a process is called a *Brownian Bridge*, or sometimes *pinned Brownian motion*. It can also be thought of as a Brownian motion conditioned on the event $X(1) = 0$. It arises (as we'll see below) in nonparametric statistical problems, and it can be used in constructing Brownian motion and related processes. From the second definition it is clear that X_2 is a Markov process, but it does not have independent increments and it is not a martingale: $\mathsf{E}[X_2(t) - X_2(s)|\mathcal{F}_s] = -\frac{t-s}{1-s}X_2(s)$.

The Kolmogorov-Smirnov Statistic

Let X_i be independent and identically distributed from some unknown distribution μ_X with distribution function $F(t) = \mathsf{P}[X_i \le t] = \mu_X\big((-\infty, t]\big)$. If called upon to guess $F(t)$ from observations of X_i we would no doubt consider the *empirical distribution function*

$$F_n(t) = \frac{\#[i \le n : X_i \le t]}{n} = \sum_{i=1}^{n} 1_{(-\infty,t]}(X_i),$$

a random function of t that starts at $F_n(-\infty) = 0$ and jumps by $1/n$ at each observation X_i. Kolmogorov and Smirnov studied the probability distribution of the quantity

$$Y_n = \sup_{-\infty < s < \infty} \sqrt{n}\big|F_n(s) - F(s)\big|,$$

the (normalized) largest deviation of the empirical distribution function from the true distribution function; it turns out that Y_n has the same probability distribution for any continuous distribution $F(t)$, and in particular is the same as that for uniformly distributed random variables with $F(t) = t$. What is the limiting distribution, as $n \to \infty$?

Regarded as a stochastic process, F_n has mean and covariance functions

$$\begin{aligned}
\mathsf{E}[F_n(t)] &= {}^1\!/\!_n \sum_{i=1}^{n} \mathsf{E}1_{(-\infty,t]}(X_i) \\
&= {}^1\!/\!_n \sum_{i=1}^{n} \mathsf{P}[X_i \le t] \\
&= F(t) \\
\mathsf{E}\Big[\Big(F_n(s) - F(s)\Big)\Big(F_n(t) - F(t)\Big)\Big] &= \mathsf{E}\Big[\Big({}^1\!/\!_n \sum_{i=1}^{n}(1_{(-\infty,s]}(X_i) - F(s))\Big)\Big({}^1\!/\!_n \sum_{j=1}^{n}(1_{(-\infty,t]}(X_j) - F(t))\Big)\Big] \\
&= n^{-2}\sum_{i=1}^{n} \mathsf{E}\Big[\big(1_{(-\infty,s]}(X_i) - F(s)\big)\big(1_{(-\infty,t]}(X_i) - F(t)\big)\Big] \\
&= n^{-1}\mathsf{E}\Big[(1_{(-\infty,s]}(X_1) - F(s))(1_{(-\infty,t]}(X_i) - F(t))\Big] \\
&= n^{-1}\mathsf{E}\Big[1_{(-\infty,s\wedge t]}(X_1) - F(s)1_{(-\infty,t]}(X_1) - 1_{(-\infty,s]}(X_1)F(t) + F(s)F(t)\Big] \\
&= n^{-1}\big[F(s\wedge t) - F(s)F(t)\big] \\
&= n^{-1}\big[F(s)\wedge F(t) - F(s)F(t)\big]
\end{aligned}$$

Thus $\sqrt{n}[F_n(t) - F(t)]$ has the same covariance function as $X_1(F(t))$ for a Brownian Bridge $X_1(s)$; by the Central Limit Theorem, the finite-dimensional distributions of $\sqrt{n}[F_n(t) - F(t)]$ converge weakly to the Normal distribution as $n \to \infty$.

It would be nice to have something stronger— to be able to assert that any continuous functional of $\sqrt{n}[F_n(t) - F(t)]$ converges weakly to a similar functional of the Brownian bridge, and in particular that the Kolmogorov-Smirnov statistic Y_n converges to $Y = \sup_{0\le t\le 1}|X_1(t)|$ in distribution. For this we need to develop the concept of the *distribution* of a stochastic process, and study weak convergence of these distributions.

Distributions

The *distribution* of an $\mathbb{R}^1$-valued random variable X on some probability space $(\Omega, \mathcal{F}, \mathsf{P})$ is just the induced measure $\mu_X(B) = \mathsf{P}[X \in B] = \mathsf{P} \circ X^{-1}$ on the Borel sets $\mathcal{B}$ of the real line; for example, X has the $N(\mu, \sigma^2)$ distribution if $\mu_X(B) = \frac{1}{\sqrt{2\pi\sigma^2}} \int_B e^{-(x-\mu)^2/2\sigma^2}\,dx$ and the Poisson distribution with mean λ if $\mu_X(B) = \sum[e^{-\lambda}\lambda^x/x! : x \in B]$.

Similarly the (joint) *distribution* of n random variables $X_1, \ldots, X_n$ is just the occupation measure μ_X of the vector $\mathcal{X} \in \mathbb{R}^n$, $\mu_X(B) = \mathsf{P}[\mathcal{X} \in B]$ on the Borel sets $\mathcal{B}^n$ in $\mathbb{R}^n$. But what about *stochastic processes*, where $n = \infty$? What is the *distribution* of Brownian motion, or of the Brownian bridge, or of the Poisson process or reflected Brownian motion?

Path spaces

If a real-valued RV takes values in $\mathbb{R}$ and a random vector in $\mathbb{R}^n$, then a real-valued stochastic process X_t defined for $t \in \mathcal{T} = [0,1]$ must take values in some set of paths $\Omega = [\omega : \mathcal{T} \to \mathbb{R}]$, and the *distribution* of X must be a probability measure μ_X on some Borel Field $\mathcal{F}$ of subsets of Ω. The simplest path space to consider is the set of *all* functions $\Omega_1 = [\omega : \mathcal{T} \to \mathbb{R}]$, and the *cylinder sets* $\mathcal{F}_1$ generated by the evaluation functionals– *i.e.*, the smallest BF containing sets of the form $[\omega : \omega(t) \in B]$ for each $t \in \mathcal{T}$ and Borel set B.

This works, after a fashion: any consistent set of finite-dimensional distributions does determine a unique measure μ_X on $\mathcal{F}_1$, and the process $X : \Omega \times \mathcal{T} \to \mathbb{R}$ defined by $X(\omega, t) = \omega_t$ does have the right probability distribution at each time t. Unfortunately some important sets of paths E are missing from $\mathcal{F}_1$, making it impossible to calculate $\mu_X[E]$; for example, $[\omega : t \mapsto \omega_t$ is continuous$]$ is not an event and even $[\omega : t \mapsto \omega_t$ is Lebesgue measurable$]$ is non-measurable. We can evaluate ω_t at fixed times t, but the quantity $Y(\omega) = \sup_{0 \le t \le 1} |\omega_t|$ is not a random variable (it's not $\mathcal{F}_1$-measurable) and so we can't calculate its probability distribution. Next time we'll look at some alternative path spaces.

Continuous Paths

For Brownian Motion and its relatives, the problem is solved by using the probability space of *continuous* functions $\Omega_2 = \mathcal{C} = [\omega : \mathcal{T} \to \mathbb{R},\ t \mapsto \omega_t$ is continuous$]$. This is a metric space in the supremum norm

$$\delta(\omega, \omega') = \sup_{0 \le t \le 1} |\omega_t - \omega'_t|$$

and so has a Borel BF $\mathcal{B} = \mathcal{F}_2$ generated by sets of the form $[\omega : \delta(\omega, \omega') < \epsilon]$ for $\epsilon > 0$ and $\omega' \in \mathcal{C}$. By the *distribution* of a path-continuous stochastic process X we will mean the measure μ_X induced on $(\mathcal{C}, \mathcal{B})$.

PATH SPACES

If a real-valued RV takes values in $\mathbb{R}$ and a random vector in $\mathbb{R}^n$, then a real-valued stochastic process X_t defined for $t \in \mathcal{T} = [0,1]$ must take values in some set of paths $\Omega = [\omega : \mathcal{T} \to \mathbb{R}]$, and the *distribution* of X must be a probability measure μ_X on some Borel Field $\mathcal{F}$ of subsets of Ω. Three possible path spaces to consider are:

1. $\Omega_1 = [\omega : \mathcal{T} \to \mathbb{R}]$	The set of *all* functions $\mathcal{T} \to \mathbb{R}$
$\mathcal{F}_1 = \sigma[\mathcal{X}_t^{-1}(B)]$	the *cylinder sets* generated by the evaluation functionals;
2. $\Omega_2 = \mathcal{C}(\mathcal{T} : \mathbb{R})$	the set of all *continuous* functions $\mathcal{T} \to \mathbb{R}$
$\mathcal{F}_2 = \mathcal{B}\big(\mathcal{C}(\mathcal{T} : \mathbb{R})\big)$	the Borel sets generated by $[\omega' \in \mathcal{C} : \sup_{0 \le s \le 1} |\omega_s - \omega'_s| < \epsilon]$
3. $\Omega_3 = \mathcal{D}(\mathcal{T} : \mathbb{R})$	the Skorohod space of all *right-continuous* functions $\mathcal{T} \to \mathbb{R}$ with left limits
$\mathcal{F}_3 = \mathcal{B}\big(\mathcal{D}(\mathcal{T} : \mathbb{R})\big)$	the Borel sets generated by Skorohod neighborhoods in $\mathcal{D}$.

The simplest one to use is Ω_1. This works, after a fashion: any consistent set of finite dimensional distributions does determine a unique probability measure μ_X on $\mathcal{F}_1$, and the process $X : \Omega \times \mathcal{T} \to \mathbb{R}$ defined by $X(\omega, t) = \omega_t$ does have the right probability distribution at each time t. Unfortunately some important sets of paths E are missing from $\mathcal{F}_1$, making it impossible to calculate $\mu_X[E]$; for example, $[\omega : t \mapsto \omega_t$ is continuous$]$ is not an event and even $[\omega : t \mapsto \omega_t$ is Lebesgue measurable$]$ is non-measurable. We can evaluate ω_t at fixed times t, but the quantity $Y(\omega) = \sup_{0 \le t \le 1} |\omega_t|$ is not a random variable (it's not $\mathcal{F}_1$-measurable) and so we can't calculate its probability distribution.

Continuous Paths

For Brownian Motion and its relatives, the problem is solved by using the probability space of *continuous* functions $\Omega_2 = \mathcal{C} = [\omega : \mathcal{T} \to \mathbb{R},\ t \mapsto \omega_t$ is continuous$]$. This is a metric space in the supremum norm

$$\delta(\omega, \omega') = \sup_{0 \le t \le 1} |\omega_t - \omega'_t|$$

and so has a Borel BF $\mathcal{B} = \mathcal{F}_2$ generated by sets of the form $[\omega : \delta(\omega, \omega') < \epsilon]$ for $\epsilon > 0$ and $\omega' \in \mathcal{C}$. By the *distribution* of a path-continuous stochastic process X we will mean the measure μ_X induced on $(\mathcal{C}, \mathcal{B})$. The space Ω_3 is suitable for processes with discontinuous paths including the Poisson process, generalized Poisson process, birth/death processes, Markov chains, the Cauchy process, and others.

Tightness and Weak Convergence

Any infinite sequence $\alpha_n \subset [0,1]$ has a limit point α_∞ in $[0,1]$, and a subsequence $\alpha_{n_k} \to \alpha_\infty$; the proof is the so-called *diagonal argument*. Start with $i = 0$, $[a_0, b_0] = [0,1]$, and $n_{0j} = j$; note that $[a_i, b_i]$ contains all of the infinite sequence n_{ij}. For each i let $[a_{i+1}, b_{i+1}]$ be $[a_i, \frac{a_i+b_i}{2}]$ if that contains infinitely-many of the n_{ij}, and otherwise let $[a_{i+1}, b_{i+1}]$ be $[\frac{a_i+b_i}{2}, b_i]$; let $n_{i=1,j}$ be the subsequence of n_{ij} which lie in $[a_{i+1}, b_{i+1}]$. Now the diagonal sequence n_{ii} must lie in each $[a_j, b_j]$ for $i \ge j$, and so must be a Cauchy sequence converging to the limit $\alpha_\infty = \cap [a_i, b_i]$ in $[0,1]$.

In $\mathbb{R}^n$ any closed and bounded set K has the property that every infinite sequence $\alpha_n \subset K$ has a limit point $\alpha_\infty \in K$; such a set K is said to be (sequentially) *compact*. A set A like $(0,1]$ whose *closure* is compact is sometimes called *precompact* or *conditionally compact*; every infinite sequence $\alpha_n \subset A$ has a limit point α_∞, but it is possible that $\alpha_\infty \notin A$. In $\mathbb{R}^n$ every bounded set is precompact, but in other metric spaces simple boundedness may not be enough; for example the functions $f_n(x) = \sin(n\pi x)$ are all elements of the space $\mathcal{C} = \mathcal{C}_b(\mathcal{T})$ of continuous bounded functions on $\mathcal{T} = [0,1]$ are all bounded by 1, but no subsequence converges uniformly on $\mathcal{T} = [0,1]$. The Arzelà-Ascoli theorem asserts that a set $A \subset \mathcal{C}$ of continuous functions is precompact if and only if the elements $\omega \in A$ are uniformly bounded and equicontinuous, *i.e.*, if and only if:

A. For some $B < \infty$, $|\omega(0)| < B$ for all $\omega \in A$;

B. For all $\epsilon > 0$ there is a $\delta > 0$ such that $\forall \omega \in A$, $\forall s, t \in \mathcal{T}$, $|s - t| < \delta \Rightarrow |\omega(s) - \omega(t)| < \epsilon$.

Now let μ_n be a sequence of probability measures on the Borel sets $\mathcal{B}$ of $(0,1]$; the numbers $\alpha_n = \mu_n((0, 1/2])$ all lie in $[0,1]$, so along some subsequence n_{1i} the numbers $\alpha_{n_{1i}}$ converge. Along a further subsequence n_{2i} the numbers $\mu_n((0, 1/4])$ and $\mu_n((1/2, 3/4])$ also converge; along subsequence subsequences n_{ki} we can insure that $\mu_n(A)$ converges for each interval $A = (\frac{j}{2^k}, \frac{j'}{2^k}]$. Finally, along the diagonal sequence $\mu_{n_{ii}}(A)$ converges for every interval with dyadic-rational endpoints. Is the limit μ_∞ a probability measure?

The surprising answer is, *maybe not.* Think about a sequence μ_n of measures each giving probability one to the single point 2^{-n}; the limit *ought* to give probability one to the limit point 0, but $0 \notin (0,1]$— and in fact the limit is $\mu(A) = 0$ for all $A \subset (0,1]$. This is the only thing that can go wrong, however:

Theorem (Prohorov). *Let μ_n be a sequence of probability measures on the Borel sets $\mathcal{F}$ of a complete separable metric space Ω. Then some subsequence μ_{n_k} converges weakly to a subprobability measure μ_∞ on $\mathcal{F}$ satisfying $0 \le \mu_\infty(\Omega) \le 1$. If for each $\epsilon > 0$ there is a compact set $K_\epsilon \subset \Omega$ satisfying $\mu_n(K_\epsilon) \ge 1 - \epsilon$ for every n, the sequence μ_n is said to be tight and necessarily $\mu_\infty(\Omega) = 1$. If every convergent subsequence converges to the same limit point μ_∞, then the entire sequence converges.*

Theorem. *A family P_n of probability measures on $(\mathcal{C}, \mathcal{B})$ is tight if and only if*

C. *For each* $\eta > 0$ *there is a* $B < \infty$ *such that* $\forall n$, $\mathsf{P}_n[\omega : |\omega(0)| > B] < \eta$;

D. *For all* $\epsilon > 0$ *and* $\eta > 0$ *there is a* $\delta > 0$ *such that* $\forall n$,

$$\mathsf{P}_n[\omega : \sup_{|s-t|<\delta} |\omega(s) - \omega(t)| > \epsilon] < \eta.$$

Corollary (Kolmogorov). *A family* P_n *of probability measures on* $(\mathcal{C}, \mathcal{B})$ *is tight if there exist numbers* $\alpha > 0$, $\beta > 0$, $B < \infty$, *and* $C < \infty$ *such that* $\forall n$,

C. $\mathsf{E}_n|\omega(0)|^\beta \le B$;

D. $\mathsf{E}_n|\omega(s) - \omega(t)|^\beta \le C|t-s|^{1+\alpha}$.

Continuous Stochastic Processes

For each finite set $J \subset \mathcal{T}$ let μ_J be a probability measure on $|J|$-dimensional Euclidean space $\mathbb{R}^J$ such that, for $J \subset J'$, the measure μ_J is the marginal for $\mu_{J'}$; call such a collection of measures a "consistent finite dimensional distribution." For example, if $m(t)$ is any function on $\mathcal{T}$ and $\gamma(s,t)$ is a (positive definite) covariance function, *i.e.*, satisfies $\sum_{i,j\le n} z_i \bar{z}_j \gamma(t_i, t_j) > 0$ for every integer n, complex $z_1, \ldots, z_n$, and times $t_i \in \mathcal{T}$, then we can construct a unique consistent finite dimensional distribution such that for each $s, t \in \mathcal{T}$, $\mu_{\{s,t\}}$ is bivariate Normal with mean vector and covariance matrix

$$\begin{pmatrix} m(s) \\ m(t) \end{pmatrix} \qquad \begin{pmatrix} \gamma(s,s) & \gamma(s,t) \\ \gamma(t,s) & \gamma(t,t) \end{pmatrix}.$$

Obviously any measure μ on $(\Omega_1, \mathfrak{F}_1)$ or on $(\mathcal{C}, \mathcal{B})$ induces a consistent family of finite dimensional distributions. Any consistent finite dimensional distribution induces a unique measure on $(\Omega_1, \mathfrak{F}_1)$, but it's harder to induce a measure on $(\mathcal{C}, \mathcal{B})$; the Poisson distributions won't work, for example, because Poisson sample-paths aren't continuous. By Kolmogorov's Corollary above,

Theorem. *Let* $\{\mu_J\}$ *be a consistent family of finite dimensional distributions. If there exist positive constants* $\alpha > 0$, $\beta > 0$, *and* $C > 0$ *such that* $\mathsf{E}|X(s) - X(t)|^\beta \le C|t-s|^{1+\alpha}$, *then* $\{\mu_J\}$ *induces a unique probability measure on* $(\mathcal{C}, \mathcal{B})$ *and, moreover,*

$$[\omega \in \mathcal{C} :\ t \mapsto \omega_t \text{ is Hölder continuous of index } \frac{\alpha}{\beta} - \epsilon]$$

has μ*-measure one for each* $\epsilon > 0$.

In particular, a Gaussian distribution satisfies this condition (with $\beta = 2\alpha$) if $m(t)$ is Hölder continuous of index $1/2$ and if $|\gamma(t,s)| \le C'|t-s|$; this condition is satisfied by Brownian motion (with or without drift) and the Brownian bridge.

THE BROWNIAN BRIDGE REVISITED

Introduction

Last time we presented Kolmogorov's Theorem, a corollary to a theorem of Prohorov:

Theorem(Kolmogorov). *A family μ_n of probability measures on $(\mathcal{C}, \mathcal{B})$ is tight if there exist numbers $\alpha > 0$, $\beta > 0$, $B < \infty$, and $C < \infty$ such that $\forall n$,*

C. $\int |\omega(0)|^\beta \, d\mu_n \le B$;

D. $\int |\omega(s) - \omega(t)|^\beta \, d\mu_n \le C|t-s|^{1+\alpha}$. *In this case any limit point μ of $\{\mu_n\}$ is a probability measure giving probability one to the set of Hölder continuous functions of index $\frac{\alpha}{\beta}$.*

Today we will use this theorem to present another construction of the Brownian Bridge; the method is quite general, and is an important tool in constructing and studying stochastic processes.

Let z_n be an iid sequence of standard $N(0,1)$ random variables and define $x_0^0 = x_1^0 = 0$. For $n \ge 0$ and $0 \le i < 2^n$ define

$$\begin{aligned} x_{2i}^{n+1} &= x_i^n \\ x_{2i+1}^{n+1} &= {}^1\!/\!_2\big(x_i^n + x_{i+1}^n + 2^{-n/2} z_{2^n+i}\big) \qquad 0 \le i < 2^n \end{aligned}$$

It's easy to verify that, with this specification, the processes

$$X_t^n = x_i^n + (2^n t - i)(x_{i+1}^n - x_i^n) \qquad \tfrac{i}{2^n} \le t < \tfrac{i+1}{2^n}$$

have the Brownian Bridge covariance $\Gamma_{st} = (s \wedge t) - st$ for $s, t \in 2^n\mathbb{N}$, inducing the Gaussian measures

$$\mu_n(B) = \mathsf{P}[X^n \in B]$$

on the Borel sets $B \in \mathcal{B}$ on $\mathcal{C}$.

For s, t not dyadic rationals in $2^n\mathbb{N}$, the covariance of X^n (or μ_n) may not quite be Γ_{st}; a tedious but straightforward calculation from the definitions shows that $\mathsf{E}X_s^n X_t^n = \Gamma_{st} = (s \wedge t) - st$ if the integer parts of 2^s and $2^n t$ differ ($\lfloor 2^n s \rfloor \ne \lfloor 2^n t \rfloor$), for example, if $|s-t| > 2^{1-n}$, while if $\lfloor 2^n s \rfloor = \lfloor 2^n t \rfloor = j$, $\mathsf{E}X_s^n X_t^n = (s \wedge t) - st - 2^{-n}(1 - (2^n t - j))(2^n s - j)$ differs from Γ_{st} by no more than 2^{-n}. It follows that $\mathsf{E}[(X_{t+\epsilon} - X_t)] = 0$ and , for small enough ϵ, that $\mathsf{E}[(X_{t+\epsilon} - X_t)^2] = \epsilon(1-\epsilon)(1-2^{-n})$; in particular, $\mathsf{E}[|X_t - X_s|^\beta] \le c_\beta |t-s|^{\beta/2}$ for every $\beta > 0$ and Kolmogorov's criteria are satisfied.

Let μ be any limit point of the family $\{\mu_n\}$; for any $0 \le s \le t \le 1$ the function $\phi(\omega) = e^{ia\omega_s + ib\omega_t}$ is continuous and bounded on $\mathcal{C}$, so $\int_\mathcal{C} \phi(\omega)\, d\mu_n$ converges to $\int_\mathcal{C} \phi(\omega)\, d\mu$; this is just the joint characteristic function of X_s and X_t, which have the bivariate Normal distribution with mean $\binom{0}{0}$ and covariance $\begin{pmatrix} s & s \\ s & t \end{pmatrix}$, so

$$\begin{aligned} \int_\mathcal{C} e^{ia\omega_s + ib\omega_t}\, d\mu_n &\to \int_\mathcal{C} e^{ia\omega_s + ib\omega_t}\, d\mu \\ &= e^{-{}^1\!/\!_2[a^2 s(1-s) + 2abs(1-t) + b^2 t(1-t)]} \end{aligned}$$

and under μ, $X_t(\omega) = \omega_t$ is a stochastic process with

C. continuous-paths;
D. normal distribution;
E. mean zero;
F. covariance $\Gamma_{st} = (s \wedge t) - st$.

This uniquely determines μ as the Brownian Bridge distribution; since $\{\mu_n\}$ has a unique limit point, necessarily $\mu_n \Rightarrow \mu$. We have succeeded in constructing the Brownian Bridge. By the way, we can now construct a Brownian Motion (or Wiener process) for all time $0 \le t < \infty$ by the formula

$$W_t = (1+t)X\big(\frac{t}{1+t}\big).$$

The technique used in this construction of the Brownian Bridge is quite powerful; the only features we used of the Brownian bridge were:

G. It has continuous-paths (otherwise, use Skorohod space $\mathcal{D}$);
H. We know how to approximate it by a sequence of processes (2.);
I. It has normally-distributed paths whose increments have mean zero (any Hölder continuous mean function would have been OK) and variance $\mathsf{E}[(X_t - X_s)^2] = \mathcal{O}(|t-s|)$;
J. We can identify the limit point: it is characterized by its covariance function, in this case.

The same technique works for many other processes (even for infinite-dimensional ones) whenever we can verify tightness and recognize the weak limit.

Gaussian Conditional Expectations

Let X be a multivariate Gaussian random vector with expectation vector μ and covariance matrix Σ. For each subset I of indices denote by X_I the random vector with components $X_i,\ i \in I$, by μ_I the expectation $\mathsf{E}[X_I]$, and by Σ_{IJ} the covariance matrix $\Sigma_{IJ} = \mathsf{E}[(X_I - \mu_I)(X_J - \mu_J)^\mathsf{T}]$. If Σ_{JJ} is nonsingular, a straightforward calculation yields

$$\begin{aligned} \mathsf{E}[X_I|X_J] &= \mu_I + \Sigma_{IJ}\Sigma_{JJ}^{-1}[X_J - \mu_J] \\ \mathsf{V}[X_I|X_J] &= \Sigma_{II} - \Sigma_{IJ}\Sigma_{JJ}^{-1}\Sigma_{JI} \end{aligned}$$

(in fact, the same formulas work even for singular Σ_{JJ} if we interpret Σ_{JJ}^{-1} as the Moore-Penrose generalized inverse). For mean-zero one-dimensional jointly normal random variables x and y, we have $\mathsf{E}[y|x] = (\mathsf{E}[xy]/\mathsf{E}[x^2])\,x$ and $\mathsf{V}[y|x] = \mathsf{E}[y^2] - \mathsf{E}[xy]^2/\mathsf{E}[x^2]$.

From these formulas we can compute the Brownian Bridge's conditional expectation for $s \le t \le u$ as

$$\begin{aligned} \mathsf{E}[X_t|X_s] &= \frac{1-t}{1-s}X_s \\ \mathsf{V}[X_t|X_s] &= \frac{(1-t)(t-s)}{1-s} \\ \mathsf{E}[X_t|X_s, X_u] &= \frac{u-t}{u-s}X_s + \frac{t-s}{u-s}X_u \\ \mathsf{V}[X_t|X_s, X_u] &= \frac{(u-t)(t-s)}{u-s} \end{aligned}$$

and, consequently,

$$\begin{aligned} \mathsf{E}[(X_{t+\epsilon} - X_t)|\mathcal{F}_t] &= \frac{-\epsilon}{1-t}X_t \\ \mathsf{V}[(X_{t+\epsilon} - X_t)|\mathcal{F}_t] &= \epsilon - \frac{\epsilon^2}{1-t} \\ &= \epsilon + \mathcal{O}(\epsilon^2) \\ \mathsf{E}[(X_{t+\epsilon} - X_t)^2|\mathcal{F}_t] &= \epsilon - \frac{\epsilon^2}{1-t} + [\frac{-\epsilon}{1-t}X_t]^2 \\ &= \epsilon + \mathcal{O}(\epsilon^2) \end{aligned}$$

Semimartingales

From this calculation $\mathsf{E}[X_{t+\epsilon}|\mathcal{F}_t] = X_t - \frac{\epsilon}{1-t}X_t + \mathcal{O}(\epsilon^2)$ it follows that X_t is not a martingale, but that the process

$$W_t = X_t + \int_0^t \frac{X_s}{1-s}\,ds$$

is a martingale, with continuous paths and a Gaussian distribution. The covariance function turns out to be $s \wedge t$, so W_t is just Brownian motion and we have the curious representation

$$X_t = X_0 + \int_0^t \frac{-X_s}{1-s}\,ds + W_t$$

of X_t as a *semimartingale*, the sum of a bounded-variation process (here $X_0 + \int_0^t \frac{-X_s}{1-s}\,ds$) and a martingale. This is our first example of a *diffusion* and of a *stochastic integral.*

INTRODUCTION TO STOCHASTIC INTEGRATION

Avner Friedman, *Stochastic Differential Equations and Applications*, *pp.*. 55—72

Introduction to Stochastic Integral Equations

Last time we constructed the distribution μ of the Brownian Bridge as a measure on the canonical space $(\mathcal{C}, \mathcal{B})$, and the B.B. itself as the canonical process $X_t(\omega) = \omega_t$ on $(\mathcal{C}, \mathcal{B}, \mu)$. Using normal distribution theory we calculated the conditional expectations $\mathsf{E}[X_{t+\epsilon}|X_t] = X_t - \frac{\epsilon}{1-t}X_t$ and $\mathsf{V}[X_{t+\epsilon}|X_t] = \epsilon - \frac{\epsilon^2}{1-t}$; since X_t is Markov, these are the same as the conditional expectations given the Borel Field generated by the entire past of the process up to time t, $\mathcal{F}_t = \sigma[X_s : s \le t]$. This led to the recognition that $W_t = X_t - \int_0^t \frac{X_s}{1-s} ds$ is a continuous-path Gaussian martingale. We calculated that the covariance function is $\mathsf{E}W_s W_t = s \wedge t$ and so recognized W_t as the Wiener process, leading to the representation

$$X_t = \int_0^t \frac{-X_s}{1-s}\,ds \;+\; W_t. \qquad (*)$$

This is our first example of a Stochastic Integral Equation (SIE); given a Brownian Motion W_t, we can try to "solve" $(*)$ for the unknown process X_t. One way to proceed is to define a sequence of processes by $X_t^0 \equiv 0$ and

$$X_t^{n+1} \;=\; \int_0^t \frac{-X_s^n}{1-s}\,ds \;+\; W_t;$$

upon subtracting,

$$X_t^{n+1} - X_t^n \;=\; \int_0^t \frac{X_s^{n-1} - X_s^n}{1-s}\,ds$$

so $\gamma_t^n = \sup_{s \le t} |X_s^{n+1} - X_s^n|$ satisfies $\gamma_t^0 = \sup_{s\le t}|W_s|$ and, for $t \le 1-\epsilon$,

$$\begin{aligned}\gamma_t^n \;&\le \int_0^t \frac{\gamma_s^{n-1}}{1-s}\,ds\\ &\le \epsilon^{-1}\int_0^t \gamma_s^{n-1}\,ds\\ &\le \epsilon^{-2}\int_0^t (t-s)\gamma_s^{n-2}\,ds\end{aligned}$$

$$\begin{aligned}
&\leq \epsilon^{-3} \int_0^t \frac{(t-s)^2}{2} \gamma_s^{n-3}\, ds \\
&\leq \epsilon^{-n} \int_0^t \frac{(t-s)^{n-1}}{(n-1)!} \gamma_s^0\, ds. \\
&\leq \frac{t^n}{\epsilon^n n!} \gamma_1^0.
\end{aligned}$$

Since $e^{t/\epsilon} \sup_{s\leq 1} |W_s| = \sum_{n=0}^{\infty} \frac{t^n}{\epsilon^n n!} \gamma_1^0 < \infty$, $\gamma_t^n \to 0$ uniformly on $t \leq 1-\epsilon$ and X_t^n converges uniformly on compact sets to a limit X_t satisfying $(*)$.

Since it's so easy to construct the Brownian Bridge by solving $(*)$, and since all we used was the conditional mean and variance of the infinitesimal increment $\mathsf{E}[X_{t+\epsilon} - X_t]$, maybe we can use a similar technique for other processes once we know the so-called infinitesimal mean and variance,

$$\mathsf{E}\big[(X_{t+\epsilon} - X_t)|\mathcal{F}_t\big] = \alpha_t \epsilon + o(\epsilon) \qquad \mathsf{E}\big[(X_{t+\epsilon} - X_t)^2|\mathcal{F}_t\big] = \beta_t \epsilon + o(\epsilon).$$

Stochastic Integrals

Stieltjes Integrals

Any finite measure μ on $(0,1]$ is determined uniquely by its distribution measure $G(t) = \mu\big((0,t]\big)$, since the Borel sets are generated by the half-open intervals and $\mu\big((a,b]\big) = G(b) - G(a)$. For any bounded and continuous function f,

$$\int_0^t f(s)\mu(ds) = \lim_{n\to\infty} 2^{-n} \sum_{j=0}^{2^n t-1} f\big(\frac{j}{2^n}\big)\big[G\big(\frac{j+1}{2^n}\big) - G\big(\frac{j}{2^n}\big)\big],$$

justifying the (18^{th}-century) Stieltjes notation $\int_0^t f(s)\, dG(ds)$. If $G(t) = \int_0^t G'(s)\, ds$, this is

$$\begin{aligned}
\int_0^t f(s)\, dG(ds) &= \lim_{n\to\infty} 2^{-n} \sum_{j=0}^{2^n t-1} f\big(\frac{j}{2^n}\big)\big[\frac{1}{2^n} G'\big(\frac{j}{2^n}\big) + o(\frac{1}{2^n}\big] \\
&= \int_0^t f(s) G'(s)\, ds;
\end{aligned}$$

whether or not $G(t)$ is differentiable, the integration-by-parts formula holds for continuously differentiable functions f (note $G(0) = 0$):

$$\int_0^t f(s)\, dG(ds) = f(t)G(t) - \int_0^t f'(s)G(s)\, ds.$$

For step functions $f(t)$ with a constant value b_i on each of n intervals $(t_i, t_{i+1}]$, the integral is just

$$\int_0^t f(s)\, dG(ds) = \sum_{i=0}^{n-1} b_i \big[G(t \wedge t_{i+1}) - G(t \wedge t_i)\big].$$

Wiener Integrals

Now let $f(t)$ be a measurable real-valued function and consider the problem of defining the Stieltjes-like integral $M_t = \int_0^t f(s)\,dW_s$ for Brownian Motion W_s; since W_s is *not* differentiable, we can't use the representation $M_t = \int_0^t f(s)\,W_s'\,ds$ as we did for differentiable functions $G(t)$ above. The other two alternatives do work, however; for continuously differentiable $f(t)$ we can define

$$\int_0^t f(s)\,dW_s = f(t)W_t - \int_0^t f'(s)W_s\,ds,$$

or for step functions we can define

$$\int_0^t f(s)\,dW_s = \sum_{i=0}^{n-1} b_i[W_{t\wedge t_{i+1}} - W_{t\wedge t_i}].$$

In either case $\int_0^t f(s)\,dW_s$ is a continuous-path Gaussian martingale with mean zero and variance

$$\begin{aligned}
\mathsf{E}\big[(\int_0^1 f(s)\,dW_s)^2\big] &= \mathsf{E}\big[(\sum_{i=0}^{n-1} b_i[W_{t_{i+1}} - W_{t_i}])(\sum_{j=0}^{n-1} b_j[W_{t_{j+1}} - W_{t_j}])\big] \\
&= \sum_{i=0}^{n-1} b_i^2 \mathsf{E}[(W_{t_{i+1}} - W_{t_i})^2] \quad \text{(by independent increments)} \\
&= \sum_{i=0}^{n-1} b_i^2 [t_{i+1} - t_i], \qquad \text{and hence } \textit{co}\text{variance} \\
\mathsf{E}\big[(\int_0^s f(u)\,dW_u)(\int_0^t g(u)\,dW_u)\big] &= \int_0^{s\wedge t} f(u)g(u)\,du.
\end{aligned}$$

ITÔ STOCHASTIC INTEGRALS AND DIFFUSIONS

Avner Friedman, *Stochastic Differential Equations and Applications*, *pp.*. 55—72

Properties of Stochastic Integrals

Last time we constructed the so-called *Wiener stochastic integral* process $M_t = \int_0^t f(s)\,dW_s$ for nonrandom square-integrable functions $f(s)$. We defined the integral first for step functions $f(t) = \sum_i b_i 1_{(t_i,t_{i+1}]}$ as $M_t = \int_0^t f(s)\,dW_s = \sum_i b_i\Big[W_{t\wedge t_{i+1}} - W_{t\wedge t_i}\Big]$, then extended by L^2 continuity. For continuously differentiable functions $f(t)$ we also defined the integral by parts as $M_t = \int_0^t f(s)\,dW_s = f(t)W_t - \int_0^t f'(s)W_s\,ds$. For continuous functions $f(t)$ we have the L^2-convergent formula

$$M_t = \int_0^t f(s)\,dW_s = \lim_{n\to\infty} \frac{1}{2^n} \sum_{j=0}^{2^n t-1} f\big(\frac{j}{2^n}\big)\big[W_{(j+1)/2^n} - W_{j/2^n}\big] \qquad (*)$$

which makes it easy to see why the increments $M_t - M_s$ have Gaussian distributions with mean zero and variance $E[(M_t - M_s)^2] = \int_s^t f^2(u)\,du$.

It was Kyoshi Itô's observation that the same construction would also work for *random* integrands $f(t)$, provided that always $f(t)$ is square-integrable and independent of $[W_{t+\epsilon} - W_t]$; since

W_t has independent increments, we can assure that independence by requiring that $f(t)$ be $\mathcal{F}_t$-measurable for each t. The function $f_t = W_t$ satisfies this condition, as does the Brownian Bridge $X_t = \int_0^t \frac{-X_s}{1-s}\,ds + W_t$ and any function $\beta_t = b_t(X_t)$ of t and X_t.

Call a stochastic process β_t (such as W_t or $b_t(X_t)$) *adapted* to the family $\{\mathcal{F}_t\}$ of BF's if β_t is $\mathcal{F}_t$-measurable for every t, *i.e.*, if $[\omega : \beta_t(\omega) \le r] \in \mathcal{F}_t$ for each $t \ge 0$ and $r \in \mathbb{R}$, and let L^2_w be the metric space of adapted processes satisfying $\mathsf{E}[\int_0^1 \beta_t^2\,dt] < \infty$. Any such process $\beta_s \in \mathcal{L}^2_2$ can be approximated by an adapted simple function, with a constant $\mathcal{F}_{t_i}$-measurable value $b_i(\omega)$ on the interval $(t_i, t_{i+1}]$, for which it is easy to calculate the stochastic integral $M_t = \int_0^t \beta_s\,dW_s = \sum_i b_i[W_{t\wedge t_{i+1}} - W_{t\wedge t_i}]$; the mean and variance of the (usually non-Gaussian) *Itô integral* M_1 are:

$$\begin{aligned}
\mathsf{E}[M_1] &= \mathsf{E}\Big[\sum_i b_i[W_{t_{i+1}} - W_{t_i}]\Big] \\
&= \sum_i \mathsf{E}\big[b_i[W_{t_{i+1}} - W_{t_i}]\big] = 0 \\
\mathsf{V}[M_1] &= \mathsf{E}\Big[\big(\sum_i b_i[W_{t_{i+1}} - W_{t_i}]\big)^2\Big] \\
&= \sum_i \mathsf{E}\big[b_i^2[W_{t_{i+1}} - W_{t_i}]^2\big] \\
&= \sum_i \mathsf{E}[b_i^2][t_{i+1} - t_i] \\
&= \int_0^1 \mathsf{E}[\beta_s^2]\,ds
\end{aligned}$$

and, more generally, $\mathsf{E}[M_t] = 0$ and $\mathsf{V}[M_t] = \mathsf{E}[M_t^2] = \int_0^t \mathsf{E}[\beta_s^2]\,ds$. Possibly more revealing is the *conditional* variance; for continuous β_s this is

$$\mathsf{V}[M_{t+\epsilon}|\mathcal{F}_t] = \int_t^{t+\epsilon} \mathsf{E}[\beta_s^2|\mathcal{F}_t]\,ds = \epsilon\beta_t^2 + o(\epsilon),$$

so $\epsilon\beta_t^2$ is just the conditional variance of $[M_{t+\epsilon} - M_t]$, to first order in ϵ.

If $\alpha_t \in L^1_w$ is a integrable adapted process the indefinite integral $\int_0^t \alpha_s\,ds$ is defined in the usual (Lebesgue or Riemann) way; for any $\mathcal{F}_0$-measurable random variable X_0, the sum

$$X_t = X_0 + \int_0^t \alpha_s\,ds + \int_0^t \beta_s\,dW_s$$

is a continuous-path adapted process whose increments have conditional mean and variance

$$\begin{aligned}
\mathsf{E}[X_{t+\epsilon} - X_t|\mathcal{F}_t] &= \int_t^{t+\epsilon} \mathsf{E}[\alpha_s|\mathcal{F}_t]\,ds \\
&= \epsilon\alpha_t + o(\epsilon), \\
\mathsf{V}[X_{t+\epsilon} - X_t|\mathcal{F}_t] &= \epsilon\beta_t^2 + o(\epsilon).
\end{aligned}$$

If we restrict our attention to *Markov* processes X_t, then the conditional mean and variance $\epsilon\alpha_t$ and $\epsilon\beta_t^2$ must be not only $\mathcal{F}_t$-measurable, but $\sigma(X_t)$-measurable— so, for some functions $a_t(x)$ and $b_t(x)$, $\alpha_t = a_t(X_t)$ and $\beta_t = b_t(X_t)$. This class of functions is called *diffusions*:

$$X_t = X_0 + \int_0^t a_s(X_s)\,ds + \int_0^t b_s(X_s)\,dW_s.$$

We have already met several examples, including:

Brownian Bridge: $X_0 = 0$, $a_t(x) = \frac{-x}{1-t}$, and $b_t(x) = 1$.

Brownian Motion with Drift: $X_0 = x_0$, $a_t(x) = \mu$, and $b_t(x) = \sigma$.

Geometric Brownian Motion: $X_0 = e^{x_0}$, $a_t(x) = x(\mu + \frac{\sigma^2}{2})$, and $b_t(x) = x\sigma$.

Reflected Brownian Motion: $X_0 = 0$, $a_s(x) = \delta(x)$, and $b_s(x) = 1$. Here $\delta(x)$ denotes Dirac's *delta function*, the (formal) derivative of the function $1_{[0,\infty)}(x)$; reflected Brownian motion is properly called a *diffusion with boundary*, and is more complicated to study than the other processes mentioned.

The remarkable and deep fact is that *all* continuous-path strong Markov processes are diffusions; see Karlin & Taylor, *A Second Course in Stochastic Processes*, chapter 15, or Stroock & Varadhan, *Multi-dimension Diffusion Processes*, or Itô & M^{c}Kean, *Diffusion Processes and their Sample Paths* (among others).

Itô's Formula

For any $\phi_t(x) \in \mathcal{C}^{1+2}(\mathbb{R}_+ \times \mathbb{R})$, Taylor's formula gives

$$\phi_{t+\epsilon}(x+\xi) = \phi_t(x) + \epsilon\frac{\partial\phi}{\partial t} + \xi\frac{\partial\phi}{\partial x} + \frac{\xi^2}{2}\frac{\partial^2\phi}{\partial x^2} \qquad + o(\epsilon) + o(\xi^2),$$

and in particular $Y_t = \phi_t(X_t)$ satisfies

$$Y_{t+\epsilon} - Y_t = \epsilon\frac{\partial\phi}{\partial t} + \Big(a_t(X_t)\epsilon + b_t(X_t)[W_{t+\epsilon} - W_t]\Big)\frac{\partial\phi}{\partial x} + \epsilon\frac{b_t^2(X_t)}{2}\frac{\partial^2\phi}{\partial x^2} \quad + o(\epsilon),$$

so $Y_t = \phi_t(X_t)$ is itself a diffusion with starting point $Y_0 = \phi_0(X_0)$ and diffusion coefficients

$$\tilde{a}_t(x) = \frac{\partial\phi}{\partial t} + a_t(x)\frac{\partial\phi}{\partial x} + \frac{b_t^2(x)}{2}\frac{\partial^2\phi}{\partial x^2} \qquad \text{and} \qquad \tilde{b}_t(x) = b_t(x)\frac{\partial\phi}{\partial x}.$$

There is a close connection between the diffusion X_t and the differential operator

$$\mathcal{L}\phi \equiv a_t(x)\frac{\partial\phi}{\partial x} + {}^1\!/\!_2\, b_t^2(x)\frac{\partial^2\phi}{\partial x^2},$$

called the *Generator* of the process; we have just seen that $\tilde{a}_t(x) = \frac{\partial}{\partial t}\phi_t(x) + \mathcal{L}\phi_t(x)$, for example, so for any $\phi_t(x)$,

$$M_t = \phi_t(X_t) - \int_0^t \big[\frac{\partial\phi}{\partial s} + a_s(Y_s)\frac{\partial\phi}{\partial x} + {}^1\!/\!_2\, b_s^2(Y_s)\frac{\partial^2\phi}{\partial x^2}\big]\, ds$$

is a martingale. This, in fact, is the modern *definition* of the Diffusion Process with coefficients $a_s(x)$ and $b_s(x)$. Note that $Y_t = \phi_t(X_t)$ is itself a martingale if ϕ satisfies the parabolic partial differential equation $\partial\phi/\partial t = -\mathcal{L}\phi$, *i.e.*,

$$0 \equiv \big[\frac{\partial\phi_t(x)}{\partial t} + a_t(x)\frac{\partial\phi_t(x)}{\partial x} + {}^1\!/\!_2\, b_t^2(x)\frac{\partial^2\phi_t(x)}{\partial x^2}\big].$$

Examples

For example, let $\phi_t(x) = x^2$; then $\mathcal{L}\phi(x) = 2a_t(x)x + b_t^2(x)$, so Itô's formula gives

$$X_t{}^2 = X_0{}^2 + \int_0^t \big(2a_s(X_s)X_s + b_s(X_s)^2\big)\,ds + \int_0^t 2X_s\,b_s(X)_s\,dW_s.$$

An interesting formula for every diffusion process X_t follows from this:

$$\begin{aligned}\int_0^t 2X_s\,dX_s &= \int_0^t 2X_s\,a_s(X_s)\,ds + \int_0^t 2X_s\,b_s(X_s)\,dW_s \\ &= \int_0^t 2X_s\,a_s(X_s)\,ds + X_t{}^2 - X_0{}^2 - \int_0^t \big(2a_s(X_s)X_s + b_s(X_s)^2\big)\,ds \\ &= X_t{}^2 - X_0{}^2 - \int_0^t b_s(X_s)^2\,ds; \qquad (*)\end{aligned}$$

for ordinary integrals we have, of course, $\int_0^t 2f(s)\,df(s) = f(t)^2 - f(0)^2$, but for stochastic integrals there is an additional term.

Inference

Suppose we observe X_s for $0 \le s \le t$, and believe that X_t is a diffusion process; of course we can observe the initial value X_0, but what can we infer about the coefficients $a_s(x)$ and $b_s(x)$? First let's consider the diffusion coefficient $b_s(x)$. There is no hope of inferring anything about $b_s(x)$ away from the observed path (unless we make additional assumptions about the form of $b_s(x)$, but if we know $a_s(x)$ and $b_s(x)$ to be sufficiently smooth in both s and x, then for small $\epsilon > 0$ the *quadratic variation* between s and $s{+}\epsilon$ is

$$\begin{aligned}Q_s^{s+\epsilon}(X) &= \lim_{n\to\infty}\sum_{i=0}^{n-1}\big(X_{s+(i+1)\frac{\epsilon}{n}} - X_{s+i\frac{\epsilon}{n}}\big)^2 \\ &= \lim_{n\to\infty}\sum_{i=0}^{n-1}\big(a_s(X_s)(\frac{\epsilon}{n}) + b_s(X_s)(W_{s+(i+1)\frac{\epsilon}{n}} - W_{s+i\frac{\epsilon}{n}})\big)^2 + o(\epsilon) \\ &= b_s(X_s)^2 \lim_{n\to\infty}\sum_{i=0}^{n-1}(W_{s+(i+1)\frac{\epsilon}{n}} - W_{s+i\frac{\epsilon}{n}})^2 + o(\epsilon) \\ &= b_s(X_s)^2 \lim_{n\to\infty}\sum_{i=0}^{n-1}(\frac{\epsilon}{n})\chi_1^2 + o(\epsilon) = \epsilon b_s(X_s)^2 \lim_{n\to\infty}\chi_n^2/n + o(\epsilon) \\ &= \epsilon b_s(X_s)^2 + o(\epsilon),\end{aligned}$$

so $b_s(X_s)^2$ is *observable* as $\lim_{\epsilon\to 0} Q_s^{s+\epsilon}(X)/\epsilon$. Since $b_s(x)$ is observable, consider diffusions with constant diffusion coefficient:

$$X_t = X_0 + \int_0^t a_s(X_s)\,ds + \sigma W_t.$$

What can we learn from the path on $0 \le s \le t$ about $a_s(x)$? Let's try to compute the *likelihood* for a. For large n set $\epsilon = t/n$ and note that

$$X_{(i+1)\epsilon} = X_{i\epsilon} + \epsilon a_{i\epsilon} + \sigma\big(W_{(i+1)\epsilon} - W_{i\epsilon}\big) + o(\epsilon).$$

The log likelihood for a upon observing only $X_{i\epsilon}$, $i = 0, \dots, n$, is

$$\ell_n(a) = c_n - \frac{n}{2}\log(2\pi\epsilon\sigma^2) - \frac{1}{2\epsilon\sigma^2}\sum_{0\le i<n}\big(X_{(i+1)\epsilon} - X_{i\epsilon} - \epsilon a_{i\epsilon}(X_{i\epsilon})\big)^2 + o(\epsilon),$$

for any constant c_n; it's convenient to choose c_n so that $\ell_n(0) = 0$, *i.e.*, $c_n = \frac{n}{2}\log(2\pi\epsilon\sigma^2) + \frac{1}{2\epsilon\sigma^2}\sum_{0\le i<n}\big(X_{(i+1)\epsilon} - X_{i\epsilon}\big)^2$, whereupon

$$\begin{aligned}
\ell_n(a) &= \frac{1}{2\epsilon\sigma^2}\sum_{0\le i<n}\Big(\big(X_{(i+1)\epsilon} - X_{i\epsilon}\big)^2 - \big(X_{(i+1)\epsilon} - X_{i\epsilon} - \epsilon a_{i\epsilon}(X_{i\epsilon})\big)^2\Big) + o(\epsilon) \\
&= \frac{1}{2\epsilon\sigma^2}\sum_{0\le i<n}\Big(2\epsilon\big(X_{(i+1)\epsilon} - X_{i\epsilon}\big)a_{i\epsilon}(X_{i\epsilon}) - \epsilon^2 a_{i\epsilon}(X_{i\epsilon})^2\Big) + o(\epsilon) \\
&= \sigma^{-2}\int_0^t a_s(X_s)\,dX_s - \frac{1}{2\sigma^2}\int_0^t a_s(X_s)^2\,ds + o(\epsilon)
\end{aligned}$$

Now we pass to the limit $n \to \infty$ (and $\epsilon \to 0$), $\ell(a) = \sigma^{-2}\big(\int_0^t a_s(X_s)\,dX_s - 1/2\int_0^t a_s(X_s)^2\,ds\big)$.

Example 1: Wiener Process

For example, if $X_t = X_0 + \mu t + \sigma W_t$ is Brownian motion with constant drift μ and diffusion rate σ^2, then $a_s(x) \equiv \mu$ and the log likelihood becomes

$$\ell(\mu) = \mu(X_t - X_0)/\sigma^2 - \mu^2 t/2\sigma^2;$$

the MLE estimate is $\hat\mu = (X_t - X_0)/t$, while the Bayesian posterior distribution for a flat prior is

$$\mu\big|\{X_s : 0 \le s \le t\} \sim N\left(\frac{X_t - X_0}{t}, \frac{\sigma^2}{t}\right).$$

Example 2: Ornstein-Uhlenbeck

Now if $X_t = X_0 - \beta\int_0^t X_s\,ds + \sigma W_t$, or $dX_t = -\beta X_t\,dt + \sigma dW_t$, then $a_s(x) \equiv -\beta x$ and

$$\begin{aligned}
\ell(\beta) &= -\frac{\beta}{\sigma^2}\int_0^t X_s\,dX_s - \frac{\beta^2}{2\sigma^2}\int_0^t (X_s)^2\,ds \\
&= -\frac{\beta}{2\sigma^2}\big({X_t}^2 - {X_0}^2 - t\big) - \frac{\beta^2}{2\sigma^2}\int_0^t (X_s)^2\,ds, && \text{(by (*))}
\end{aligned}$$

so for a uniform prior we would have

$$\beta\big|\{X_s : 0 \le s \le t\} \sim N\left(\frac{{X_t}^2 - {X_0}^2 - t}{2\int_0^t (X_s)^2\,ds}, \frac{\sigma^2}{\int_0^t (X_s)^2\,ds}\right).$$

Testing Hypotheses

The likelihood function provides the basis for testing hypotheses like $H_0 : X_t$ **is Browian motion** (with no drift) against alternatives like $H_1 : X_t$ **is a Wiener process** (with constant drift) or $H_2 : X_t$ **is an O-U process** (with linear drift), by finding either P-values or posterior probabilities.

RANDOM MEASURES

Nonparametric Statistics and Random Measures

Let $(\mathcal{X}, \mathcal{B})$ be a measurable space and $(\Omega, \mathcal{F}, \mathsf{P})$ a probability space; denote by $\mathcal{M} = \mathcal{M}(\mathcal{X}, \mathcal{B})$ the vector space of σ-finite signed measures on $(\mathcal{X}, \mathcal{B})$. If we observe a random variable $X \in \mathcal{X}$, what can we say about its probability distribution $\mu_X \in \mathcal{M}$? The fundamental problem of statistics is making inference about μ_X on the basis of observation of X. In a *parametric* analysis we postulate that μ_X lies in a small family of distributions $\mu_X \in \{P_\theta : \theta \in \Theta\}$ (*e.g.*, if $\mathcal{X} = \mathbb{R}^n$, we might postulate that μ_X lies in the multivariate normal family with constant mean vector and covariance matrix $\Sigma = \sigma^2 \mathbf{I}$) indexed by a parameter θ lying in a low-dimensional space Θ (*e.g.*, $\Theta = \{(\mu, \sigma^2)\} \subset \mathbb{R}^2$). If some σ-finite measure $\nu(dx)$ dominates all the $P_\theta(dx)$, then the Radon-Nikodym derivative (or density) $L(\theta, x) = P_\theta(dx)/\nu(dx)$ is called the *likelihood function* and inference often proceeds either by

1. seeking the value of $\theta \in \Theta$ that maximizes $L(\theta, X)$, and studying its properties (the Frequentist approach); or by
2. specifying a "prior" probability measure $\pi(d\theta)$ on Θ, calculating the conditional "posterior" distribution $\pi(d\theta|X)$, and studying its properties (the Bayesian approach).

In the *nonparametric* approach no finite-dimensional Θ is postulated: all probability measures $\mu \in \mathcal{M}$ are regarded as possible distributions for X, and analysis proceeds either by

1. seeking the value of $\mu \in \mathcal{M}$ that maximizes some analogue of the likelihood like $\mu(dx)/\nu(dx)$ for a reference measure ν on $\mathcal{X}$, and studying its properties (the Frequentist approach); or by
2. specifying a "prior" probability measure on the possible distributions $\mu \in \mathcal{M}$, calculating the conditional "posterior" distribution, and studying its properties (the Bayesian approach). We can think of μ as a "random measure," first under a prior distribution and later under a posterior. We now turn to the study of random measures.

A *random measure* can be thought of in at least three different ways:

1. A function $\mu : \mathcal{B} \times \Omega \to \mathbb{R}$, mapping $(B, \omega) \mapsto \mu(B, \omega) \in \mathbb{R}$;
2. A function $\mu : \Omega \to \mathcal{M}$, mapping $\omega \mapsto \mu(\cdot, \omega) \in \mathcal{M}$;
3. A function $\mu : \mathcal{B} \to L^1(\Omega, \mathcal{F}, \mathsf{P})$, mapping $B \mapsto \mu(B, \cdot) \in L^1(\Omega, \mathcal{F}, \mathsf{P})$.

We omit the ω and denote the value by $\mu(B)$ in all three cases. The second perspective represents μ simply as a random variable, taking values in some abstract space $\mathcal{M}$; sometimes that's useful in technical arguments, but it is usually easier to think about random measures from the third perspective, as a family of ordinary random variables indexed by the Borel sets $B \in \mathcal{B}$.

Examples

Example 1: Wiener Measure

For *any* set T, any mean function $\mu : T \to \mathbb{R}$, and any real positive-definite covariance function $\rho : T \times T \to \mathbb{R}$, there exists a probability space $(\Omega, \mathcal{F}, \mathsf{P})$ and a Gaussian process X_t indexed by $t \in T$ with $\mathsf{E}[X_t] = \mu_t$ and $\mathsf{E}[(X_s - \mu_s)(X_t - \mu_t)] = \rho_{st}$. In particular we can take $T = \mathcal{B}$, the Borel sets in $\mathcal{X} = \mathbb{R}_+$; $\mu_B = 0$ for all $B \in \mathcal{B}$; and $\rho_{AB} = \lambda(A \cap B)$, the Lebesgue measure of the intersection. In this case the "cumulative distribution function" (or Stieltjes function) $W(t) = \mu\big((0, t]\big)$ associated with the random measure μ is just the standard Wiener process, and integrals $\int f(t)\mu(dt)$ of simple or even L^2 functions are just the same as Wiener integrals $\int f(t)\, dW_t$. The construction is not limited to $\mathbb{R}^1$, however, and just as easily leads to n-dimensional Gaussian measures and Wiener integrals, and the n-parameter analogue of the Wiener process sometimes called the "Brownian Sheet."

Example 2: Brownian Bridge

If we now take $T = \mathcal{B}((0,1])$, the Borel sets in the unit interval $\mathcal{X} = (0,1]$; $\mu_B = 0$ for all $B \in \mathcal{B}$; and $\rho_{AB} = \lambda(A \cap B) - \lambda(A)\lambda(B)$, the cumulative distribution function $B(t) = \mu\big((0,t]\big)$ is just the standard Brownian Bridge process, and integrals of simple or L^2 functions can be written in terms of Wiener integrals as $\int f(t)\mu(dt) = \int f(t)\,dW_t - W_1 \int f(t)\,dt$ for any Wiener process $W_t = B_t + tZ$, $Z \sim N(0,1)$ independent of B_t.

Example 3: The Gamma Process

Preliminaries: Gamma, Beta, and Dirichlet Distributions

If $X \sim \text{Ga}(\alpha, 1)$ and $Y \sim \text{Ga}(\beta, 1)$ are independent Gamma random variables, then X and Y have joint density function

$$f(x,y)\,dx\,dy = \frac{x^{\alpha-1}e^{-x}}{\Gamma(\alpha)}\;\frac{y^{\beta-1}e^{-y}}{\Gamma(\beta)}\;1_{\mathbb{R}_+}(x)\,1_{\mathbb{R}_+}(y)\;dx\,dy$$

so $W = X + Y$ and $Z = \frac{X}{X+Y}$ have joint distribution

$$\begin{aligned} f(w,z)\,dw\,dz &= \frac{(zw)^{\alpha-1}e^{-zw}}{\Gamma(\alpha)}\;\frac{((1-z)w)^{\beta-1}e^{-(1-z)w}}{\Gamma(\beta)}\;1_{\mathbb{R}_+}(w)\,w\,dw\,1_{[0,1]}(z)\,dz \\ &= \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}(z)^{\alpha-1}(1-z)^{\beta-1}\,1_{[0,1]}(z)\,dz\;\frac{w^{\alpha+\beta-1}e^{-w}}{\Gamma(\alpha+\beta)}\;1_{\mathbb{R}_+}(w)\,dw \end{aligned}$$

It follows that W and Z are independent with $\text{Ga}(\alpha+\beta, 1)$ and $\text{Be}(\alpha, \beta)$ distributions, respectively; thus the conditional distribution of X, given $X + Y = W$, is that of W times an independent $\text{Be}(\alpha, \beta)$ variable. We will need this for $\alpha = \beta = 1/2^n$.

The Construction

Let α be a σ-finite nonnegative measure on the space $(\mathcal{X}, \mathcal{B})$, and let $(\Omega, \mathcal{F}, \mathsf{P})$ be a probability space; the *Gamma Process with mean* α is a random measure $\nu : \mathcal{B} \times \Omega \to \mathbb{R}$ which assigns independent Gamma random variables $\nu(\Lambda_i) \sim \Gamma(\alpha_i, 1)$ to disjoint sets $\Lambda_i \in \mathcal{B}$ with finite measures $\alpha(\Lambda_i) = \alpha_i < \infty$. Here is an explicit construction of ν for $\mathcal{X} = \mathbb{R}_+$:

Let z_i^n be a doubly-indexed independent family of random variables with the Beta distribution $\text{Be}(\frac{1}{2^n}, \frac{1}{2^n})$; the z_i^0 are independent with uniform distributions, the z_i^1 have the $\text{Be}(1/2, 1/2)$, *etc.* Define a stochastic process X_t at integer times $t \in \mathbb{N}$ by $X_t = x_t^0$ where

$$x_t^0 = \sum_{i=1}^{t} -\log(z_i^0),$$

so X_t has independent increments $[X_t - X_s]$ with the $\Gamma(t-s, 1)$ distribution for integers s, t. For successive n define X_t at dyadic rational times recursively by $X_t = x_i^n$, $t = i/2^n$, where

$$\begin{aligned} x_{2i}^{n+1} &= x_i^n \\ x_{2i+1}^{n+1} &= x_i^n + (x_{i+1}^n - x_i^n)z_i^{n+1} \end{aligned}$$

This defines X_t for all dyadic rational t; by our preliminary observation above, the increments $[X_t - X_s]$ are independent with the $\text{Ga}(t-s, 1)$ distribution. The process X_t is nonnegative and nondecreasing, so we can extend the definition to all of $\mathbb{R}_+$ by requiring right-continuity: $X_t = \inf[x_i^n : t \le i/2^n]$. We will see below that right-continuity is the best we can hope for, *i.e.*, that the process X_t does *not* have continuous sample paths (in fact, it has infinitely many jumps in every open interval $(t, t+\epsilon)$ almost surely!) For both rational and irrational $s < t$, the increments $[X_t - X_s]$ are independent with the Gamma $\Gamma\big((t-s), 1\big)$ distributions, and hence with finite means $\mathsf{E}[X_t - X_s] = (t-s)$ and variances $\mathsf{V}[X_t - X_s] = (t-s)$.

Given any σ-finite measure α on $\mathbb{R}_+$, define a right-continuous function $A(x) = \alpha\big((0,x]\big)$ and a random measure ν by:

$$\nu\big((s,t]\big) = X_{A(t)} - X_{A(s)}$$

for the standard Gamma process X_t defined above; we extend by additivity to the field generated by the half-open intervals $(s,t]$, and by continuity to the Borel sets with finite α-measure, upon noting that

$$\begin{aligned} \mathsf{E}\nu\big((s,t]\big) &= [A(t) - A(s)] = \alpha\big((s,t]\big) \\ \mathsf{V}\nu\big((s,t]\big) &= [A(t) - A(s)] = \alpha\big((s,t]\big), \end{aligned}$$

so (by L^2 continuity) $\mathsf{E}[\nu(B)] = \mathsf{V}[\nu(B)] = \alpha(B)$ for all $B \in \mathcal{B}$. We will see below that, almost surely, ν is a discrete measure concentrated on a (random) countable set of points $\tau_i(\omega)$.

Example 4: The Dirichlet Process

Now let α be a *finite* nonnegative measure on $(\mathcal{X}, \mathcal{B})$ and let $\nu(dx)$ be a Gamma process random measure with mean $\mathsf{E}[\nu(dx)] = \alpha(dx)$; since $\alpha(\mathbb{R}_+) < \infty$, $\nu(\mathbb{R}_+)$ is a well-defined random variable and we can construct

$$\mu(A) = \frac{\nu(A)}{\nu(\mathbb{R}_+)}$$

for all $A \in \mathcal{B}$. Each random variable $\mu(A)$ has a Beta $\mathrm{Be}(\alpha(A), \alpha(A^c))$ distribution, and for any partition Λ_i of $\mathcal{X}$ into n disjoint sets, the n-variate random variables $X_i = \mu(\Lambda_i)$ have the Dirichlet $D(\alpha_1, \cdots, \alpha_n)$ distribution with parameters $\alpha_i = \alpha(\Lambda_i)$. Just as the Gamma process ν was almost-surely concentrated on a countable set of points τ_i, so too is the Dirichlet process μ... in fact, it is the *same* set $\{\tau_i\}$! The Dirichlet process is, almost surely, a discrete distribution.

The Dirichlet Process is an important example, because of its use in nonparametric Bayesian statistics. The principal result is this:

Theorem. *Let $\mu \sim \mathrm{Dir}(\alpha_o)$ for some finite measure α_o and let $X_1, X_2, \ldots, X_n$ be independent observations all with distribution μ_ω. Then, conditional on $X_1, \ldots, X_n$, $\mu \sim \mathrm{Dir}(\alpha_n)$ for the measure $\alpha_n(dx) = \alpha_o(dx) + \sum_{i=1}^n \delta(x - X_i)\,dx$ equal to $\alpha_o(dx)$ plus a unit point mass at each observed X_i.*

Corollary. *Under the same conditions, the predictive distribution for X_{n+1} assigns mass $\frac{1}{n+\alpha(\mathbb{R})}$ to $x = X_i$ for each $1 \le i \le n$ and the rest of the mass $\frac{\alpha(\mathbb{R})}{n+\alpha(\mathbb{R})}$ to the prior mean, $\frac{\alpha(dx)}{\alpha(\mathbb{R})}$.*

Note that, from the corollary, the probability of a *tie* among the first n variables is

$$1 - \prod_{i=1}^{n} \left(\frac{\alpha(\mathbb{R})}{\alpha(\mathbb{R}) + i - 1} \right) = 1 - \frac{\alpha(\mathbb{R})^n \Gamma\big(\alpha(\mathbb{R})\big)}{\Gamma\big(\alpha(\mathbb{R}) + n\big)},$$

arbitrarily close to 1 for large enough n; if the X_i "really" come from any continuous distribution, no ties will be observed no matter how large n might be. If $f(x)$ is *any density at all* and $\epsilon > 0$, the posterior for a Bayesian prior giving probability ϵ to $\mu_X(dx) = f(x)\,dx$ and $1-\epsilon$ to $\mu \sim \mathrm{Dir}(\alpha_o)$ will eventually be concentrated on $f(x)$. This proves that Bayesian analysis can be inconsistent. See me for more references if you're interested in this point.

Path Discontinuity and SII Processes

A celebrated theorem of Lèvy and Khinchine asserts that every stationary, independent increment (SII) process X_t has a characteristic function of the form:

$$\mathsf{E}[e^{i\lambda X_t}] = e^{i\lambda x_0 + it\lambda m - t\frac{\sigma^2\lambda^2}{2} + t\int_{\mathbb{R}} \left(e^{i\lambda u} - 1\right)\nu(du)}$$

for some initial value x_0, drift m, diffusion constant σ^2, and jump rate ("Lèvy") measure ν. If $\nu(du) \equiv 0$ then X_t is simply Brownian motion with drift, $X_t = x_0 + mt + \sigma W_t$; $\nu(E)$ is the rate at which the process jumps by amounts $u \in E$. The total jump rate $\nu(\mathbb{R})$ need not be *finite*, but ν must satisfy $\int [1 \wedge |u|]\nu(du) < \infty$; it's OK to have infinitely many tiny jumps if they're small enough to have a finite sum, almost surely. If $\mu(\mathbb{R}) < \infty$ then we can interpret the process as one with exponentially distributed waiting times (with means $1/\mu(\mathbb{R})$) between successive jumps, which are randomly distributed with distribution $\mu(du)/\mu(\mathbb{R})$. With a little more work it's possible to make sense of processes with jump measures satisfying only the weaker condition $\int [1 \wedge u^2]\nu(du) < \infty$, but the argument gets more subtle. Ask if you need references.

A standard Poisson process $N(t)$, for example, has characteristic function

$$\begin{aligned}\mathsf{E}[e^{i\lambda X_t}] &= \sum_{k=0}^{\infty} e^{-t}\frac{t^k}{k!}e^{i\lambda k} \\ &= e^{t(e^{i\lambda}-1)},\end{aligned}$$

corresponding to $x_0 = 0$, $m = 0$, $\sigma^2 = 0$, and $\nu(du) = \delta(u-1)\,du$. A *generalized Poisson process* is a sum $X_t = \sum_i u_i N_i(r_i t)$ of re-scaled independent Poissons, which takes jumps of size u_i at rate r_i; it has measure $\nu(du) = \sum_i r_i \delta(u_i - u)\,du$. Any SII process can be approximated by the sum of Brownian motion with drift and a generalized Poisson Process. In particular, its paths will be continuous if and only if it is Brownian, and if not we can find the rate of jumps by identifying the Lèvy measure ν.

For example, the characteristic function of the standard Gamma Process $X_t \sim \mathrm{Ga}(t, 1)$ is $\mathsf{E}[e^{i\lambda X_t} = (1 - i\lambda)^{-t}$; it has no drift or diffusion part, and has Lèvy measure $\nu(du)$ satisfying

$$-t\log(1 - i\lambda) = t\int_{\mathbb{R}} \left(e^{iu\lambda} - 1\right)\nu(du)$$

or, after differentiating with respect to λ,

$$\frac{it}{1 - i\lambda} = t\int_{\mathbb{R}} iue^{iu\lambda}\,\nu(du)$$

But $it\int_0^\infty e^{-u(1-i\lambda)}du = it(1 - i\lambda)^{-1}$, so

$$\begin{aligned} ite^{-u}1_{(0,\infty)}(u)\,du &= tiu\,\nu(du) \\ \nu(du) &= u^{-1}e^{-u}\,du \qquad\qquad (u > 0). \end{aligned}$$

This is not a finite measure, so the Gamma process jumps infinitely often in every time interval; the rate of jumps bigger than ϵ is $\int_\epsilon^\infty e^{-u}/u\,du$, finite for every $\epsilon > 0$, and the mean sum of all jumps in time t is $t\int_0^\infty ue^{-u}/u\,du = t$.

As interesting exercises, try to find:

1. The Lèvy measure $\nu(du)$ for the Cauchy process with characteristic function $\mathsf{E}[e^{i\lambda X_t}] = e^{-t|\lambda|}$;
2. The joint distribution for the largest jump of the Gamma process $X_t \sim \mathrm{Ga}(\alpha(dt))$ in the time interval $(0, t]$ and the time τ at which it occurs (hint: do $\alpha(dt) = dt$ first);
3. The joint distribution for the largest jump of the Dirichlet process $X_t \sim \mathrm{Dir}(\alpha(dt))$ in the time interval $(0, t]$, and the time τ at which it occurs.