MA8109 Stochastic processes and differential equations

Fall 2015 – Synopsis

This note is intended to provide a *synopsis* of the course: What has been covered, basic definitions and important results, etc.

The note will keep on growing as the lectures move ahead. Ideally, a new version will be posted every week.

Notation

Here I summarize some notation used throughout.

- $\mathbb{N}$, $\mathbb{Z}$, $\mathbb{Q}$, $\mathbb{R}$, $\mathbb{C}$ are the sets of natural numbers (starting at 1), integers, rational numbers, real numbers, and complex numbers respectively. Also, $\mathbb{N}_0 = \{0\} \cup \mathbb{N}$, and $\overline{\mathbb{R}} = [-\infty, \infty]$.
- I write $\underline{\lim}$ and $\overline{\lim}$ instead of the more common $\liminf$ and $\limsup$.
- $\mathcal{B}$ is the σ-algebra of *Borel sets* on $\mathbb{R}$ or $\overline{\mathbb{R}}$ (depending on context).
- I use := to mean "is defined as", and =: if the term being defined is on the right.
- A^{c} is the *complement* $\Omega \setminus A$. The "universal" set Ω needs to be understood.
- $A \sqcup B$ is the union $A \cup B$ of two *disjoint* sets A and B.
- $\bigsqcup_{n=1}^{\infty} A_n$ is the union of a sequence of *pairwise disjoint* sets.
- Y^X, where X and Y are sets, is the set of functions $X \to Y$.
- As a special case, $Y^{\mathbb{N}}$ is the set of all sequences $(y_1, y_2, \ldots)$ in Y.
- $[S]$ equals 1 if the statement S is true, 0 otherwise (*indicator bracket*).
- $[A]$ is the *indicator function* of the set A, defined by $[A](x) = [x \in A]$.
- If $a \in \overline{\mathbb{R}}$ we write $a^+ := \max(a, 0)$ and $a^- := (-a)^+ = -\min(a, 0)$. Then $a^\pm \geq 0$, $a^+ a^- = 0$, $a = a^+ - a^-$, and $|a| = a^+ + a^-$.
- If f is a function, define $f^\pm$ by $f^\pm(x) = f(x)^\pm$.

First week (W34)

A recurring example is **coin tossing space** $\Omega = \{0,1\}^{\mathbb{N}}$, consisting of all infinite sequences of zeroes and ones, representing coin tosses (zero for tails, one for head) if you wish.

An *algebra on* Ω, (or perhaps more precisely, an algebra of subsets of Ω) is a set $\mathcal{A}$ of subsets of Ω so that

- $\emptyset \in \mathcal{A}$
- $A \in \mathcal{A}$ implies $A^{\mathsf{c}} \in \mathcal{A}$
- $A, B \in \mathcal{A}$ implies $A \cup B \in sA$

For each $n \in \mathbb{N}$, there is an algebra $\mathcal{F}_n$ of subsets of Ω, defined as the events *determined by* $(\omega_1,\dots,\omega_n)$: Thus $A \in \mathcal{F}$ if and only if whenever $\omega \in A$ and $\omega' \in \Omega$ $\omega_k = \omega'_k$ for $k = 1,\dots,n$ implies $\omega' \in A$. Or put differently, if $\pi_n\colon \Omega \to \{0,1\}^n$ is the projection map onto the first n coordinates, the members of $\mathcal{F}_n$ are the inverse images of sets $B \subseteq \{0,1\}^n$. Thus $\mathcal{F}_n$ has 2^{2^n} members.

If we think of independent coin tosses with an unbiased coin, elementary probability theory dictates a probability $P(\pi_n^{-1}(B)) = 2^{-n}\#B$ when $B \subseteq \{0,1\}^n$ (here $\#B$ is the number of members of B).

The algebras $\mathcal{F}_n$ form an increasing sequence of algebras, and so their union

$$\mathcal{F}_* := \bigcup_{n=1}^{\infty} \mathcal{F}_n$$

is an algebra too: It consists of all *finitely determined* events.

The *strong law of large numbers* implies that

$$\lim_{n\to\infty} \frac{1}{n}\sum_{k=1}^{n} \omega_k = \frac{1}{2} \quad \text{a. s.},$$

where "a. s." stands for "*almost surely*", meaning "with probability 1".

Note that we are unable to even give this statement a precise meaning within our current framework so far, since it is a statement regarding an event not in $\mathcal{F}_*$ (worse, it is utterly independent of *any finite* number of cointosses ω_k).

Our next task is to remedy this.

σ-algebras and measures

1 Definition. A *σ-algebra on Ω* (or perhaps more precisely, a σ-algebra of subsets of Ω) is a set $\mathcal{F}$ of subsets of Ω so that

- $\emptyset \in \mathcal{F}$
- $A \in \mathcal{F}$ implies $A^c \in \mathcal{F}$
- $A_n \in \mathcal{F}$ for $n = 1, 2, \ldots$ implies $\bigcup_{k=1}^{\infty} A_k \in \mathcal{F}$

Because any intersection of σ-algebras is itself a σ-algebra, there exists a smallest σ-algebra $\mathcal{F} := \sigma(\mathcal{F}_*)$ containing $\mathcal{F}_*$, called the σ-algebra *generated by* $\mathcal{F}_*$.

We want to extend P to a *probability measure* on $\mathcal{F}$.

2 Definition. A *measure* on $\mathcal{F}$ is a map $\mu\colon \mathcal{F} \to [0, \infty]$ satisfying

– $\mu(\emptyset) = 0$

– $A_n \in \mathcal{F}$ pairwise disjoint for $n \in \mathbb{N}$ implies $\mu\Big(\bigsqcup_{n=1}^{\infty} A_n\Big) = \sum_{n=1}^{\infty} \mu(A_n)$

3 Definition.

– A *measurable space* is a pair $(\Omega, \mathcal{F})$ where Ω is a set and $\mathcal{F}$ a σ-algebra on Ω.

– A *measure space* is a triple $(\Omega, \mathcal{F}, \mu)$ where $(\Omega, \mathcal{F})$ is a measurable space and μ a measure on $\mathcal{F}$.

– A *probability space* is a measure space $(\Omega, \mathcal{F},)$ where P is a probability measure.

4 Definition. A *monotone class* is a set $\mathcal{M}$ of subsets of Ω satisfying

– If $A_n \in \mathcal{M}$ and $A_n \subseteq A_{n+1}$ for all $n \in \mathbb{N}$, then $\bigcup_{n=1}^{\infty} A_n \in \mathcal{M}$,

– If $A_n \in \mathcal{M}$ and $A_n \supseteq A_{n+1}$ for all $n \in \mathbb{N}$, then $\bigcap_{n=1}^{\infty} A_n \in \mathcal{M}$.

5 Lemma (Monotone Class Lemma) *If $\mathcal{A}$ is an algebra on Ω and $\mathcal{M}$ is a montone class with $\mathcal{A} \subseteq \mathcal{M}$, then $\mathcal{A} \subseteq \sigma(\mathcal{M})$.*

From this we get

6 Theorem (Uniqueness of extension) *Let $\mathcal{A}$ be an algebra. Any two finite measures which agree on all members of $\mathcal{A}$, also agree on all members of $\sigma(\mathcal{A})$.*

Returning to cointossing space $(\Omega, \mathcal{F})$ with $\Omega = \{0, 1\}^{\mathbb{N}}$, we conclude that there cannot be more than one probability measure on this space extending the function P defined previously on $\mathcal{F}_*$.

That there in fact *exists* such a measure is non-trivial, but true. Thanks to the uniqueness theorem, we do not need worry too much about which of several possible methods of construction we use; they must all produce the same measure.

Lebesgue measure

This is another measure of great importance. It is defined on the σ-algebra $\mathcal{B}$ of *Borel subsets* of $\mathbb{R}$, which is the σ-algebra generated by the set of intervals (or equivalently, open intervals – or closed intervals – or half open intervals $(a,b]$ – or open sets – or closed sets – or ...). We shall write λ for Lebesgue measure. It is the unique Borel measure (meaning a measure on $\mathcal{B}$) so that $\lambda((a,b]) = b - a$ for all $a \le b$. (These do not form an algebra, so the uniqueness theorem does not apply directly – but the set of all finite unions of such integrals does, if we also include intervals of the form $(-\infty, a]$ and (a, ∞).)

Second week (W35)

7 Definition. A *measurable function* on a measurable space $(\Omega, \mathcal{F})$ is a function $f: \Omega \to \overline{\mathbb{R}}$ so that $f^{-1}(-\infty, a[)-\infty, a] \in \mathcal{F}$ for all $a \in \overline{\mathbb{R}}$. (Then $f^{-1}(B) \in \mathcal{F}$ for all Borel sets B, because the sets B satisfying the condition is a σ-algebra.)

A *random variable* (R. V.) on a probability space $(\Omega, \mathcal{F}, P)$ is a measurable function on $(\Omega, \mathcal{F})$. (We usually use uppercase letters such as X for random variables.)

8 Lemma *If a sequence of measurable functions converges pointwise to some limit, then the limit is measurable.*

We can now define a random variable U on coin tossing space:

$$U(\omega) = \sum_{n=1}^{\infty} \omega_n 2^{-n}.$$

Think of it as using the coin tosses ω_n as the digits in the binary expansion of $U(\omega) \in [0,1]$.

We write $P(U \le u)$ as shorthand notation for $P(\{\omega \in \Omega\colon U(\omega) \le u\})$.

It turns out that $P(U \le u) = u$ for all $u \in [0,1]$. (Easily proved for *dyadic rational* u, that is, $u = m/2^k$ for integers m, k; then it follows for all u, beacuse $P(U \le u) = u$ is a monotone function of u.) In other words, U is *uniformly distributed* on the interval $[0,1]$. We shall call such a random variable a *standard uniform variable.* From it, we can build random variables of any desired distribution.

9 Definition. The *distribution* of a random variable X on $(\Omega, \mathcal{F}, P)$ is the Borel measure μ_X given by

$$\mu_x(B) = P(X \in B) = P(X^{-1}(B)).$$

It is uniquely determined by the *cumulative distribution function*

$$F_X(x) = \mu_X([-\infty, x]) = P(X \le x).$$

In particular, the distribution of a standard uniform variable U is Lebesgue measure on $[0,1]$:

$$\mu_U(B) = \lambda(B \cap [0,1]) \qquad (B \in \mathcal{B}).$$

Integration

We define the integral for certain *measurable* functions f on a measure space $(\Omega, \mathcal{F}, \mu)$:

10 Definition. A *simple function* is a measurable function which takes only a finite number of values. Such a function can be written

$$\varphi = \sum_{k=1}^{n} a_k \, [A_k]$$

with $a_k \in \overline{\mathbb{R}}$ and $A_k \in \mathcal{F}$. We can always choose the a_k to be distinct and nonzero and the A_k to be nonempty and mutually disjoint. This may be called the *canonical* representation of φ. It is unique up to permutation of the indices. For a non-canonical representation, we must take care not to subtract infinities. So we disallow $a_j = -\infty$, $a_k = \infty$ and $A_j \cap A_k \neq \emptyset$.

11 Definition. The *integral of a simple function* such as above is

$$\int_\Omega \varphi \, d\mu = \sum_{k=1}^{n} a_k \mu(A_k).$$

Here and elsewhere we use the convention that $0 \cdot (\pm\infty) = 0$. If the sum contains terms equal to both $-\infty$ and $+\infty$, we do not define the integral of φ. Note that the integral of a *nonnegative* simple function is *always* defined. Its value may be ∞.

12 Definition. The *integral of a nonnegative measurable function* f is

$$\int_\Omega f \, d\mu = \sup\Bigl\{ \int_\Omega \varphi \, d\mu \colon \varphi \text{ is simple and } 0 \le \varphi \le f \Bigr\}.$$

13 Theorem (The Monotone Convergence Theorem (MCT)) *If f_n is a measurable function and $0 \le f_n \le f_{n+1}$ for $n \in \mathbb{N}$ then*

$$\int_\Omega \lim_{n\to\infty} f_n \, d\mu = \lim_{n\to\infty} \int_\Omega f_n \, d\mu.$$

(Note that both limits exist by monotonicity, and the limit function on the left is measurable.)

14 Theorem (Fatou's lemma) *If $f_n \ge 0$ is measurable for all $n \in \mathbb{N}$ then*

$$\int_\Omega \underline{\lim}_{n\to\infty} f_n \, d\mu \le \underline{\lim}_{n\to\infty} \int_\Omega f_n \, d\mu.$$

After showing that any nonnegative measurable function is a pointwise limit of a non-decreasing sequence of nonnegative simple function, we have no difficulty using MCT to show that the integral is *additive*, and in the end, we get an integral that is linear, given by

15 Definition. The *integral of a measurable function f* is defined to be

$$\int_\Omega f\,d\mu = \int_\Omega f^+\,d\mu - \int_\Omega f^-\,d\mu.$$

If both integrals on the right have infinite value, we do not define the integral. If they are both finite, we call f *integrable*.

16 Theorem (The Dominated Convergence Theorem (DCT)) *If f_n is measurable and $|f_n| \le g$ for all $n \in \mathbb{N}$ where g is integrable, and if the sequence converges pointwise, then*

$$\int_\Omega \lim_{n\to\infty} f_n\,d\mu = \lim_{n\to\infty}\int_\Omega f_n\,d\mu.$$

The Riemann integral (or the Darboux integral – the two are equivalent, even though they are constucted in slightly different ways) is the integral you learned in basic calculus based on Riemann sums. Any Riemann integrable function is Lebesgue integrable, and the Riemann integral equals the Lebesgue integral. (I am sure you are much relieved.) But it is trivial to find Lebesgue integrable functions which are not Riemann integrable: The indicator function $[\mathbb{Q}]$ of the rational numbers is one example. Note that $\mathbb{Q}$ is countable, and so $\lambda(\mathbb{Q}) = 0$, since the Lebesgue measure of any singleton set is zero. But $[\mathbb{Q}]$ is discontinuous everywhere, whereas Riemann integrable functions are continuous *almost everywhere* (i.e., except on a set of measure zero).

Third week (W36)

The expectation of a random variable $X\colon \Omega \to \overline{\mathbb{R}}$ is simply its integral with respect to probability measure:

$$E[X] := \int_\Omega X\,dP.$$

If $g\colon \overline{\mathbb{R}} \to \overline{\mathbb{R}}$ is any Borel measurable function, then $g(X)$ is another random variable. (Strictly speaking, we should write it as a function composition $g \circ X$, since we are really talking about the function $\omega \mapsto X(g(\omega))$, but common convention suggest hiding ω as much as possible.)

Recalling the definition of the *distribution* μ_X of X, we find

$$E[g(X)] = \int_{\overline{\mathbb{R}}} g\,d\mu_X.$$

(This is almost trivial when g is a simple function, and the general case follows by the bootstrapping procedure, noting that a nonnegative measurable g is the limit of an increasing sequence of simple functions and employing MCT.)

Thus we recover the usual formula from elementary probability.

We can create a random variable X with any given distribution μ, and corresponding cumulative distribution $F(x) := \mu\big([-\infty, x]\big)$ by letting U be a standard uniform variable and setting

$$X(\omega) = \tilde{F}\big(U(\omega)\big), \qquad \tilde{F}(u) = \min\{x \in \overline{\mathbb{R}}\colon F(x) \ge u\}.$$

In particular, we can create a standard Gaussian variable in this way, using the standard Gaussian density function φ and cumulative distribution Φ:

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}, \quad \Phi(x) = \int_{-\infty}^{x} \varphi(t)\,dt.$$

Note that if X is a standard Gaussian variable, then $\sigma X + \mu$ is a Gaussian variable with variance σ^2 and expectation (mean) μ.

Stochastic independence: We generalize the notion of independence from events: algebras $\mathcal{A}_k$ with $k = 1, \ldots, n$ are called *independent* when

$$P\Big(\bigcap_{k=1}^{n} A_k\Big) = \prod_{k=1}^{n} P(A_k) \qquad \text{whenever } A_k \in \mathcal{A}_k \text{ for } k = 1, \ldots, n,$$

and an infinite collection of algebras is called independent if every finite subcollection is independent.

The σ-algebra generated by a random variable X is the set of events $\{X \in B\} = X^{-1}(B)$ where $B \subseteq \overline{\mathbb{R}}$ is a Borel set. A collection of random variables is called independent if the σ-algebras they generate are independent.

In coin tossing space, all the algebras corresponding to a single coin toss, $\mathcal{F}_n = \{\emptyset, \{\omega_n = 0\}, \{\omega_n = 1\}, \Omega\}$, are independent by construction.

With a bit of work, we can also conclude that the σ-algebras corresponding to disjoint sets of coin tosses, such as

$$\sigma\Big(\bigcup_{k=1}^{\infty} \mathcal{F}_{(2k-1)2^n}\Big), \qquad \text{with } n = 0, 1, 2, \ldots,$$

are independent. It follows that the random variables

$$U_n := \sum_{k=1}^{\infty} \omega_{(2k-1)2^n} 2^{-n}$$

are independent.

In proving the above, the following is useful:

17 Lemma *If algebras $(\mathcal{A}_i)_{i\in I}$ are independent, then the generated σ-algebras $\sigma(\mathcal{A}_i)$ are also independent.*

The proof is by showing that you can replace the $\mathcal{A}_i$ by $\sigma(\mathcal{A}_i)$ one by one without destroying independence, by noting that the set of sets A which are independent of all the $\mathcal{A}_j$ for $j \neq i$ is a monotone class, and using the monotone class lemma. Since the main condition for independence involves only a finite number of algebras at a time, this is sufficient.

Characteristic functions

The *characteristic function* of a stochastic variable $X\colon \Omega \to \mathbb{R}^n$ is the function of $\xi \in \mathbb{R}^n$ given by the expectation $E[e^{i\xi\cdot X}]$ where $\cdot$ denotes the ordinary scalar product. We calculate

$$E[e^{i\xi\cdot X}] = \int_\Omega e^{i\xi\cdot X} = \int_{\mathbb{R}^n} e^{i\xi\cdot x}\, d\mu_X(x) = \hat{\mu}_X(\xi),$$

where $\hat{\mu}_X$ is the Fourier transform of μ_X.

The conventions for Fourier transforms vary, of course – here we have chosen to drop the factor $(2\pi)^{-n/2}$ that is commonly included, and we also use the plus sign in the exponent where a minus sign is quite common. But the present definition matches the conventional definition of characteristic function.

From the theory of distributions (in analysis, not probability – also called generalized functions) we can learn the important fact that *two distributions with the same characteristic function are in fact identical.*

Differentiating under the integral sign yields important formulas like

$$E[X_j] = \frac{\partial}{\partial \xi_j} \hat{\mu}_X(0)$$

and higher analogues such as

$$E[X_j X_k] = \frac{\partial^2}{\partial\xi_j \partial\xi_k} \hat{\mu}_X(0)$$

and so on.

These can be proved directly from the definition of derivative, using DCT – provided that X_j and $X_j X_k$ are integrable.

Gaussian families

(Note: We stick to Gaussian variables with expectation zero for now.)

The characteristic function of a single standard Gaussian is

$$\hat{\mu}_N(\xi) = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} e^{-x^2/2+i\xi}\,dx = \frac{e^{-\xi^2/2}}{\sqrt{2\pi}} \int_{\mathbb{R}} e^{-(x-i\xi)^2/2}\,dx = e^{-\xi^2/2}.$$

The second integral above can be evaluated by using Cauchy's integral theorem around a rectangular contour with corners at $\pm M$ and $\pm M - i\xi$ and letting $M \to \infty$.

We generalize this to m linear combinations of n independent standard Gaussians N_k:

$$X_j = \sum_{k=1}^{n} a_{jk} N_k$$

which we can write as a matrix equation $X = AN$ where the X_j form a column vector X, the N_k form a column vector N, and A is an $m \times n$ matrix with real entries. We calculate

$$\hat{\mu}_X(\xi) = E(e^{i\xi^\top AN}) = e^{-\xi^\top AA^\top \xi/2} = e^{-\xi^\top C\xi/2},$$

where $C = AA^\top$ is called the *covariance matrix*, since its j,k entry is in fact $E(X_j X_k)$ (as is seen by differentiating with respect to ξ_j and ξ_k).

Let us *define* that a variable $X\colon \Omega \to \mathbb{R}^n$ is *Gaussian* with covariance matrix C if its characteristic function is the one above. Here C is symmetric and non-negative definite, which means $\xi^\top C\xi \ge 0$ for all $\xi \in \mathbb{R}^n$.

Next, a possibly infinite collection of random variables is called a *Gaussian family* (with expectation zero) if any finite collection of them forms a Gaussian n-dimensional variable.

The linear span of a Gaussian family is again a Gaussian family. And if you wish to include variables with a non-zero expectation, just throw the constant functions into the mix and take more linear combinations.

Gaussian families and Hilbert spaces

$L^2(\Omega, \mathcal{F}, \mu)$ is the set of *square integrable* functions, which in terms of expectations means that $X \in L^2$ if and only if $E(X^2) < \infty$.

If $X, Y \in L^2$ then XY is integrable too, and the *Cauchy–Schwarz* inequality holds:

$$|E(XY)| \le E(X^2)^{1/2} E(Y^2)^{1/2}.$$

We define the L^2 norm $\|\cdot\|_2$ and inner product $\langle\cdot,\cdot\rangle$ by

$$\|X\|_2 = E(X^2)^{1/2}, \quad \langle X, Y\rangle = E(XY), \qquad X, Y \in L^2.$$

It should be clear that this defines a real inner product space. Less obvious, but still true, is that it is *complete*, so it is in fact a *real Hilbert space*.

You may be more familiar with the theory of complex Hilbert spaces. Real Hilbert space theory is mostly the same, except that you don't need to worry about complex conjugation.

There is one small problem, though: The axioms of normed spaces require that $\|X\| = 0$ only if $X = 0$. But $\|X\|_2 = 0$ only yields $X = 0$ *almost surely*. So to really get a proper normed space, we need to consider the elements of the space to be equivalence classes of random variables, where X and Y are considered equivalent if $X = Y$ a.s.

Clearly, any Gaussian family is contained in L^2. It turns out that n members of a Gaussian family are independent if and only if they are *mutually orthogonal*. This is remarkable because pairwise independence does not imply independence of n variables in general, but in a Gaussian family this implication does hold. Also, it allows us to bring the whole Hilbert space theory with orthogonal projections, etc., to bear on problems in Gaussian families.

Brownian motion

A *stochastic process* is just a family $(X_t)_{t\in T}$ of stochastic variables, where T can be any set.

In practice for us, T will usually be the interval $[0,\infty)$ or an initial segment of that interval. But in many applications such as spatially distributed random fields, T will be a subset of $\mathbb{R}^n$ instead.

The process is called *Gaussian* if the variables form a Gaussian family.

Brownian motion is a Gaussian stochastic process $(B_t)_{t\geq 0}$ with expectation zero and stationary, independent increments, and normalized so that $E(B_1^2) = 1$.

Equivalently (and this we shall adopt as the definition): It is a Gaussian process with expectation zero satisfying

$$E(B_s B_t) = s \wedge t$$

for all $s, t \geq 0$. Here $s \wedge t := \min\{s, t\}$.

We can *construct Brownian motion* on coin tossing space by starting with a countably infinite collection of independent standard Gaussian variables on this space, indexed as $N_{k,n}$.

To make a long story short, we begin by setting

$$B_n = \sum_{j=1}^{n} N_{0,j}$$

and noting that these do satisfy the requirements of a Brownian motion restricted to integer t.

By induction, assume we have defined $B_{n/2^k}$ for all $n \in \mathbb{N}_0$ and some $k \in \mathbb{N}_0$, then we interpolate and add some randomness to the middle points:

$$B_{(2n+1)/2^{k+1}} = \frac{1}{2}(B_{n/2^k} + B_{(n+1)/2^k}) + 2^{-(k+2)/2} N_{k+1,n}$$

The motivation for this is a small bit of Hilbert space geometry.

Finally, we define $B_{k,t}$ by setting $B_{k,n/2^k} = B_{n/2^k}$ and interpolating linearly between these points, and we take the limit as $k \to \infty$.

The result is not only Brownian motion as defined above; but also, the above series will almost surely converge uniformly on bounded intervals, so that the limit function is continuous.

In summary, *this version of Brownian motion has continous paths* (almost surely).

This construction is known as Lévy's construction. But it is less well known than it deserves to be.

Fourth week (W37)

We finished the Lévy construction of Brownian motion. Along the way, we used

18 Lemma (Borel–Cantelli) *If* (A_n) *is a sequence of events with* $\sum_{n=1}^{\infty}$*, then* $P(A_n$ *i.o.*$) = 0$.

Here "i.o." stands for "infinitely often", and the event in question is

$$\bigcap_{n=1}^{\infty}\bigcup_{k=n}^{\infty} A_k.$$

The proof consists of noting that $P\big(\bigcup_{k=n}^{\infty} A_k\big) \le \sum_{k=n}^{\infty} P(A_k) \to 0$ when $n\to\infty$, because of the assumed convergence.

Here is a useful *scaling law* for standard Brownian motion: If $(B_t)_{t\ge0}$ is a standard Brownian motion and $\tilde{B}_t = \sqrt{a}B_{t/a}$ where $a > 0$ is a constant, then $(\tilde{B}_t)_{t\ge0}$ is a standard Brownian motion as well.

We also have a simple *restarting law*: If $(B_t)_{t\ge0}$ is a standard Brownian motion and $t_0 > 0$ is fixed, then $\tilde{B}_t = B_{t-t_0} - B_{t_0}$ defines another standard Brownian motion.

Quadratic variation

To begin with, note that if $0 = t_0 < t_1 < t_2 < \cdots < t_n = t$, then (from a fairly trivial calculation)

$$E\Big(\sum_{k=0}^{n-1}(B_{t_{k+1}} - B_{t_k})^2\Big) = t.$$

A bit more work shows that in fact

$$\sum_{k=0}^{n-1}(B_{t_{k+1}} - B_{t_k})^2 \to t$$

in L^2 norm as the mesh size of the partition goes to zero, and so the above convergence holds a.s. for *some* sequence of partitions with mesh size going to zero.

We may define the quadratic variation of a function $f\colon [a,b]\to\mathbb{R}$ as

$$\mathrm{QV}(f;[a,b]) = \overline{\lim}\sum_{k=0}^{n-1}\big(f(t_{k+1}) - f(t_k)^2\big)$$

where the limit superior is defined by taking the supremum over all partitions with mesh size $< \delta$ and then taking the limit $\delta\to0$.

Then it follows that for Brownian motion, $\mathrm{QV}(B_t;[0,t]) \ge t$ a.s.

This is in stark contrast to functions of *bounded* (linear) *variation*, which have zero quadratic variation. In particular, differentiable functions do have bounded variation, so the paths of Brownian motion are almost surely nowhere differentiable.

First steps toward the Itô integral: An example

We start out very naïvely, trying to make sense of the integral $\int_0^t B_s\,dB_s$. Remember that we found

$$\sum_{k=0}^{n-1}(B_{t_{k+1}}-B_{t_k})^2 \to t$$

in L^2 norm, i.e.,

$$\sum_{k=0}^{n-1} B_{t_{k+1}}(B_{t_{k+1}}-B_{t_k}) - \sum_{k=0}^{n-1} B_{t_k}(B_{t_{k+1}}-B_{t_k}) \to t$$

However, *both* sums on the left are reasonable candidates for an approximation to $\int_0^t B_s\,dB_s$!

The first sum is a *Stratonovich* sum, and can be used to define the Stratonovich integral. The second sumis an *Itô* sum, and can be used to define the Itô integral.

We can easily evaluate the sum of the two sums:

$$\sum_{k=0}^{n-1} B_{t_{k+1}}(B_{t_{k+1}}+B_{t_k}) - \sum_{k=0}^{n-1} B_{t_k}(B_{t_{k+1}}-B_{t_k}) = \sum_{k=0}^{n-1}\left(B_{t_{k+1}}^2 - B_{t_k}^2\right) = B_t^2 - B_0^2 = B_t^2.$$

And so we find that

$$\sum_{k=0}^{n-1} B_{t_{k+1}}(B_{t_{k+1}}-B_{t_k}) \to \tfrac{1}{2}(B_t^2+t) \quad \text{(Stratonovich)},$$

$$\sum_{k=0}^{n-1} B_{t_k}(B_{t_{k+1}}-B_{t_k}) \to \tfrac{1}{2}(B_t^2-t) \quad \text{(Itô)}.$$

Let $\mathcal{F}_t$ be the smallest σ-algebra for which B_s is measurable for all $s \le t$.

A stochastic process $(X_t)_{t\ge 0}$ is called *adapted* if X_t is $\mathcal{F}_t$-measurable for all t. We call it $(\mathcal{B}\times\mathcal{F})$-measurable on $[S,T]$ if the function $(t,\omega)\mapsto X_t(\omega)$ is measurable with respect to the σ-algebra $\mathcal{B}\times\mathcal{F}$, where $\mathcal{B}$ is the Borel σ-algebra on $[S,T]$. And we say the process is an L^2 process if

$$E\Big(\int_S^T X_t^2\,dt\Big) = \int_S^T E(X_t^2)\,dt < \infty.$$

Note that the first equality is just Tonelli's theorem.

An *elementary process* has the form

$$t \mapsto \sum_{k=0}^{n-1} X_k[t_k \le t < t_{k+1}],$$

where $S = t_0 < t_1 < \cdots < t_n = T$.

It is adapted if and only if X_k is $\mathcal{F}_{t_k}$-measurable for all k; then it is clearly $(\mathcal{B}\times\mathcal{F})$-measurable, and it is in L^2 if and only if $E(X_k^2) < \infty$ for all k. In this case, we define the Itô integral:

$$\int_S^T \sum_{k=0}^{n-1} X_k[t_k \le t < t_{k+1}]\,dB_t = \sum_{k=0}^{n-1} X_k(B_{t_{k+1}} - B_{t_k}).$$

Notice that if $X_k = B_{t_k}$, this is an Itô sum for $\int_S^T B_t\,dB_t$. To get a Stratonovich sum, we would have to put $X_k = B_{t_{k+1}}$, but then the corresponding elementary process is not adapted.

The Itô integral turns out to be an *isometry* of one L^2 space into another:

$$E\Big(\Big(\int_S^T X_t\,dB_t\Big)^2\Big) = \int_S^T E(X_t^2)\,dt$$

for any elementary, adapted L^2 process X. Therefore, the Itô integral can be extended by continuity to the L^2 *closure* of the space of elementary adapted processes; and this closure turns out to be the space of *all* adapted, $(\mathcal{B}\times\mathcal{F})$-measurable L^2 processes.

Elementary properties of the Itô integral include linearity, additivity ($\int_S^T + \int_T^U = \int_S^U$), and

$$E\Big(\int_S^T X_t\,dB_t\Big) = 0.$$

Further, the integral is $\mathcal{F}_T$-measurable, meaning in particular that the stochastic process

$$\Big(\int_0^t X_s\,dB_s\Big)_{t\ge 0}$$

is an adapted process.

Fifth week (W38)

Noted the Lebesgue–Radon–Nikodym theorem, of which we mainly need the Radon–Nikodym part, that if μ and ν are finite (or σ-finite) measures with $\nu \ll \mu$, then there is a unique function called the *Radon–Nikodym derivative* and written $d\nu/d\mu$ so that

$$\nu(A) = \int_A \frac{d\nu}{d\mu}\, d\mu$$

for all measurable sets A.

The notation is meant to encourage the highly illegal practice of cancelling the $d\mu$ factors, after which the resulting equality is trivially true.

Conditional expectation

Recall the definition of conditinal probability: $P(A|B) = P(A \cap B)/P(B)$. Clearly, the function $A \mapsto P(A \mid B)$, which we may also write $P(\cdot \mid B)$, is itself a probability measure.

The expectation of a random variable X with respect to this probability measure is its conditional expectation. It is given by

$$E(X \mid B) = \frac{1}{P(B)} \int_B X\, dP.$$

Next, if we partition Ω into disjoint pieces, as in

$$\Omega = \bigsqcup_{k=1}^{n} B_k,$$

we can associate $E(X \mid B_k)$ with the piece B_k. Make a piecewise constant function:

$$Y(\omega) = \sum_{k=1}^{n} E(X \mid B_k)\, [\omega \in B_k].$$

This is measurable with respect to the σ-algebra $\mathcal{G}$ generated by the sets B_k, $k = 1, \ldots, n$, and you may verify that

$$\int_A Y\, dP = \int_A X\, dP \qquad \text{for all } A \in \mathcal{G}$$

(for it is true when $A = B_k$, and any $A \in \mathcal{G}$ is a disjoint union of some of the sets B_k). Moreover Y is the *only* $\mathcal{G}$-measurable function satisfying this property. This motivates

19 Definition. Let X be a random variable with $E(|X|) < \infty$ (i.e., $X \in L^1$), and $\mathcal{G} \subseteq \mathcal{F}$ a σ-algebra. Then the *conditional expectation* of X with respect to $\mathcal{G}$ (or we may say *given* $\mathcal{G}$) is the unique $\mathcal{G}$-measurable L^1-function $E(X \mid \mathcal{G})$ satisfying

$$\int_A E(X \mid \mathcal{G})\, dP = \int_A X\, dP \qquad \text{for all } A \in \mathcal{G}.$$

The proof idea is to note that if $X \geq 0$ then $A \mapsto \int_A X\,dP$ is a measure on $\mathcal{G}$ which is (trivially) absolutely continuous with respect to P (restricted to $\mathcal{G}$), and then $E(X \mid \mathcal{G})$ is just the Radon–Nikodym derivative of this measure with respect to P (restricted to $\mathcal{G}$).

One simple consequence of the definition is that

$$E\bigl(E(X \mid \mathcal{G})\,Y\bigr) = E(XY)$$

for every bounded $\mathcal{G}$-measurable variable Y. (This holds *by definition* if Y is the indicator function of a set $A \in \mathcal{G}$, and the rest is a standard approximation argument.) Put differently,

$$E\bigl((X - E(X \mid \mathcal{G}))\,Y\bigr) = 0$$

for all such Y, which looks like the definition of an *orthogonal projection.*

Indeed, if $X \in L^2$ then $E(X \mid \mathcal{G})$ is in fact the orthogonal projection of X in the subspace $L^2(\Omega, \mathcal{G}, P|_{\mathcal{G}}|)$.

Martingales

For this, we need the concept of *filtration*, which is simply a family $(\mathcal{M}_t)_{t\geq 0}$ of σ-algebras where $s < t$ implies $\mathcal{M}_s \subseteq \mathcal{M}_t$. (The obvious example is $\mathcal{F}_t$, associated with Brownian motion.)

20 Definition. A stochastic process $(M_t)_{t\geq 0}$ is called a *martingale* if $E(M_t \mid \mathcal{M}_s) = M_s$ for all $t \geq s \geq 0$.

The terminology comes from gaming. Assuming M_t is your accumulated winnings at time t, the martingale property says that the game is *fair* in the sense that your future expected winnings given your winnings at time s are the same as your current winnings.

21 Proposition *The Itô integral $\int_0^t X_t\,dB_t$, where X_t is an adapted, measurable L^2 process, is a martingale.*

To prove this, do it for an elementary adapted process first. The rest is just approximation.

To be continued . . .