We start with some technical details regarding the Ito integral and Ito's lemma. Then we do the whole thing over for general *diffusion processes* (definition below). The things we use about Brownian motion W_t to define and understand the Ito integral mostly carry over to general diffusions using more or less the same reasoning.

We return to the Ito integral with respect to Brownian motion

$$X_t \;=\; \int_0^t F_s \, dW_s \; . \tag{1}$$

We assume, possibly without remembering to state every time, that F_t is measurable with respect to $\mathcal{F}_t$. Recall the notation for the approximation: Δt is a small time step, $t_k = k\Delta t$, and the approximation of (1) is the Ito-Riemann sum

$$X_t^{\Delta t} \;=\; \sum_{t_k < t} F_{t_k} \left(W_{t_{k+1}} - W_{t_k} \right) \; . \tag{2}$$

As a reminder, it is crucial here that $\Delta W = W_{t_{k+1}} = W_{t_k}$ is in the future of $\mathcal{F}_{t_k}$ where F_{t_k} is defined. At time t_k, the value F_{t_k} is completely know (i.e. known exactly with no random error) but ΔW is completely unknown (i.e. independent of the information we have so far).

We want to understand the convergence as $\Delta t \to 0$ of the approximation (2). For the purpose of defining the Ito integral (1) it suffices to consider a specific sequence $\Delta t_n \to 0$ as $n \to \infty$. The question of convergence for values of Δt not in this sequence need not be addressed and is harder. I use the sequence is $\Delta t_n = 2^{-n}$. For convenience, I simplify the notation to $X_t^n = X_t^{2^{-n}}$. The rapid convergence of Δt_n to zero makes it hard for certain fluctuations to mess up the convergence of X_t^n. The specific choice makes $\Delta t_{n+1} = \frac{1}{2}\Delta t_n$ and makes it easy to compare X^{n+1} to X^n.

The convergence proof establishes that the limit exists by showing that the differences $Y^n = X^{n+1} - X^n$ converge to zero fast enough. The specific argument is a version of the *Borel Cantelli lemma* as it usually is applied in probability, but with two aspects of it combined. Any appropriate probability book will have the original Borel Cantelli argument that is not being followed here.

The proof uses the phrase *almost surely*. This does not mean that non-convergence does not exist, only that it cannot happen. More precisely, there are outcomes in which $\lim_{n\to\infty} X^n$ does not exist, but if A is the event consisting of all these outcomes, then $\Pr(A) = 0$. You can repeat an experiment as often as you want. If you understand the situation correctly, an event of probability zero will never happen. The event A may not be empty, but *almost surely*, $\omega \notin A$.

One can show that event A has probability zero by finding a random variable S so that $S \geq 0$ and $E[S] < \infty$ and $A = \{S = \infty\}$. We have $X^m = X^1 + \sum_{n=1}^{m-1} Y^n$. Therefore

$$\lim_{m\to\infty} X^m \;=\; X^0 + \lim_{m\to\infty} \sum_{n=1}^{m} Y^n \; .$$

The limit on the left exists if the sum on the right exists. The limit on the right is, by definition of infinite sums,

$$\lim_{m\to\infty} \sum_{n=1}^{m} Y^n \;=\; \sum_{n=1}^{\infty} Y^n \;.$$

Now *recall*[1] that the infinite sum on the right exists if the sum is *absolutely convergent*. This means that

$$S \;=\; \sum_{n=1}^{\infty} |Y^n| < \infty \;.$$

We are going to establish this by showing that $E[S] < \infty$. What we have to show is

$$E[S] \;=\; \sum_{n=0}^{\infty} E\big[\,\big|X_t^{n+1} - X_t^n\big|\,\big] \;<\; \infty \;.$$

In fact, we will do a calculation that shows that

$$E\big[\,\big|X_t^{n+1} - X_t^n\big|\,\big] \;<\; C\alpha^n \;, \tag{3}$$

with a positive $\alpha < 1$.

Evaluating $E[|Y|]$ by a calculation may be harder than evaluating $E\big[Y^2\big]$. Fortunately, the *Cauchy Schwarz inequality* implies that

$$E[|Y|] \;\le\; \big(E\big[Y^2\big]\big)^{1/2} \;. \tag{4}$$

This follows from the usual Cauchy Schwarz inequality, which states that if X and Y are any pair of random variables, then

$$E\left[|XY|\right] \;\le\; \big(E\left[X^2\right] E\left[Y^2\right]\big)^{1/2} \;. \tag{5}$$

If you don't "remember" this but you have a training in probability, you probably know that the correlation coefficient cannot be larger than one. That is, if U and V are two random variables, then $\mathrm{cov}(U,V) \le (\mathrm{var}(U)\mathrm{var}(V))^{1/2}$. The proof of this is to take $X = U - \mu_U$ and $Y = V - \mu_V$ and apply (5). You can get (4) from (5) by the well known (to those who know it well) trick of taking $X = 1$ in (5). If all this isn't enough proof, you can make a proof in discrete probability by minimizing $\sum y_k p_k^2$ with the constraint that $\sum y_k p_k$ is fixed. Here the numbers y_k are the possible values of Y, which we may assume are non-negative, and p_k are the corresponding probabilities. The method of Lagrange multipliers gives the optimality condition as $y_k p_k = \lambda p_k$ for some Lagrange multiplier λ. This implies that the y_k are equal, which makes the two sides of (4) equal. Otherwise, the right side is strictly larger than the left.

[1] If you have not taken a class on mathematical analysis that covered this fact, you can look it up or accept it as very plausible.

The calculation will show that

$$E\Big[\big(X^{n+1} - X^n\big)^2\Big] \;\leq\; C\beta^n \;, \tag{6}$$

where β is a non-negative constant and $\beta < 1$. This gives

$$E\big[\big|X^{n+1} - X^n\big|\big] \;\leq\; \sqrt{C\beta^n} \;=\; C\left(\sqrt{\beta}\right)^n \;.$$

Note that C in the middle is different from C on the right. It is a common convention in proofs like this that C just means "some constant", which can be different from place to place. This proves the desired inequality (3), with $\alpha = \sqrt{\beta}$. Of course, if $\beta < 1$ then $\sqrt{\beta} < 1$ as well.

After this buildup, we finally come to the business of comparing X^n to X^{2n}. I use the following more or less standard notations: $t_k = k\Delta t$ is generalized to $t_{k+1/2} = (k+\frac{1}{2})\Delta t$, time subscripts are abbreviated, as $F_k = F_{t_k}$, $W_{k+1/2} = W_{t_{k+1/2}} = W_{t_k+\Delta t/2}$, etc. One piece of the sum (2) is $F_k(W_{k+1} - W_k)$. This piece comes from the time interval $[t_k, t_{k+1}]$. The corresponding sum for X^{2n} has this broken into two equal sub-intervals. The corresponding terms in the sum are $F_k(W_{k+1/2} - W_k) + F_{k+1/2}(W_{k+1} - W_{k+1/2})$. Subtracting gives

$$\Big[F_k(W_{k+1/2} - W_k) + F_{k+1/2}(W_{k+1} - W_{k+1/2})\Big] - \Big[F_k(W_{k+1} - W_k)\Big]$$
$$= \; (F_{k+1/2} - F_k)(W_{k+1} - W_{k+1/2}) \;.$$

Summing over t_k gives (except for a possible extra term (which I ignore and which does not change the final result (6), and see explanations below)

$$\begin{aligned}
E\Big[\big(X_t^{2n} - X_t^n\big)^2\Big] &= \sum_{t_k<t} E\big[(F_{k+1/2} - F_k)^2(W_{k+1} - W_{k+1/2})^2\big] \\
&= \sum_{t_k<t} E\big[(F_{k+1/2} - F_k)^2\big] \; E\big[(W_{k+1} - W_{k+1/2})^2\big] \\
E\Big[\big(X_t^{2n} - X_t^n\big)^2\Big] &= \tfrac{1}{2}\Delta t \sum_{t_k<t} E\big[(F_{k+1/2} - F_k)^2\big] \;. \qquad (7)
\end{aligned}$$

Explanations:

- The top sum on the right is a single sum over k because the cross terms $E\big[(F_{j+1/2} - F_j)(W_{j+1} - W_{j+1/2})(F_{k+1/2} - F_k)(W_{k+1} - W_{k+1/2})\big]$ are zero if $j \neq k$. This is because one of the two terms ΔW_j or ΔW_k is in the future of all other terms and therefore independent of them.

- The middle line follows from the top line because the ΔW interval, $[t_{k+1}, t_{k+1/2}]$, is in the future of both F_k and $F_{k+1/2}$. Therefore, the ΔW^2 factor is independent of the ΔF factor.

- The bottom line uses the Brownian motion formula $E\big[(W_{k+1} - W_{k+1/2})^2\big] = (t_{k+1}, t_{k+1/2})$.

Unfortunately, some fudging starts now. We need to concoct a reason for the terms $E\left[(F_{k+1/2} - F_k)^2\right]$ to go to zero as $\Delta t \to 0$. If they do not, the overall sum might be

$$C \cdot \Delta t \cdot (\textit{number of terms}) \;=\; Ct \;,$$

which does not go to zero with Δt. I propose the following *regularity* hypothesis on F_t:

$$|F_{t'} - F_t| \;\leq\; C\left[\,|W_{t'} - W_t| \,+\, |t' - t|\,\right] \;.$$

This is motivated by thinking of F as a function of W_t and t, as it is in many applications of the Ito integral. This implies the simple bound

$$E\left[(F_{k+1/2} - F_k)^2\right] \;\leq\; C\Delta t \;. \tag{8}$$

I proceed now assuming that F satisfies this bound. I will give a few comments later (in a later class?) about what to do if this bound does not hold.

Assuming (8), we get from (7)

$$E\left[\left(X_t^{2n} - X_t^n\right)^2\right] \;\leq\; C\Delta t^2 \cdot (\textit{number of terms}) = C\Delta t \cdot t \;=\; C_t\left(\tfrac{1}{2}\right)^n \;.$$

But this is an estimate of the form (6), which was the last thing needed to finish the proof.

We next come to the important properties of the Ito integral. Two of the most basic properties are the Ito isometry formula and the part of Ito's lemma informally written as $(dW_t)^2 = dt$. These have the same source, which is that $E\left[(\Delta W)^2\right] = \Delta t$, and that fluctuations from this expectation value cancel out in the manner of the law of large numbers. This cancellation is the key, in view of the fact that

$$E\left[\,\left|\Delta W^2 - \Delta t\right|\,\right] \;\geq\; \Delta t \;.$$

The deviations of $(\Delta W)^2$ from Δt are not small relative to Δt. But they cancel each other enough to ignore them because their expected value is zero.

The formula $(dW_t)^2 = dt$ is true not in the *pointwise* sense that the numbers are equal for each t. Instead, integrals of them,.

$$\begin{aligned} U_t &= \int_0^t F_s^2\, ds \;, \\ V_t &= \int_0^t F_s^2\,(dW_s)^2 \;, \end{aligned}$$

are equal. More precisely, we give definitions of U_t and V_t as limits and do a calculation that shows the limits are the same. The natural approximations are

$$\begin{aligned} U_t^n &= \sum_{t_k < t} F_k^2 \Delta t \\ V_t^n &= \sum_{t_k < t} F_k^2\,(W_{k+1} - W_k)^2 \;. \end{aligned}$$

Arguments like the one above for the Ito integral show that the limits

$$\left.\begin{aligned} \lim_{n\to\infty} U_t^n &= U_t \\ \lim_{n\to\infty} V_t^n &= V_t \end{aligned}\right\} \tag{9}$$

exist (the left sides are the definitions of the right sides). The question is: are they the same?

If they are the same, it is the result of cancellation. Computing the expected square is a great way to find cancelation. If

$$E\Big[(U_t^n - V_t^n)^2\Big] \;\to\; 0 \quad \text{as } n \to \infty,$$

then[2] $U_t = V_t$. The difference is

$$V_t^n - U_t^n \;=\; R_t^n \;=\; \sum_{t_k < t} F_k^2 \left[(W_{k+1} - W_k)^2 \;-\; \Delta t\right] \;.$$

The usual three facts lead to the cancelation: (i) the term in square brackets $[\cdots]$ has mean value zero, (ii) it is independent of the random factor that multiplies it (F_k^2), (iii) cross terms have expected value zero (one of them is in the future). Therefore we calculate (we did similar calculations before)

$$E\left[\left\{(W_{k+1} - W_k)^2 \;-\; \Delta t\right\}^2\right] \;=\; 2\Delta t^2 \;,$$

and use this to get

$$E[R_t^n] \;=\; 2\Delta t \left(\sum_{t_k < t} E\left[F_k^4\right] \Delta t \right) \;.$$

But this is Δt multiplied by an approximation to a Riemann integral $\int_0^t F_s^4 ds$, so it goes to zero with Δt. This, at last, shows that $U_t = V_t$. This is the sense in which $(dW_t)^2 = dt$.

This calculation is the main step in one approach to Ito's lemma. Recall the point of the lemma is to define $du(W_t, t)$, where $u(w, t)$ is a differentiable function (up to second or third partial derivatives). The definition of du should have the property that

$$u(W_t, t) \;=\; u(W_0, 0) \;+\; \int_0^t du(W_s, s) \;. \tag{10}$$

[2]Hopefully this is plausible. A mathematical proof, for those who know the relevant measure theory, is $E[(U_t - V_t)^2] \leq \liminf_{n\to\infty} E[(U_t^n - V_t^n)^2]$ (Fatou's lemma), under the hypothesis that $U_t^n \to U_t$ and $V_t^n \to V_t$ almost surely. If the right side goes to zero, then the left side must be zero and $U_t = V_t$ almost surely.

The Ito's lemma formula uses the partial derivatives $\partial_w u(w,t)$, $\partial_w^2(w,t)$, and $\partial_t u(w,t)$:

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

The actual meaning of this formula comes from (10):

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

When we prove Ito's lemma, this is integral formula is what prove.

We start the proof with the least interesting and most technical part of it, remainders in Taylor series. Suppose $f(x)$ has bounded derivatives up to order 3, then $f(x+\Delta x) - f(x) = f'(x)\Delta x + \frac{1}{2}f''(x)\Delta x^2 + R$, where R is the *remainder* of the series. It satisfies (find a *good* calculus book) $|R| \leq \frac{1}{6}\max|f'''|\,|\Delta x|^3$. There is a clever approach to multi-variable Taylor series that uses this one variable theorem and the chain rule. We have $\Delta u = u(w+\Delta w, t+\Delta t) - u(w,t)$. Now (this is the trick) define $f(x) = u(w + x\Delta w, t + x\Delta t) - u(w,t)$. This f satisfies $f(0) = 0$, and $f(1) = \Delta u$. We will apply the one variable theorem above with $\Delta x = 1$. The chain rule gives (leaving out arguments for simplicity)

$$\begin{aligned}
f' &= \partial_w u \Delta w \;+\; \partial_t u \Delta t \\
f'' &= \partial_w^2 u \,(\Delta w)^2 \;+\; 2\partial_w\partial_t u\,(\Delta w \Delta t) \;+\; \partial_t^2 u\,(\Delta t)^2 \\
f''' &= \partial_w^3 u\,(\Delta w)^3 \;+\; 3\partial_w^2\partial_t u\left(\Delta w^2 \Delta t\right) \;+\; 3\partial_w\partial_t^2 u\left(\Delta w \Delta t^2\right) \;+\; \partial_t^3 u\,(\Delta t)^3 \;.
\end{aligned}$$

For Ito's lemma, we need both terms in f' and only the first term of f''. Note that if $|\Delta w| < 1$ and $\Delta t < 1$ (Δt is positive but Δw could be positive or negative), then $\Delta w^2 \Delta t < |\Delta w|\,\Delta t$ and $\Delta t^3 < \Delta t^2$. Therefore, if all the partial derivatives on the right are bounded, we conclude that

$$\left|\Delta u \;-\; \left(\partial_w u \Delta w \;+\; \partial_t u \Delta t \;+\; \tfrac{1}{2}\partial_w^2 u\right)\right| \;\leq\; C\left(|\Delta w|^3 \;+\; |\Delta w|\,\Delta t \;+\; \Delta t^2\right) \;.$$

The term in parentheses on the left is the Taylor series approximation needed for Ito's lemma. The quantity on the right is the remainder bound for this approximation.

A new thing in stochastic calculus is that we take the Taylor series approximation

$$u(W_{k+1}, t_{k+1}) - u(W_k, t_k) \;\approx\; \partial_w u(W_k,t_k)\Delta W_k + \partial_t u(W_k,t_k)\Delta t + \tfrac{1}{2}\partial_w^2 u(W_k,t_k)\Delta W_k^2 \;.$$

We use both first derivative terms but only one of the three second derivative terms. This is because we need an approximation whose error is an order of magnitude smaller than Δt. The overall error in making an error at each time step is the sum of all the individual errors. If the individual errors are on the order of Δt, then their sum does not go to zero as $\Delta t \to 0$. If the individual terms go to zero faster than Δt, then even their sum goes to zero with Δt. We have seen that $E[\Delta W^2] = \Delta t$. This implies that typical ΔW values have $|\Delta W| = O(\Delta t^{1/2})$. Squaring this gives $\left|\Delta W^2\right| = O(\Delta t)$, which we knew. Going further, $\left|\Delta W^3\right| = O(\Delta t^{3/2})$ and $|\Delta W \Delta t| = O(\Delta t^{3/2})$. Adding them up, we expect that

$$\sum_{t_k < t} \left|\Delta W_k^3\right| \;\leq\; C \sum_{t_k \leq t} \Delta t^{1/2} \Delta t \;\leq\; \Delta t^{1/2}\, Ct \;.$$

Adding up the individual errors from each time step gives a total error on the order of $\Delta t^{1/2}$ which tends to zero as $\Delta t \to 0$.

It is time to assemble the pieces. We have

$$u(W_t, t) \;-\; W(0,0) \;=\; \sum_{t_k < t} \Delta u_k \;+\; \Delta u_* \;,$$

where

$$\Delta u_k \;=\; u(W_{k+1}, t_{k+1}) - u(W_k, t_k) \;,$$

and

$$\Delta u_* \;=\; u(W_t, t) - u(W_{n^*(t)}, t_{n^*(t)}) \;.$$

The last term is the difference in u values between the actual t and the largest $t_k < t$. You may just ignore it in the first reading. Using the Taylor expansions above and adding them up gives

$$\begin{aligned}
u(W_t, t) \;-\; W(0,0) \;&=\; \sum_{t_k < t} \left(\partial_w u(W_k, t_k)\right) \Delta W_k \\
&+\; \sum_{t_k < t} \left(\partial_t u(W_k, t_k)\right) \Delta t \\
&+\; \frac{1}{2} \sum_{t_k < t} \left(\partial_w^2 u(W_k, t_k)\right) \Delta W_k^2 \\
&+\; \sum_{t_k < t} R_k \;.
\end{aligned}$$

Here R_k is the remainder in the Taylor approximation. The first sum on the right converges to the Ito integral. The second sum converges to the Riemann integral. The third sum also converges to the Riemann integral, as we showed a few pages ago. The expected value of the last term is

$$\begin{aligned}
E\left[\sum_{t_k < t} R_k\right] \;&\leq\; C \sum_{t_k < t} E\left[|\Delta W_k|\right] \Delta t \;+\; C \sum_{t_k < t} E\left[|\Delta W_k|^{3/2}\right] \\
&\leq\; C \Delta t^{1/2} \sum_{t_k < t} \Delta t \;=\; C_t \sqrt{\Delta t} \;.
\end{aligned}$$

This is the proof of Ito's lemma.

The Ito isometry formula is based on similar ideas but simpler calculations. It says that if X_t is the Ito integral (1), then

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

The proof is a calculation using the definition (2).

$$E\left[\left(X_t^{\Delta t}\right)^2\right] \;=\; \sum_{t_k<t} E\Big[F_{t_k}^2 \left(W_{t_{k+1}} - W_{t_k}\right)^2\Big] \;.$$

We use the tower property to evaluate the terms on the right. The tower property tells us that if U is any random variable, then

$$E[\,E[\,U \mid \mathcal{F}_k\,]\,] \;=\; E\,[U] \;.$$

We assumed that F_{t_k} is known at time t_k, so given the information in $\mathcal{F}_{t_k}$, F_{t_k} is a known constant. Therefore

$$\begin{aligned} E\Big[F_{t_k}^2 \left(W_{t_{k+1}} - W_{t_k}\right)^2\Big] &= E\Big[E\Big[F_{t_k}^2 \left(W_{t_{k+1}} - W_{t_k}\right)^2 \mid \mathcal{F}_{t_k}\Big]\Big] \\ &= E\Big[\,F_{t_k}^2\, E\Big[\left(W_{t_{k+1}} - W_{t_k}\right)^2 \mid \mathcal{F}_{t_k}\Big]\Big] \\ &= \Delta t\, E\big[\,F_{t_k}^2\,\big] \;. \end{aligned}$$

We use this calculation to get

$$E\left[\left(X_t^{\Delta t}\right)^2\right] \;=\; \sum_{t_k<t} E\big[F_{t_k}^2\big]\, \Delta t \;,$$

which converges to the right side of (13) as $\Delta t \to 0$.

One application of the isometry formula (13) is to the understanding of stochastic integrals for small t. If the integrand F_t is a continuous function of t, the approximation is

$$\int_0^t F_s\, dW_s \;\approx\; \int_0^t F_0\, dW_s \;=\; F_0 W_t \;.$$

This means that for small t, X_t is approximately a Gaussian (because W_t is Gaussian) with mean zero and variance $F_0^2 t$ (because F_0 is known at time zero (i.e. now) so it is a constant, and W_t has variance t. You might be uneasy with this admittedly vague argument, partly because F_t is random and you don't know how far F_t might be from F_0. A more solid if less intuitive argument would be to assume that $v(t) \;=\; E\big[F_t^2\big]$ is a differentiable function of t, so $|v(t) - v(0)| \leq Ct$. Then

$$E\big[X_t^2\big] \;=\; \int_0^t v(s)\, ds \;\approx\; v(0) \int_0^t ds \;,$$

with a more concrete error bound

$$\left| \int_0^t \left(v(s) - v(0) \right) \, ds \right| \;\leq\; \int_0^t \left| v(s) - v(0) \right| \, ds \;\leq\; C \int_0^t s \, ds \;=\; Ct^2 \; .$$

At the end I indulged in the mathematicians' habit of taking C to mean "just some constant" without implying that two instances of C have the same value. Yes, one of the C values above is twice the other, but nevermind. This reasoning shows that

$$E\left[X_t^2\right] \;=\; F_0^2 t \,+\, O\left(t^2\right) \; . \tag{14}$$

Let us pause here to explain the *big O* notation. Suppose $f(t)$ is some function defined for small t and $g(t)$ is another function with $g(t) > 0$ for $t > 0$. We say $f(t) = O(g(t))$ if there is a constant C so that $|f(t)| \leq Cg(t)$ for all sufficiently small t. *Sufficiently small* means that there is a positive t_0 so that if t is in the range $0 < t \leq t_0$, then $|f(t)| \leq Cg(t)$. This mathematical definition may have more or less practical significance depending on how big C might be or how small t_0 might be. You might see (indeed, have seen in (14)) this used in the sense that $f(t) = h(t) + O(g(t))$ implies that $|f(t) - h(t)| \leq Cg(t)$ for all $t \in (0, t_0)$. The big O notation often is used combined with powers of t. Powers of t are the mathematician's version of *orders of magnitude* in science. Where an order of magnitude is a power if ten (roughly) for a scientist, it is a power of t to a mathematician. The formula (14) states that the difference between $E\left[X_t^2\right]$ and $v(0)t$ is *on the order of* t^2, which is an order of magnitude smaller than the order of $v(0)t$ (assuming $v(0) \neq 0$). It says that the error in the approximation $E\left[X_t^2\right] \approx v(0)t$ is an order of magnitude smaller than the quantity being approximated.

Coming back to (14), this tells us that, for small t, typical values of $|X_t|$ are on the order of $\sqrt{t}$, because typical values of X_t^2 are on the order of t. Now suppose

$$X_t \;=\; \int_0^t F_s \, dW_s \,+\, \int_0^t G_s \, ds \; .$$

Then typical values of the first integral on the right are on the order of $t^{1/2}$ and typical values of the second integral are on the order of t (it's just a Riemann integral). Therefore the Ito integral typically is an order of magnitude larger than the Riemann integral. In particular

$$E\left[\, X_t^2 \,\right] \;\approx\; F_0^2 t \; .$$

Of course, the expected value of the Ito integral is zero, so

$$E\left[\, X_t \right] \;\approx\; G_0 t \; .$$