Chapter 2

Price Process Characterization

In these notes we ask which stochastic processes are available to model a price, and find the answer far more restrictive than it looks. Brownian motion turns out to be the only continuous noise there is, which forces the diffusion term of the previous chapter's equation rather than merely permitting it — the licence to write the equation at all comes from chapter 4, and the two halves are worth keeping distinct. Dropping continuity admits jumps, and we work the Poisson process through in enough detail to see how a default is modelled and why that case is so much more tractable than the general one.

2.1 Why This Form and No Other

Everything from chapter 5 onwards models a price as an Itô diffusion. Before that becomes a habit, ask how much of a restriction that is. The equation has a specific shape — a drift proportional to dt, a random part proportional to dW, and nothing else. Why should any price be of that shape?

The answer is that most of the shape is forced, but by two different things, and they should be kept apart from the start.

That the random part is proportional to dW — that there is no other continuous noise to choose from — is what this chapter proves, and it needs continuity together with the martingale property. That a price may be written in this shape *at all*, with a drift accumulating at a rate alongside the noise, is a separate question and continuity does not answer it: it needs the price to be a semimartingale, which is a consequence of no-arbitrage rather than of smooth paths. Chapter 4 supplies that half and states the combined result.

So what follows pins the noise, and the drift is somebody else's problem. The theorem that does it is the following.

Write $\langle M\rangle_t$ for the quadratic variation defined in chapter 1, the limit of the sums of squared increments over refining partitions, which for Brownian motion is t.

Theorem 2.1 (Lévy's characterization)**.** *Let M be a continuous local martingale with $M_0 = 0$. Then M is a standard Brownian motion if and only if $\langle M\rangle_t = t$ for all t.*

Sketch. Fix $\lambda \in \mathbb{R}$ and consider the exponential

$$Z_t = e^{\lambda M_t - \frac{1}{2}\lambda^2 t}.$$

Applying Itô's lemma, and remembering that the second order term carries $d\langle M\rangle_t$ rather than dt,

$$dZ_t = Z_t\Big(\lambda\, dM_t - \tfrac{1}{2}\lambda^2\, dt + \tfrac{1}{2}\lambda^2\, d\langle M\rangle_t\Big).$$

Now impose the hypothesis $\langle M\rangle_t = t$. The last two terms cancel exactly, leaving

$$dZ_t = Z_t\lambda\, dM_t,$$

which has no drift, so Z is a martingale. Therefore $\mathbb{E}[Z_t \mid \mathcal{F}_s] = Z_s$, and rearranging,

$$\mathbb{E}\big[e^{\lambda(M_t - M_s)} \,\big|\, \mathcal{F}_s\big] = e^{\frac{1}{2}\lambda^2(t-s)}. \tag{2.1}$$

The right hand side is the moment generating function of a $N(0, t-s)$ variable, so the increment is Gaussian with variance $t-s$. And the right hand side is a *number*: it does not depend on $\mathcal{F}_s$ at all, so the increment is independent of everything that has happened. Gaussian, independent increments, variance $t-s$ — that is Brownian motion. □

Remark (Where the independence came from)**.** Look again at which step produced it. Equation (2.1) has a deterministic right hand side for exactly one reason: $d\langle M\rangle_t$ cancelled against dt. Without the hypothesis there is nothing to cancel against, and the exponential that stays a martingale is the one that carries its own quadratic variation in the exponent, $\tilde{Z}_t = e^{\lambda M_t - \frac{1}{2}\lambda^2\langle M\rangle_t}$, which gives

$$\mathbb{E}\Big[e^{\lambda(M_t-M_s) - \frac{1}{2}\lambda^2(\langle M\rangle_t - \langle M\rangle_s)} \,\Big|\, \mathcal{F}_s\Big] = 1.$$

This is all that survives, and the two factors cannot be separated: the increment and the variance it accumulates are in general dependent, and in chapter 10 that dependence is the leverage correlation which produces the skew. Only when $\langle M\rangle_t - \langle M\rangle_s$ is deterministic can the second factor be taken out of the expectation, and it is that step, and no other, which turns the identity into a statement about the law of the increment alone. Independence of increments is not a general feature of continuous martingales; it is what a deterministic quadratic variation buys.

One honest caveat on the sketch: using the real exponential requires an integrability argument to promote Z from a local martingale to a true one. The standard proof uses $e^{i\lambda M_t + \frac{1}{2}\lambda^2 t}$ instead, whose modulus is bounded so that the promotion is free, and reads off the characteristic function rather than the moment generating function. Nothing else changes.

What the theorem says is: Brownian motion is not one choice of noise among many. It is the *only* continuous local martingale whose quadratic variation accumulates at a constant rate, and any other candidate is either not continuous, not a martingale, or a rescaling of this one in disguise. The next proposition makes "in disguise" precise.

Theorem 2.2 (A continuous local martingale is an integral against Brownian motion)**.** *Let M be a continuous local martingale with $M_0 = 0$ whose quadratic variation is absolutely continuous,*

$$\langle M\rangle_t = \int_0^t \sigma_s^2\, ds,$$

with $\sigma_s > 0$. Then there is a Brownian motion W such that

$$M_t = \int_0^t \sigma_s\, dW_s.$$

Proof. Define W by integrating the increments of M scaled down by σ:

$$W_t = \int_0^t \frac{1}{\sigma_s}\, dM_s.$$

It is a stochastic integral against a local martingale, so it is itself a continuous local martingale starting from zero. Its quadratic variation is

$$\langle W\rangle_t = \int_0^t \frac{1}{\sigma_s^2}\, d\langle M\rangle_s = \int_0^t \frac{1}{\sigma_s^2}\,\sigma_s^2\, ds = t,$$

using that the quadratic variation of $\int H\, dM$ is $\int H^2\, d\langle M\rangle$. So W satisfies the hypotheses of Theorem 2.1 and is a Brownian motion. And integrating it back,

$$\int_0^t \sigma_s\, dW_s = \int_0^t \sigma_s \frac{1}{\sigma_s}\, dM_s = M_t. \qquad \square$$

Remark. The trick recurs in chapter 6. A process whose quadratic variation grows at a varying rate is a Brownian motion running on a distorted clock, and dividing by σ resets the clock. Lévy's theorem is what certifies that the reset produces a genuine Brownian motion rather than merely something with the right variance.

Two conditions in the statement are doing work. If σ vanishes on a set of times of positive measure, M is simply not moving there and the construction needs an independent Brownian motion added on that set to fill the gap; the conclusion survives, on a slightly enlarged probability space. And if $\langle M\rangle$ is not absolutely continuous — if the process accumulates variance in bursts concentrated on a set of measure zero — then M is still a time-changed Brownian motion, by the Dambis-Dubins-Schwarz theorem, but not an integral against one. Such processes are unusual in practice and this is the only place they will be mentioned.

The general case is proved next because the argument is the same idea as the one above with the integral removed, and because the picture it produces — the clock of chapter 1 — is used repeatedly later.

Theorem 2.3 (Dambis, Dubins and Schwarz)**.** *Let M be a continuous local martingale with $M_0 = 0$ and $\langle M\rangle_\infty = \infty$. Set*

$$\tau_u = \inf\{t : \langle M\rangle_t > u\}.$$

Then $B_u = M_{\tau_u}$ is a Brownian motion with respect to the filtration $\mathcal{G}_u = \mathcal{F}_{\tau_u}$, and

$$M_t = B_{\langle M\rangle_t}.$$

Proof. One lemma first. *If $\langle M\rangle$ is constant on an interval then so is M.* For if $\langle M\rangle_b = \langle M\rangle_a$ with $a < b$, then the local martingale $t \mapsto M_{a+t} - M_a$ has vanishing quadratic variation on $[0, b-a]$, and a continuous local martingale starting from zero with vanishing bracket is identically zero — the supermartingale argument of chapter 1. So M does not move where its clock does not run, which is what one would want: the bracket measures the accumulated randomness, and no accumulation means nothing happened.

Now the properties of τ. Because $\langle M\rangle$ is continuous, non-decreasing and increases to infinity, τ_u is finite for every u, non-decreasing, right-continuous, and satisfies

$$\langle M\rangle_{\tau_u} = u.$$

It fails to be continuous exactly where $\langle M \rangle$ is flat: an interval on which the clock does not advance is crossed by τ in a single jump. By the lemma M is constant across such an interval, so $B_u = M_{\tau_u}$ does not jump there, and B is continuous. This is the step the theorem's proof exists to handle, and it is why the hypothesis is about $\langle M \rangle_\infty$ rather than about M.

B is adapted to $\mathcal{G}$ by construction. It is a local martingale because optional stopping says so: for $v \geq u$, applying it to M between the stopping times τ_u and τ_v gives $\mathbb{E}[M_{\tau_v} \mid \mathcal{F}_{\tau_u}] = M_{\tau_u}$, which is $\mathbb{E}[B_v \mid \mathcal{G}_u] = B_u$. The integrability that optional stopping wants is supplied by localising, exactly as elsewhere: stop at $\tau_u \wedge n$ and let n grow.

Its bracket is the time change of the bracket, $\langle B \rangle_u = \langle M \rangle_{\tau_u} = u$, since the sums of squared increments defining $\langle B \rangle$ over a partition of $[0, u]$ are the sums defining $\langle M \rangle$ over the image partition of $[0, \tau_u]$ — the same increments of the same path, relabelled. So $\langle B \rangle_u = u$, and Theorem 2.1 applies: B is a Brownian motion.

Finally the identity. Fix t and put $u = \langle M \rangle_t$. Then $\tau_u \geq t$, with strict inequality only when $\langle M \rangle$ is flat on $[t, \tau_u]$ — and by the lemma M is then constant there, so $M_{\tau_u} = M_t$ either way. That is $B_{\langle M \rangle_t} = M_t$, which is the square. □

Structure (Quadratic variation is a clock)**.** Read Theorem 2.3 as a picture rather than as a construction, because the picture is what gets used.

Every continuous local martingale is a Brownian motion watched on a clock of its own making, and $\langle M \rangle_t$ is the reading of that clock at calendar time t. Volatility is the speed of the clock: high volatility is not a different noise but the same noise consumed faster. That is why chapter 1 could define the bracket and not say what it meant — its meaning is this theorem.

Several later statements are this one in disguise. Lévy's characterization is the case where the clock runs at calendar speed, and Theorem 2.2 the case where it has a rate. A stochastic volatility model (chapter 10) is a model of the clock, with v_t the rate at which it turns. A skew is what appears when the clock is correlated with the motion: if time speeds up on the way down, the terminal distribution cannot be symmetric. And chapter 4's doubling strategy is a clock deliberately built to reach infinity before the calendar reaches one.

Remark (Why this is not circular)**.** The proof used Lévy's characterization to conclude that B is a Brownian motion, and Lévy's characterization is a statement about a process whose clock runs at calendar speed. There is no circle: the time change was performed precisely to manufacture that hypothesis. M does not satisfy it and B does, by construction, because τ was defined as the inverse of M's own clock. It is the same manoeuvre as dividing by σ in Theorem 2.2, carried out when there is no σ to divide by.

Structure (The theorems above are a classification)**.** Read together, Lévy's characterization and the representation theorem do something stronger than supply two facts. They classify.

Up to a time change, there is only one continuous local martingale. Dambis-Dubins-Schwarz says every one of them is $B_{\langle M \rangle_t}$ for a Brownian motion B; Lévy says B is unique in law; and the representation theorem is the special case where the clock is absolutely continuous, so the time change can be written as an integral.

A classification beats a list because it tells you when to stop looking. There is no exotic continuous martingale waiting to be discovered and used to model a price differently — the class is exhausted by one object and a choice of clock. All the modelling freedom that remains lives in the clock, which is to say in the volatility, and that is why every chapter after this one is about volatility rather than about finding a better noise.

Remark (Where the deterministic quadratic variation went)**.** Lévy's theorem required $\langle M \rangle_t = t$, which is about as deterministic as a quadratic variation gets. A stochastic volatility model has $\langle M \rangle_t = \int_0^t \sigma_s^2\, ds$ with σ random, which is not deterministic at all. How is the theory available to such a model?

The answer is that Lévy's theorem was never applied to the price. Read the proof again and watch which process it is applied to. The hypothesis of Theorem 2.2 allows σ_s to be any adapted process — as random as one likes, driven by its own noise, depending on M's whole history. Nothing deterministic is assumed. What the proof then does is compute

$$\langle W \rangle_t = \int_0^t \frac{1}{\sigma_s^2}\, d\langle M \rangle_s = \int_0^t \frac{1}{\sigma_s^2}\, \sigma_s^2\, ds = t,$$

and the randomness cancels identically. Not approximately, and not on average: the σ_s^2 that arrived from $d\langle M \rangle_s$ is the same σ_s^2 that was divided by, whatever it happened to be on that path.

So the deterministic quadratic variation Lévy needs is *manufactured* rather than assumed. Dividing by σ is precisely the operation that removes the randomness from the variation, and W is the process it is removed from. The price keeps its random variation and its dependent increments; W takes the deterministic variation and the independence. The two theorems apply to two different processes, and confusing which is which is what makes the objection seem to bite.

In the language of chapter 1 the price is B read on a random clock, and reading a Brownian motion at random times is what destroys the independence: knowing the past tells you how fast the clock is currently running, and therefore how large the next increment is likely to be. Lévy's hypothesis is the statement that the clock cannot surprise you.

Remark (Checking which process is which)**.** Both claims are measurable, on the same simulated paths, in `variation_spread`. Take the realised variance over $[0, 1]$ and record how much it varies from path to path, as the sampling grid is refined:

	400 **steps**	1600 **steps**
Price, constant σ	0.071	0.035
Price, 100% vol-of-vol	2.57	2.31
Recovered W, either case	0.071	0.035

The figures are standard deviations across paths as a fraction of the mean. Read the rows against each other. With constant volatility the price's own variation concentrates as the grid is refined, halving as the step count quadruples — it is converging to a constant. With random volatility it does not: concentration would have taken 2.57 to 1.29 and it went to 2.31, which is the estimate of a heavy-tailed quantity wobbling rather than a distribution collapsing. Its limit is the random variable $\int_0^1 \sigma_s^2\, ds$, and refining the grid measures that random number more precisely rather than making it less random.

And the recovered W concentrates in both cases, at exactly the rate $\sqrt{2/n}$ that a chi-square average must. Its variation is converging to the constant 1 however wildly the volatility behaved, which is Lévy's hypothesis being satisfied by construction. That is the whole content of the previous two remarks, in numbers.

So the shape of the Itô diffusion is not chosen. Any continuous local martingale, given only that it accumulates variance at some rate, *must* be $\int \sigma\, dW$; and adding a finite-variation drift then gives the general form. Chapter 4 completes the argument by explaining why a price has to be a local martingale in the first place.

Remark (The representation does not hand down independent increments)**.** Theorem 2.2 is easy to over-read. It says every continuous local martingale is an integral against a Brownian motion, and Brownian motion has independent increments, so it is tempting to conclude that M inherits them. It does not, and the reason is that σ_s is an *adapted* process — it is allowed to depend on everything that has happened up to time s, including M itself.

The independence lives in W, not in M. What the theorem provides is a Brownian motion out of which M can be rebuilt; all of M's memory has been pushed into the integrand. A local volatility model $\sigma_s = \sigma(M_s, s)$ is the plainest case: its increments obviously depend on the past, since they depend on where the process currently is, and it is nonetheless exactly of the form $\int \sigma \, dW$.

The previous remark already gives the sharp version. For a continuous local martingale,

$$M \text{ has independent (and Gaussian) increments} \iff \langle M \rangle_t \text{ is deterministic}$$

which for $M = \int \sigma \, dW$ means σ deterministic. Anything else — any model in which volatility is itself random — has dependent increments by construction. That is the whole of chapters 10 to 12.

Remark (How restrictive independent increments would be)**.** Very little would survive the requirement. A continuous process with stationary independent increments must be $\mu t + \sigma W_t$ for constants μ and σ — Brownian motion with drift, and nothing else. Relaxing stationarity but keeping independence buys only deterministic $\mu(t)$ and $\sigma(t)$.

Remark (Checking it)**.** The claim is concrete enough to measure. Take $M = \int \sigma \, dW$ on $[0, 1]$, split the interval in half, and correlate the realised variance of the first half against the second across many paths. Increments independent of the past would force this to be zero, since the two windows are disjoint. Running it in `variance_clustering`:

Volatility of volatility	**Correlation**
0 (constant σ)	−0.00
50%	+0.67
100%	+0.54

Zero when the volatility is constant, as it must be, and emphatically not zero otherwise — and M is a perfectly good continuous local martingale in every row, since the volatility is driven by its own independent Brownian motion. Under its market name the second row is volatility clustering, which is among the least controversial facts about returns.

The non-monotonicity in the table is real and is not the dependence weakening. It is the linear correlation losing its ability to see the dependence, because the realised variances become more extremely lognormal — the same effect chapter 18 derives as a hard bound on how correlated two lognormals can be. It is an early warning that a correlation is a poor summary of a dependence.

2.2 Dropping Continuity

Every statement in the previous section carried the word *continuous*, and it is fair to ask how much it was doing. Prices in fact jump: a bond that defaults, a stock through an earnings announcement, a currency whose peg breaks. If continuity were a technical convenience the theory would extend and nobody would mind. It is not, and the cheapest way to see that is a counterexample.

Definition 2.4 (Poisson process)**.** A Poisson process N_t with intensity λ starts at zero, has independent increments, and $N_t - N_s$ is Poisson distributed with mean $\lambda(t-s)$. Its paths are constant except for jumps of size 1, arriving at rate λ.

Writing the Poisson law into the definition makes it look like a choice among many, and it is not. The distribution is forced by the same two hypotheses that force everything else in this chapter.

Theorem 2.5 (Counting with stationary independent increments leaves no choice)**.** *Let N be a counting process — starting at zero, non-decreasing, moving only by integer jumps — with stationary independent increments, and suppose two jumps do not arrive at once, in the sense that*

$$\mathbb{P}(N_h \geq 2) = o(h) \qquad \text{as } h \downarrow 0. \tag{2.2}$$

Then N is a Poisson process: there is a $\lambda \geq 0$ with N_t Poisson distributed of mean λt.

Proof. Everything comes from the probability of *nothing happening*. Put $p(t) = \mathbb{P}(N_t = 0)$. The event that nothing happens on $[0, t+s]$ is the intersection of nothing happening on $[0,t]$ and nothing happening on $(t, t+s]$, and those are independent by hypothesis and identically distributed to $[0,s]$ by stationarity. So

$$p(t+s) = p(t)\,p(s),$$

which is Cauchy's exponential equation. A monotone solution — and p is non-increasing, since nothing happening on a longer interval is a smaller event — must be $p(t) = e^{-\lambda t}$ for some $\lambda \in [0,\infty]$. This is the same functional equation, and the same conclusion, that chapter 12 meets when asking which volatility structures admit a finite-dimensional state; there it forces an exponential in maturity, here in time.

That fixes the law of the first arrival: $\mathbb{P}(\tau_1 > t) = e^{-\lambda t}$, so the waiting time is exponential, and by stationarity and independence each subsequent wait is an independent copy. A sum of n independent exponentials is Gamma distributed, and $\{N_t \geq n\} = \{\tau_1 + \cdots + \tau_n \leq t\}$, so

$$\mathbb{P}(N_t \geq n) = \int_0^t \frac{\lambda^n u^{n-1} e^{-\lambda u}}{(n-1)!}\, du,$$

and differencing consecutive values of n gives $\mathbb{P}(N_t = n) = e^{-\lambda t}(\lambda t)^n / n!$.

Condition (2.2) is what rules out the alternatives. Without it the jumps need not have size one — a process that jumps by two at Poisson times has stationary independent increments and is not a Poisson process — so (2.2) is precisely the statement that the counting is of single events. The degenerate endpoints are $\lambda = 0$, where nothing ever happens, and $\lambda = \infty$, where $p(t) = 0$ for every t and infinitely many arrivals occur immediately, which the right-continuity of a counting path excludes. $\square$

Remark (What the theorem is for)**.** Read alongside Lévy's characterization it completes a pair. There, continuity plus a constant rate of accumulated variance forced Brownian motion and left no freedom. Here, counting plus stationary independent increments forces the Poisson process and leaves one number. So each of the two building blocks of (2.10) below is not a modelling choice but the unique object satisfying its description, which is why a general process with stationary independent increments can be decomposed into them and nothing else.

It is also the step the Lévy-Itô proof leans on hardest, and now it need not be waved at: applying the theorem to the events "a jump of size at least ϵ occurred" gives that those arrivals are Poisson, with a rate that must be finite because a right-continuous path cannot have infinitely many jumps of a fixed size in a bounded interval.

It is not a martingale — it only ever goes up — but subtracting its mean makes one.

Calculation 2.6 (The compensator)**.** Look for a deterministic function $a(t)$ making $N_t - a(t)$ a martingale. Splitting the increment and using independence,

$$\begin{aligned}
\mathbb{E}\big[N_t - a(t) \,\big|\, \mathcal{F}_s\big] &= \mathbb{E}\big[N_s + (N_t - N_s) \,\big|\, \mathcal{F}_s\big] - a(t) \\
&= N_s + \mathbb{E}\big[N_t - N_s\big] - a(t) && \text{(independent increments)} \\
&= N_s + \lambda(t-s) - a(t).
\end{aligned}$$

For this to equal $N_s - a(s)$ we need $a(t) - a(s) = \lambda(t-s)$ for all $s < t$, so $a(t) = \lambda t$ up to a constant.

Definition 2.7 (Compensated Poisson process)**.** $M_t = N_t - \lambda t$, which by the calculation above is a martingale.

The compensator λt is doing the same job as the drift correction elsewhere: it removes the part of the motion that was predictable and leaves only the surprise. Note how little was used — only that the increments are independent with mean $\lambda(t-s)$ — which is why the same subtraction works for any process with independent increments.

Now its two quadratic variations, which is where the jump case starts to diverge from everything in the previous chapter.

Calculation 2.8 (Realised quadratic variation)**.** Take any partition and refine it. Between jumps the path is flat, so those increments contribute nothing. Once the partition is fine enough that no two jumps share an interval, each interval containing a jump contributes $(\Delta M)^2 = (\pm 1)^2 = 1$ — the $-\lambda\, dt$ part of the increment contributes at order dt^2 and vanishes in the limit. Counting,

$$[M]_t = \sum_{s \le t} (\Delta M_s)^2 = N_t.$$

So the realised quadratic variation is the number of jumps: an integer, a staircase, and *random*.

Calculation 2.9 (Predictable quadratic variation)**.** Now find the deterministic $b(t)$ making $M_t^2 - b(t)$ a martingale. Split the increment again, using that M is a martingale so the cross term drops:

$$\begin{aligned}
\mathbb{E}\big[M_t^2 \,\big|\, \mathcal{F}_s\big] &= \mathbb{E}\big[\big(M_s + (M_t - M_s)\big)^2 \,\big|\, \mathcal{F}_s\big] \\
&= M_s^2 + 2M_s \underbrace{\mathbb{E}\big[M_t - M_s \,\big|\, \mathcal{F}_s\big]}_{=\,0} + \mathbb{E}\big[(M_t - M_s)^2 \,\big|\, \mathcal{F}_s\big] \\
&= M_s^2 + \operatorname{Var}\big(N_t - N_s\big) && \text{(the increment has mean zero)} \\
&= M_s^2 + \lambda(t-s),
\end{aligned}$$

using that a Poisson variable has variance equal to its mean. So $b(t) = \lambda t$ works, and

$$\langle M \rangle_t = \lambda t.$$

Set $\lambda = 1$ and compare the two answers: $\langle M\rangle_t = t$, exactly Brownian motion's, while $[M]_t = N_t$, which is not t and is not even deterministic. That is the counterexample below, already assembled.

Example 2.1 (Why Lévy's theorem needs continuity)**.** Take $\lambda = 1$, so $M_t = N_t - t$. Then:

- M is a martingale, hence a local martingale, and $M_0 = 0$;
- $\mathrm{Var}(N_t) = \lambda t = t$, so M accumulates variance at exactly the rate Brownian motion does, $\langle M\rangle_t = t$;
- M is emphatically not a Brownian motion. Its paths are flat except at a countable set of times where they jump by one, and its increments are Poisson rather than normal.

Every hypothesis of Lévy's characterization holds except continuity, and the conclusion fails completely. So continuity is not a regularity condition attached to make a proof go through; it is the entire content of the theorem.

Remark (Two quadratic variations, which agree only in the continuous case)**.** The example also exposes a distinction the continuous theory lets us ignore. There are two natural notions of accumulated variance:

- the *realised* quadratic variation $[M]_t$, the limit of $\sum(\Delta M)^2$ along a refining partition, which is what chapter 1's figure measured; and
- the *predictable* quadratic variation $\langle M\rangle_t$, the compensator — the process that must be subtracted from M^2 to leave a martingale.

For a continuous local martingale the two coincide, which is why we have been writing one symbol without comment — and the coincidence is not an accident of notation but something chapter 1 already proved without saying so.

Its theorem on the bracket of a continuous square integrable martingale did three things: it showed the sums of squares converge, which constructs $[M]$; it showed $M^2 - [M]$ is a martingale; and it showed $[M]$ is the *only* continuous increasing adapted process starting at zero with that property. The last clause is the one that matters here, because the predictable bracket is by definition the unique *predictable* increasing process compensating M^2 — and a continuous adapted process is automatically predictable, its value at any time being the limit of its values strictly before. So $[M]$ is a candidate for the compensator, uniqueness makes it the compensator, and $[M] = \langle M\rangle$. Localisation carries it from square integrable martingales to local ones.

Continuity is doing the whole of the work in that argument, and the Poisson case shows exactly where it fails. $[M]_t = N_t$, since every jump contributes $(\Delta M)^2 = 1$, and N is adapted and increasing — but it is not predictable: its jumps arrive at the totally inaccessible times of §2.3, so no sequence announcing them exists and its value cannot be known an instant beforehand. So $[M]$ is disqualified as a compensator, and the compensator is something else: $\langle M\rangle_t = t$. The realised variance is itself random, and equal to the number of jumps.

That is a genuine difference in kind. In the continuous world, variance accumulates at a rate you can know in advance. In the jump world it arrives in lumps, and how much arrived is not known until it has.

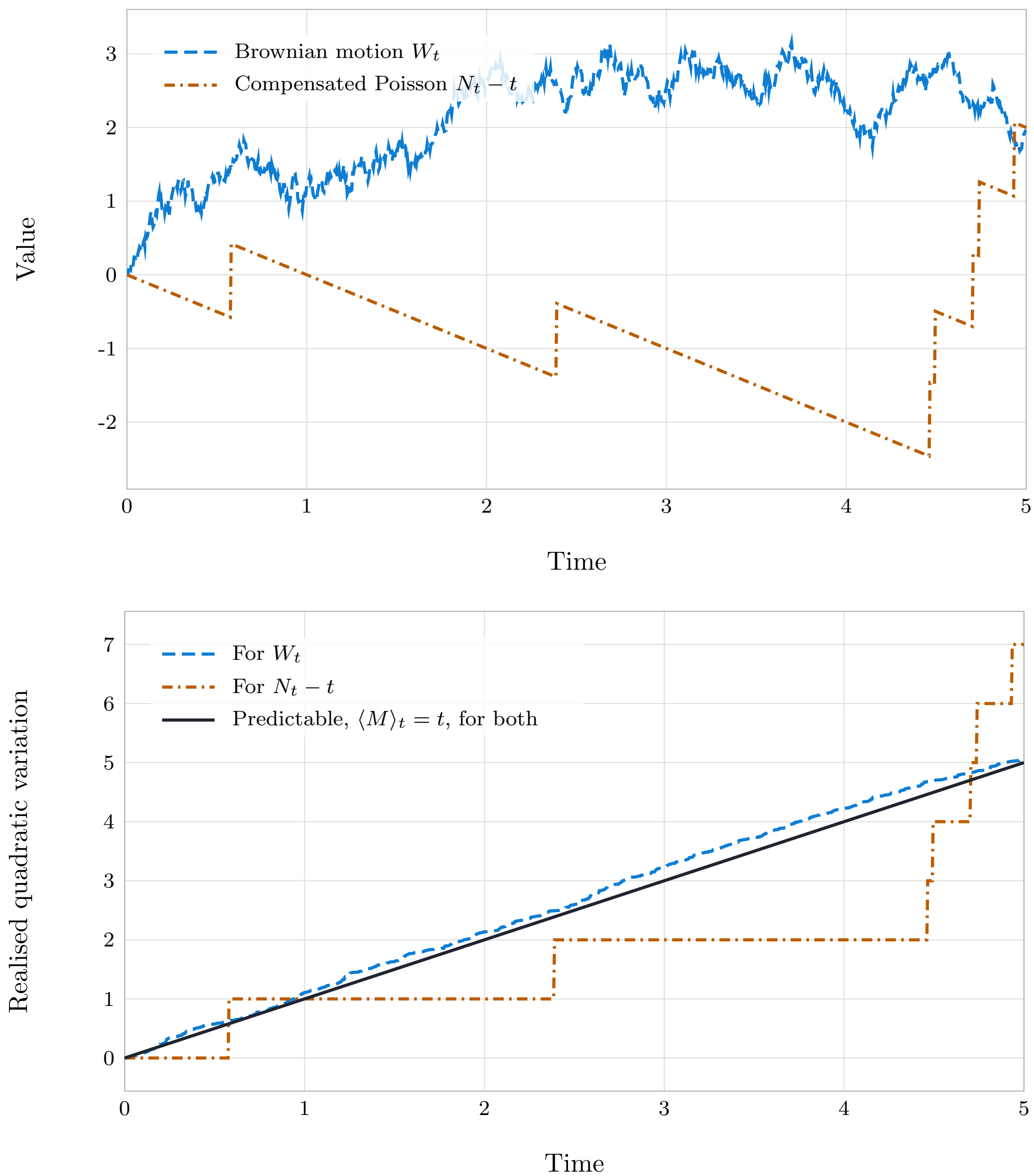

Figure 2.1: Above: a Brownian motion and a compensated Poisson process of intensity one, over five years. Both are martingales, both start at zero, and both accumulate variance at exactly rate one — so both satisfy every hypothesis of Lévy's characterization except continuity, and only one of them is a Brownian motion. Below: their realised quadratic variations, against the straight line that is the predictable variation of both. For the Brownian motion the realised and the predictable agree. For the jump process the realised variation is a staircase that counts the jumps, and where it happens to sit relative to the line is not known in advance.

Drawn by `poisson_arrivals` and `compensated_poisson`.

Once continuity is abandoned the classification changes shape rather than breaking down. The general statement is that a process with stationary independent increments is a drift, plus a Brownian motion, plus jumps, and nothing else:

Theorem 2.10 (Lévy-Itô decomposition)**.** *Any Lévy process X can be written as*

$$X_t = b\,t + \sigma W_t + \underbrace{\int_{|x|<1} x\,\tilde{N}(t,dx)}_{\textit{small jumps, compensated}} + \underbrace{\int_{|x|\geq 1} x\,N(t,dx)}_{\textit{large jumps}},$$

where N counts the jumps of each size and $\tilde{N}$ is that count less its mean.

Structure of the proof. The pieces are not found by inspiration; each is peeled off by a step that is forced.

The large jumps. Fix $\epsilon > 0$ and count the jumps of size at least ϵ. Stationary independent increments make the arrival times of these a Poisson process, by Theorem 2.5 applied to the events "a jump of size at least ϵ occurred", and its intensity $\nu(\{|x| \geq \epsilon\})$ must be finite, because a path with infinitely many jumps of size ϵ in a bounded interval is not right-continuous. So for each ϵ the large jumps form a compound Poisson process, which can be subtracted. What is left is a Lévy process with jumps smaller than ϵ.

The small jumps. Now let $\epsilon \downarrow 0$. The measure ν need not be finite near the origin — infinitely many tiny jumps per unit time is allowed — so the sum of them may diverge. What does not diverge is the sum with its mean removed. The compensated sums over successive shells $\epsilon_{k+1} \leq |x| < \epsilon_k$ are martingales, and they are orthogonal, because disjoint jump sizes are independent. So their partial sums are a martingale whose variance is the sum of the variances, $\int_{|x|<1} x^2\,\nu(dx)$, and that integral is finite for every Lévy measure. An orthogonal series with summable variances converges in L^2; the limit is the compensated integral, and the convergence is the whole content of the small-jump term.

What remains. Subtracting both jump parts leaves a Lévy process with continuous paths. It has stationary independent increments and no jumps, so its increments are Gaussian — this is the same rigidity that Lévy's characterization exploits — and a continuous Lévy process is therefore $bt + \sigma W_t$.

Two of these steps are the ones to look at closely if the argument is to be reconstructed rather than read. That the jump counts are Poisson is Theorem 2.5, which uses stationarity and independence and nothing else, and is where the structural hypothesis is spent. That the compensated small-jump sum converges is an L^2 orthogonality argument of exactly the kind used to build the Itô integral in chapter 1 — a series of orthogonal martingale increments with summable variances — which is why the finiteness of $\int_{|x|<1} x^2\nu(dx)$ appears in the definition of a Lévy measure rather than being derived later. The remaining work, and it is real work, is in the interchanges of limits that these two steps require, and in showing the three pieces are independent of each other. □

Remark (What the small jump term is)**.** It is a *purely discontinuous martingale* with infinitely many jumps. If ν is an infinite measure near the origin — which is allowed, and is the interesting case — then in every interval, however short, the path jumps infinitely often. There is no first jump and no way to list them.

That has a consequence for what can be written down. "The sum of the jumps" is not a sum one can take, because $\sum |\Delta X_s|$ diverges whenever $\int_{|x|<1} |x|\,\nu(dx) = \infty$. Only the compensated sum converges, and that is why the term appears in the theorem with its mean subtracted rather than as a plain integral against N. The compensation is not tidiness; without it there is nothing to converge to.

So there are two regimes, separated by a single integral:

- *Finite variation*, when $\int_{|x|<1} |x|\,\nu(dx) < \infty$. The jumps can be added up directly, the compensator is a finite drift that may be absorbed into b, and the process is a difference of two increasing pure-jump processes.

- *Infinite variation*, when that integral diverges while $\int_{|x|<1} x^2\nu(dx)$ — which is finite for every Lévy measure by definition — does not. Here the paths wander an infinite distance in any interval, and only the centred sum exists.

Structure (Finite quadratic variation, infinite length)**.** The second regime deserves a closer look: it is Brownian motion's signature reproduced with no continuous part at all.

A Brownian path has infinite length and finite quadratic variation: chapter 1 built the whole of Itô calculus on the second half of that sentence. The infinite-variation small jump process has exactly the same pair of properties. Its quadratic variation is

$$[X]_t = \sum_{s\le t} (\Delta X_s)^2, \qquad \mathbb{E}[X]_t = t\int_{|x|<1} x^2\,\nu(dx) < \infty,$$

finite because squares of small numbers are summable where the numbers themselves are not — and its first variation is infinite for the same reason in reverse.

So "rough enough to need Itô calculus" does not require continuity. What it requires is that the squares converge and the absolute values do not, and a process can arrange that with jumps just as well as a Brownian motion arranges it without them. The α-stable processes make the boundary explicit: with $\nu(dx) \propto |x|^{-1-\alpha}dx$ the small-jump integral $\int_0^1 x^{1-\alpha}dx$ converges exactly when $\alpha < 1$, so those are the finite variation cases and $\alpha \in [1, 2)$ are not. As $\alpha \to 2$ the jump activity concentrates ever more tightly near zero and the process approaches Brownian motion, which is the $\alpha = 2$ member of the family and the only one with no jumps at all.

Two consequences follow, and the second is the one that costs money. The Brownian part and the jump part do not interact: a purely discontinuous martingale has zero covariation with a continuous one, which is why the decomposition splits into pieces that can be treated separately rather than merely into terms that happen to add up. And a process made of small jumps can be made to look very like a diffusion — same finite quadratic variation, same infinite variation, paths that are visually indistinguishable at any resolution a data set offers. Whether a price is a diffusion or an infinite-activity jump process is therefore not settled by looking at it, and chapter 20 is where that becomes an identification problem rather than a philosophical one.

Remark (The cutoff at one is arbitrary, and b is not intrinsic)**.** One piece of bookkeeping, because it explains an otherwise puzzling feature of how Lévy processes are quoted.

Nothing distinguishes $|x| = 1$ as the place to split. Moving the cutoff moves jumps between the compensated term and the uncompensated one, and since the two are centred differently, the drift b changes to absorb the difference. So b is not a property of the process; it is a property of the process *and* the truncation chosen to state it. The diffusion coefficient σ^2 and the Lévy measure ν are intrinsic, and the triplet (b, σ^2, ν) is only meaningful once the truncation is named — which is why the literature carries a truncation function through statements that would otherwise be cleaner without it.

In the finite variation case the awkwardness disappears, since the jumps need no compensation at all and there is a genuine drift to quote.

The two jump terms are split for a reason. Large jumps are rare, so there are finitely many in any interval and they can simply be added up. Small jumps may arrive infinitely often, and the sum of them need not converge — but the sum of them *minus their mean* does, which is what compensating achieves. It is the same manoeuvre that turned N into a martingale, applied to each jump size at once.

Read against the previous section, the decomposition says the earlier result was the continuous corner of a larger classification. Brownian motion is still the only continuous piece available. What dropping continuity buys is the jump terms, and it buys nothing else.

Remark (Itô's lemma acquires a term that is not a derivative)**.** The chain rule changes in an instructive way. For a process with both parts,

$$f(X_t) = f(X_0) + \int_0^t f'(X_{s-})\,dX_s + \frac{1}{2}\int_0^t f''(X_{s-})\,d\langle X^c\rangle_s + \sum_{s\le t}\Big[f(X_s) - f(X_{s-}) - f'(X_{s-})\Delta X_s\Big],$$

where X^c is the continuous part and X_{s-} the value just before any jump.

The first three terms are the Itô formula already derived. The sum is new, and the important thing about it is that it is not a Taylor expansion. Across a jump the process does not move a little, it moves a lot, so no expansion in ΔX is legitimate; the bracket is the *exact* change $f(X_s)-f(X_{s-})$, less the amount the first-order term already counted for. For a continuous process every bracket is zero and the formula collapses to the one we have.

This is why the second-order term in Itô's lemma is a statement about continuous paths specifically. It came from $(\Delta W)^2$ surviving the limit while higher powers vanished, and that balance is a property of a process whose increments are of order $\sqrt{dt}$. A jump has no such scaling and has to be handled whole.

Remark (A third direction)**.** Jumps are one way out of the continuous world. There is another that gets asked about often, and the answer is not the obvious one.

Fractional Brownian motion B^H is the natural generalisation of Brownian motion in which increments are no longer independent: a Hurst parameter $H > 1/2$ gives a process that tends to continue in the direction it was going, and $H < 1/2$ one that tends to reverse. It is continuous. So one might expect it to slot into the classification above as a continuous alternative to W.

It does not, and the quadratic variation says why immediately.

Calculation 2.11 (The quadratic variation of fractional Brownian motion)**.** The defining property is that increments have variance

$$\mathbb{E}\big[(B^H_t - B^H_s)^2\big] = |t-s|^{2H},$$

which at $H = 1/2$ is the $|t-s|$ of ordinary Brownian motion. Chop $[0,t]$ into n equal intervals of width t/n and take expectations of the sum of squared increments:

$$\mathbb{E}\Big[\sum_{i=0}^{n-1}\big(\Delta_i B^H\big)^2\Big] = n\cdot\Big(\frac{t}{n}\Big)^{2H} = t^{2H}\,n^{1-2H}.$$

Everything now turns on the sign of the exponent $1-2H$ as $n \to \infty$:

$$t^{2H} n^{1-2H} \longrightarrow \begin{cases} 0 & \text{if } H > 1/2, \\ t & \text{if } H = 1/2, \\ \infty & \text{if } H < 1/2. \end{cases}$$

Read the middle line: the two powers of n cancel exactly, and only at $H = 1/2$. The whole of Itô calculus lives on that cancellation. A rougher process ($H < 1/2$) accumulates squared increments faster than the partition refines them away, and a smoother one ($H > 1/2$) more slowly; in neither case is there a finite non-zero limit for a second-order term to be built from.

Neither of the other cases can be a semimartingale. A semimartingale has finite quadratic variation, which rules out $H < 1/2$ outright; and a continuous process with *zero* quadratic variation has no martingale part at all, so it would have to be of finite variation, which fractional Brownian motion is not. So for every $H \neq 1/2$ it fails to be a semimartingale.

By the theorem quoted in chapter 4, that is fatal for a *price*: a non-semimartingale price admits a free lunch with vanishing risk, and for fractional Brownian motion the arbitrage can be written down explicitly. It is instructive that these arbitrages have the same character as the doubling strategy — they require trading continuously and arbitrarily fast, and they disappear once a minimum delay between trades, or any transaction cost, is imposed.

None of which rules out roughness in finance, because prices are not the only processes in a model. The volatility σ_t is not a traded asset, and nothing in this chapter's representation theorem asked it to be a semimartingale — only to be adapted, and to make $\int \sigma^2 ds$ finite. So one may take σ to be as rough as the data demands while the price it drives remains an ordinary continuous semimartingale, and every result here still applies. That is exactly what rough volatility models do, and chapter 10 takes up why anyone would want to.

2.3 Two Kinds of Jump

Before leaving the subject, two things the word "jump" runs together need separating: they are different mathematical objects, modelled by different means, and confusing them is a live source of error.

A jump happens at a random time τ, and stopping times come in two kinds.

Definition 2.12 (Predictable and totally inaccessible times)**.** A stopping time τ is *predictable* if there is a sequence $\tau_n \uparrow \tau$ with $\tau_n < \tau$ — an announcing sequence, so the time can be seen coming. It is *totally inaccessible* if it coincides with no predictable time: $\mathbb{P}(\tau = \sigma) = 0$ for every predictable σ.

The distinction is exactly the one between a scheduled event and a surprise, and both occur in markets.

A rate moves when the central bank meets, and the meeting dates are published a year ahead. A stock gaps on its earnings date. These are jumps at times that are not merely predictable but *deterministic*: what is random is the size, not the timing.

A company defaults. Nobody knows when, and no announcing sequence exists — the whole character of the event is that it is not visible in advance. This is a totally inaccessible time.

Remark (Only one of the two has an intensity)**.** The consequence that matters is what the compensator looks like. For a totally inaccessible time the compensator is continuous, and can be written as $\int_0^t \lambda_s\, ds$ for an *intensity* λ — a probability of the event per unit time, exactly as for the Poisson process above. For a jump at a scheduled date the compensator has an atom at that date, and there is no intensity: a rate of occurrence per unit time is meaningless for an event that happens at a known moment.

So the two are modelled by entirely different means. A default is given an intensity. A meeting date is given a distribution for the size of the move on that date, and no intensity at all. Practitioners do the second by putting a discrete lump of variance on the event date rather than spreading variance evenly through calendar time, which is why an implied volatility term structure has visible kinks around known event dates and why interpolating a surface straight through them is wrong.

The inaccessible case deserves more attention: it turns out to be far more tractable than the general theory of jumps would suggest, and it is how the entire credit market is modelled.

Definition 2.13 (Doubly stochastic, or Cox, default)**.** Let $\lambda_t \geq 0$ be an adapted intensity. Conditional on the whole path of λ, let N be an inhomogeneous Poisson process with that intensity, and let the default time τ be its first jump.

Calculation 2.14 (Survival is a discount factor)**.** Conditional on the path of λ, the first jump of an inhomogeneous Poisson process arrives after T with probability $\exp(-\int_t^T \lambda_s ds)$. Averaging over the paths,

$$\mathbb{P}\big(\tau > T \,\big|\, \mathcal{F}_t,\ \tau > t\big) = \mathbb{E}\Big[e^{-\int_t^T \lambda_s\, ds}\,\Big|\,\mathcal{F}_t\Big]. \tag{2.3}$$

Compare that with the bond price of chapter 7, $P(t,T) = \mathbb{E}[e^{-\int_t^T r_s ds} \mid \mathcal{F}_t]$. They are the same formula. The survival probability is a discount factor in which the intensity plays the part of the short rate, and everything that was ever proved about one transfers immediately to the other:

- a defaultable zero coupon bond paying nothing on default is worth $\mathbb{E}[e^{-\int_t^T (r_s+\lambda_s)ds}]$, so default risk simply adds the intensity to the discount rate;
- a credit curve is bootstrapped from credit default swaps exactly as chapter 7 bootstraps a rate curve from par swaps;
- and if λ is given affine dynamics, chapter 12's Riccati equations deliver survival probabilities in closed form, because (2.3) is the object those equations were derived for.

So the most important jump in finance is modelled by machinery these notes develop anyway, wearing different labels — which rules out concluding that continuous models leave credit untouched.

Remark (Incompleteness is not the end of the argument)**.** Chapter 4 will observe that jumps make a market incomplete, because a move of random size at a random time cannot be hedged by a position taken beforehand. What do markets do about it: they trade the jump.

A credit default swap pays precisely on the default event, so holding one spans exactly the risk that the underlying bond could not hedge. Adding it to the traded set restores completeness for that risk, and the price stops being a range. The same pattern recurs — variance swaps span variance risk, and an option written on an event date spans the event.

The honest statement is therefore not that incompleteness makes jump models unusable, but that it tells you which instrument you are missing. An incomplete model is a shopping list.

These notes stay with continuous models from here on, and chapter 4 gives the reason: continuity is what lets the underlying and the money market replicate everything, which is what makes the price unique rather than a range, and that uniqueness is what the rest of the theory is built on.

Remark (So should a trade be modelled with jumps?)**.** The question the previous paragraph invites, and it does not have a single answer. It has a criterion: does the payoff care whether the path is continuous?

For a great many trades it does not. A payoff that depends on variance accumulated over months — a vanilla option held to expiry, a variance swap — is insensitive to whether that variance arrived smoothly or in a few lumps, because a sum of many small contributions and a sum containing a few large ones look alike once they are added up. There a diffusion calibrated to the right prices will also give the right hedge, and the jumps are absorbed into the volatility with no harm done.

For others it is the whole trade:

- anything with a *barrier*, since a diffusion cannot cross a level without touching it and a jump can gap straight over — the two models disagree about the one event the payoff is defined by;
- short-dated options far from the money, whose entire value is the probability of a large move over a short horizon, which is precisely what a diffusion says is negligible and a jump model says is not;
- anything on an asset that can default, where the jump is the risk;
- and the *hedge* of any of the above, since a delta computed from a model that does not believe in gaps cannot protect against one.

So the criterion is not the asset, it is the payoff. The same stock may be modelled as a diffusion for a one year at-the-money option and must not be for a one week barrier struck nearby.

There is a cost on the other side, and it is the running theme of these notes rather than a new objection. A jump model has more parameters — an intensity, a distribution of jump sizes — and those parameters are poorly identified by liquid vanilla quotes, in exactly the way chapter 10's β and chapter 11's mixing weight are. Fitting the surface does not pin them down, so adding jumps removes one model risk and introduces another, to be marked by judgement and reserved against as chapter 24 describes.

Which is why common practice is neither extreme. Most things are priced with a continuous model, with a lump of variance added on known event dates and an explicit reserve held for gap risk on the products where continuity is doing real work — reaching for a full jump model when the payoff makes the jump unavoidable, rather than as a general upgrade.

References

- Rogers, L. C. G. (1997). Arbitrage with fractional Brownian motion. *Mathematical Finance*, 7(1), 95–105.

- Cheridito, P. (2003). Arbitrage in fractional Brownian motion models. *Finance and Stochastics*, 7(4), 533–553.

- Kunita, H., & Watanabe, S. (1967). On square integrable martingales. *Nagoya Mathematical Journal*, 30, 209–245.

- Follmer, H., & Sondermann, D. (1986). Hedging of non-redundant contingent claims. In *Contributions to Mathematical Economics*, North-Holland, 205–223.

- Schweizer, M. (2001). A guided tour through quadratic hedging approaches. In *Option Pricing, Interest Rates and Risk Management*, Cambridge University Press, 538–574.

- Dubins, L. E., & Schwarz, G. (1965). On continuous martingales. *Proceedings of the National Academy of Sciences*, 53(5), 913–916.

- Mandelbrot, B. B., & Van Ness, J. W. (1968). Fractional Brownian motions, fractional noises and applications. *SIAM Review*, 10(4), 422–437.