From Riemann to Itô: Integrals, Variation, and Brownian Motion

Notes by Zack Polaski

October 23, 2025

Contents

1 Motivation: When Classical Integration Fails

In classical calculus, integration is intimately tied to the idea of *smooth accumulation*. If a function $f(t)$ represents a rate of change and $x(t)$ represents a smooth input or path, then the Riemann integral

$$\int_0^T f(t)\, dx(t)$$

is interpreted as the limit of sums $\sum f(t_i)\, \Delta x_i$, where $\Delta x_i = x(t_{i+1}) - x(t_i)$. This construction implicitly assumes that $x(t)$ does not fluctuate "too violently" - that its total variation over the interval $[0, T]$ is finite.

Riemann's vision: integration as area accumulation

The Riemann integral partitions the domain $[0, T]$ into small subintervals of length Δt and forms rectangles of height $f(t_i)$ and width Δt. In the limit of finer partitions, these rectangles converge to the exact area under the curve f. Mathematically:

$$\int_0^T f(t)\, dt = \lim_{|\Pi| \to 0} \sum_i f(t_i)\, (t_{i+1} - t_i),$$

where $\Pi = \{0 = t_0 < t_1 < \cdots < t_n = T\}$ is a partition of the time axis and $|\Pi| = \max_i (t_{i+1} - t_i)$ is its mesh size.

If we wish to integrate with respect to another function $g(t)$ rather than t itself, the natural extension is the *Riemann–Stieltjes* integral:

$$\int_0^T f(t)\, dg(t) = \lim_{|\Pi| \to 0} \sum_i f(t_i)\, [g(t_{i+1}) - g(t_i)].$$

Here, the increments $dg(t)$ replace the uniform increments dt and act as a custom "ruler" along the integration axis. The limit exists if g is continuous and of *bounded variation* - that is, if the total accumulated magnitude of its oscillations is finite.

Enter Brownian motion

A Wiener process W_t (or Brownian motion) breaks these assumptions completely: it is continuous everywhere, but differentiable nowhere. Over any finite interval $[0, T]$, the total variation

$$V(W; [0, T]) = \sup_{\Pi} \sum_i |W_{t_{i+1}} - W_{t_i}|$$

is infinite with probability one. Intuitively, as we refine the partition Π, the Brownian path exhibits more and more small oscillations, so that the sum of their magnitudes grows without bound. In this sense, W_t is "too rough" to serve as a classical integrator.

Note on the Supremum over Partitions

The total variation

$$V(g;[a,b]) = \sup_{\Pi} \sum_i |g(t_{i+1}) - g(t_i)|$$

is defined as a *supremum*, not a maximum. This distinction matters: there may be no single partition that achieves the upper bound.
For well-behaved (bounded-variation) functions, the sums stabilize and the supremum is finite—sometimes even attained by a particular partition. For rough paths such as Brownian motion, finer partitions always reveal additional oscillations that increase the total distance traveled. Hence no finite partition gives the bound, but the supremum (possibly $+\infty$) remains meaningful as the least upper limit over all refinements.

The problem is not just technical: the Riemann–Stieltjes construction fundamentally requires control over the total variation of the integrator g. Without that control, the Riemann sums $\sum f(t_i)\,\Delta g_i$ fail to converge—the partial sums oscillate indefinitely as we take finer partitions.

Conceptual failure of the classical approach

Thus, for Brownian motion we cannot even define

$$\int_0^T f(t)\,dW_t$$

in the usual pathwise sense. The integrator has *unbounded local variation*, and the notion of area accumulation by rectangles no longer applies. The breakdown of the Riemann paradigm motivates a new definition of integration—one that measures convergence not by pointwise limits of sums, but by convergence of random variables in *mean square*. This is the foundation of Itô's stochastic calculus.

Key question

Why can't we define $\int_0^T f(t)\,dW_t$ in the Riemann sense?

Answer: Because Brownian paths have infinite total variation, the Riemann sums $\sum f(t_i)\,\Delta W_i$ do not converge as the partition is refined. Classical integration depends on bounded variation; stochastic integration replaces that requirement with convergence in mean square.

2 Riemann and Riemann–Stieltjes Integrals

2.1 Riemann sums and partitions

The construction of the Riemann integral begins with the idea of approximating an area by summing narrow rectangles. To formalize this, we introduce a *partition* of the interval $[a, b]$:

$$\Pi = \{a = t_0 < t_1 < t_2 < \cdots < t_n = b\},$$

which divides the domain into n subintervals $[t_i, t_{i+1}]$. The *mesh size* (or norm) of the partition,

$$|\Pi| = \max_i (t_{i+1} - t_i),$$

quantifies how fine the subdivision is. As $|\Pi| \to 0$, the partition becomes arbitrarily dense.

For a chosen sample point t_i in each subinterval, the corresponding *Riemann sum* is

$$S_\Pi = \sum_{i=0}^{n-1} f(t_i)\,[x_{i+1} - x_i].$$

Each term $f(t_i)\,(x_{i+1} - x_i)$ represents the area of a thin rectangle of height $f(t_i)$ and width $\Delta x_i = x_{i+1} - x_i$.

Definition (Riemann Integral)

We say that f is *Riemann integrable* on $[a, b]$ if there exists a limit I such that

$$\lim_{|\Pi| \to 0} S_\Pi = I,$$

independent of the choice of sample points within each subinterval. In that case, we write

$$I = \int_a^b f(x)\,dx.$$

Interpretation

The convergence of S_Π in the classical Riemann sense relies on the regularity of the *integrand* f, since the integrator x itself is perfectly well-behaved. In other words, f must not oscillate too wildly on $[a, b]$. Later, in the Riemann–Stieltjes setting, the focus of this requirement shifts from the integrand f to the integrator g, whose variation will control whether the integral is well-defined. TO BE CLEAR: this does not mean that f's regularity stops mattering from this point forward, but rather that once we fix f to be "reasonable," the *primary* obstacle to defining the integral shifts to the integrator g (or W_t as we will see later in the Ito setting).

2.2 Riemann–Stieltjes Generalization

The Riemann integral uses the ordinary distance increments $(x_{i+1} - x_i)$ as the measure of "width." In many settings, however, one wishes to accumulate f not with respect to uniform distance along the x-axis, but with respect to another function $g(x)$ that increases according to some intrinsic scale.

Definition (Riemann–Stieltjes Integral)

Let f and g be real-valued functions on $[a, b]$, and let

$$\Pi = \{a = t_0 < t_1 < \cdots < t_n = b\}$$

be a partition with mesh $|\Pi|$. The corresponding *Riemann–Stieltjes sum* is

$$S_\Pi = \sum_{i=0}^{n-1} f(t_i)\,[g(t_{i+1}) - g(t_i)].$$

If there exists a finite limit

$$\lim_{|\Pi|\to 0} S_\Pi = I,$$

independent of the choice of sample points t_i, then we say the *Riemann–Stieltjes integral* of f with respect to g exists, and we write

$$I = \int_a^b f(x)\,dg(x).$$

Interpretation

The increments $dg(x)$ act as a *custom measuring stick* that replaces the uniform ruler dx. Intuitively, the integral measures the accumulation of $f(x)$ weighted by the rate at which $g(x)$ changes. When $g(x) = x$, the construction reduces to the ordinary Riemann integral.

Geometric intuition

If g is increasing rapidly over a region, that region carries more "measure" in the integral; if g is flat, it contributes nothing. The Riemann–Stieltjes integral therefore generalizes the idea of area under a curve to accumulation with respect to an arbitrary monotone scale.

Existence condition: bounded variation of g

For the Riemann–Stieltjes integral $\int_a^b f\,dg$ to exist for all continuous f, it suffices that g be of *bounded variation* on $[a, b]$:

$$V(g; [a, b]) = \sup_{\Pi} \sum_i |g(t_{i+1}) - g(t_i)| < \infty.$$

This ensures that the increments dg do not oscillate uncontrollably and that the Riemann sums remain bounded and converge as the partition is refined.

Key idea

Bounded variation of the integrator g guarantees that the "total movement" of g is finite. The classical Riemann–Stieltjes integral breaks down precisely when this property fails—as it does for Brownian motion.

3 Total Variation and Boundedness

3.1 Definition

The *total variation* of a function $g : [a, b] \to \mathbb{R}$ measures how much the function "moves" up and down over the interval. Formally, for any partition $\Pi = \{a = t_0 < t_1 < \cdots < t_n = b\}$, define the variation along that partition as

$$V(g; \Pi) = \sum_{i=0}^{n-1} |g(t_{i+1}) - g(t_i)|.$$

The total variation of g on $[a, b]$ is then the supremum of these partial sums over all possible partitions:

$$V(g; [a, b]) = \sup_{\Pi} \sum_{i} |g(t_{i+1}) - g(t_i)|.$$

- If $V(g; [a, b]) < \infty$, we say that g has **finite total variation** (or is of **bounded variation**).
- If $V(g; [a, b]) = \infty$, we say that g has **unbounded variation**.

Intuition

The total variation captures the cumulative distance traveled by the function's graph, counting both upward and downward movements. A smooth, monotone, or piecewise monotone function travels a finite distance overall and therefore has finite total variation. By contrast, functions that oscillate infinitely often—such as Brownian motion or $x \sin(1/x)$ near the origin—have infinite total variation, meaning that the "total distance traveled" diverges as the partition is refined.

Conceptual summary

Total variation quantifies the cumulative amount of motion in a function. A bounded-variation function moves a finite distance overall, while an unbounded-variation function exhibits endless small oscillations whose total length diverges. The distinction determines whether Riemann–Stieltjes integration with respect to that function can be defined.

3.2 Examples

To develop intuition for total variation, it helps to examine a few representative cases ranging from smooth to highly irregular functions.

- **Linear function:** For $g(x) = x$ on $[0, 1]$,

$$V(g; [0, 1]) = \sup_{\Pi} \sum_i |x_{i+1} - x_i| = 1.$$

  The variation simply equals the total increase of the function, since g is monotone. Every monotone function, in fact, has finite total variation equal to $|g(b) - g(a)|$.

- **Trigonometric function:** For $g(x) = \sin x$ on $[0, 2\pi]$, the function moves up from 0 to 1, then down to -1, then back up to 0, giving a total travel distance of $1 + 2 + 1 = 4$. Hence $V(g; [0, 2\pi]) = 4 < \infty$.

- **Highly oscillatory function:** For $g(x) = x \sin(1/x)$ on $(0, 1]$, the amplitude of oscillations decays to zero, but their frequency increases without bound as $x \to 0$. The cumulative length of these oscillations diverges, so $V(g; (0, 1]) = \infty$. This function provides a classical example of a continuous function of *unbounded variation*.

- **Brownian motion:** A Wiener process W_t is continuous everywhere but nowhere differentiable. On any interval $[0, T]$, the expected partial variation along a uniform partition of n subintervals satisfies

$$\mathbb{E}\left[\sum_i |W_{t_{i+1}} - W_{t_i}|\right] = \sqrt{\frac{2T}{\pi}}\,\sqrt{n} \;\to\; \infty.$$

  Thus $V(W; [0, T]) = \infty$ almost surely: Brownian paths are *continuous yet infinitely rough*.

Observation

Smooth or piecewise monotone functions have finite total variation, while functions with infinitely many small oscillations—such as $x \sin(1/x)$ or Brownian motion—accumulate infinite total variation as the partition is refined. This infinite variation is precisely what breaks the classical Riemann–Stieltjes framework for stochastic paths.

4 Lebesgue–Stieltjes Integration

4.1 From symbolic dg to a measure μ_g

The Riemann–Stieltjes integral

$$\int_a^b f(x)\,dg(x)$$

can be viewed as a sum of the form $\sum f(t_i)[g(t_{i+1})-g(t_i)]$, where the increments $dg(x)$ serve as a custom "measuring stick." Lebesgue's framework formalizes this idea by treating those increments as the action of a genuine *measure* on subsets of $\mathbb{R}$.

Definition (Lebesgue–Stieltjes measure)

If g is a monotone increasing function on $\mathbb{R}$, we can associate to it a measure μ_g defined on half-open intervals by

$$\mu_g((a,b]) = g(b) - g(a),$$

and then extend μ_g uniquely to all Borel sets by countable additivity.

The induced integral

The *Lebesgue–Stieltjes integral* of f with respect to g is then defined as the Lebesgue integral of f with respect to the measure μ_g:

$$\int_a^b f(x)\,d\mu_g(x).$$

This generalizes the Riemann–Stieltjes construction by replacing the symbolic "increment" $dg(x)$ with an actual measure μ_g that encodes how g accumulates across the domain.

> **Intuition**
>
> When g is monotone, the difference $g(b) - g(a)$ represents the "mass" assigned to the interval $(a,b]$. The function g thus acts as a *cumulative distribution function* for a measure μ_g. Integrating f with respect to μ_g means summing f over $\mathbb{R}$, weighted by how g distributes mass, rather than by uniform distance dx.

Special case

If $g(x) = x$, then $\mu_g((a, b]) = b - a$ is the ordinary Lebesgue (length) measure, and the Lebesgue–Stieltjes integral reduces to the familiar Lebesgue integral $\int f(x)\,dx$.

4.2 Monotonicity and Signed Measures

The construction of a Lebesgue–Stieltjes measure μ_g works most cleanly when g is *monotone increasing*, since in that case $g(b) - g(a) \geq 0$ for all $a < b$, ensuring that μ_g assigns nonnegative "mass" to every interval.

- If g is **monotone increasing**, then

  $$\mu_g((a, b]) = g(b) - g(a) \geq 0,$$

  and μ_g defines a **positive measure**. In this setting, g behaves analogously to a cumulative distribution function: μ_g represents how total "mass" is distributed along the real line.

- If g has **bounded variation** but is not monotone (for example, $g(x) = \sin x$), then g can be decomposed as the *difference of two monotone increasing functions*:

  $$g = g_1 - g_2,$$

  where g_1 and g_2 capture the cumulative upward and downward movements of g, respectively. Each of g_1 and g_2 induces a positive measure μ_{g_1} and μ_{g_2}, and the corresponding Lebesgue–Stieltjes measure for g is their difference:

  $$\mu_g = \mu_{g_1} - \mu_{g_2}.$$

  The result is a **signed measure** that can assign both positive and negative "mass" to sets, reflecting the alternating direction of g's variation.

Interpretation

Monotonicity of g ensures that its associated measure μ_g is positive and additive. When g oscillates but still has bounded total movement, it can be represented as the net effect of two positive measures—one corresponding to its upward variation and one to its downward variation. This is the measure-theoretic version of the *Jordan decomposition*.

4.3 Example: $g(x) = x^2$ on $[0, 2]$

Consider $g(x) = x^2$ on the interval $[0, 2]$. Since g is continuous and strictly increasing, it induces a positive Lebesgue–Stieltjes measure μ_g defined by

$$\mu_g((a, b]) = g(b) - g(a) = b^2 - a^2.$$

Induced measure interpretation

This measure assigns to each interval $(a, b]$ a "mass" equal to the change in x^2 across that interval. Because $g'(x) = 2x$, the density of this measure relative to ordinary Lebesgue measure increases linearly with x; more "weight" is concentrated toward the right end of the interval.

Example integral

To compute a simple case,

$$\int_0^2 1 \, d\mu_g = \mu_g([0, 2]) = g(2) - g(0) = 4 - 0 = 4.$$

Thus, integrating the constant function $f(x) = 1$ with respect to μ_g returns the total "mass" of μ_g over $[0, 2]$.

Geometric meaning

In the ordinary Riemann integral $\int_0^2 1 \, dx$, each unit interval of x contributes equally to the total area. In the Lebesgue–Stieltjes integral $\int_0^2 1 \, d\mu_g$, each region's contribution is weighted by how much $g(x)$ increases there. Because $g(x) = x^2$ grows faster for large x, the measure μ_g assigns greater weight near the right endpoint.

4.4 Example: Jump Function

Let

$$g(x) = \mathbf{1}_{x \geq 1} = \begin{cases} 0, & x < 1, \\ 1, & x \geq 1. \end{cases}$$

The function g is monotone increasing and exhibits a single jump of size 1 at the point $x = 1$. Its associated Lebesgue–Stieltjes measure μ_g is therefore concentrated entirely at that jump:

$$\mu_g((a,b]) = \begin{cases} 0, & b \le 1, \\ 1, & a < 1 < b, \\ 0, & a \ge 1. \end{cases}$$

Measure interpretation

The measure μ_g assigns a "mass" of 1 to any interval that includes the point $x = 1$ and zero elsewhere. Hence, μ_g is precisely the **Dirac measure at** $x = 1$, denoted δ_1:

$$\mu_g = \delta_1, \qquad \mu_g(A) = \begin{cases} 1, & 1 \in A, \\ 0, & 1 \notin A. \end{cases}$$

Integration with respect to μ_g

For any measurable function f,

$$\int_{\mathbb{R}} f(x)\, d\mu_g(x) = \int_{\mathbb{R}} f(x)\, d\delta_1(x) = f(1).$$

The entire "mass" of the integral is concentrated at the jump point.

Interpretation

A jump of size Δg at a point x_0 corresponds to a discrete unit of measure located at x_0. Integrating with respect to such a g "picks out" the value of f at the jump. This connects step functions to atomic (Dirac) measures and illustrates how Lebesgue–Stieltjes integration naturally accommodates discrete as well as continuous contributions.

5 Jordan Decomposition Example: $g(x) = \sin x$

5.1 Total Variation

To illustrate the connection between bounded variation and the construction of signed measures, consider the function

$$g(x) = \sin x \quad \text{on } [0, 2\pi].$$

Although g is not monotone, it has *finite total variation.*

Computation of $V_g(x)$

By definition,

$$V_g(x) = \sup_{\Pi} \sum_i |g(t_{i+1}) - g(t_i)|.$$

The function $\sin x$ oscillates between -1 and 1 over $[0, 2\pi]$:

$$\sin 0 = 0, \quad \sin \frac{\pi}{2} = 1, \quad \sin \pi = 0, \quad \sin \frac{3\pi}{2} = -1, \quad \sin 2\pi = 0.$$

The cumulative "distance traveled" by the graph across one full cycle is obtained by summing the absolute changes between successive extrema:

$$|1 - 0| + |0 - 1| + |-1 - 0| + |0 - (-1)| = 1 + 1 + 1 + 1 = 4.$$

Hence, the total variation of $\sin x$ over $[0, 2\pi]$ is finite and equal to

$$V_g([0, 2\pi]) = 4.$$

Observation

The function $g(x) = \sin x$ rises and falls symmetrically, covering a total "vertical distance" of 4 units in one full oscillation. Because this total movement is finite, $\sin x$ is of **bounded variation**, even though it is not monotone. This property will allow g to be represented as the difference of two monotone increasing functions in the Jordan decomposition.

5.2 Decomposition

Since $g(x) = \sin x$ is of bounded variation on $[0, 2\pi]$, the **Jordan decomposition theorem** guarantees that g can be written as the difference of two monotone increasing functions:

$$g = g_1 - g_2.$$

Intuitively, g_1 accumulates the *upward* variation of $\sin x$, while g_2 accumulates the *downward* variation.

Constructive formulas

A convenient explicit form for these monotone components is

$$g_1(x) = \frac{V_g(x) + g(x)}{2}, \qquad g_2(x) = \frac{V_g(x) - g(x)}{2},$$

where $V_g(x)$ denotes the cumulative total variation of g from 0 up to the point x:

$$V_g(x) = \sup_{\Pi_x} \sum_i |g(t_{i+1}) - g(t_i)|.$$

For $g(x) = \sin x$ on $[0, 2\pi]$, this cumulative variation increases by 1 each time the function moves from a zero crossing to an adjacent extremum, so that

$$V_g(0) = 0, \quad V_g(\tfrac{\pi}{2}) = 1, \quad V_g(\pi) = 2, \quad V_g(\tfrac{3\pi}{2}) = 3, \quad V_g(2\pi) = 4.$$

Verification at key points

Using the above formulas:

x	0	$\frac{\pi}{2}$	π	$\frac{3\pi}{2}$	2π
$\sin x$	0	1	0	-1	0
$V_g(x)$	0	1	2	3	4
$g_1(x)$	0	1	1	1	2
$g_2(x)$	0	0	1	2	2

Both g_1 and g_2 are monotone increasing, as required. Their difference reconstructs the original function:

$$g_1(x) - g_2(x) = \sin x.$$

Geometric intuition

- g_1 increases during the upward motions of $\sin x$ and stays flat during the downward ones.
- g_2 increases during the downward motions and remains flat during the upward ones.

Each is a nondecreasing "cumulative counter" of how much total vertical motion has occurred in one direction.

Interpretation

The Jordan decomposition expresses any bounded-variation function as the difference of two monotone increasing ones. For $g(x) = \sin x$, the upward and downward movements are stored separately in g_1 and g_2, making it possible to define a signed measure $\mu_g = \mu_{g_1} - \mu_{g_2}$ and to integrate with respect to g within the Lebesgue–Stieltjes framework.

5.3 Signed Measure Interpretation

Once a function g of bounded variation is decomposed as

$$g = g_1 - g_2,$$

with g_1 and g_2 monotone increasing, we can define corresponding Lebesgue–Stieltjes measures

$$\mu_{g_1}((a,b]) = g_1(b) - g_1(a), \qquad \mu_{g_2}((a,b]) = g_2(b) - g_2(a).$$

Each of these is a positive measure, since the defining functions are monotone. The measure associated with g itself is then given by the difference

$$\mu_g = \mu_{g_1} - \mu_{g_2}.$$

Meaning of a signed measure

The object μ_g is called a **signed measure** because it can assign positive or negative "mass" to sets, depending on whether the corresponding variation in g was upward or downward. In this sense, μ_g encodes both the direction and magnitude of g's cumulative motion.

Integration with respect to a signed measure

For any measurable function f, we define

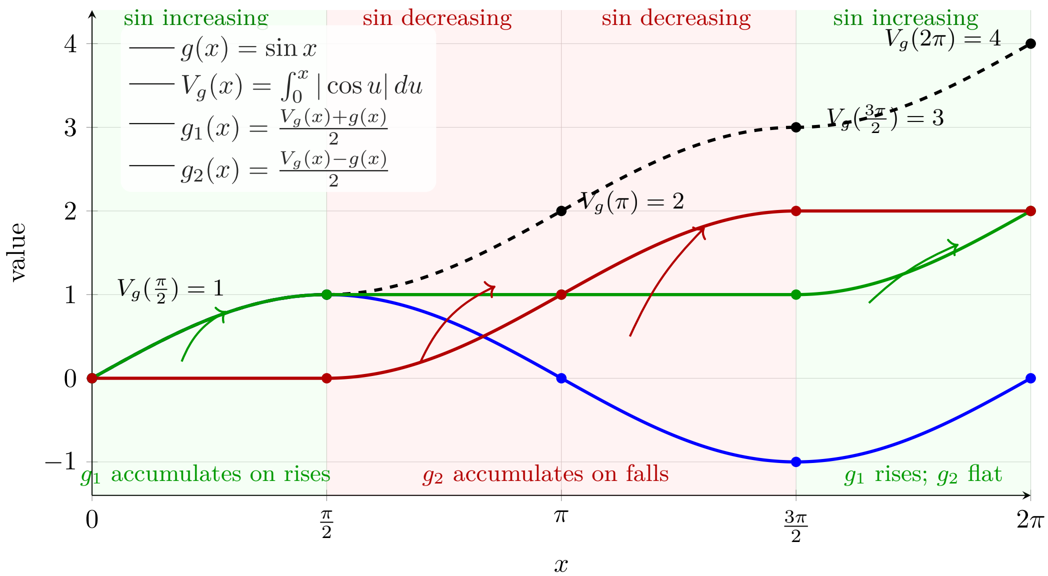
$$\int f(x)\, d\mu_g(x) = \int f(x)\, d\mu_{g_1}(x) - \int f(x)\, d\mu_{g_2}(x),$$

where both terms on the right are ordinary Lebesgue integrals with respect to positive measures.

Intuition

Every bounded-variation function g induces a *signed measure* μ_g, representing the net accumulation of its upward and downward movements. Integrating f with respect to dg therefore means taking the difference between the accumulation of f over the increasing parts of g and its accumulation over the decreasing parts. This viewpoint allows integration to remain well-defined even when g oscillates, provided its total variation is finite.

Jordan Decomposition of $g(x) = \sin x$ on $[0, 2\pi]$



g_1 increases on up-moves, flat on down-moves.
g_2 increases on down-moves, flat on up-moves.
$g = g_1 - g_2, \quad V_g = g_1 + g_2.$

6 Wiener Processes and Infinite Variation

6.1 Why Brownian Paths Fail for Riemann–Stieltjes

A standard Wiener process W_t (Brownian motion) provides the canonical example of a continuous function whose total variation is infinite on every finite interval. Although its paths are continuous almost surely, they fluctuate too violently to admit a classical Riemann–Stieltjes integral.

Expected total variation along a partition

Consider a uniform partition of the interval $[0,T]$ into n subintervals of length $\Delta t = T/n$. The total variation along that partition is approximated by

$$V(W;\Pi_n) = \sum_{i=0}^{n-1} |W_{t_{i+1}} - W_{t_i}|.$$

Each increment $W_{t_{i+1}} - W_{t_i}$ is independent and normally distributed:

$$W_{t_{i+1}} - W_{t_i} \sim N(0,\, \Delta t).$$

Therefore, by linearity of expectation,

$$\mathbb{E}[V(W;\Pi_n)] = n\,\mathbb{E}|N(0,\Delta t)| = n\sqrt{\frac{2\Delta t}{\pi}} = \sqrt{\frac{2T}{\pi}}\,\sqrt{n}.$$

As the partition is refined ($n \to \infty$), this expectation diverges:

$$\mathbb{E}\,V(W;[0,T]) = \lim_{n\to\infty} \mathbb{E}[V(W;\Pi_n)] = \infty.$$

Detailed derivation: $\mathbb{E}[V(W;\Pi_n)] = \sqrt{\frac{2T}{\pi}}\,\sqrt{n}$

Setup (uniform partition). Fix $T > 0$ and divide $[0,T]$ into n equal pieces:

$$\Pi_n : \quad 0 = t_0 < t_1 < \cdots < t_n = T, \qquad t_{i+1} - t_i = \Delta t := \frac{T}{n}.$$

Define the (partition-level) variation sum

$$V(W;\Pi_n) \;:=\; \sum_{i=0}^{n-1} \big|W_{t_{i+1}} - W_{t_i}\big|.$$

Key fact about Brownian increments. For Brownian motion W, each increment is Gaussian with mean 0 and variance equal to the time step:

$$W_{t_{i+1}} - W_{t_i} \ \sim\ N(0, \Delta t).$$

(Independence of increments is true, but *not* needed for the next step; linearity of expectation does not require independence.)

Step 1: Linearity of expectation. By linearity,

$$\mathbb{E}\big[V(W;\Pi_n)\big] = \mathbb{E}\left[\sum_{i=0}^{n-1} \big|W_{t_{i+1}} - W_{t_i}\big|\right] = \sum_{i=0}^{n-1} \mathbb{E}\big[\big|W_{t_{i+1}} - W_{t_i}\big|\big]\,.$$

Because all n increments are identically distributed $N(0, \Delta t)$,

$$\mathbb{E}\big[V(W;\Pi_n)\big] = n\,\mathbb{E}\big[\,\big|N(0,\Delta t)\big|\,\big]\,.$$

Step 2: Scaling a normal variable. Recalling that if $Z \sim N(0,1)$ (a standard normal) then for any real a, $aZ \sim N(0, a^2)$, we can write $N(0, \Delta t)$ as $\sqrt{\Delta t}\,Z$ with $Z \sim N(0,1)$. Then

$$\mathbb{E}\big[\,\big|N(0,\Delta t)\big|\,\big] = \mathbb{E}\Big[\,\big|\sqrt{\Delta t}\,Z\big|\,\Big] = \sqrt{\Delta t}\,\mathbb{E}[\,|Z|\,]\,.$$

So

$$\mathbb{E}\big[V(W;\Pi_n)\big] = n\,\sqrt{\Delta t}\,\mathbb{E}[\,|Z|\,]\,.$$

Step 3: Compute $\mathbb{E}[|Z|]$ for $Z \sim N(0,1)$. Let $\phi(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$ be the standard normal density. By symmetry,

$$\mathbb{E}[|Z|] = \int_{-\infty}^{\infty} |x|\phi(x)\,dx = 2\int_0^{\infty} x\,\phi(x)\,dx.$$

Use the identity $\phi'(x) = -x\phi(x)$, i.e. $x\phi(x) = -\phi'(x)$. Then

$$\int_0^{\infty} x\,\phi(x)\,dx = \int_0^{\infty} -\phi'(x)\,dx = \big[-\phi(x)\big]_0^{\infty} = 0 - \big(-\phi(0)\big) = \phi(0) = \frac{1}{\sqrt{2\pi}}.$$

Therefore

$$\mathbb{E}[|Z|] = 2 \cdot \frac{1}{\sqrt{2\pi}} = \sqrt{\frac{2}{\pi}}.$$

Step 4: Put it all together. Plugging $\mathbb{E}[|Z|] = \sqrt{2/\pi}$ into Step 2:

$$\mathbb{E}\big[V(W;\Pi_n)\big] = n\,\sqrt{\Delta t}\,\sqrt{\frac{2}{\pi}} = n\,\sqrt{\frac{T}{n}}\,\sqrt{\frac{2}{\pi}} = \sqrt{\frac{2T}{\pi}}\,\sqrt{n}.$$

Conclusion. As $n \to \infty$ (finer partitions), $\mathbb{E}[V(W;\Pi_n)] \sim C\sqrt{n}$ with $C = \sqrt{2T/\pi}$, hence it diverges. This growth already signals that the total variation (the supremum over all partitions) is infinite almost surely.

Almost sure divergence

This growth of the expected variation implies that, with probability one, the total variation $V(W;[0,T])$ itself is infinite. Intuitively, as we inspect the Brownian path on finer and finer scales, we discover ever more oscillations—each small in amplitude but collectively unbounded in total length.

> **Key consequence**
>
> Brownian motion W_t is continuous but of **unbounded total variation** on every finite interval. Hence, $\int_0^T f(t)\,dW_t$ cannot be defined in the Riemann–Stieltjes sense, even for smooth integrands f. This failure motivates the development of Itô integration, which replaces finite total variation with finite *quadratic variation*.

6.2 But finite quadratic variation

While Brownian motion has infinite *total* variation, its *quadratic* variation is finite and, in fact, deterministic. Let $\Pi = \{0 = t_0 < t_1 < \cdots < t_n = T\}$ be any partition and define

$$Q_\Pi := \sum_{i=0}^{n-1} \bigl(W_{t_{i+1}} - W_{t_i}\bigr)^2.$$

As the mesh $|\Pi| = \max_i(t_{i+1} - t_i) \to 0$, we have

$$Q_\Pi \;\longrightarrow\; T,$$

in L^2 (hence in probability). This limit is called the *quadratic variation* of Brownian motion and is denoted $[W]_T = T$.

Computation under a uniform partition

Take $t_{i+1} - t_i = \Delta t = T/n$. Using independent Gaussian increments,

$$W_{t_{i+1}} - W_{t_i} \sim N(0,\Delta t), \qquad \mathbb{E}\bigl[(W_{t_{i+1}} - W_{t_i})^2\bigr] = \Delta t, \qquad \operatorname{Var}\bigl((W_{t_{i+1}} - W_{t_i})^2\bigr) = 2(\Delta t)^2.$$

Therefore,

$$\mathbb{E}[Q_\Pi] = \sum_{i=0}^{n-1} \Delta t = T, \qquad \operatorname{Var}(Q_\Pi) = \sum_{i=0}^{n-1} 2(\Delta t)^2 = 2n(\Delta t)^2 = \frac{2T^2}{n} \xrightarrow[n\to\infty]{} 0.$$

Hence $Q_\Pi \to T$ in L^2, so $Q_\Pi \to T$ in probability and along subsequences almost surely (by standard subsequence arguments).

Convergence in L^2 and What the "L" Means

Meaning of L^p spaces. The letter L stands for *Lebesgue*—these spaces were first defined in Lebesgue's theory of integration. For a random variable X (or a measurable function) on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, the space $L^p(\Omega)$ consists of all random variables whose p-th absolute moment is finite:

$$L^p(\Omega) = \big\{X : \ \mathbb{E}[|X|^p] < \infty\big\}.$$

Each such space carries the norm $\|X\|_{L^p} = (\mathbb{E}[|X|^p])^{1/p}$.
When $p = 2$, the space $L^2(\Omega)$ is a **Hilbert space**, meaning it has an inner product

$$\langle X, Y\rangle = \mathbb{E}[XY]$$

and is *complete*: every Cauchy sequence (in this norm) converges to a unique limit within L^2.

What convergence in L^2 means. A sequence of random variables X_n *converges in L^2* to X if

$$\mathbb{E}[(X_n - X)^2] \longrightarrow 0.$$

This is also called *mean-square convergence*. It measures closeness in the sense of expected squared error, not pathwise agreement.

Applying this to the quadratic variation sum. For

$$Q_\Pi = \sum_{i=0}^{n-1} (W_{t_{i+1}} - W_{t_i})^2,$$

we computed

$$\mathbb{E}[Q_\Pi] = T, \qquad \mathrm{Var}(Q_\Pi) = \frac{2T^2}{n}.$$

Because $\mathbb{E}[Q_\Pi]$ already equals T,

$$\mathbb{E}[(Q_\Pi - T)^2] = \mathrm{Var}(Q_\Pi) + (\mathbb{E}[Q_\Pi] - T)^2 = \mathrm{Var}(Q_\Pi) = \tfrac{2T^2}{n} \longrightarrow 0.$$

Hence $Q_\Pi \to T$ in L^2.

Interpretation. Convergence in L^2 means that the sequence of random variables becomes *mean-square close* to the limit: the expected squared deviation shrinks to zero. Because $L^2(\Omega)$ is complete, this guarantees the limit exists and is unique as an element of that space.

Why Convergence in L^2 Implies Convergence in Probability

1. Formal Statement. Let X_n, X be random variables with finite second moments. If

$$\mathbb{E}\big[(X_n - X)^2\big] \to 0,$$

then $X_n \to X$ **in probability**.

2. Proof Sketch (Chebyshev's Inequality). For any $\varepsilon > 0$,

$$\mathbb{P}(|X_n - X| > \varepsilon) \le \frac{\mathbb{E}[(X_n - X)^2]}{\varepsilon^2}.$$

If $\mathbb{E}[(X_n - X)^2] \to 0$, then the right-hand side tends to 0, so

$$\mathbb{P}(|X_n - X| > \varepsilon) \to 0,$$

which is exactly the definition of convergence in probability.

3. Intuitive Summary. Convergence in L^2 (mean-square) controls not only how often large deviations occur, but also their expected *size*. Thus it is a *stronger* form of convergence:

$$L^2 \;\Rightarrow\; L^1 \;\Rightarrow\; \text{in probability} \;\Rightarrow\; \text{in distribution}.$$

Each implication is one-way only.

General (nonuniform) partitions

For any partition Π,

$$\mathbb{E}[Q_\Pi] = \sum_i (t_{i+1} - t_i) = T, \qquad \mathrm{Var}(Q_\Pi) = 2 \sum_i (t_{i+1} - t_i)^2 \;\le\; 2\,|\Pi| \sum_i (t_{i+1} - t_i) = 2T\,|\Pi| \;\to\; 0,$$

as $|\Pi| \to 0$. Thus $Q_\Pi \to T$ in L^2 for *any* refining sequence of partitions, not just uniform ones.

Key Message

Brownian motion has **finite quadratic variation** $[W]_T = T$ but **infinite total variation**. Quadratic variation is the correct "roughness scale" for stochastic calculus: Itô integration and Itô's lemma are built around this property, replacing pathwise bounded variation by L^2 control of squared increments.

Intuitive contrast. The sum of absolute increments $\sum |\Delta W_i|$ diverges because the path contains infinitely many small oscillations—each contributing positively to the total "distance traveled." When these same infinitesimal fluctuations are *squared*, however, their magnitude shrinks much faster than their frequency grows. The explosion of tiny increments is thus tamed by the squaring operation, and the cumulative effect stabilizes at a finite limit. In this sense, quadratic variation captures the "right" level of roughness: fine enough to feel every infinitesimal jiggle, yet coarse enough for the total to remain finite.

7 From Riemann to Itô: Mean-Square Integration

The failure of the Riemann–Stieltjes integral for Brownian motion forces us to redefine what it means for an "integral" to exist. Instead of demanding that Riemann sums converge *pathwise* for every realization of the integrator, we require convergence *in mean square*—that is, in the L^2 sense of random variables. This shift from geometric convergence to probabilistic convergence gives rise to the **Itô integral**.

7.1 Riemann-type stochastic sums

Given a partition $\Pi = \{0 = t_0 < t_1 < \cdots < t_n = T\}$, define the stochastic Riemann sum

$$S_\Pi = \sum_{i=0}^{n-1} f(t_i)\,[W_{t_{i+1}} - W_{t_i}],$$

where $f(t_i)$ is evaluated at the *left endpoint* of each interval. Each term is a random variable, since the Brownian increments $W_{t_{i+1}} - W_{t_i}$ are random.

We seek a random variable I (to be interpreted as the stochastic integral) such that these sums converge to I in *mean square*:

$$\lim_{|\Pi|\to 0} \mathbb{E}\big[(S_\Pi - I)^2\big] = 0.$$

When this limit exists, we define

$$I = \int_0^T f(t)\,dW_t.$$

What "mean-square limit" means

For random variables X_n and X, convergence in mean square (or L^2) means

$$\mathbb{E}\big[(X_n - X)^2\big] \longrightarrow 0.$$

This notion measures convergence in the *average squared distance* sense, not for each individual sample path. In stochastic integration, this weaker but probabilistically natural convergence replaces the pathwise limit used by Riemann.

Conceptual shift

Riemann: requires pointwise convergence of deterministic sums $\sum f(t_i)\,\Delta x_i$ for all partitions.
Itô: requires convergence of random sums $\sum f(t_i)\,\Delta W_i$ in the mean-square sense:

$$\mathbb{E}\big[(S_\Pi - I)^2\big] \to 0.$$

This relaxation is crucial: although Brownian paths are too irregular for pathwise limits, their increments are "regular enough on average" for L^2 convergence to hold.

Interpretation

The Itô integral $\int_0^T f(t)\,dW_t$ is thus defined as a random variable that captures the limiting contribution of infinitesimal products $f(t)\,dW_t$ in the mean-square sense. It preserves many of the linearity and isometry properties of the classical integral, but its existence depends on the integrand f being *square-integrable* rather than continuous. This forms the probabilistic foundation for modern stochastic calculus.

Square-Integrable Random Variables

A random variable X is said to be **square-integrable** if

$$\mathbb{E}[X^2] < \infty.$$

This means X lies in the space $L^2(\Omega, \mathcal{F}, \mathbb{P})$, the set of all random variables with finite second moment.

Why it matters: Square-integrability guarantees that X has a finite mean and finite variance, and allows us to define inner products

$$\langle X, Y\rangle = \mathbb{E}[XY],$$

which makes L^2 a *Hilbert space*. In stochastic calculus, the Itô integral is built precisely on such L^2 spaces.

Pathwise vs. Mean-Square Convergence

The problem. For Brownian motion, the Riemann-type sums

$$S_\Pi = \sum f(t_i)\,\Delta W_i$$

do not converge for individual sample paths as the partition is refined. Each trajectory of W_t is continuous but of infinite local variation: as

$|\Pi| \to 0$, the increments ΔW_i become smaller in size but vastly more numerous, and their random signs cause wild, persistent oscillations. There is simply no stable *pathwise* limit.

The partial exception. The squared increments do behave better:

$$Q_\Pi = \sum (\Delta W_i)^2 \;\to\; T,$$

and this convergence holds even pathwise (almost surely). But this describes only the *energy* of the noise—the cumulative size of fluctuations—not the mixed product $f(t_i)\Delta W_i$ needed for an integral.

Why we need the mean-square framework. The Itô integral cannot be constructed geometrically, path by path. Instead, we treat it as a limit of random variables that stabilizes *on average* across all realizations of the process:

$$\lim_{|\Pi|\to 0} \mathbb{E}\big[(S_\Pi - I)^2\big] = 0.$$

This criterion demands that the expected squared error between the approximation S_Π and the limiting random variable I vanish. Even though each trajectory of W_t remains infinitely jagged, the ensemble of all trajectories has predictable second-moment structure:

$$\mathbb{E}[(\Delta W_i)^2] = \Delta t, \qquad \mathbb{E}[\Delta W_i \, \Delta W_j] = 0 \text{ for } i \neq j.$$

These properties make the random sums S_Π converge in the L^2 sense, giving a well-defined stochastic integral.

Intuitive picture. Think of Brownian motion as an infinitely rough coastline. Its total length (absolute increments) diverges no matter how fine your ruler, so no pathwise integral can exist. But if you measure not the literal length, but the *average squared displacement* contributed by each small wave, the explosion of tiny wiggles is tamed. Mean-square convergence captures this idea: each single coastline is chaotic, but the *average energy across all coastlines* is stable.

Bottom line. Itô integration replaces the impossible geometric limit (pathwise convergence) with a probabilistic one (mean-square convergence). The integral exists not because each individual path behaves nicely, but because the entire ensemble of paths has a regular and computable second-moment structure.

Why Mean-Square Convergence Works

The squared Brownian increments $(\Delta W_i)^2$ converge pathwise because they are always positive and reflect the cumulative *energy* of the noise.

However, the actual stochastic integral involves the *mixed products* $f(t_i)\Delta W_i$, where each ΔW_i can take random positive or negative values and oscillate infinitely often on any finite interval.

Pathwise failure. For a single Brownian trajectory ω, the random signs of $\Delta W_i(\omega)$ prevent the Riemann-type sums $\sum f(t_i)\Delta W_i(\omega)$ from stabilizing. As the partition refines, the number of terms explodes while their individual magnitudes shrink, producing endless cancellation and amplification—a chaotic sequence with no pathwise limit.

Statistical salvation. Across the entire ensemble of Brownian paths, these oscillations are perfectly balanced:

$$\mathbb{E}[\Delta W_i] = 0, \qquad \mathbb{E}[(\Delta W_i)^2] = \Delta t, \qquad \mathbb{E}[\Delta W_i\, \Delta W_j] = 0 \text{ for } i \neq j.$$

Thus, while no individual trajectory is well-behaved, the *distribution of trajectories* has precise second-moment structure. By measuring convergence in mean-square,

$$\mathbb{E}\big[(S_\Pi - I)^2\big] \to 0,$$

we smooth out the pathwise chaos and capture the predictable average behavior of the noise.

The key idea. The Itô integral exists not because each path of W_t behaves nicely, but because the collection of all paths is statistically regular. Mean-square convergence harnesses this symmetry: random fluctuations that are wild pointwise become perfectly controlled when viewed through expectation.

Analogy. Each Brownian path is an infinitely jagged coastline; its detailed length is meaningless. But if we measure not the literal length, but the *average squared displacement* across all coastlines, the result converges to a finite, stable quantity. This is the spirit of Itô integration.

7.2 Itô Isometry

The Itô integral is defined first for *simple adapted* (step) processes and then extended by L^2–closure. The cornerstone identity controlling this limit is the **Itô isometry**.

Step 1: Define the integral for simple adapted processes

Fix a partition $0 = t_0 < t_1 < \cdots < t_n = T$ and let

$$H(t) = \sum_{k=0}^{n-1} \xi_k \, \mathbf{1}_{(t_k, t_{k+1}]}(t), \qquad \text{with } \ \xi_k \ \mathcal{F}_{t_k}\text{-measurable (adapted).}$$

Define

$$\int_0^T H_t \, dW_t \ := \ \sum_{k=0}^{n-1} \xi_k \, (W_{t_{k+1}} - W_{t_k}).$$

Step 2: Prove the isometry for simple processes

Use that $W_{t_{k+1}} - W_{t_k}$ is independent of $\mathcal{F}_{t_k}$, mean 0, variance $(t_{k+1} - t_k)$, and that nonoverlapping increments are independent:

$$\mathbb{E}\left[\left(\int_0^T H_t \, dW_t\right)^2\right] = \mathbb{E}\left[\left(\sum_{k=0}^{n-1} \xi_k \, \Delta W_k\right)^2\right] = \sum_{k=0}^{n-1} \mathbb{E}\big[\xi_k^2 \, (\Delta W_k)^2\big]$$

(cross terms vanish by independence/zero mean)

$$= \sum_{k=0}^{n-1} \mathbb{E}\big[\xi_k^2 \, \mathbb{E}\big[(\Delta W_k)^2 \mid \mathcal{F}_{t_k}\big]\big] = \sum_{k=0}^{n-1} \mathbb{E}\big[\xi_k^2\big] \ (t_{k+1} - t_k)$$

$$= \mathbb{E}\left[\int_0^T H_t^2 \, dt\right].$$

This identity is the *Itô isometry* for simple H.

Step 3: Extend by L^2–closure (general predictable f)

Let

$$\mathcal{H}^2 := \left\{ f \text{ predictable} : \ \ \mathbb{E}\int_0^T f(t)^2 \, dt < \infty \right\}.$$

Simple adapted processes are dense in $\mathcal{H}^2$ under the norm $\|f\|_{\mathcal{H}^2}^2 = \mathbb{E}\int_0^T f^2 \, dt$. For $f \in \mathcal{H}^2$, choose simple $H^{(m)} \to f$ in $\mathcal{H}^2$ and *define*

$$\int_0^T f \, dW \ := \ L^2\text{-}\lim_{m \to \infty} \int_0^T H^{(m)} \, dW.$$

The isometry passes to the limit:

$$\boxed{\mathbb{E}\left[\left(\int_0^T f(t) \, dW_t\right)^2\right] \ = \ \mathbb{E}\left[\int_0^T f(t)^2 \, dt\right].}$$

Immediate consequences

- **Zero mean:** $\mathbb{E}\Big[\int_0^T f\,dW\Big] = 0$ for $f \in \mathcal{H}^2$.
- **Variance formula:** $\mathrm{Var}\Big(\int_0^T f\,dW\Big) = \mathbb{E}\int_0^T f^2\,dt$.
- **Martingale property:** $M_t := \int_0^t f\,dW$ is an L^2-martingale with quadratic variation $[M]_t = \int_0^t f(s)^2\,ds$.
- **Orthogonality of increments:** If $0 \le s \le t \le u \le v \le T$, then $\int_s^t f\,dW$ is independent of $\int_u^v f\,dW$ *given* $\mathcal{F}_t$ when f is predictable (in particular, disjoint-increment covariances vanish).

Why the isometry matters

The Itô isometry replaces geometric control (bounded variation) with *second-moment* control: to integrate f against Brownian motion, it suffices that f be square-integrable in time (predictable). This L^2 identity both *defines* the stochastic integral (via closure) and quantifies its size. In effect,

$$\| \int_0^T f\,dW \|_{L^2(\Omega)} = \| f \|_{L^2([0,T]\times\Omega)}.$$

L^p–Martingales and Why L^2 Is Special

Definition. A stochastic process $(M_t)_{t\ge 0}$ adapted to a filtration $(\mathcal{F}_t)$ is an *L^p–martingale* if:

$$\mathbb{E}[|M_t|^p] < \infty, \qquad \mathbb{E}[M_t \mid \mathcal{F}_s] = M_s \quad \text{for all } s < t.$$

Hierarchy. L^p spaces form a nesting:

$$L^p \subseteq L^q \quad \text{for } p > q,$$

so every L^2–martingale is also an L^1–martingale, but not vice versa.
Why L^2 is special.

- $L^2(\Omega)$ is a *Hilbert space* with inner product $\langle X, Y\rangle = \mathbb{E}[XY]$.
- This allows geometric tools: orthogonality, projections, and the Itô isometry:
$$\mathbb{E}\left[\left(\int_0^T f\,dW\right)^2\right] = \mathbb{E}\int_0^T f^2\,dt.$$
- Key theorems (Doob's inequalities, martingale representation, predictable quadratic variation) all exploit the Hilbert structure.

Examples.

$$W_t \text{ and } M_t = \int_0^t f(s)\,dW_s$$

are both L^2–martingales since $\mathbb{E}[W_t^2] = t < \infty$ and $\mathbb{E}[M_t^2] = \mathbb{E}\int_0^t f(s)^2\,ds < \infty$.

Summary table:

| Type | Integrability | Space | Key Use | |:–|:–|:–|:–| | L^1 | $\mathbb{E}[|M_t|] < \infty$ | Banach | General martingale theory | | L^2 | $\mathbb{E}[M_t^2] < \infty$ | Hilbert | Itô calculus, isometries | | L^p, $p > 2$ | $\mathbb{E}[|M_t|^p] < \infty$ | Banach | Strong moment bounds |

7.3 Example

For the constant integrand $f(t) \equiv 1$, the Itô integral reduces to the net Brownian increment over $[0, T]$:

$$\int_0^T 1\,dW_t \;=\; W_T.$$

Derivation via stochastic Riemann sums

Take any partition $\Pi = \{0 = t_0 < \cdots < t_n = T\}$ and form

$$S_\Pi = \sum_{i=0}^{n-1} 1 \cdot \left(W_{t_{i+1}} - W_{t_i}\right) = W_T - W_0 = W_T,$$

since the increments telescope. Thus S_Π is *identically* W_T for every partition, so the mean-square limit is immediate:

$$\lim_{|\Pi|\to 0} \mathbb{E}\big[(S_\Pi - W_T)^2\big] = 0.$$

Hence, by definition of the Itô integral,

$$\boxed{\int_0^T 1\,dW_t = W_T.}$$

Checks via Itô isometry

The isometry gives

$$\mathbb{E}\left[\left(\int_0^T 1\,dW_t\right)^2\right] = \mathbb{E}\left[\int_0^T 1^2\,dt\right] = T,$$

which matches $\mathrm{Var}(W_T) = T$ and $\mathbb{E}[W_T] = 0$, confirming the result.

Stochastic Integrals Yield Random Variables

Key observation. In Itô calculus, the result of an integral like

$$\int_0^T f(t)\, dW_t$$

is *not* a deterministic number—it is a **random variable**. More precisely, it is a function of the underlying probability space $(\Omega, \mathcal{F}, \mathbb{P})$ that depends on the particular sample path of the Brownian motion.

Example revisited. For the constant integrand $f(t) \equiv 1$,

$$\int_0^T 1\, dW_t = W_T.$$

Here W_T is itself a random variable:

$$W_T : \Omega \to \mathbb{R}, \qquad W_T(\omega) \sim N(0, T).$$

Each ω represents a particular Brownian path; along that path, the integral simply returns the final displacement of the noise. Thus the integral's "value" changes randomly from one trajectory to another.

Why this differs from classical calculus. In classical (Riemann or Lebesgue) integration:

- both f and the variable of integration x are deterministic;
- the integral $\int f(x)\, dx$ is a single real number representing the signed area under a curve.

In Itô integration:

- $W_t(\omega)$ is random—each ω gives a different, infinitely rough path;
- $f(t, \omega)$ may also depend on the random history of W_t;
- $\int_0^T f(t)\, dW_t$ is therefore a *random variable* (depending on ω), or equivalently, a new stochastic process $I_t = \int_0^t f(s)\, dW_s$.

Instead of computing a deterministic area, we are now constructing a new random quantity whose distribution and variance can be analyzed statistically.

What makes this definition meaningful. The integral exists not because the Brownian path is smooth (it is not), but because the *ensemble of all paths* has a perfectly regular second-moment structure:

$$\mathbb{E}[(\Delta W_i)^2] = \Delta t, \qquad \mathbb{E}[\Delta W_i\, \Delta W_j] = 0 \text{ for } i \neq j.$$

Even though each individual trajectory oscillates infinitely, those oscillations are statistically balanced. When we define the integral in the *mean-square* sense,

$$\mathbb{E}\big[(S_\Pi - I)^2\big] \to 0,$$

the randomness "averages out," and we obtain a stable random variable I with well-defined expectation and variance.

Why this still deserves to be called an integral. The Itô integral retains the core structural properties of integration, but expressed in probabilistic rather than geometric terms:

$$\mathbb{E}\left[\left(\int_0^T f\,dW\right)^2\right] = \mathbb{E}\left[\int_0^T f^2(t)\,dt\right],$$

(the *Itô isometry*) and

$$\mathbb{E}\left[\int_0^T f\,dW\right] = 0.$$

It is linear, satisfies an analogue of the fundamental theorem of calculus for stochastic processes, and defines a martingale in t:

$$M_t = \int_0^t f(s)\,dW_s.$$

So although its outcome is random, its behavior is as mathematically stable and structured as any classical integral—only in the space of random variables.

Big picture intuition.

- In deterministic calculus, $\int f\,dx$ measures *accumulated area.*
- In stochastic calculus, $\int f\,dW$ measures *accumulated noise.*

The former is geometric; the latter is probabilistic. If the classical integral outputs a number summarizing a curve, the Itô integral outputs a random variable summarizing a distribution of curves.

Metaphor. Each Brownian path is like one coastline—infinitely jagged and unpredictable. A Riemann integral would attempt to measure the literal length of that coastline and fail. The Itô integral instead measures the *expected squared displacement* contributed by all possible coastlines together. The output is not a single length but a random displacement whose statistics we can compute exactly.

From Random Variables to Numbers: The Monte Carlo Connection

The Itô integral produces a random variable. For any integrand f, the quantity

$$I_T = \int_0^T f(t)\, dW_t$$

is a *random variable*, not a fixed number. Its value depends on the particular Brownian trajectory $W_t(\omega)$ taken by the stochastic process. Each realization of W_t gives one sample value of I_T. In this sense, Itô integration defines a *distribution* of possible outcomes rather than a single deterministic quantity.

Example (constant integrand). For $f(t) = 1$ we obtained

$$I_T = \int_0^T 1\, dW_t = W_T, \qquad W_T \sim N(0, T).$$

Each path produces one random value $W_T(\omega)$. Its expected value and variance are $\mathbb{E}[W_T] = 0$ and $\mathrm{Var}(W_T) = T$. The integral's *output* is therefore a Gaussian random variable whose mean and spread we can compute exactly.

How we get a number out of it. To extract a deterministic quantity, we apply a statistical operation such as the *expectation*:

$$\mathbb{E}[I_T] = \mathbb{E}\left[\int_0^T f(t)\, dW_t\right].$$

This gives a single real number summarizing the average outcome across all Brownian paths. In applications—especially in finance or physics—this expectation represents the "observable" quantity such as an average payoff or expected displacement.

Monte Carlo interpretation. Monte Carlo simulation is the numerical embodiment of this idea.

- Simulate M independent Brownian paths $W_t^{(m)}$.
- For each path, compute a discrete approximation $I_T^{(m)} \approx \sum f(t_i)\, [W_{t_{i+1}}^{(m)} - W_{t_i}^{(m)}]$.
- Estimate expectations by averaging:

$$\mathbb{E}[I_T] \approx \frac{1}{M} \sum_{m=1}^{M} I_T^{(m)}.$$

Each $I_T^{(m)}$ is one random sample of the Itô integral; their average converges (by the law of large numbers) to the true expectation.

Putting it all together.

Viewpoint	**Object**	**Deterministic?**	**Interpretation / Output**
Itô integral itself	$I_T = \int_0^T f\,dW_t$	No	Random variable with known distribution
Expectation	$\mathbb{E}[I_T]$	Yes	Mean (average) value across all paths
Variance / moments	$\mathbb{E}[I_T^2]$, $\mathrm{Var}(I_T)$	Yes	Measures dispersion / uncertainty
Monte Carlo estimate	$\frac{1}{M}\sum_{m=1}^{M} I_T^{(m)}$	Yes	Numerical estimate of expectation

Intuitive picture. The Itô integral gives you a *random outcome* for each random path of Brownian motion. Monte Carlo simulation samples many such paths and averages them to reveal the deterministic structure hidden underneath the randomness—typically the expectation or some function of it.

Summary.

- The integral $\int f\,dW$ is a random variable (or stochastic process).
- Deterministic quantities arise by taking expectations or other moments of that random variable.
- Monte Carlo simulation approximates those deterministic quantities by averaging over many simulated paths.

Thus: Riemann integration gives a number by summing over one curve. Itô integration gives a *distribution* by summing over infinitely many possible curves, and Monte Carlo recovers a number by averaging over them.

7.4 Quadratic Variation in Itô Calculus

Finite *quadratic variation* is the structural feature that makes Itô calculus work and is precisely what generates the *Itô correction term* in Itô's lemma. For Brownian motion,

$$[W]_t \;=\; \lim_{|\Pi|\to 0} \sum (W_{t_{i+1}} - W_{t_i})^2 \;=\; t,$$

whereas the *total* variation is infinite. More generally, for a (scalar) Itô process

$$X_t \;=\; X_0 \;+\; \int_0^t b_s\,ds \;+\; \int_0^t \sigma_s\,dW_s,$$

its quadratic variation is

$$[X]_t \;=\; \int_0^t \sigma_s^2\,ds.$$

Heuristic differential rules

The presence of finite quadratic variation gives the informal multiplication table

$$(dW_t)^2 = dt, \qquad dt\, dW_t = 0, \qquad (dt)^2 = 0,$$

and, more generally, for an Itô process X_t with diffusion coefficient σ_t,

$$(dX_t)^2 = \sigma_t^2\, dt.$$

These rules encode the non-negligible contribution of squared random increments—the essence of quadratic variation.

Itô's lemma (one dimension)

Let $f \in C^{1,2}([0,T] \times \mathbb{R})$ and X_t be as above. Then

$$\boxed{df(t,X_t) = f_t(t,X_t)\, dt + f_x(t,X_t)\, dX_t + \tfrac{1}{2} f_{xx}(t,X_t)\, d[X]_t.}$$

Substituting $dX_t = b_t\, dt + \sigma_t\, dW_t$ and $d[X]_t = \sigma_t^2\, dt$ yields

$$df(t,X_t) = \Big(f_t + b_t f_x + \tfrac{1}{2}\sigma_t^2 f_{xx}\Big) dt + \sigma_t f_x\, dW_t.$$

The term $\frac{1}{2} f_{xx}\, d[X]_t$ is the Itô correction, arising solely from quadratic variation; it has no analogue in classical (pathwise) calculus where $d[x]_t \equiv 0$ for smooth paths.

Example: $f(x) = x^2$ for Brownian motion

Take $X_t = W_t$, so $b \equiv 0$, $\sigma \equiv 1$, and $[W]_t = t$. With $f(x) = x^2$,

$$df(W_t) = 2W_t\, dW_t + \tfrac{1}{2} \cdot 2\, d[W]_t = 2W_t\, dW_t + dt.$$

Integrating $0 \to T$ and taking expectations,

$$\mathbb{E}[W_T^2] - \mathbb{E}[W_0^2] = \mathbb{E}\int_0^T 2W_t\, dW_t + \int_0^T dt = 0 + T,$$

so $\mathbb{E}[W_T^2] = T$. This computation *comes from* the quadratic variation term dt.

Quadratic covariation and multivariate Itô

For two continuous semimartingales X, Y,

$$[X,Y]_t \;=\; \lim_{|\Pi|\to 0} \sum \big(X_{t_{i+1}} - X_{t_i}\big)\big(Y_{t_{i+1}} - Y_{t_i}\big).$$

If $X_t = \int_0^t \sigma_s^X \, dW_s^X$ and $Y_t = \int_0^t \sigma_s^Y \, dW_s^Y$ with $\mathrm{Cov}(dW_t^X, dW_t^Y) = \rho_t \, dt$, then

$$[X, Y]_t = \int_0^t \rho_s \, \sigma_s^X \sigma_s^Y \, ds.$$

In d dimensions, Itô's lemma adds the full correction $\frac{1}{2} \sum_{i,j} f_{x_i x_j} \, d[X^i, X^j]_t$.

Why the correction appears

Classical Taylor expansions neglect squared differentials because smooth paths have zero quadratic variation. Brownian paths do not: $(\Delta W)^2$ sums to a finite limit (t). When you expand $f(t, X_t)$ and retain terms up to order dt, the $(dX_t)^2$ term *survives* and contributes $\frac{1}{2} f_{xx} \, d[X]_t$. This surviving second-order term is exactly the Itô correction and is the mathematical fingerprint of finite quadratic variation in stochastic calculus.

8 Stratonovich Integration

8.1 Definition via symmetric (midpoint) sums

The **Stratonovich** integral is obtained by replacing the left-endpoint sampling in Itô sums with *symmetric* (midpoint) sampling. For a partition $\Pi = \{0 = t_0 < \cdots < t_n = T\}$, define the Riemann-type sums

$$S_\Pi^\circ \;=\; \sum_{i=0}^{n-1} f\Big(\tfrac{t_i+t_{i+1}}{2},\, W_{\frac{t_i+t_{i+1}}{2}}\Big)\,\big(W_{t_{i+1}} - W_{t_i}\big).$$

If S_Π° converges in probability as $|\Pi| \to 0$, the limit is the *Stratonovich* integral $\displaystyle\int_0^T f \circ dW$.

8.2 Relation to the Itô integral (conversion formula)

For sufficiently smooth f (e.g. $f \in C^1$ with suitable growth), the Stratonovich and Itô integrals are related by the classical conversion:

$$\boxed{\int_0^T f(W_t) \circ dW_t \;=\; \int_0^T f(W_t)\, dW_t \;+\; \frac{1}{2}\int_0^T f'(W_t)\, dt.}$$

More generally, for an Itô process X_t with $[X]_t$ its quadratic variation,

$$\int_0^T f(X_t) \circ dX_t \;=\; \int_0^T f(X_t)\, dX_t \;+\; \frac{1}{2}\int_0^T f'(X_t)\, d[X]_t.$$

8.3 Where the $\frac{1}{2}$-correction comes from: quadratic variation

The midpoint sampling can be written heuristically as

$$f\Big(\tfrac{W_{t_i}+W_{t_{i+1}}}{2}\Big)\,\Delta W_i = f(W_{t_i})\Delta W_i \;+\; \tfrac{1}{2} f'(W_{t_i})\,(\Delta W_i)^2 \;+\; o\big((\Delta W_i)^2\big).$$

Summing over i and passing to the limit,

$$\sum_i f(W_{t_i})\Delta W_i \;\to\; \int_0^T f(W_t)\, dW_t \quad \text{(Itô)},$$

while

$$\sum_i \tfrac{1}{2} f'(W_{t_i})\,(\Delta W_i)^2 \;\to\; \tfrac{1}{2}\int_0^T f'(W_t)\, d[W]_t \;=\; \tfrac{1}{2}\int_0^T f'(W_t)\, dt,$$

because the quadratic variation of Brownian motion satisfies $[W]_t = t$. Thus midpoint (Stratonovich) sums differ from left-point (Itô) sums by exactly the quadratic-variation correction.

8.4 Chain rule comparison

A key feature of Stratonovich integration is that it satisfies the *classical* chain rule:

$$d\,\phi(W_t) \;=\; \phi'(W_t)\circ dW_t,$$

for smooth ϕ. In Itô form, the same differential carries the Itô correction:

$$d\,\phi(W_t) \;=\; \phi'(W_t)\,dW_t \;+\; \tfrac{1}{2}\phi''(W_t)\,dt.$$

The difference between these two rules is precisely the term generated by the nonzero quadratic variation of W.

Interpretation

Stratonovich integration uses symmetric sampling, which "builds in" half of the quadratic-variation effect at the Riemann-sum level. As a result, it obeys the ordinary chain rule and is invariant under smooth changes of variables—features that align with classical calculus and many physical models. Itô integration, by contrast, samples at the left endpoint, and the missing symmetric contribution reappears as the $\frac{1}{2}$-correction proportional to $d[W]_t$.

9 Summary Table: The Hierarchy of Integrals

Integral type	Measure element	Variation property	Convergence mode
Riemann	dx (distance)	Finite total variation	Pathwise
Riemann–Stieltjes	$dg(x)$	Finite total variation	Pathwise
Lebesgue–Stieltjes	μ_g (measure from g)	Finite total variation (possibly signed)	Measure-theoretic
Itô	dW_t (martingale increments)	Finite quadratic variation	Mean-square (L^2)

Why mean-square convergence makes the Itô integral well-defined

1. The space. The Itô integral is constructed in the Hilbert space $L^2(\Omega)$, where $\|X\|_{L^2}^2 = \mathbb{E}[X^2]$. This space is complete: every Cauchy sequence of random variables converges to a unique limit in L^2.
2. The convergence. Mean-square convergence, $\mathbb{E}[(S_\Pi - I)^2] \to 0$, is exactly convergence in this L^2 norm. Hence if the Riemann-type sums S_Π are Cauchy in L^2, a unique limit I exists.
3. The isometry. For simple adapted processes,

$$\mathbb{E}\left[\left(\sum_i \xi_i \Delta W_i\right)^2\right] = \sum_i \mathbb{E}[\xi_i^2]\Delta t_i,$$

proving an exact isometry $\|I(H)\|_{L^2}^2 = \mathbb{E}\int_0^T H_t^2\, dt$. This makes the mapping $H \mapsto I(H)$ linear, continuous, and norm-preserving.
4. The extension. Since simple adapted processes are dense in $\mathcal{H}^2 = \{f : \mathbb{E}\int f^2 < \infty\}$, the operator I extends uniquely to all of $\mathcal{H}^2$. This extension defines $\int_0^T f\, dW$ for any $f \in \mathcal{H}^2$.
Conclusion. Mean-square convergence is not a handwave—it leverages the completeness and geometry of $L^2(\Omega)$. Within this space, convergence is strong enough to guarantee both *existence* and *uniqueness* of the stochastic integral, giving a mathematically rigorous foundation to Itô calculus.

Intuitive analogy. Think of Riemann integration like working in $\mathbb{R}$: you need your sums to converge numerically (pointwise). In stochastic calculus, each Riemann sum is a *random variable*—so you need them to converge in a space where random variables can be compared meaningfully. That's $L^2(\Omega)$. And the miracle is that Itô's construction gives you an exact "Pythagorean" structure:

Variance of integral = Integral of variance density.

This identity pins everything down—no ambiguity, no guesswork.