1 IEOR 6712: Notes on Brownian Motion II: Introduction to stochastic integration

1.1 Functions of bounded variation

A real-valued function $f = f(t)$ on $[0, \infty)$ is said to be of *bounded variation* if the $y-$axis distance (e.g., up-down distance) it travels in any finite interval of time is finite. As we shall soon see, the paths of Brownian motion do not have this property leading us to conclude, among other things, that in fact the paths are not differentiable. If f is differentiable, then during an infinitesimal time interval of length dt, near t, the $y-$axis distance traversed is $|df(t)| = |f'(t)|dt$, and so the total distance traveled over an interval $[a, b]$ is $\int_a^b |f'(t)|dt$, as we know from calculus.

To make this precise, consider an interval $[a, b]$ and a partition Π of the interval of the form $a = t_0 < t_1 < \cdots < t_n = b$. Define the variation over $[a, b]$ with respect to Π by

$$V_\Pi(f)[a, b] = \sum_{k=1}^{n} |f(t_k) - f(t_{k-1})|. \tag{1}$$

As the partition gets finer and finer, this quantity approaches the *total variation* of f over $[a, b]$: Leting $|\Pi| = \max_{1\le k\le n}\{t_k - t_{k-1}\}$ denote the maximum interval length of a partition, the total variation of f over $[a, b]$ is defined by

$$V(f)[a, b] = \lim_{|\Pi|\to 0} V_\Pi(f)[a, b], \tag{2}$$

If $V(f)[a, b] < \infty$, then f is said to be of bounded variation over $[a, b]$. If $V(f)[0, t] < \infty$ for all $t \ge 0$, then f is said to be of bounded variation.

A common example of such a collection of partitions with $|\Pi| \to 0$, indexed by integers $n \ge 1$, is given by dividing $[a, b]$ up into n subintervals of length $(b-a)/n$; $t_k = a+k(b-a)/n,\ 0 \le k \le n$. Then $|\Pi| = (b - a)/n \to 0$ as $n \to \infty$.

If f is a differentiable function, then from the mean value theorem from calculus, we know that $f(t_k) - f(t_{k-1}) = f'(t_k^*)(t_k - t_{k-1})$ for some value $t_k^* \in [t_{k-1}, t_k]$ yielding the representation

$$V_\Pi(f)[a, b] = \sum_{k=1}^{n} |f'(t_k^*)|(t_k - t_{k-1}).$$

But this is a Reimann sum for the integral $\int_a^b |f'(x)|dx$; thus when f is differentiable,

$$V(f)[a, b] = \lim_{|\Pi|\to 0} V_\Pi(f)[a, b] = \int_a^b |f'(x)|dx < \infty.^1$$

Thus differentiable functions are of bounded variation.

[1] More precisely, for the Reimann integral $\int_a^b f'(x)dx$ to exist we need that $f'(x)$ be bounded on $[a, b]$ and continuous except at a set of points of Lebesgue measure 0.

Also note that *if f is of bounded variation over $[a,b]$, then $f(x)$ is bounded over $[a,b]$*; $\sup_{x\in[a,b]}|f(x)|<\infty$: For $x\in[a,b]$, and any partition Π of $[a,x]$, $t_0=a<t_1<\cdots<t_n=x$,

$$\begin{aligned}|f(x)-f(a)| &= \Big|\sum_{i=1}^{n}(f(t_i)-f(t_{i-1})\Big| \\ &\le \sum_{i=1}^{n}|f(t_i)-f(t_{i-1}|=V_{\Pi}(f)[a,x].\end{aligned}$$

Thus, letting $|\Pi|\to 0$ yields $|f(x)-f(a)|\le V(f)[a,x]\le V(f)[a,b]<\infty$. Thus, for all $x\in[a,b]$, $|f(x)|=|f(x)-f(a)+f(a)|\le|f(x)-f(a)|+|f(a)|\le V(f)[a,b]+|f(a)|$

The following Proposition identifies the class of functions that are of bounded variation:

Proposition 1.1 *A function f is of bounded variation over $[a,b]$ if and only if $f(x)$ can be re-written as $f(x)=M_1(x)-M_2(x)$ where M_1 and M_2 are bounded monotone increasing functions on $[a,b]$. (This composition is not unique.)*

Proof : Suppose f is of bounded variation over $[a,b]$ and let $v(x)=V(f)[a,x]$, $x\in[a,b]$. Then as is easily verified, $M_1(x)=v(x)+f(x)/2$, $M_2(x)=v(x)-f(x)/2$ yields the desired bounded monotone functions. For example: for $x\in[a,b)$ and $h>0$ small enough so that $x+h\le b$, note that $M_1(x+h)-M_1(x)\ge 0$ (monotonicity) is proved as follows: Since $M_1(x+h)-M_1(x)=V(f)[x,x+h]+(f(x+h)-f(x))/2$, we must show that

$$V(f)[x,x+h]+(f(x+h)-f(x))/2\ge 0. \tag{3}$$

But for any partiton of $[x,x+h]$, $x=t_0<t_1<\cdots<t_n=x+h$, we have

$$\begin{aligned}|f(x+h)-f(x)| &= \Big|\sum_{i=1}^{n}(f(t_i)-f(t_{i-1})\Big| \\ &\le \sum_{i=1}^{n}|f(t_i)-f(t_{i-1})|=V_{\Pi}(f)[x,x+h].\end{aligned}$$

Thus, letting $|\Pi|\to 0$ yields that $V(f)[x,x+h]\ge|f(x+h)-f(x)|$.

Thus we have

$$V(f)[x,x+h]+(f(x+h)-f(x))/2\ge|f(x+h)-f(x)|+(f(x+h)-f(x))/2. \tag{4}$$

If $f(x+h)\ge f(x)$, then (3) holds automatically; otherwise
if $f(x+h)<f(x)$, then $|f(x+h)-f(x)|=f(x)-f(x+h)>0$ and the RHS of (4) becomes

$$|f(x+h)-f(x)|+(f(x+h)-f(x))/2=(f(x)-f(x+h))/2>0.$$

Thus (3) holds as was to be shown. $M_1(x)$ is a bounded function since (as we pointed out above before the statement of this Proposition) $f(x)$ itself is a bounded function over $[a,b]$, and so $M_1(x)\le M_1(b)=V(f)[a,b]+f(b)/2<\infty$, $x\in[a,b]$. Thus indeed $M_1(x)$ is bounded and monotone increasing. The proof for $M_2(x)$ is the same.

The converse is immediate since for a bounded monotone increasing function $M(x)$, $V(M)[a,b]=M(b)-M(a)<\infty$, and thus If $f(x)=M_1(x)-M_2(x)$, we have $V(f)[a,b]\le V(M_1)[a,b]+V(M_2)[a,b]=M_1(b)-M_1(a)+M_2(b)-M_2(a)<\infty$. ■

Brownian motion has paths of unbounded variation

It should be somewhat intuitive that a typical Brownian motion path can't possibly be expresssed as the difference of monotone functions as in Proposition 1.1; the paths seem to exhibit rapid infinitesimal movement up and down in any interval of time. This is most apparent when recalling how BM can be obtained by a limiting (in n) procedure by scaling a simple symmetric random walk (scaling space by $1/\sqrt{n}$ and scaling time by n): In any time interval, no matter how small, the number of moves of the random walk tends to infinity, while the size of each move tends to 0. This intuition turns out to be correct:

Proposition 1.2 *With probability* 1, *the paths of Brownian motion* $\{B(t)\}$ *are not of bounded variation;* $P(V(B)[0,t] = \infty) = 1$ *for all fixed* $t > 0$.

We will prove Proposition 1.2 in the next section after we introduce the so-called squared (quadratic) variation of a function, and prove that BM has finite non-zero quadratic variation.

1.2 Quadratic variation

For a function f, we define the squared variation (also called the quadratic variation) by squaring the terms in the definition of total variation:

$$Q_\Pi(f)[a,b] = \sum_{k=1}^{n} |f(t_k) - f(t_{k-1})|^2. \tag{5}$$

$$Q(f)[a,b] = \lim_{|\Pi|\to 0} Q_\Pi(f)[a,b]. \tag{6}$$

If f is differentiable, then it is easy to see that $Q(f)[a,b] = 0$:

$$\begin{aligned}
\sum_{k=1}^{n} |f(t_k) - f(t_{k-1})|^2 &= \sum_{k=1}^{n} |f'(t_k^*)|^2 |t_k - t_{k-1}|^2 \text{ mean value theorem} \\
&\le |\Pi| \sum_{k=1}^{n} |f'(t_k^*)|^2 |t_k - t_{k-1}| \\
&\to 0, \text{ as } |\Pi| \to 0,
\end{aligned}$$

because $\sum_{k=1}^{n} |f'(t_k^*)|^2 |t_k - t_{k-1}| \to \int_a^b |f'(x)|^2 dx$.

Remark 1.1 In general, we can define for any $p > 0$, the p^{th} variation via

$$V_\Pi^{(p)}(f)[a,b] = \sum_{k=1}^{n} |f(t_k) - f(t_{k-1})|^p. \tag{7}$$

$$V^{(p)}(f)[a,b] = \lim_{|\Pi|\to 0} V_\Pi^{(p)}(f)[a,b]. \tag{8}$$

If f is a continuous function, and for some $p > 0$, $V^{(p)}(f)[a,b] < \infty$, then $V^{(p+\epsilon)}(f)[a,b] = 0$, for all $\epsilon > 0$ as can be seen as follows:

$$\sum_{k=1}^{n} |f(t_k) - f(t_{k-1})|^{p+\epsilon} \le M(\epsilon) V_\Pi^{(p)}(f)[a,b],$$

where $M(\epsilon) = \max_{1\le i\le n} |f(t_k) - f(t_{k-1})|^\epsilon \to 0$, as $|\Pi| \to 0$ by uniform continuity of the function f over the closed interval $[a,b]$.

1.2.1 BM has finite non-zero quadratic variation

In this section we will prove that BM satisfies $Q(B)[0,t]=t$; in particular the paths can't be differentiable.[2] This leads to the intuitive differential interpretation $(dB(t))^2 = dt$, a somewhat mind boggling conclusion.

Proposition 1.3 *The paths of Brownian motion $\{B(t)\}$ satisfy*

$$\lim_{|\Pi|\to 0} Q_\Pi(B)[0,t] = Q(B)[0,t] = t, \tag{9}$$

where convergence is in L^2: For each $t>0$, $E(Q_\Pi(B)[0,t]-t)^2 \to 0$. Thus (via Chebychev's inequality $(P(|X|>\epsilon) \le E(X^2)/\epsilon^2)$) convergence holds in probability as well: For each $t>0$, it holds that for all $\epsilon>0$, as $|\Pi|\to 0$,

$$P(|Q_\Pi(B)[0,t]-t|>\epsilon) \to 0.$$

In fact, for sufficiently refined sequences of partitions $\{\Pi_n : n\ge 1\}$ with $|\Pi_n|\to 0$ as $n\to\infty$, one can obtain convergence with wp1; (9) can be made to occur wp1, for each $t>0$.

Proof :
Noting that $t=\sum_{k=1}^n t_k - t_{k-1}$, and letting $\Delta_k = B(t_k)-B(t_{k-1})$, and $Y_k = \Delta_k^2 - (t_k - t_{k-1})$ we see that $\sum_{k=1}^n Y_k = -t + \sum_{k=1}^n \Delta_k^2$; so we need to prove that $E[\sum_{k=1}^n Y_k]^2 \to 0$ as $|\Pi|\to 0$. To this end, since the Y_k are independent (hence uncorrelated) and have mean 0, it follows that

$$E[\sum_{k=1}^n Y_k]^2 = \sum_{k=1}^n E(Y_k^2).$$

But $Y_k^2 = \Delta_k^4 + |t_k - t_{k-1}|^2 - 2\Delta_k^2|t_k-t_{k-1}|$ and $E(\Delta_k^2) = |t_k - t_{k-1}|$ and $E(\Delta_k^4) = 3|t_k-t_{k-1}|^2$, where we are using the fact that the 4^{th} moment of a $N(0,\sigma^2)$ rv X is given by $E(X^4)=3\sigma^4$.
Thus $E(Y_k^2) = 2|t_k - t_{k-1}|^2$ yielding

$$E[\sum_{k=1}^n Y_k]^2 = 2\sum_{k=1}^n |t_k - t_{k-1}|^2 \to 0, \text{ as } |\Pi|\to 0,$$

where we identify $\sum_{k=1}^n |t_k - t_{k-1}|^2$ as the squared variation, with respect to our partition Π, of the differentiable function $f(t)=t$; hence it tends to 0 as $|\Pi|\to 0$.
To prove convergence with probability 1 for a refined enough sequence of partitions:
Choose partitions Π_n (indexed by n) of $[0,t]$ by dividing into 2^n subintervals of length $t2^{-n}$. For then

$$E[\sum_{k=1}^{2^n} Y_k]^2 = 2\sum_{k=1}^{2^n} |t_k - t_{k-1}|^2 = 2\cdot 2^n t^2 2^{-2n} = 2t^2 2^{-n},$$

and Chebychev's inequality $(P(|X|>\epsilon) \le E(X^2)/\epsilon^2)$ then yields

$$P\Big(\Big|\sum_{k=1}^{2^n} Y_k\Big| > \epsilon)\Big) \le 2t^2 2^{-n} e^{-2}.$$

[2]It can be proved that with probability 1, the sample paths are *nowhere* differentiable.

Since $\sum_{n=1}^{\infty} 2t^2 2^{-n} e^{-2} = 2t^2 e^{-2} \sum_{n=1}^{\infty} 2^{-n} < \infty$, we deduce from the Borel-Cantelli lemma (using events $A_n = \{\left|\sum_{k=1}^{2^n} Y_k\right| > \epsilon)\},\ n \geq 1$) that wp1, only for finitely many values of n does it hold that $\left|\sum_{k=1}^{2^n} Y_k\right| > \epsilon$, or equivalently that wp1, $\left|\sum_{k=1}^{2^n} Y_k\right| \leq \epsilon$ for all n sufficiently large.

Since ϵ is arbitrary, we conclude by letting $\epsilon \downarrow 0$ that wp1,

$$\sum_{k=1}^{2^n} Y_k \to 0, \quad \text{as } n \to \infty$$

. ■

Remark 1.2 In Proposition 1.3, convergence in L^2, hence in probability, always holds for any collection of partitions with $|\Pi| \to 0$; but wp1 convergence depends on having $|\Pi|$ tend to 0 sufficiently fast. A sufficient condition is to have a sequence of partitions $\{\Pi_n : n \geq 1\}$–such as the $t2^{-n}$ ones used in our proof of Proposition 1.3–such that

$$\sum_{n=1}^{\infty} |\Pi_n| < \infty.$$

Proof of Proposition 1.2: $V(B)[0,t] = \infty$, wp1, for all $t > 0$.

Proof :

$$Q_\Pi(B)[0,t] = \sum_{k=1}^{n} |B(t_k) - B(t_{k-1})|^2 \leq \max_k |B(t_k) - B(t_{k-1})| \sum_{k=1}^{n} |B(t_k) - B(t_{k-1})| = MV_\Pi(B)[0,t],$$

where $M = M(\Pi) = \max_k |B(t_k) - B(t_{k-1})| \to 0$, wp1, as $|\Pi| \to 0$ since the paths of BM are continuous, hence *uniformly continuous*[3] over any interval $[0,t]$. If $V(B)[0,t] < \infty$, wp1, for all $t > 0$, then we would conclude that $MV_\Pi(B)[0,t] \to 0$, wp1, as $|\Pi| \to 0$ and hence conclude that $Q(B)[0,t] = 0$, wp1, for all $t > 0$ contradicting Proposition 1.3 (convergence wp1 implies convergence in probability). Hence indeed $V(B)[0,t] = \infty$, wp1, for all $t > 0$. ■

1.3 Stochastic integration: the Ito integral

Our objective here is to make sense of integrals of the form $\int_0^t X(s)dB(s)$, where $\{B(t)\}$ is standard BM, and $\{X(t)\}$ is a stochastic process. The final product is called an Ito integral, and the integration is referred to as Ito integration, after the Japanese mathematician Kiyoshi Itō (1915–2008) who developed much of this theory over 50 years ago.

When a function f has paths of bounded variation, then one can define integrals $\int_0^t g(s)df(s)$ by using partitions Π of $[0,t]$, forming Reimann-Stieltjes sums, $\sum_{k=1}^{n} g(t_k^*)(f(t_k) - f(t_{k-1}))$ and letting $|\Pi| \to 0$ as is done in elementary calculus. Proposition 1.1 makes this clear, since using a monotone function M for integration increments $M(t_k) - M(t_{k-1})$ is only a minor extension of using the monotone function $f(t) = t$ that yields integrals like $\int_0^t g(s)ds$. Also, if f is differentiable, then the integral would be given by $\int_a^b g(s)f'(s)ds$ since $df(s) = f'(s)ds$.

[3]A function $f = f(x)$ is said to be uniformly continuous over a finite closed interval $[a,b]$ if for all $\epsilon > 0$ there exists a $\delta > 0$ such that whenever $|x - y| < \delta$ is holds that $|f(x) - f(y)| < \epsilon$. Any continuous function over a finite closed interval $[a,b]$ is automatically uniformly continuous.

Such integrals are also independent of what partitions are used when letting $|\Pi| \to 0$; any collection is to be allowed. Since BM does not have paths of bounded variation, we need to define $\int_0^t X(s)dB(s)$ differently, we can not merely do so path by path in the Reimann-Stieltjes sense.

How will we define the integral? The plan

Given the interval $[0,t]$ and a partition Π, $0 = t_0 < t_1 < \cdots < t_n = t$, if the sample paths of X on $[0,t]$ are simple functions of the form $X(s) = \sum_{k=1}^n c_k I\{t_{k-1} \le s < t_k\},\ s \in [0,t]$, with the c_k random variables, e.g., they are piecewise constant, then we call X a *simple process* and define $\int_0^t X(s)dB(s) = \sum_{k=1}^n c_k(B(t_k) - B(t_{k-1}))$.

The trick now is to choose a suitable framework that allows a general process X to be approximated by simple processes X_n, e.g., $X_n \to X$ as $n \to \infty$ so that we can then define $\int_0^t X(s)dB(s) = \lim_{n\to\infty} \int_0^t X_n(s)dB(s)$. The type of convergence in the limit must be specified and the framework has to be such that the resulting integral does not depend on the choice of approximating functions. We present this framework next.

The framework

We now assume that $\{\mathcal{F}_t : t \ge 0\}$ is a filtration of information (e.g. a collection of events) such that

1. $\mathcal{F}_s \subset \mathcal{F}_t$ if $s \le t$: All events in $\mathcal{F}_s$ are contained in $\mathcal{F}_t$.
2. $B(t) \in \mathcal{F}_t,\ t \ge 0$ and $X(t) \in \mathcal{F}_t,\ t \ge 0$: $B(t)$ and $X(t)$ are completely determined by the events in $\mathcal{F}_t$. (We say that $\{X(t)\}$ and $\{B(t)\}$ are both *adapted* to the filtration.)
3. For any $t \ge 0$, the future increments $\{B(t+h) - B(t) : h \ge 0\}$ are independent of the past $\mathcal{F}_t$.
4. $\{X(t)\}$ is a square integrable process:

$$E\Big(\int_0^t X^2(s)ds\Big) < \infty,\ t \ge 0.$$

The reader should realize that a special case of an appropriate filtration is when $\mathcal{F}_t = \sigma\{B(s) : 0 \le s \le t\}$, e.g., when $\mathcal{F}_t$ is exactly the information contained in $\{B(s) : 0 \le s \le t\}$. But then $\{X(t)\}$ would be restricted to only be allowed to be determined by the Brownian motion. The more general framework above thus allows the process $\{X(t)\}$ to be determined by further information, but as long as it does not destroy the independent increments property of the Brownian motion (*3*). Further information might include events that are entirely independent of the Brownian motion, for example.

Ito integral for simple processes

We now fix a $t > 0$ and wish to define an integral

$$\int_0^t X(s)dB(s)$$

under our assumed framework.

Consider the set of *simple* processes of the form

$$X(s) = \sum_{k=1}^{n} c_k I\{t_{k-1} < s \le t_k\},\ 0 \le s \le t,$$

where $0 = t_0 < t_1 < \cdots < t_n = t$ is a deterministic partition of $[0,t]$ and the c_k are random variables adapted to the filtration in the predictable sense $c_k \in \mathcal{F}_{t_{k-1}}$, and are square integrable, $E(c_k^2) < \infty$.

Now we define, in the obvious manner, the Ito integral $I(X) = I(X)_t$ over $[0,t]$ as

$$I(X)_t \stackrel{\text{def}}{=} \sum_{k=1}^{n} c_k(B(t_k) - B(t_{k-1}). \tag{10}$$

More generally, we define the integral over all intervals of the form $[0,u],\ u \le t$ via

$$I(X)_u \stackrel{\text{def}}{=} \sum_{k=1}^{n} c_k(B(t_k \wedge u) - B(t_{k-1} \wedge u),\ u \in [0,t], \tag{11}$$

which is a nice succinct way of re-writing the more obvious

$$I(X)_u \stackrel{\text{def}}{=} c_{k+1}(B(u) - B(t_k)) + \sum_{i=1}^{k} c_i(B(t_i) - B(t_{i-1}),\ u \in (t_k, t_{k+1}],\ 0 \le k \le n-1. \tag{12}$$

From (11) it is immediate that $\{I(X)_u : 0 \le u \le t\}$ forms a stochastic process over $[0,t]$ with continuous sample paths. Moreover, since $\{B(t)\}$ is a martingale what we have actually done is to construct yet another martingale since (recall Lecture Notes on martingales in discrete time) in discrete time we know that

$$(H \cdot X)_n = \sum_{i=1}^{n} H_i(X_i - X_{i-1}),\ n \ge 1,$$

forms a martingale when $H_n \in \mathcal{F}_{n-1}$ (e.g., H is a predictable process, a gambling strategy) and $X_i - X_{i-1}$ are the increments of a martingale.

Thus our Ito integral for a simple process can be viewed in the same way, with $I(X)_u$ the total winnings that you have at time $u \le t$ when betting according to the Brownian increments and stakes c_k. Note further that by construction, $I(X)_u \in \mathcal{F}_u,\ u \le t$; this process is adapted to the underlying filtration. We summarize:

Proposition 1.4 *For a simple processes X over $[0,t]$, $\{I(X)_u : 0 \le u \le t\}$ is a martingale with continuous sample paths.*

An important fact is that the second moment of $I(X)$ is equal to $E(\int_0^t X^2(s)ds)$:

Proposition 1.5 *[Ito isometry] For a simple process X over $[0,t]$,*

$$E(I(X)_u^2) = E(\int_0^u X^2(s)ds),\ u \in [0,t].$$

Proof : We prove this for $u = t$, the $u < t$ case being similar. In our framework, $c_k \in \mathcal{F}_{t_{k-1}}$ and $B(t_k) - B(t_{k-1})$ is independent of $\mathcal{F}_{t_{k-1}}$; thus c_k is independent of $B(t_k) - B(t_{k-1})$. From this we see that the cross terms when considering $I(X)^2$ have mean 0 (they are uncorrelated). (In fact this is a reiteration of the fact that $I(X)$ is a martingale and martingale increments are uncorrelated.) Thus

$$\begin{aligned}
E(I(X)^2) &= E\Big(\sum_{k=1}^{n} c_k^2(B(t_k) - B(t_{k-1})^2\Big) \\
&= \sum_{k=1}^{n} E(c_k^2(B(t_k) - B(t_{k-1}))^2) \\
&= \sum_{k=1}^{n} E(c_k^2)E(B((t_k) - B(t_{k-1}))^2) \\
&= \sum_{k=1}^{n} E(c_k^2)(t_k - t_{k-1}).
\end{aligned}$$

On the other hand due to the fact that the indicators $I\{t_{k-1} < s \le t_k\}$ are disjoint and that the square of an indicator is itself, we have

$$X^2(s) = \sum_{k=1}^{n} c_k^2 I\{t_{k-1} < s \le t_k\},\ 0 \le s \le t,$$

so that

$$\begin{aligned}
E(\int_0^t X^2(s)ds) &= E\Big(\sum_{k=1}^{n} \int_0^t c_k^2 I\{t_{k-1} < s \le t_k\}ds\Big) \\
&= E\Big(\sum_{k=1}^{n} \int_{t_{k-1}}^{t_k} c_k^2 ds\Big) \\
&= E\Big(\sum_{k=1}^{n} c_k^2(t_k - t_{k-1})\Big) \\
&= \sum_{k=1}^{n} E(c_k^2)(t_k - t_{k-1});
\end{aligned}$$

the equality is established. ■

Another elementary result (proof left to the reader) is:

Lemma 1.1 (Linearity of the Ito integral) *If both X and Y are simple processes over $[0, t]$ then $I(X + Y) = I(X) + I(Y)$, and $I(aX) = aI(X)$, for any constant a.*

Finally:

Proposition 1.6 *The quadratic variation of $I(X)$ is given by $Q(I(X))[0,t] = \int_0^t X(s)^2 ds,\ t \geq 0$. In other words, if $Y(t) = I(X)_t$, then for each $t \geq 0$,*

$$\lim_{|\Pi|\to 0} \sum_{k=1}^{n} |Y(t_k) - Y(t_{k-1})|^2 = \int_0^t X(s)^2 ds.$$

(Convergence here, as $|\Pi| \to 0$, is in probability, wherebye we say $X_n \to X$ in probability, as $n \to \infty$, if for all $\epsilon > 0$, it holds that $P(|X_n - X| > \epsilon) \to 0$, as $n \to \infty$.)

Proof : Note that for a simple process $X(s) = \sum_{j=1}^{m} c_j I\{s_{j-1} < s \leq s_j\},\ 0 \leq s \leq t$, where $0 = s_0 < s_1 < \cdots < s_m = t$, we can compute (directly) the quadratic variation by doing so over each subinterval $(s_{j-1}, s_j]$ and summing up. But for $s \in (s_{j-1}, s_j]$, $X(s) = c_j$ and thus for any partition Π of $(s_{j-1}, s_j]$ (e.g., $s_{j-1} = t_0 < t_1 < \cdots < t_n = s_j$), we have $Q_\Pi(I(X))[s_{j-1}, s_j] = c_j^2 \sum_{k=1}^{n} (B(t_{k-1} - B(t_k))^2$ which of course tends to $c_j^2(s_j - s_{j-1})$, and this can be re-written as $c_j^2(s_j - s_{j-1}) = \int_{s_{j-1}}^{s_j} c_j^2 ds = \int_{s_{j-1}}^{s_j} X^2(s) ds$. Thus, summing up over all m subintervals yields $Q(I(X))[0,t] = \int_0^t X(s)^2 ds$. ■

Extending the integral from simple processes to general ones

We will now show that if a general process X (within our assumed framework) over $[0,t]$ could be approximated by simple processes X_n in an L^2 sense, e.g, $E \int_0^t |X_n(s) - X(s)|^2 ds \to 0$ as $n \to \infty$, then we can define

$$I(X) = \lim_{n\to\infty} I(X_n), \quad \text{in an } L^2 \text{ sense}, \tag{13}$$

e.g., there exists a rv $I(X)$ such that $E(I(X) - I(X_n))^2 \to 0$ as $n \to \infty$.

This is because of Proposition 1.5 and the linearity of the integral:

If $E \int_0^t |X_n(s) - X(s)|^2 ds \to 0$, then so does $E \int_0^t |X_n(s) - X_m(s)|^2 ds \to 0$, as $m, n \to \infty$, (formally we are asserting that a convergent sequence is also a *Cauchy* sequence), and hence (via Proposition 1.5) $E(I(X_n) - I(X_m))^2 = E \int_0^t |X_n(s) - X_m(s)|^2 ds \to 0$, as $m, n \to \infty$. This implies that $I(X_n)$ is a Cauchy sequence, hence it must converge (formally we are asserting that a Cauchy sequence of rvs in L^2 must converge); we denote this limit by $I(X)$. It can be shown that this limit is insensitive to which approximating sequence is used: If X'_n is another sequence, with $E \int_0^t |X'_n(s) - X(s)|^2 ds \to 0$, then $I(X'_n)$ converges to the same limit as does $I(X_n)$; the limit is unique.

To make this work, we need the following crucial result:

Lemma 1.2 *If X is a general process (within our framework), then over each $[0,t]$ there exists a sequence of simple processes X_n (within our framework), such that $E \int_0^t |X_n(s) - X(s)|^2 ds \to 0$ as $n \to \infty$.*

Proof :[sketch] Suppose that X (in our framework) is bounded and has continuous sample paths over $[0,t]$. Now choose the partions indexed by n, $t_k = kt/n,\ 0 \leq k \leq n$, and define

$$X_n(s) = \sum_{k=1}^{n} X(t_{k-1}) I\{t_{k-1} < s \leq t_k\},\ 0 \leq s \leq t. \tag{14}$$

In other words we choose the left endpoint value, $c_k = X(t_{k-1})$, over the interval $[t_{k-1}, t_k)$. This ensures that $c_k \in \mathcal{F}_{t_{k-1}}$ as is required in our framework. Note that

$$\int_0^t (X(s) - X_n(s))^2 ds \to 0, \text{ wp1, as } n \to \infty,$$

because of the continuous sample paths, e.g., $(X(s) - X_n(s))^2 \to 0$ for each s by continuity and thus so does the integral since the integrand is non-negative and bounded. Thus the bounded convergence theorem implies that $E\int_0^t |X_n(s) - X(s)|^2 ds \to 0$. (Even if X is not bounded, the approximating functions defined in (14) typically do the trick; for example they work when $X(t) = B(t)$.)

In the general case, we use finer partions (indexed by n) of the form $t_k = tk2^{-n}, \ 0 \le k \le 2^n$. Then we average the process over intervals,

$$c_k \stackrel{\text{def}}{=} \frac{1}{(t_{k-1} - t_{k-2})} \int_{t_{k-2}}^{t_{k-1}} X(s)ds,$$

(Note how $c_k \in \mathcal{F}_{t_{k-1}}$ because $\mathcal{F}_{t_{k-2}} \subset \mathcal{F}_s \subset \mathcal{F}_{t_{k-1}}$ for all $s \in [t_{k-2}, t_{k-1}]$.) and define (with $c_1 = X(0)$)

$$X_n(s) \stackrel{\text{def}}{=} \sum_{k=1}^{n} c_k I\{t_{k-1} < s \le t_k\}, \ 0 \le s \le t.$$

One can show (with some work of course) that indeed, $E\int_0^t |X_n(s) - X(s)|^2 ds \to 0$ as $n \to \infty$. ■

1.4 Further extension

For a simple process X, we directly defined $I(X)$ as a stochastic process over an interval $[0, t]$, but for a general process what we have done is only define, for each t, a random variable $I(X)_t$. It turns out that we could (with more work) not only define $I(X)$ as a stochastic process over $[0, t]$ but consider $I(X)_t = \int_0^t X(s)dB(s), \ t \ge 0$ all at once as a stochastic process and with continuous sample paths. Such a construction can be done, but the details are beyond the scope of these lecture notes. We shall assume this extension has been done throughout the rest of these notes.

1.5 General properties of the Ito integral

From the way we define $I(X)$, it is immediate that it inherits all the nice properties (linearity, etc.) of the simple process case. For the most part, the results below are proved using Lemma 1.2 by first proving the result for the stochastic process $\{I(X_n)_t : t \ge 0\}$ and then taking the limit as $n \to \infty$ to obtain the result for the stochastic process $\{I(X)_t : t \ge 0\}$

Theorem 1.1

1. *$I(X)$ is a martingale with respect to the filtration $\mathcal{F}_t, \ t \ge 0$, and it has (with probability 1) continuous sample paths.*

2. *Ito isometry:*

$$E(I(X)_t^2) = E(\int_0^t X^2(s)ds), \ t \ge 0.$$

3. *$I(X+Y) = I(X) + I(Y)$, and $I(aX) = aI(X)$, for any constant a for any two processes X, Y.*

4. *The quadratic variation of $I(X)$ is given by*
   $Q(I(X))[0,t] = \int_0^t X(s)^2 ds,\ t \geq 0$. In other words, if $Y(t) = I(X)_t$, then for each $t \geq 0$,

$$\lim_{|\Pi|\to 0} \sum_{k=1}^{n} |Y(t_k) - Y(t_{k-1})|^2 = \int_0^t X(s)^2 ds.$$

   (Convergence here, as $|\Pi| \to 0$, is in probability, wherebye we say $X_n \to X$ in probability, as $n \to \infty$, if for all $\epsilon > 0$, it holds that $P(|X_n - X| > \epsilon) \to 0$, as $n \to \infty$.)

1.6 Ito's formula part I

We first directly compute

$$\int_0^t B(s) dB(s).$$

In this case we use approximation functions

$$X_n(s) = \sum_{k=1}^{n} B(t_{k-1}) I\{t_{k-1} < s \leq t_k\},\ 0 \leq s \leq t,$$

with Π_n the standard t/n partion. (It is easily seen that $X_n \to B$ in L^2 over $[0,t]$.)

Using the identity

$$B(t_{k-1})(B(t_k) - B(t_{k-1})) = \frac{1}{2}(B^2(t_k) - B^2(t_{k-1})) - \frac{1}{2}(B(t_k) - B(t_{k-1}))^2$$

yields

$$\begin{aligned} I(X_n) &= \sum_{k=1}^{n} B(t_{k-1})(B(t_k) - B(t_{k-1}) \\ &= \frac{1}{2}B^2(t) - \frac{1}{2}\sum_{k=1}^{n}(B(t_k) - B(t_{k-1})^2 \\ &\to \frac{1}{2}B^2(t) - \frac{1}{2}t, \text{ as } n \to \infty, \end{aligned}$$

where we recognize $\sum_{k=1}^{n}(B(t_k) - B(t_{k-1})^2$ as converging to t, the quadratic variation. Thus $I(B)_t = \int_0^t B(s)dB(s) = \frac{1}{2}B^2(t) - \frac{1}{2}t,\ t \geq 0$.

Note that if f is a differentiable function with $f(0) = 0$ then $\int_0^t f(s)df(s) = \int_0^t f(s)f'(s)ds = f^2(t)/2$. Ito integration is different due to the non-zero quadratic variation; a second term gets included.

Also note (we know this from previous lectures), that indeed $\{\frac{1}{2}B^2(t) - \frac{1}{2}t\}$ is a martingale.

Ito's formula for $f = f(x)$

Let $f = f(x)$ be a real-valued function ($f : \mathbf{R} \to \mathbf{R}$) that is continuous.
Then $\{f(B(t)) : t \geq 0\}$) defines a new stochastic process with continuous sample paths. If f is also differentiable, then we would like to compute the differential, $df(B(t)$. Ito's formula tells us how to do just that.

Theorem 1.2 *If $f = f(x)$ is twice differentiable and f'' is continuous, then (differential form)*

$$df(B(t)) = f'(B(t))dB(t) + \frac{1}{2}f''(B(t))dt,$$

which in integral form is

$$f(B(t)) = f(0) + \int_0^t f'(B(s))dB(s) + \frac{1}{2}\int_0^t f''(B(s))ds.$$

Note how the first integral is an Ito integral[4] while the second is (path wise) a Reimann integral; both are well defined. It is the appearance of the second integral which distinguishes Ito calculus from regular calculus.

As an example, take $f(x) = x^2/2$ yielding

$$\frac{1}{2}B^2(t) = \int_0^t B(s)dB(s) + \frac{1}{2}\int_0^t ds,$$

which in turn yields $\int_0^t B(s))dB(s) = \frac{1}{2}B^2(t) - \frac{1}{2}t$, a fact that we know from direct calculation in our example above.

More generally, consider $f(x) = x^n/n$, for $n \geq 2$, to obtain

$$\frac{1}{n}B^n(t) = \int_0^t B^{n-1}(s)dB(s) + \frac{1}{2}\int_0^t (n-1)B^{n-2}(s)ds,$$

yielding

$$\int_0^t B^{n-1}(s)dB(s) = \frac{1}{n}B^n(t) - \frac{1}{2}\int_0^t (n-1)B^{n-2}(s)ds.$$

Thus we see that Ito's formula is a powerful tool for computing Ito integrals, without having to use the definition involving limits. This is analogous to how we compute Reimann integrals: We do not take limits with Reimann sums, we instead use the fact that

$$f(t) = f(0) + \int_0^t f'(s)ds.$$

But what is more is that we end up creating martingales that might not be obvious if we were to look at them directly. To make our point, consider $f(x) = \sin(x)$. Then, as the reader can check, we conclude from Ito's formula that

$$\int_0^t \cos(B(s))dB(s) = \sin(B(t)) + \frac{1}{2}\int_0^t \sin(B(s))ds.$$

In particular, the right hand side forms a martingale; something not obvious to a casual reader.

[4] Assuming of course that $\{f'(B(t))\}$ is a square integrable process, which we shall assume. This requirement can be loosened.

Sketch of a proof of Ito's formula:

The proof of Ito's formula relies on using the Taylor's series expansion

$$f(x+h) - f(x) = f'(x)h + \frac{1}{2}h^2 f''(x) + R(x,h),$$

where $R(x,h)$ is the remainder.

We partition $[0,t]$ and use Taylor's formula on each subinterval;

$$\begin{aligned} f(B(t)) - f(0)) &= \sum_{k=1}^{n}(f(B(t_k) - f(B(t_{k-1})) \\ &= \sum_{k=1}^{n} f'(B(t_{k-1}))(B(t_k) - B(t_{k-1})) + \frac{1}{2}\sum_{k=1}^{n} f''(B(t_{k-1}))(B(t_k) - B(t_{k-1}))^2 + ERROR. \end{aligned}$$

As $|\Pi| \to 0$, the first sum converges to $\int_0^t f'(B(s))dB(s)$, the second to $\frac{1}{2}\int_0^t f''(B(s))ds$ (recall "$(dB(s))^2 = ds$ ") while the ERROR tends to 0 since it involves the p^{th} variations of BM for $p > 2$, all of which are 0 (recall Remark 1.1); we have derived Ito's formula.

Note in passing that when $f''(x) = 0,\ x \geq 0$, Ito's formula becomes $f(B(t)) = f(0) + \int_0^t f'(B(s))dB(s)$ implying that $\{f(B(t)) : t \geq 0\}$ is a (mean $f(0)$) martingale (since it is an Ito integral). But the only way this can happen is if f is of the form $f(x) = a + bx$; e.g., $f(B(t)) = a + bB(t)$ is just another BM, but with initial position a and variance term b^2. Later, in our next set of Notes, we will consider Ito's formula for functions of two variables $f = f(t,x)$, and see that in this case, non-trival new martingales can be obtained via forming a new process $f(t, B(t))$ when $\frac{\partial f}{\partial t} + \frac{1}{2}\frac{\partial^2 f}{\partial x^2} = 0$.