ma414l7.tex

Lecture 7. 23.2.2012

Quadratic Variation.

A non-negative right-continuous submartingale is of class (D). So it has a Doob-Meyer decomposition. We specialize this to X^2, with $X \in c\mathcal{M}^2$:

$$X^2 = X_0^2 + M + A,$$

with M a continuous martingale and A a continuous (so predictable) and increasing process. We write

$$\langle X \rangle := A$$

here, and call $\langle X \rangle$ the *quadratic variation* of X. We shall see later that this is a crucial tool for the stochastic integral. For BM, we shall later identify this with a quadratic analogue of the (finite) variation (FV) of I.4.

Quadratic Covariation.

We write $\langle M, M \rangle$ for $\langle M \rangle$, and extend $\langle . \rangle$ to a bilinear form $\langle ., . \rangle$ with two different arguments by the polarization identity:

$$\langle M, N \rangle := \frac{1}{4}(\langle M + N, M + N \rangle - \langle M - N, M - N \rangle.$$

(The polarization identity reflects the Hilbert-space structure of the inner product $\langle ., . \rangle$.) If N is of finite variation, $M \pm N$ has the same quadratic variation as M, so $\langle M, N \rangle = 0$.

Where there is a Hilbert-space structure, one can use the language of projections, of Pythagoras' theorem etc., and draw diagrams as in Euclidean space. The right way to treat the Linear Model of statistics is in such terms (analysis of variance = ANOVA, sums of squares etc.)

8. Brownian motion

Brownian motion originates in work of the botanist Robert Brown in 1828. It was introduced into finance by Louis Bachelier in 1900, and developed in physics by Albert Einstein in 1905 (see the handout for background and references).

The fact that Brownian motion *exists* is quite deep, and was first proved by Norbert WIENER (1894–1964) in 1923. In honour of this, Brownian motion is also known as the *Wiener process*, and the probability measure generating it – the measure P^* on $C[0, 1]$ (one can extend to $C[0, \infty)$) by

$$P^*(A) = P(W_. \in A) = P(\{t \to W_t(\omega)\} \in A)$$

for all Borel sets $A \in C[0,1]$ – is called *Wiener measure.*

Definition and Existence

Definition. A stochastic process $X = (X(t))_{t\geq 0}$ is a standard (one-dimensional) Brownian motion, BM or $BM(\mathbf{R})$, on some probability space $(\Omega, \mathcal{F}, \mathcal{P})$, if
(i) $X(0) = 0$ a.s.,
(ii) X has *independent increments*: $X(t+u)-X(t)$ is independent of $\sigma(X(s) : s \leq t)$ for $u \geq 0$,
(iii) X has *stationary increments*: the law of $X(t+u) - X(t)$ depends only on u,
(iv) X has *Gaussian increments*: $X(t+u) - X(t)$ is normally distributed with mean 0 and variance u, $X(t+u) - X(t) \sim N(0,u)$,
(v) X has *continuous paths*: $X(t)$ is a continuous function of t, i.e. $t \to X(t,\omega)$ is continuous in t for all $\omega \in \Omega$.

The path continuity in (v) can be relaxed by assuming it only a.s.; we can then get continuity by excluding a suitable null-set from our probability space.

We shall henceforth denote standard Brownian motion $BM(\mathbf{R})$ by $W = (W(t))$ (W for Wiener), though $B = (B(t))$ (B for Brown) is also common. Standard Brownian motion $BM(\mathbf{R}^d)$ in d dimensions is defined by $W(t) := (W_1(t), \ldots, W_d(t))$, where $W_1, \ldots, W_d$ are independent standard Brownian motions in one dimension (independent copies of $BM(\mathbf{R})$).

We turn next to Wiener's theorem, on existence of Brownian motion. The proof (cf. [BK], 5.3.1) is a streamlined version of the classical one due to Lévy in his book of 1948 and Cieselski in 1961.

Theorem (Wiener, 1923). Brownian motion exists.

Covariance.

Before addressing existence, we first find the covariance function. For $s \leq t$, $W_t = W_s + (W_t - W_s)$, so as $E(W_t) = 0$,

$$cov(W_s, W_t) = E(W_s W_t) = E(W_s^2) + E[W_s(W_t - W_s)].$$

The last term is $E(W_s)E(W_t - W_s)$ by independent increments, and this is zero, so

$$cov(W_s, W_t) = E(W_s^2) = s \qquad (s \leq t): \qquad cov(W_s, W_t) = \min(s,t).$$

A Gaussian process (one whose finite-dimensional distributions are Gaussian) is specified by its mean function and its covariance function, so among centered (zero-mean) Gaussian processes, the covariance function $\min(s,t)$ serves as the signature of Brownian motion.

Finite-dimensional Distributions.

For $0 \leq t_1 < \ldots < t_n$, the joint law of $X(t_1), X(t_2), \ldots, X(t_n)$ can be obtained from that of $X(t_1), X(t_2) - X(t_1), \ldots, X(t_n) - X(t_{n-1})$. These are jointly Gaussian, hence so are $X(t_1), \ldots, X(t_n)$: the finite-dimensional distributions are *multivariate normal*. Recall that the multivariate normal law in n dimensions, $N_n(\mu, \Sigma)$ is specified by the mean vector μ and the covariance matrix Σ (non-negative definite). So to check the finite-dimensional distributions of BM – stationary independent increments with $W_t \sim N(0,t)$ – it suffices to show that they are multivariate normal with mean zero and covariance $cov(W_s, W_t) = \min(s,t)$ as above.

Construction of BM.

It suffices to construct BM for $t \in [0,1]$). This gives $t \in [0,n]$ by dilation, and $t \in [0,\infty)$ by letting $n \to \infty$. First, take $L^2[0,1]$, and any complete orthonormal system (cons) (ϕ_n) on it. Now L^2 is a Hilbert space, under the inner product

$$\langle f, g \rangle = \int_0^1 f(x)g(x)dx \qquad (\text{or } \int fg),$$

so norm $\|f\| := (\int f^2)^{1/2})$. By Parseval's identity,

$$\int_0^1 fg = \sum_{n=0}^{\infty} \langle f, \phi_n \rangle \langle g, \phi_n \rangle$$

(where convergence of the series on the right is in L^2, or in mean square: $\|f - \sum_0^n \langle f, \phi_k \rangle \phi_k\| \to 0$ as $n \to \infty$). Now take, for $s, t \in [0,1]$,

$$f(x) = I_{[0,s]}(x), \qquad g(x) = I_{[0,t]}(x).$$

Parseval's identity becomes

$$\min(s,t) = \sum_{n=0}^{\infty} \int_0^s \phi_n(x)dx \int_0^t \phi_n(x)dx.$$

Now take (Z_n) independent and identically distributed $N(0,1)$ (recall that we can construct these, indeed from one $X \sim U[0,1]$), and write

$$W_t = \sum_{n=0}^{\infty} Z_n \int_0^t \phi_n(x)dx.$$

This is a sum of independent zero-mean random variables. By Kolmogorov's theorem on random series, it converges a.s. if the sum of the variances converges. This is $\sum_{n=0}^{\infty}(\int_0^t \phi_n(x)dx)^2$, $= t$ by above. So the series above converges a.s., and by excluding the exceptional null set from our probability space (as we may), everywhere.

The Haar System. Define

$$H(t) := 1 \text{ on } [0, \frac{1}{2}), \quad -1 \text{ on } [\frac{1}{2}, 1], \quad 0 \text{ else.}$$

Write $H_0(t) \equiv 1$, and for $n \geq 1$, express n in dyadic form as $n = 2^j + k$ for a unique $j = 0, 1, \ldots$ and $k = 0, 1, \ldots, 2^j - 1$. Using this notation for n, j, k throughout, write

$$H_n(t) := 2^{j/2} H(2^j t - k)$$

(so H_n has support $[k/2^j, (k+1)/2^j]$). So if m, n ($m \neq n$) have the same j, $H_m H_n \equiv 0$, while if m, n have different js, one can check that $H_m H_n$ is $2^{(j_1+j_2)/2}$ on half its support, $-2^{(j_1+j_2)/2}$ on the other half, so $\int H_m H_n = 0$. Also H_n^2 is 2^j on $[k/2^j, (k+1)/2^j]$, so $\int H_n^2 = 1$. Combining:

$$\int H_m H_n = \delta_{mn},$$

and (H_n) form an orthonormal system, called the *Haar system.* For completeness: the indicator of any dyadic interval $[k/2^j, (k+1)/2^j]$ is in the linear span of the H_n (difference two consecutive H_ns and scale). Linear combinations of such indicators are dense in $L^2[0, 1]$. Combining: the Haar system (H_n) is a complete orthonormal system in $L^2[0, 1]$.

The Schauder System. We obtain the *Schauder system* by integrating the Haar system. Consider the triangular function (or 'tent function')

$$\Delta(t) := 2t \text{ on } [0, \frac{1}{2}), \quad 2(1-t) \text{ on } [\frac{1}{2}, 1], \quad 0 \text{ else.}$$

Define the *Schauder functions* by $\Delta_0(t) := t$, $\Delta_1(t) := \Delta(t)$,

$$\Delta_n(t) := \Delta(2^j t - k) \qquad (n = 2^j + k \geq 1).$$

Note that Δ_n has support $[k/2^j, (k+1)/2^j]$ (so is 'localized' on this dyadic interval, which is small for n, j large). We find that

$$\int_0^t H(u)du = \frac{1}{2}\Delta(t), \qquad \int_0^t H_n(u)du = \lambda_n \Delta_n(t),$$

where $\lambda_0 = 1$ and for $n \geq 1$,

$$\lambda_n = \frac{1}{2} \times 2^{-j/2} \mathcal{A}(\backslash = \in^{|} + \| \geq \infty).$$

The Schauder system (Δ_n) is again a cons on $L^2[0,1]$.

Theorem (Paley-Wiener-Zygmund, 1932). For $(Z_n)_0^\infty$ independent $N(0,1)$ random variables, λ_n, Δ_n as above,

$$W_t := \sum_{n=0}^{\infty} \lambda_n Z_n \Delta_n(t)$$

converges uniformly on $[0,1]$, a.s. The process $W = (W_t : t \in [0,1])$ is Brownian motion.

Lemma. For Z_n independent $N(0,1)$,

$$|Z_n| \leq C\sqrt{\log n} \qquad \forall n \geq 2,$$

for some random variable $C < \infty$ a.s.

Proof of the Lemma. For $x > 1$,

$$P(|Z_n| \geq x) = \frac{2}{\sqrt{2\pi}} \int_x^\infty e^{-u^2/2} du \leq \sqrt{2/\pi} \int_x^\infty u e^{-u^2/2} du = \sqrt{2/\pi} e^{-x^2/2}.$$

So for any $a > 1$,

$$P(|Z_n| > \sqrt{2a \log n}) \leq \sqrt{2/\pi} \exp\{-a \log n\} = \sqrt{2/\pi} n^{-a}.$$

Since $\sum n^{-a} < \infty$ for $a > 1$, the Borel-Cantelli lemma gives

$$P(|Z_n| > \sqrt{2a \log n} \text{ for infinitely many } n) = 0.$$

So

$$C := \sup_{n \geq 2} \frac{|Z_n|}{\sqrt{\log n}} < \infty \qquad a.s.$$

Proof of the Theorem.
1. Convergence. Choose J and $M \geq 2^J$; then

$$\sum_{n=M}^{\infty} \lambda_n |Z_n| \Delta_n(t) \leq C \sum_M^\infty \lambda_n \sqrt{\log n} \Delta_n(t).$$

The right is majorized by

$$C \sum_{J}^{\infty} \sum_{k=0}^{2^j-1} \frac{1}{2} 2^{-j/2} \sqrt{j+1} \Delta_{2^j+k}(t)$$

(perhaps including some extra terms at the beginning, using $n = 2^j + k < 2^{j+1}, \log n \leq (j+1) \log 2$, and $\Delta_n(.) \geq 0$, so the series is absolutely convergent). In the inner sum, only one term is non-zero (t can belong to only one dyadic interval $[k/2^j, (k+1)/2^j)$), and each $\Delta_n(t) \in [0, 1]$. So

$$LHS \leq C \sum_{j=J}^{\infty} \frac{1}{2} 2^{-j/2} \sqrt{j+1} \qquad \forall t \in [0, 1],$$

and this tends to 0 as $J \to \infty$, so as $M \to \infty$. So the series $\sum \lambda_n Z_n \Delta_n(t)$ is absolutely and uniformly convergent, a.s. Since continuity is preserved under uniform convergence and each $\Delta_n(t)$ is continuous, W_t is continuous in t.
2. Covariance. By absolute convergence and Fubini's theorem,

$$E(W_t) = E\left(\sum_0^{\infty} \lambda_n Z_n \Delta_n(t)\right) = \sum \lambda_n \Delta_n(t) E(Z_n) = \sum 0 = 0.$$

So the covariance is

$$E(W_s W_t) = E\left[\sum_m Z_m \int_0^s \phi_m \times \sum_n Z_n \int_0^t \phi_n\right] = \sum_{m,n} E[Z_m Z_n] \int_0^s \phi_m \int_0^t \phi_n,$$

or as $E[Z_m Z_n] = \delta_{mn}$, the Parseval calculation above gives

$$\sum_n \int_0^s \phi_m \int_0^t \phi_n = \min(s, t).$$

3. Joint Distributions. Take $t_1, \ldots, t_m \in [0, 1]$; we have to show that $(W(t_1), \ldots, W(t_n))$ is multivariate normal, with mean vector 0 and covariance matrix $(\min(t_i, t_j))$. The multivariate characteristic function is

$$E\left(\exp\left\{i \sum_{j=1}^m u_j W(t_j)\right\}\right) = E\left(\exp\left\{i \sum_{j=1}^m u_j \sum_{n=0}^{\infty} \lambda_n Z_n \Delta_n(t)\right\}\right),$$

which by independence of the Z_n is

$$\prod_{n=0}^{\infty} E\left(\exp\left\{i \lambda_n Z_n \sum_{j=1}^m u_j \Delta_n(t_j)\right\}\right).$$

Since each Z_n is $N(0,1)$, the right-hand side is

$$\prod_{n=0}^{\infty} \exp\left\{-\frac{1}{2}\lambda_n^2\left(\sum_{j=1}^{m} u_j\Delta_n(t_j)\right)^2\right\} = \exp\left\{-\frac{1}{2}\sum_{n=0}^{\infty}\lambda_n^2\left(\sum_{j=1}^{m} u_j\Delta_n(t)\right)^2\right\}.$$

The sum in the exponent on the right is

$$\sum_{n=0}^{\infty}\lambda_n^2\sum_{j=1}^{m}\sum_{k=1}^{m} u_j u_k\Delta_n(t_j)\Delta_n(t_k) = \sum_{j=1}^{m}\sum_{k=1}^{m} u_j u_k\sum_{n=0}^{\infty}\int_0^{t_j} H_n(u)du\int_0^{t_k} H_n(u)du,$$

giving

$$\sum_{j=1}^{m}\sum_{k=1}^{m} u_j u_k \min(t_j, t_k),$$

by the Parseval calculation, as (H_n) are a cons. Combining,

$$E\left(\exp\left\{i\sum_{j=1}^{m} u_j W(t_j)\right\}\right) = \exp\left\{-\frac{1}{2}\sum_{j=1}^{m}\sum_{k=1}^{m} u_j u_k \min(t_j, t_k)\right\}.$$

This says that $(W(t_1), \ldots, W(t_n))$ is multinormal with mean 0 and covariance function $\min(t_j, t_k)$ as required. This completes the construction of BM, and the proof of the Theorem. //

9. Quadratic Variation of Brownian Motion

Recall that a $N(\mu, \sigma^2)$ distributed random variable ξ has moment-generating function

$$M(t) := E\left(\exp\{t\xi\}\right) = \exp\left\{\mu t + \frac{1}{2}\sigma^2 t^2\right\}.$$

We take $\mu = 0$ below; we can recover the general case by adding μ back on. So, for ξ $N(0, \sigma^2)$ distributed,

$$M(t) = \exp\left\{\frac{1}{2}\sigma^2 t^2\right\} = 1 + \frac{1}{2}\sigma^2 t^2 + \frac{1}{2!}\left(\frac{1}{2}\sigma^2 t^2\right)^2 + O(t^6)$$

$$= 1 + \frac{1}{2!}\sigma^2 t^2 + \frac{3}{4!}\sigma^4 t^4 + O(t^6).$$

As the Taylor coefficients of the moment-generating function are the moments (hence the name moment-generating function!), $E(\xi^2) = var(\xi) = \sigma^2$,

$E(\xi^4) = 3\sigma^4$, so $var(\xi^2) = E(\xi^4) - [E(\xi^2)]^2 = 2\sigma^4$. For W Brownian motion on $\mathbf{R}$, this gives

$$E(W(t)) = 0, \qquad var(W(t)) = E((W(t)^2) = t, \qquad var(W(t)^2) = 2t^2.$$

In particular, for $t > 0$ small, this shows that the variance of $W(t)^2$ is negligible compared with its expected value. Thus, the randomness in $W(t)^2$ is negligible compared to its mean for t small. This suggests that if we take a fine enough partition $\mathcal{P}$ of $[0,t]$ – a finite set of points $0 = t_0 < t_1 < \ldots < t_n = t$ with grid mesh $\|\mathcal{P}\| := \max|t_i - t_{i-1}|$ small enough – then writing $\Delta W(t_i) := W(t_i) - W(t_{i-1})$ and $\Delta t_i := t_i - t_{i-1}$,

$$\sum_{i=1}^{n} (\Delta W(t_i))^2$$

will closely resemble

$$\sum_{i=1}^{n} E((\Delta W(t_i))^2) = \sum_{i=1}^{n} \Delta t_i = \sum_{i=1}^{n} (t_i - t_{i-1}) = t.$$

This is in fact true:

$$\sum_{i=1}^{n} (\Delta W(t_i))^2 \to \sum_{i=1}^{n} \Delta t_i = t \qquad \text{in probability} \qquad (\max|t_i - t_{i-1}| \to 0).$$

This limit is called the *quadratic variation* of W over $[0,t]$.

Start with the formal definitions. A *partition* π_n of $[0,t]$ is a finite set of points t_{ni} such that $0 = t_{n0} < t_{n1} < \ldots < t_{n,k(n)} = t$; the *mesh* of the partition is $|\pi_n| := \max_i(t_{ni} - t_{n,(i-1)})$, the maximal subinterval length. We consider *nested* sequences (π_n) of partitions (each refines its predecessors by adding further partition points), with $|\pi_n| \to 0$. Call (writing t_i for t_{ni} for simplicity)

$$\pi_n B := \sum_{t_i \in \pi_n} (W(t_{i+1}) - W(t_i))^2$$

the *quadratic variation* of W on (π_n). The following classical result is due to Lévy (in his book of 1948); see e.g. [P], I.3.

Theorem (Lévy). The quadratic variation of a Brownian path over $[0,t]$ exists and equals t, in mean square (and hence in probability):

$$\langle W \rangle_t = t.$$