This section discusses the Ito integral. For this purpose, standard Brownian motion will be called[1] W_t. Recall that this is a Gaussian process with the independent increments property and continuous sample paths. The *standard* Brownian motion, the one we use all the time unless there is some explicit statement otherwise, has initial position $W_0 = 0$, zero mean, and variance equal to t: $E[W_t] = 0$ and $E\left[W_t^2\right] = t$ for all $t > 0$.

We let $\mathcal{F}_t$ be the corresponding filtration. That means that $\mathcal{F}_t$ contains all information about the path W_s for $0 \leq s \leq t$, or $\mathcal{F}_t$ is generated by the path up to time t, which may be denoted $W_{[0,t]}$. The increment of Brownian motion between time t and $t' > t$ is $W_{t'} - W_t$. The independent increments property of Brownian motion states that this increment is independent of $\mathcal{F}_t$. A consequence of independent increments is

$$E\left[W_{t'} - W_t \mid \mathcal{F}_t\right] \;=\; 0 \;. \tag{1}$$

Since $W_{t'} - W_t$ is independent of $\mathcal{F}_t$, conditioning on $\mathcal{F}_t$ does not change the expected value, which is zero. The formula (1) makes Brownian motion a *martingale*.

Let X_t be another stochastic process that is *adapted* to the same filtration $\mathcal{F}_t$. Adapted means that the value of X_t is known if the path $W_{[0,t]}$ is known. Therefore, if we write $X_t = X_t(W)$, then there is a function (also called X_t) $X_t(W_{[0,t]})$. Stochastic processes with this property also are called *progressively measurable*, or *non-anticipating*. Strictly speaking, these three terms have slightly different meanings, a difference even most theoretical probabilists pay little attention to. We say that the process X_t is a martingale if

$$E\left[X_{t'} \mid \mathcal{F}_t\right] \;=\; X_t \;. \tag{2}$$

The formula (1) simply states that Brownian motion is a martingale.

Now let F_t be another random process adapted to the filtration $\mathcal{F}_t$. The Ito integral is written

$$X_t \;=\; \int_0^t F_s \, dW_s \;. \tag{3}$$

This defines a stochastic process X_t, which also turns out to be adapted to $\mathcal{F}_t$. The Ito integral allows us to define stochastic processes with specified properties in terms of Brownian motion. For example, the solution to a *stochastic differential equation* involves an Ito integral. Stochastic differential equations are the important way to model continuous time dynamical systems subject to random noise, if the noise is continuous.

There are many adapted functions that come up in applications. A simple one is $F_t = W_t$. More complicated is

$$F_t \;=\; \max_{0 \leq s \leq t} W_s \;.$$

[1]The mathematics of Brownian motion was worked out by Norbert Wiener, so Brownian motion is also known as the Wiener process.

Option contracts based on this are *lookback* options. Another possibility is

$$F_t = \int_0^t W_s \, ds \; .$$

In general, integrating an adapted function (Ito or Riemann integral) gives another adapted function. Options that depends on such integrals are *Asian* options. In each case, the value F_t is determined by $W_{[0,t]}$.

The Ito integral (3) is defined as a limit of *Ito-Riemann sums* much in the way the Riemann integral is defined using Riemann sums. We define a small *time step*, Δt, and anticipate the limit $\Delta t \to 0$. The uniformly spaced discrete times are $t_k = k\Delta t$. The approximation to (3) is

$$X_t^{\Delta t} = \sum_{t_k \leq t} F_{t_k} \left(W_{t_{k+1}} - W_{t_k} \right) \; . \tag{4}$$

It is crucial in this definition that the Brownian motion increment $\Delta W_k = W_{t_{k+1}} - W_{t_k}$ is in the future of t_k. This makes the increment ΔW_k independent of F_{t_k}. That, in turn, makes $X_t^{\Delta t}$ into a martingale.

To see this we have to check the martingale property (2) for the process (4). If $t' > t$, then

$$X_{t'}^{\Delta t} - X_t^{\Delta t} = \sum_{t < t_k \leq t'} F_{t_k} \left(W_{t_{k+1}} - W_{t_k} \right) \; .$$

The martingale property asks us to take the conditional expectation, conditioned on $\mathcal{F}_t$. The terms on the right are easier to understand conditioned on $\mathcal{F}_{t_k}$. Therefore, we use the tower property, which is predicated on $t_k \geq t$ (which is true here). Let Y_k be the conditional expectation given more information:

$$Y_k = E\left[F_{t_k} \left(W_{t_{k+1}} - W_{t_k} \right) \mid \mathcal{F}_{t_k} \right] \; .$$

The tower property implies that

$$E\left[F_{t_k} \left(W_{t_{k+1}} - W_{t_k} \right) \mid \mathcal{F}_t \right] = E\left[Y_k \mid \mathcal{F}_t \right] \; .$$

On the other hand, F_{t_k} is known at time t_k, so it is a constant in the definition of Y_k. That implies that

$$Y_k = F_{t_k} \, E\left[W_{t_{k+1}} - W_{t_k} \mid \mathcal{F}_{t_k} \right] \; .$$

But $t_{k+1} > t_k$, so the Brownian increment $W_{t_{k+1}} - W_{t_k}$ is in the future of $\mathcal{F}_{t_k}$ and therefore is independent of $\mathcal{F}_{t_k}$. Therefore (1) implies that $E\left[W_{t_{k+1}} - W_{t_k} \mid \mathcal{F}_{t_k} \right] = 0$ and also that $Y_k = 0$. Putting all these zeros together with the tower property gives that

$$E\left[X_{t'}^{\Delta t} - X_t^{\Delta t} \mid \mathcal{F}_t \right] = 0 \; ,$$

which is the martingale property for $X_t^{\Delta t}$.

To summarize, using the forward looking finite difference in (4) has the effect of making the approximations martingales. For that reason, the Ito integral is always defined using the forward looking difference. I must ask you to believe without proof that this makes the limiting process X_t also a martingale. We will see in examples that using the backward looking finite difference instead gives a limit that definitely is not a martingale.

The fact that (3) gives a martingale sometimes is called *Doob's theorem* or the Doob *martingale theorem*. The problem is that there are many important Doob theorems about martingales, so it is not clear which one "the" Doob martingale theorem would refer to. There is an interpretation of the martingale property that is useful in finance and economics. In that interpretation, W_t represents that value at time t of some asset and the integrand F_t represents a trading strategy on that asset. At time t you place a "bet" of size F_t on the future increment $W_{t+dt} - W_t$. In the approximation (4), you think of Δt as a small trading delay and you place bets at time t_k that lasts time Δt. The size of this bet is F_{t_k}. It (the betting strategy) can use any information on the Brownian motion path up to time t_k, but nothing beyond. Then X_t, or $X_t^{\Delta t}$ represents the total return from that "investment" (i.e. betting) strategy. Doob's theorem states that this is a martingale. There is no betting strategy based on a martingale that returns anything other than another martingale.

The Doob *stopping time theorem* is a special case of the martingale theorem. A *stopping time* is a random time $\tau \geq 0$ that is determined by the path W so that you know at time t whether $\tau \leq t$. That is to say that $\{\tau \leq t\} \in \mathcal{F}_t$. Some examples are

$$\begin{aligned} \tau_1 &= \min \{ \, t \mid W_t = 1 \} \ , \\ \tau_2 &= \min \{ \, t > \tau_1 \mid W_t = 0 \} \ . \end{aligned}$$

In both cases, you know at time t whether the event has happened in the time interval $[0, t]$. An example that is not a stopping time is $\tau_3 = \max \{t < 3 \mid W_t \leq 0\}$. Indeed, suppose that you are at time $t = 2$ and $W_2 > 0$. You do not know at that time whether or not W_t will return to zero before time 3. Many examples of stopping times (including those above) are *hitting times* (the first time a random process hits a given set) or *first passage* times (same definition, but you might call τ_2 a second passage time). You can interpret a stopping time as a betting strategy. You stay in the game until you hit a specified stopping time. To stop at time t, you must know at time t whether to stop then.

Suppose τ is a *bounded* stopping time and we set $F_t = 1$ if $t \leq \tau$ and $F_t = 0$ for $t \geq \tau$. Bounded means that there is a fixed T with $\tau \leq T$ – the random stopping time τ does not exceed the deterministic ending time T. If τ is a stopping time then F_t is an adapted process, because you know at time t whether $F_t = 0$ or $F_t = 1$ (the only possible values). Then (integrate dW_s up to the stopping time)

$$\int_0^T F_s \, dW_s \ = \ \int_0^\tau dW_s \ = \ W_\tau \ .$$

The martingale property implies that

$$E\left[W_\tau\right] \;=\; 0\;. \tag{5}$$

This is the stopping time theorem. A bounded but non-anticipating stopping strategy cannot produce positive expected value from a martingale.

There is a paradox related to this called the *gambler's ruin* or the *Saint Petersburg* problem, or the *stopping time paradox*. We know $W_0 = 0$. It is a theorem that the hitting time τ_1 above is finite "almost surely". More precisely, let M be the event $\{X_t < 1 \text{ for all } t > 0\}$. This is the event informally written $\{\tau_1 = \infty\}$. We will show in some future class that $\Pr(M) = 0$. An event A is *almost sure* if $\Pr(A) = 1$, which is the same as saying $\Pr(\text{not } A) = 0$. Of course, $E\left[W_{\tau_1}\right] = 1 \neq 0$. The paradox is that this seems to violate the stopping time theorem (5). Of course, the stopping time formula was for bounded stopping times. Not only is τ_1 not bounded, but $E\left[\tau_1\right] = \infty$.

To understand the (non) paradox more clearly, look at a " cutoff" bounded time $\tau_T = \min(\tau_1, T)$. If $W_t = 1$ for some $t \leq T$, then $W_{\tau_T} = 1$. Otherwise, $W_{\tau_T} = W_T$. We will see in some future class that the most likely way to avoid hitting $W = 1$ before time T is to go far in the negative direction. Therefore $E\left[W_{\tau_T}\right] = 0$ is a combination of the very likely event that $W_{\tau_T} = 1$ and the unlikely event that W_{τ_T} is a large negative number. The small probability and the large negative value lead to the overall expected value zero. Now, suppose you are a fund manager. You can bet big on W_{τ_T}. Most of the time you will look smart and produce a profit. If you do this strategy year after year, there is a good chance you will produce a profit year after year. You could even write an Op/Ed in the Wall Street Journal claiming that the Doob stopping time theorem is a liberal hoax perpetrated by Paul Krugman and Bill Ayers. But there is a chance you will lose big. As a fund manager, all that happens to you is that you get fired. You get to keep your fancy house. The investors lose their money, and possibly their houses.

Let us return to the Ito integral (3) and compute the limit of (4) in the case $F_t = W_t$. That is, we compute

$$X_t \;=\; \int_0^t W_s\, dW_s\;. \tag{6}$$

But before coming to the right answer, here is the wrong answer. The calculations are unjustified and the answer is incorrect. Suppose you use ordinary calculus and suppose W_s is a differentiable function of s Then $dW_s = \frac{dW}{ds}ds$ and, using the chain rule of ordinary calculus,

$$W_s dW_s \;=\; W_s \frac{dW}{ds} ds \;=\; \frac{1}{2}\left(\frac{d}{ds} W_s^2\right) ds\;.$$

Substituting this into (6) gives

$$X_t \;=\; \frac{1}{2}\int_0^t \left(\frac{d}{ds}\, W_s^2\right) ds \;=\; \frac{1}{2}\, W_t^2\;.$$

But we know this formula is wrong because of Doob's theorem, which implies $E\left[X_t\right] = 0$, while $E\left[\frac{1}{2}W_t^2\right] = \frac{1}{2}t \neq 0$.

We can get the correct answer for (6) using the approximation formula (4). Start with the identity $a(b-a) = \frac{1}{2}(b+a)(b-a) - \frac{1}{2}(b-a)(b+a)$. This implies that

$$\begin{aligned} W_{t_k}\left(W_{t_{k+1}} - W_{t_k}\right) &= \frac{1}{2}\left(W_{t_{k+1}} + W_{t_k}\right)\left(W_{t_{k+1}} - W_{t_k}\right) - \frac{1}{2}\left(W_{t_{k+1}} - W_{t_k}\right)^2 \\ &= \frac{1}{2}\left(W_{t_{k+1}}^2 - W_{t_k}^2\right) - \frac{1}{2}\left(W_{t_{k+1}} - W_{t_k}\right)^2 \ . \end{aligned}$$

Of course, then

$$\begin{aligned} X_t^{\Delta t} &= \sum_{t_k \leq t} W_{t_k}\left(W_{t_{k+1}} - W_{t_k}\right) \\ &= \left[\frac{1}{2}\sum_{t_k \leq t}\left(W_{t_{k+1}}^2 - W_{t_k}^2\right)\right] - \left[\frac{1}{2}\sum_{t_k \leq t}\left(W_{t_{k+1}} - W_{t_k}\right)^2\right] \ . \end{aligned}$$

The first sum on the right is a *telescoping* sum, which means it has the form $(a-b) + (b-c) + \cdots + (y-z) = a - z$. The long sum collapses to just $a - z$ like the collapsable telescopes sailors used to use. Let $n_*(t) = \min\left\{k \mid t_k > t\right\}$. Then (recall again that $W_0 = 0$)

$$\sum_{t_k \leq t}\left(W_{t_{k+1}}^2 - W_{t_k}^2\right) = W_{t_{n^*(t)}}^2 \ .$$

But $t_{n^*(t)} \to t$ as $\Delta t \to 0$, so $W_{t_{n^*(t)}}^2 \to W_t^2$ (W_t is a continuous function of t). Therefore, the first sum has the limit

$$\lim_{\Delta t \to 0}\left[\frac{1}{2}\sum_{t_k \leq t}\left(W_{t_{k+1}}^2 - W_{t_k}^2\right)\right] = \frac{1}{2}W_t^2 \ .$$

This is the incorrect answer we got using ordinary calculus. It must be that the second sum has a non-zero limit as $\Delta t \to 0$, and one that restores the martingale property.

The second sum is the sum of a large number of independent (by the independent increments property) random variables. A typical term is $Y_k = \left(W_{t_{k+1}} - W_{t_k}\right)^2$, with $E\left[Y_k\right] = t_{k+1} - t_k = \Delta t$, and[2] $\mathrm{var}\left[Y_k\right] = 2\Delta t^2$. The number of terms in the sum is $n^*(t)$ as before. Therefore

$$E\left[\sum_{t_k \leq t}\left(W_{t_{k+1}} - W_{t_k}\right)^2\right] = n^*(t)\Delta t = t_{n^*(t)} \ .$$

[2]If V is a Gaussian random variable with expected value zero and variance σ^2, then $\mathrm{var}\left[V^2\right] = E\left[V^4\right] - E\left[V^2\right]^2 = 3\sigma^4 - \sigma^4 = 2\sigma^4$. Apply this to the increment $V = W_{t_{k+1}} - W_{t_k}$.

As before, $t_{n^*(t)} \to t$ as $\Delta t \to 0$. Moreover, the variance of the sum is the sum of the variances of the terms, so

$$\mathrm{var}\left[\sum_{t_k \le t} \left(W_{t_{k+1}} - W_{t_k} \right)^2 \right] \;=\; 2n^*(t)\Delta t^2 \;=\; 2t_{n^*(t)}\Delta t \to 0 \quad \text{as } \Delta t \to 0.$$

The variance going to zero and the expected value having a limit does not prove completely that the sum has a finite limit. But it almost does, particularly if you take the rapidly decreasing sequence $\Delta t = \frac{1}{2}, \frac{1}{4}, \ldots$ rather than allowing, say, $\Delta t = \frac{1}{2}, \frac{1}{3}, \frac{1}{4}, \frac{1}{5}$, etc. We will come back to this point in a future class. For now, please accept the claim that the limit is what it seems to be.

Putting these results together gives the final result

$$X_t \;=\; \int_0^t W_s \, dW_s \;=\; \lim_{\Delta t \to 0} X_t^{\Delta t} \;=\; \tfrac{1}{2} W_t^2 \;-\; \tfrac{1}{2} t \; . \tag{7}$$

Let us check that this is a martingale. That means checking that $E\left[X_t' \mid \mathcal{F}_t\right] = X_t$ if $t' > t$. For this, write

$$X_{t'} \;=\; \tfrac{1}{2}\left[(W_{t'} - W_t) + W_t\right]^2 \;-\; \tfrac{1}{2}\left[(t' - t) + t\right] \; .$$

Take the conditional expectation, multiply out the square, and use the properties of Brownian motion:

$$\begin{aligned}
E[X_{t'} \mid \mathcal{F}_t] \;&=\; \tfrac{1}{2} E\Big[\left(W_{t'} - W_t \right)^2 \mid \mathcal{F}_t \Big] \;+\; E[\, W_{t'} - W_t \mid \mathcal{F}_t]\, W_t \;+\; \tfrac{1}{2} W_t^2 \\
&\quad -\tfrac{1}{2}\left(t' - t\right) \;-\; \tfrac{1}{2} t \\
&=\; \tfrac{1}{2} W_t^2 \;-\; \tfrac{1}{2} t \\
&=\; X_t \; .
\end{aligned}$$

In the top line we pulled W_t out of the conditional expectation because W_t is known in $\mathcal{F}_t$. Also, $E[\, W_{t'} - W_t \mid \mathcal{F}_t] = 0$ by the independent increments property. The correct formula for the Ito does indeed give a martingale as Doob says is should.

The relation (3) may be written informally as

$$dX_t \;=\; F_t \, dW_t \; .$$

More generally, the formula

$$dX_t \;=\; F_t \, dW_t \;+\; G_t \, dt \tag{8}$$

means that

$$X_t \;=\; X_0 \;+\; \int_0^t F_s \, dW_s \;+\; \int_0^t G_s \, ds \; . \tag{9}$$

In that expression, the first integral on the right is an Ito integral while the second is a Riemann integral. The integrand G_s may be random, but the Riemann integral is defined even for random functions, as long as they are continuous. The expression (8) is called the *Ito differential*. It is used in modeling random processes. For example, the formula

$$dX_t \;=\; a(X_t)\,dt \,+\, b(X_t)\,dW_t \tag{10}$$

is an Ito *stochastic differential equation*. The integral expression (9) gives it a precise mathematical meaning, once we have defined the Ito integral.

The theorem called *Ito's lemma* tells us how to calculate the Ito differential of functions of Brownian motion. Suppose $u(w,t)$ is a sufficiently differentiable function of its arguments. In particular, suppose $\partial_w u$, $\partial_w^2 u$ and $\partial_t u$ make sense (are continuous, possible differentiable themselves, ...). Then $X_t = u(W_t,t)$ is an adapted stochastic process. Ito's lemma is the formula

$$d\,u(W_t,t) \;=\; \partial_w u(W_t,t)\,dW_t \,+\, \partial_t u(W_t,t)\,dt \,+\, \tfrac{1}{2}\partial_w^2 u(W_t,t)\,dt\;. \tag{11}$$

The first two terms on the right are the chain rule from ordinary calculus. The last term is new to the Ito stochastic calculus. We will prove (sort of) Ito's lemma next class. What we actually prove is the integral version

$$u(W_t,t) \,-\, u(0,0) \;=\; \int_0^t \partial_w u\,dW_s \,+\, \int_0^t \left(\partial_t u \,+\, \tfrac{1}{2}\partial_w^2 u\right)\,ds\;. \tag{12}$$

Let us check Ito's lemma using the function $u(w,t) = \frac{1}{2}w^2 - \frac{1}{2}t$. This has $\partial_w u = w$, $\partial_w^2 u = 1$, and $\partial_t u = -1$. Ito's lemma (11) then gives

$$d\left(\tfrac{1}{2}W_t^2 - \tfrac{1}{2}t\right) \;=\; W_t\,dW_t \,-\, \tfrac{1}{2}dt \,+\, \tfrac{1}{2}dt \;=\; W_t\,dW_t\;.$$

The integral version (12) becomes in this example

$$\tfrac{1}{2}W_t^2 - \tfrac{1}{2}t \;=\; \int_0^t W_s\,dW_s\;.$$

This example shows that Ito's lemma can play the role of the fundamental theorem of calculus. In ordinary calculus you try not to evaluate integrals by taking the limits of Riemann sums. Instead you use the fundamental theorem of calculus and look for a function whose derivative is the integrand. Ordinary calculus would not be much use to ordinary people without theorems like the fundamental theorem. In stochastic calculus, you look for a function $u(w,t)$ so that du, as given by (11) is the integrand.

Another important fact about the Ito integral is the *Ito isometry formula*:

$$E\left[\left(\int_0^t F_s\,dW_s\right)^2\right] \;=\; \int_0^t E\left[F_s^2\right]\,ds\;. \tag{13}$$

The left side is the variance of the random variable that is the Ito integral. On the right is a Riemann integral of a function that is not random, $g(s) = E\left[F_s^2\right]$. The isometry formula makes it possible to calculate or estimate the variance of an Ito integral. Recall that the expected value is zero (Doob's theorem). Since a general Ito integral is not Gaussian (Our example gave $(W_t^2 - t)/2$, which is not Gaussian.), the mean and variance do not determine it completely. Still, it is useful information.

Let us verify (13) in our example of $\int_0^t W_s \, dW_s$. On the left we have[3]

$$\begin{aligned} E\left[\left\{\tfrac{1}{2}\left(W_t^2 - t\right)\right\}^2\right] &= \tfrac{1}{4}E\left[W_t^4\right] - \tfrac{1}{2}tE\left[W_t^2\right] + \tfrac{1}{4}t^2 \\ &= \tfrac{1}{4}3t^2 - \tfrac{1}{2}t \cdot t + \tfrac{1}{4}t^2 \\ &= \tfrac{1}{2}t^2 \; . \end{aligned}$$

On the right we have to integrate $E\left[F_s^2\right] = E\left[W_s^2\right] = s$. The result is

$$\int_0^t s \, ds = \tfrac{1}{2}t^2 \; .$$

Thus, the left and right sides of (13) agree in this example.

[3]Memorize the formula $E\left[Y^4\right] = 3\sigma_Y^4$ for a mean zero Gaussian Y.