Stochastic Calculus Notes, Lecture 6

Last modified October 31, 2002

1 Integration with respect to Brownian Motion

While integrals of functions of Brownian motion paths are not hard to define, integrals with respect to Brownian motion do give trouble. In fact, there is some ambiguity about what the integral should be. The Ito integral is really just a convention to choose one of the several possibilities. The Ito convention is that the "stochastic integral" with respect to Brownian motion should be a martingle.

Many financial models take the form of stochastic differential equations (SDE). The definition of the solution of an SDE has the same ambiguity as the stochastic integral with respect to Brownian motion. We again choose the Ito convention that the solution as far as possible should be a martingale. In fact, the solution of an Ito SDE is defined in terms of the Ito integral.

1.1. Integrals involving a function of t only: The stochastic integral with respect to Brownian motion is an integral in which dX_t (whatever that means) plays the role of dt in the Riemann integral. The simplest case involves just a function of t:

$$Y_g = \int_0^T g(t)dX_t \ . \tag{1}$$

This integral is defined in somewhat the same way the Riemann integral is defined. We choose n and $\Delta t = T/n$ and take

$$Y_g^{(n)} = \sum_{k=0}^{n-1} g(t_k)\Delta X_k \ , \tag{2}$$

where $\Delta X_k = X_{t_{k+1}} - X_{t_k}$, and $t_k = k\Delta t$. Since $Y_g^{(n)}$ is a sum of gaussian random variables, it is also gaussian. Clearly $E[Y_g^{(n)}] = 0$. We will understand the limit as $\Delta t \to 0$ (including whether it exists) if we calculate the limit of $\mathrm{var}(Y_g^{(n)} = E[Y_g^{(n)2}$. Since the ΔX_k are independent normals with mean zero and variance Δt, the variance of the sum is

$$\mathrm{var}(Y_g^{(n)}) = \sum_{k=0}^{n-1} g(t_k)^2 \Delta t \ .$$

The right side is the standard Riemann approximation to the integral $\int_0^T g(t)^2 dt$, so letting $\Delta t \to 0$ gives

$$E[Y_g^2] = \mathrm{var}(Y_g) = \int_0^T g(t)^2 dt \ . \tag{3}$$

This may not have seemed so subtle, and it was not. Every reasonable definition of (1) gives the same answer.

1.2. Different kinds of convergence: In the abstract setting we have a probability space, Ω, and a family of random variables $Y_n(\omega)$. We want to take the limit as $n \to \infty$. The limit above is the limit "in distribution". The probability density for Y_n converges to the probability density of a random variable Y. The cental limit theorem is of this kind: the probability density converges to a gaussian. Another kind of convergence as "pointwise", asking that, for each ω (or almost every ω) the limit $\lim_{n\to\infty} Y_n(\omega) = Y(\omega)$ should exist. The difference between these notions is that one gives an actual (function of a) random variable, $Y(\omega)$, while the other just gives a probability density without necessarily saying which Y goes with a particular ω. Proving convergence in distribution for gaussian random variables is easy, just calculate the mean and variance. Note that this does not depend on the joint distribution of Y_n and Y_{n+1}. Proving pointwise convergence requires you to understand the differences $Y_{n+1} - Y_n$, which do depend on the joint distributions.

1.3. Proving pointwise convergence: The abstract setting has a probability space, Ω, with a probability measure, and a sequence of random variables, $X_n(\omega)$. The X_n could be just numbers (what we usually call random variables), or vectors (vector values random variables), or even functions of another variable (say, t). In any of these cases, we have a norm, $\|X\|$. For the case of a number, we just use the absolute value, $|X|$. For vectors, we can use any vector norm. For functions, we can also use any norm, such as the "sup" norm, $\|X(\omega)\| = \max_{0\leq t\leq T} |X(\omega,t)|$, or the L^2 norm, $\|X\|^2 = \int_{0\leq t\leq T} X(\omega,t)^2 dt$. A theorem in analysis says that the limit $\lim_{n\to\infty} X_n(\omega)$ exists if

$$S(\omega) = \sum_{n=1}^{\infty} \|X_{n+1}(\omega) - X_n(\omega)\| < \infty \; . \tag{4}$$

This is easy to understand. The limit exists if and only if the sum, $X_1(\omega) + \sum_n X_{n+1}(\omega) - X_n(\omega)$, converges. The condition (4) just says that this sum converges absolutely. It is possible that the limit exists even though S is infinite. For example (forgetting ω) if $X_n = (-1)^n/n$.

The limit will exist for (almost) every ω if $S(\omega) < \infty$ for (almost) every ω. We know that $S < \infty$ almost surely if $E[S] = \int S(\omega)dP(\omega) < \infty$. The expected value criterion is useful because we might be able to calculate the expected value, particularly in a Stochastic Calculus class that is devoted mostly to such calculations. Of course, it is possible that $S < \infty$ almost surely even though $E[S] = \infty$. For example, suppose $S = 1/Z^2$ where Z is a standard normal $S \sim \mathcal{N}(0,1)$ (OK, not likely for (4), but that does not change this point). In jargon, we would say these criteria are not "sharp"; it is possible to fail these tests and still converge. As far as I can tell, a sharp criterion would be much more complicated, and unnecessary here. From (4), the criterion $E[S] < \infty$

may be stated:

$$\sum_{n=1}^{\infty} E\Big[\|X_{n+1} - X_n\|\Big] < \infty \; . \tag{5}$$

We continue our succession of convenient but not sharp criteria. It is often easier to calculate $E[Y^2]$ than $E[|Y\|]$. Fortunately, there is the "Cauchy Schwartz" inequality: $E[|Y|] < E[Y^2]^{1/2}$ (proof left to the reader). If we define (and hope to calculate)

$$s_n^2 = E\Big[\|X_{n+1} - X_n\|^2\Big] \; ,$$

then $E\Big[\|X_{n+1} - X_n\|\Big] < s_n$, so (5) holds if

$$\sum_{n=1}^{\infty} s_n < \infty \; . \tag{6}$$

1.4. The integral as a function of X: We apply the above criteria to showing that the limit (2) exists for (almost) any Brownian motion path, X. Pointwise convergence does two things for us. First, it shows that Y_g is a function of X, i.e., a random variable defined on the probability space of Brownian motion paths. Second, it shows that shows that if we use the approximation (2) on the computer, we will get an approximation to the right $Y_g(X)$, not just a random variable with (approximately) the right distribution. Whether that is important is a subject of heated debate, with me heatedly on one of the sides. We will see that it is much easier to compare $Y_g^{(n)}$ with $Y_g^{(2n)}$ than with $Y_g^{(n+1)}$. To translate from our situation to the abstract, the abstract X_n will be our $Y_g^{(2^L)}$, the abstract n our L, and the abstract ω our X. That is, we seek to show that the limit

$$\lim_{L\to\infty} Y_g^{(2^L)}(X) = Y_g(X)$$

exists for (almost) every Brownian motion path, X. We will do this by calculating (bounding would be a more apt term) $E[(Y_g^{(2n)} - Y_g^{(n)})^2]$ with $n = 2^L$.

1.5. Comparing the Δt and $\Delta t/2$ approximations: (See Assignment 6, question 2 for a slightly different version of this notation). We will fix g and stop writing it. We have $Y^{(n)}$ based on $\Delta t = T/n$ and $Y^{(2n)}$ based on $\Delta t/2 = T/(2n)$. We take $t_k = k\Delta t$, which is appropriate for $Y^{(n)}$. The contribution to $Y^{(n)}$ from the interval (t_k, t_{k+1}) is $g(t_k)(X_{t_{k+1}} - X_{t_k})$. For $Y^{(2n)}$, the interval (t_k, t_{k+1}) is divided into two subintervals $(t_k, t_{k+1/2})$ and $(t_{k+1/2}, t_{k+1})$, using the notation $t_{k+1/2} = t_k + \Delta t/2 = (k + 1/2)\Delta t$. The the contribution to $Y^{(2n)}$ from these two intervals added is

$$g(t_k)(X_{t_{k+1/2}} - X_{t_k}) + g(t_{k+1/2})(X_{t_{k+1}} - X_{t_{k+1/2}}) \; .$$

Define ΔY_k to be the difference between the single $Y^{(n)}$ contribution and the two $Y^{(2n)}$ contributions from the interval (t_k, t_{k+1}), so that $Y^{(2n)} - Y^{(n)} = \sum_{k=0}^{n-1} \Delta Y_k$. A calculation gives

$$\Delta Y_k = (g(t_{k+1}) - g(t_{k+1/2}))(X_{t_{k+1}} - X_{t_{k+1/2}}) \ .$$

Only the X values are random, and increments $X_{t_{k+1}} - X_{t_{k+1/2}}$ from distinct intervals are independent. Therefore

$$E\left[(Y^{(2n)} - Y^{(n)})^2\right] = \sum_{k=0}^{n-1} E[\Delta Y_k^2] = \sum_{k=0}^{n-1} \Delta g_k^2 \Delta t/2 \ ,$$

where we have used the notation $\Delta g_k = g(t_{k+1}) - g(t_{k+1/2})$ and the fact that $E[(X_{t_{k+1}} - X_{t_{k+1/2}})^2] = \Delta t/2$.

Now suppose that $|g'(t)| \leq r$ for all t. Then $\Delta g_k \leq \frac{r\Delta t}{2}$ (an interval of length $\frac{\Delta t}{2}$) so

$$E\left[(Y^{(2n)} - Y^{(n)})^2\right] \leq n\frac{r^2\Delta t^2}{4}\frac{\Delta t}{2} \ .$$

Simplifying using the relationship $n\Delta t = T$ gives

$$E\left[(Y^{(2n)} - Y^{(n)})^2\right] \leq T\frac{r^2\Delta t^2}{8} \ .$$

Finally, take n to be of the form 2^L, write $Y_L = Y^{(2^L)}$, and see that we have shown $s_L^2 \leq \mathit{Const} \cdot \Delta t^2$, so $s_L \leq \mathit{Const} \cdot 2^-L$, and the criterion (6) is easily satisfied.

1.6. Unanswered theoretical questions: Here are some questions that would be taken up in a more theoretical course and their answers, without proof. *Q1*: This defines Y_g only for functions $g(t)$ that are differentiable. What about other functions? *A1*: Because $E[Y_g^2] = \int_0^T g(t)^2 dt$, we can "extend" the mapping $g \mapsto Y_g$ to any g with $\int g^2 < \infty$, as we do for the Fourier transform. *Q2*: What happens if we let $n \to \infty$ but not by powers of 2? *A2*: This can be done in at least two ways, eigher using a more sophisticated argument and higher than second order moments, or by using a uniqueness theorem for the limit. Even without this, we met our primary goal of showing that Y_g is a well defined function of X.

1.7. White noise: White noise is something of an idealization, like the $\delta-$function. Imagine a function, $W(t)$ that is gaussian with mean zero and has $W(t)$ independent of $W(s)$ for $t \neq s$. Also imagine that the strength of the noise is independent of time. This is a common model for fluctuations. For example, in modeling phone calls, we may think that the rate of new calls being initiated fluctuates from its mean but that fluctuations at different times are independent. Suppose we try to integrate white noise over intervals of time: $Y_{[a,b]} = \int_a^b W(t)dt$. We can determine how the variance $\sigma_{[a,b]}^2 = E[Y_{[a,b]}^2]$

depends on the interval by noting that Y variables for disjoint intervals should be independent. In particular, if $a < b < c$ we have $Y_{[a,c]} = Y_{[a,b]} + Y_{[b,c]}$, so $\sigma^2_{[a,c]} = \sigma^2_{[a,b]} + \sigma^2_{[b,c]}$. The only this can happen, and have, for any offset, d, $\sigma^2_{[a,b]} = \sigma^2_{[a+d,b+d]}$ (homogeneous in time) is for $\sigma^2_{[a,b]} = Const \cdot (b-a)$. The "standard" white noise has $\sigma^2_{[a,b]} = b - a$.

1.8. White noise is not a function: White noise is too rough to be a function, even a random function, in the usual sense. To see this, consider an interval $(0, \epsilon)$. Since $\int_0^\epsilon W(t)dt$ has variance ϵ, it's standard deviation, which is the order of magnitude of a typical $Y_{[0,\epsilon]}$, is $\sqrt{\epsilon}$. In order to have $\int_0^\epsilon W(t)dt \sim \sqrt{\epsilon}$, we must have $W(t) \sim 1/\sqrt{\epsilon}$ in at least over a reasonable fraction of the interval. Letting $\epsilon \to 0$, we see that $W(t)$ should have infinite values almost everywhere, not much of a function. Just as the $\delta-$function is defined in an abstract way as a measure, there are abstract definitions that allow us to make sense of white noise.

Another way to see this is to try to define $Y_T = \int_0^T W(t)^2 dt$. Since we already think white noise paths are discontinuous, it is natural to try to define the Riemann sum using averages over small intervals rather than values $W(t_k)$. We call the averages

$$W_{k,n} = \frac{1}{\Delta t} \int_{t_k}^{t_{k+1}} W(t)dt \; .$$

The approximation to Y_Y is

$$Y_T^{(n)} = \Delta t \sum_{k=0}^{n-1} W_{k,n}^2 \; .$$

The random variables $W_{k,n}$ are independent gaussians with mean zero and variance $\frac{1}{\Delta t^2}\Delta t = \frac{1}{\Delta t}$. Therefore the $W_{k,n}^2$ are independent with mean $\frac{1}{\Delta t}$ and variance $\frac{2}{\Delta t^2}$ (as the reader should verify). Therefore, $Y_T^{(n)}$ has mean $T/\Delta t$ and standard deviation $\sqrt{2T/\Delta t}$. Clearly, as $n \to \infty$, $Y_T^{(n)} \to \infty$. In other words, $\int_0^T W(t)^2 dt = \infty$ by the most reasonable definition.

1.9. White noise and Brownian motion: The integrals $Y_{[a,b]}$ of white noise have the same statistical properties as the increments of Brownian motion. The joint distribution of $Y_{[a_1,b_1]}$, ..., $Y_{[a_n,b_n]}$ is the same as the joint distribution of the increments $X_{b_1} - X_{a_1}$, ..., $X_{b_n} - X_{a_n}$ (assuming, though this is not necessary, that $a_1 \leq b_1 \leq a_2 \cdots \leq b_n$): both are multivariate normal with zero covariances and variances $b_k - a_k$. If $W(t)$ were a function, this would lead us to write the three relationships

$$X_t = \int_0^t W(s)ds \;\; , \quad \frac{dX_t}{dt} = W(t) \;\; , \quad dX_t = W(t)dt \; . \tag{7}$$

Any of these may be taken as the definition of white noise. This is probably the main reason most people (who are interested) are interested in Brownian motion, that it gives a mathematically rigorous and systematic way to make sense of white noise.

1.10. Correlations of integrals with respect to Brownian motion: It seems clear that two integrals with respect to Brownian motion should be jointly gaussian with some covariance we can calculate. In fact, if $Y_f = \int_0^T f(t)dX_t$ and $Y_g = \int_0^T g(t)dX_t$, then the approximations $Y_f^{(n)}$ and $Y_g^{(n)}$ are jointly normal and have covariance $\sum_{k=0}^{n-1} f(t_k)g(t_k)\Delta t$. Taking the limit $\Delta t \to 0$ gives

$$\mathrm{cov}(Y_f, Y_g) = \int_{t=0}^{T} f(t)g(t)dt \; . \tag{8}$$

1.11. δ correlated white noise: The correlation formula (8) has an interpretation used by 90% of the interested world, not including most mathematicians. If we write $Y_g = \int_{s=0}^{T} g(s)dX_s$ and formally interchange the order of integration, we get, since dX_t and dX_s are the only random variables,

$$\begin{aligned} E[Y_f Y_g] &= E\left[\int_{t=0}^{T} f(t)dX_t \int_{s=0}^{T} g(s)dX_s\right] \\ &= \int_{t\in[0,T]} \int_{s\in[0,T]} f(t)g(s)E[dX_t dX_s] \; . \end{aligned}$$

We get (8) with the rule

$$E[dX_t dX_s] = \delta(t-s)dt \; . \tag{9}$$

This is the first instance of the informal Ito rule $dX^2 = dt$. It is equivalent to (7) with the rule $E[W(t)W(s)] = \delta(t-s)$, which is another indication that white noise is not a normal function. If we write $W(t)dt = dX_t$ to write $Y_f = \int f(t)W(t)dt$, the formula (**??**) follows.

A useful approximation to white noise with a time step Δt is

$$W^{(\Delta t)}(t) = \sum_k Z_k \mathbf{1}_{I_k}(t) \tag{10}$$

where I_k is the interval $[t_k, t_{k+1}]$ and the Z_k are independent gaussians with the proper variance $\mathrm{var}(Z_k) = \frac{1}{\Delta t}$. For example, this gives

$$\int_0^T f(t)W^{(\Delta t)}dt = \sum_k \int_{I_k} f(t)dt Z_k \; ,$$

which is a random variable practically identical to the approximation (2). The difference is that $\Delta t f(t_k)$ is replaced by $\int_{I_k} f(t)dt = f(t_k)\Delta t + o(\Delta t)$. We identifly the random variables ΔX_k and $\Delta t Z_k$ because they are both multivariate normal and have the same mean ($E[] = 0$), variance, and covariances.

2 Ito Integration

2.1. Forward dX_t: We want to define stochastic integrals such as

$$Y_T = \int_0^T V(X_t) dX_t \; . \tag{11}$$

The Ito convention is that $E[dX_t \mid \mathcal{F}_t] = 0$. When we make $\Delta t = T/n$ approximations to (11), we always do it in a way that makes the anologue of dX_t have conditional expectation zero. For example, we might use

$$Y_T^{(n)} = \sum_{k=0}^{n-1} V(X_{t_k})(X_{t_{k+1}} - X_{t_k}) \; . \tag{12}$$

The specific choice $\Delta X_k = X_{t_{k+1}} - X_{t_k}$ gives $E[\Delta X_k \mid \mathcal{F}_{t_k}] = 0$, which is in keeping with the Ito convention. We will soon show that the limit $Y_T^{(n)}(X) \to Y_T(X)$ exists. This limit is the Ito integral.

2.2. Example 1: This example illustrates the convergence of the approximations, the way in which the Ito integral differs from an ordinary integral, and the fact that other approximations of dX_t lead to different limits. Take

$$Y_T = \int_0^T X_t dX_t \; ,$$

and use the approximation

$$Y_t^{(n)} = \sum_{k=0}^{n-1} X_{t_k}(X_{t_{k+1}} - X_{t_k}) \; .$$

The trick (see any book on this) is to write

$$X_{t_k} = \frac{1}{2}(X_{t_{k+1}} + X_{t_k}) - \frac{1}{2}(X_{t_{k+1}} - X_{t_k}) \; .$$

Now,

$$(X_{t_{k+1}} + X_{t_k})(X_{t_{k+1}} - X_{t_k}) = X_{t_{k+1}}^2 + X_{t_k}^2 \; ,$$

so, using $t_n = T$ and $X_0 = 0$,

$$\begin{aligned} \sum_{k=0}^{n-1} X_{t_k}(X_{t_{k+1}} - X_{t_k}) &= \frac{1}{2}\sum_{k=0}^{n-1}\left(X_{t_{k+1}}^2 - X_{t_k}^2\right) + \frac{1}{2}\sum_{k=0}^{n-1}\left(X_{t_{k+1}} - X_{t_k}\right)^2 \\ &= \frac{1}{2}X_T^2 + \frac{1}{2}\sum_{k=0}^{n-1}\left(X_{t_{k+1}} - X_{t_k}\right)^2 \end{aligned}$$

The second term on the right is the sum of a large number of independent terms with the same distribution, and mean $\frac{1}{2}E[\Delta X_k^2] = \frac{\Delta t}{2}$. Thus, the second term is approximately $\frac{n\Delta t}{2} = \frac{T}{2}$. Letting $\Delta t \to 0$, we get

$$\int_0^T X_t dX_t = \frac{1}{2}X_t^2 - \frac{1}{2}T \ .$$

This is one of the martingales we saw earlier. The Ito integral (11) always gives a martingale, as we will see.

2.3. Other definitions of the stochastic integral give different answers: A sensible person might suggest other approximations to (11). With $I_k = [t_k, t_{k+1}]$, we approximated $\int_{I_k} V(X_t)dX_t$ by $V(X_{t_k})(X_{t_{k+1}} - X_{t_k})$, which seems like the rectangle rule for ordinary integration. What would happen if we try the trapezoid rule,

$$\text{(Wrong!)} \qquad \int_{I_k} V(X_t)dX_t \approx \frac{1}{2}[V(X_{t_k}) + V(X_{t_{k+1}})](X_{t_{k+1}} - X_{t_k}) \ \ ?$$

The reader should check that in the example $V(x) = x$ above this would give

$$\text{(Wrong!)} \qquad \int_0^T X_t dX_t = \frac{1}{2}X_t^2 \ \ .$$

Also, if X_t were a differentiable function of t, with derivative $\frac{dX_t}{dt} = W(t)$, we could write

$$\text{(Wrong!)} \qquad \int_0^T X_t dX_t = \int_0^T X_t \frac{dX_t}{dt} dt = \frac{1}{2}\int_0^T \frac{dX_t^2}{dt} dt = X_T^2/2 \ .$$

From this it seems that the Ito calculus is different from ordinary calculus because the function X_t is not differentiable in the ordinary sense. The derivative, white noise, is not a function in the ordinary sense.

2.4. Convergence and existance of the integral (11): We show that the approximation (12) converges to something as $\Delta t \to 0$ (really $\Delta t = T/2^k$, $k \to \infty$), assuming that V is "Lipschitz continuous": $|V(x) - V(x')| \leq C\,|x - x'|$. For example, $V(x)$ would be Lipschitz continuous if V' were a bounded function. The convergence is again "pointwise"; the event that the approximations do not converge has probability zero. As in paragraph 1.5, we compare the contributions from interval $I_k = [k\Delta t, (k+1)\Delta t]$ when we have Δt, corresponding to $n = 2^L$ subintervals, and $\Delta t/2$ corresponding to $2n = 2^{L+1}$ intervals. For Δt there is just

$$\int_{t\in I_k} V(X_t)dX_t \approx V(X_k)(X_{k+1} - X_k) \ .$$

We use the shorthand X_k for X_{t_k}, and below, $X_{k+1/2}$ for $X_{(k+1/2}\Delta t$. For $\Delta t/2$ there are two contributions:

$$\int_{t\in I_j} V(X_t)dX_t \approx V(X_k)(X_{k+1/2} - X_k) + V(X_{k+1/2})(X_{k+1} - X_{k+1/2}) \ .$$

The difference between these is

$$D_j = V(X_{k+1/2} - V(X_k))(X_{k+1} - X_{k+1/2}) \ .$$

Therefore, using the old double summation trick,

$$\begin{aligned} s_L^2 &= E\left[\left(Y_T^{(2n)} - Y_T^{(n)}\right)^2\right] \\ &= E\left[\left(\sum_{k=0}^{n-1} D_k\right)^2\right] \\ &= \sum_{j=0}^{n-1}\sum_{k=0}^{n-1} E[D_j D_k] \ . \end{aligned}$$

The terms with $j \neq k$ are zero. Suppose, for example, that $k > j$. Then $E[(X_{k+1} - X_{k+1/2}) \mid \mathcal{F}_{k+1/2}] = 0$, so $E[D_j, D_k] = 0$. When V is Lipschitz continuous,

$$E[D_k^2] \leq C^2 E[(X_{k+1} - X_{k+1/2})^2 (X_{k+1/2} - X_k)^2] = C^2 \Delta t^2/2.$$

since there are $n = 2^L$ terms, this gives $s_L^2 \leq n\Delta t \leq C^2\Delta t/4 = C^2 T^2 2^{-L}/4$, so $\sum_L s_L < \infty$, which implies pointwise convergence.

2.5. How continuous are Brownian motion paths: We know that Brownian motion paths are continuous but not differentiable. The total variation, the total distance travelled (not the net distance $|X_{t'} - X_t|$), is infinite for any interval. To understand the accuracy and convergence of approximations like (12), we would like some positive quantitative measure of continuity of Brwonian motion paths. One positive statement is "Hölder continuity". The function $f(t)$ is Hölder continuous with exponent α if there is some C so that

$$|f(t') - f(t)| \leq C \left|t' - t\right|^\alpha \ ,$$

for any t and t'. Only exponents between larger than zero and not more than one are relevent. Exponent $\alpha = 1$ is for Lipschitz continuous functions. A larger α means a more regular function. Besides Brownian motion, fractals such as the Koch snowflake and the space filling curve are other examples of natural Hölder continuous functions. The function $f(t) = -1/\log(t)$ is continuous at $t = 0$ but not Hölder continuous there. The exponent $\alpha = 1/2$ seems natural for Brownian motion because (see the discussion of total variation and quadratic variation)

$$E[|X_{t'} - X_t|] \sim \left|t' - t\right|^{1/2} \ .$$

Actually, this is just slightly optimistic. It is possible to prove, using the Brownian bridge construction (upcoming) that Brownian motion paths are Hölder continuous with any positive exponent less than 1/2:

Lemma: For any positive $\alpha < 1/2$, every $T > 0$, and (almost) every Brownian motion path X_t, there is a C_X so that

$$|X_{t'} - X_t| \leq C_X \left| t' - t \right|^\alpha \ .$$

for all $t \leq T$ and $t' \leq T$. Furthermore, $E[C_X] < \infty$.
Remark: The proof of this lemma is really a calculation of (an upper bound for) $E[C_X]$.

2.6. Ito integration with nonanticipating functions: Ito wants to integrate more general functions than $V(X_t)$ with respect to Brownian motion. For example, he might want to calculate

$$\int_0^T \left(\max_{s<t} X_s \right) dX_t \ ,$$

or the iterated integral

$$\int_0^T \left(\int_0^t X_s^2 dX_s \right) dX_t \ .$$

Therefore, we consider the more general Ito integral

$$Y_T = \int_0^T V_t dX_t \ , \tag{13}$$

where, for each t, V_t is measurable with respect to $\mathcal{F}_t$. Such functions are called "adapted" or "nonanticipating" or "causal" (possible subtle distinctions between these notions go unmentioned here). Nonanticipating functions are important in studying stochastic decision problems; we are supposed to make decisions at time t based on information in $\mathcal{F}_t$. Martha Stewart can explain the consequences of violating this rule, or appearing to do so. The examples above have

$$V_t = \max_{s<t} X_s$$

and

$$V_t = \int_0^t X_s^2 dX_s$$

respectively, both measurable in $\mathcal{F}_t$. Of course, V_t is a function of X also (ω in the abstract description), but as usual we do not indicate that explicitly.

We can show that integrals as general as (13) exist by showing that approximations

$$Y_T^{(n)} = \sum_{j=0}^{n-1} V_{t_j} (X_{j+1} - X_j) \tag{14}$$

converge as $\Delta t \to 0$. The argument in paragraph 2.4 works fine for this purpose if you assume that V_t is a Hölder continuous function of t (with $E[C^2] < \infty$, C being the Hölder exponent). Because we might want $V_t = X_t$ (the case of paragraph 2.4), we should allow Hölder exponents less than 1/2. As before, the difference between the Δt and $\Delta t/2$ approximations is

$$Y_T^{(2n)} - Y_T^{(n)} = \sum_{k=0}^{n-1} D_k \ ,$$

with (in the same shorthand notation)

$$D_k = \left(V_{k+1/2} - V_k\right)\left(X_{k+1} - X_{k+1/2}\right) \ .$$

Again, because V is nonanticipating, $E[D_i D_j] = 0$ if $i \neq j$. Also,

$$E[D_k^2] \leq E[C^2](\Delta t/2)^{2\alpha}\Delta t/2 \ ,$$

which proves convergence as before.

2.7. Further extension, the Ito isometry: A mapping is an isometry if distances are the same before and after the mapping is applied. For example, rigid rotations of three dimensional space are isometries; the distance between a pair of points is the same before and after the transformatin is applied. The formula (3) is the first shows that the mapping $g \mapsto Y_g$ is an isometry in the sense that if the distance between g_1 and g_2 is

$$\|g_1 - g_2\|^2 = \int_0^T (g_1(t) - g_2(t))^2 dt \ ,$$

and the distance between random variables (functions of a random variable) $Y_1(\omega)$ and $Y_2(\omega)$ is

$$\|Y_1 - Y_2\|^2 = E[(Y_1 - Y_2)^2] = \int_\Omega (Y_1(\omega) - Y_2(\omega))^2 dP(\omega) \ ,$$

then we have the isometry (which is just a restatement of (3))

$$\|Y_{g_1} - Y_{g_2}\|^2 = \|g_1 - g_2\|^2 \ .$$

Since the mapping is linear, this is the same (just take $g = g_1 - g_2$) as showing that

$$\|Y_g\|^2 = \|g\|^2 \ .$$

Ito showed that his stochastic integral is an isometry in the same sense. The left side is the same, and the right side is related to $\int_0^T V_t^2 dt$. The difference is that the latter integral is random. The final Ito isometry is, using $Y_t(V)$ to indicate that Y_T depends on the function V:

$$E\left[\left(Y_t(V)\right)^2\right] = \int_0^T E[V_t^2]dt \ . \tag{15}$$

It is easy to verify this identity using the approximations (14) as usual. The approximations (14) might converge to something as $\Delta t \to 0$ even when V_t is not nonanticipatint (i.e. anticipating?), but it is very unlikely that the limit would satisfy the Ito isometry.

The isometry formula is useful in practical calculations (see assignment 7). It also has several applications in the theory. One theoretical application is in showing that the mapping $V_t \mapsto Y_T(V)$ may be defined for any nonanticipating V so that the right side of (15) is finite. For any such V and any ϵ, we must find a $V^{(\epsilon)}$ so that $V^{(\epsilon)}$ is Hölder continuous in the sense we need, and so that $\int_0^T E[(V_t - V_t^{(\epsilon)})^2] \leq \epsilon$. The Ito isometry formula then shows that, if Y_T were to exist, $E[(Y_T - Y^{(\epsilon)})^2] \leq \epsilon$, where $Y_T^{(\epsilon)}$ is the Ito integral of $V^{(\epsilon)}$. From this, it is possible to show that the $Y_T^{(\epsilon)}$ do have a limit as $\epsilon \to 0$ (in a certain sense), which is the desired Y_T.

2.8. Martingale property: As a function of T, the Ito integral is a martingale. We can see this from the approximations (12). If we fix Δt and let T vary, it is clear that $Y_T^{(n)}$ is a martingale, since each of the increments, $V_{t_n}\left(X_{t_{n+1}} - X_{t_n}\right)$, has mean zero when projected onto funcions measurable in $\mathcal{F}_{t_n}$. Actually, I'm cheating a bit here since Δt was supposed to depend on T, but hopefully the idea is clear. The Ito isometry formula is an expression of the martingale property. If Z_n is a discrete time martingale with "martingale differences" $W_n = Z_n - Z_{n-1}$, then (with the convention that $W_0 = Z_0$)

$$Z_n = \sum_{k=0}^{n} W_k \ . \tag{16}$$

The martingale property is that $E[W_k \mid \mathcal{F}_j] = 0$ if $k > j$. Therefore, $E[W_k W_j] = 0$ for $k \neq j$ (we may as well suppose $k > j$, why?). Thus $E[Z_n^2] = \sum_{k=0}^n E[W_k^2]$. In the Ito integral may be thought of as a continuous time version of (16), with $V_t dX_t$ playing the role of W_k, and the integral playing the role of the sum. Corresponding to $E[W_k W_j] = 0$, we have $E[V_s dX_s V_t dX_t] = \delta(t-s) E[V_t^2]$, which leads to the Ito isometry formula.