Stochastic processes and SDEs

Maksim Levental

May 2016

Stochastic Processes

Definition

A *stochastic process* is a collection of random variables $X_t \triangleq X(\omega, t)$ indexed by an index set T such that $|T| \geq |\mathbb{N}|$ such that every finite dimensional joint distribution is specified[a]. For fixed ω, $X_t(\omega)$ is a *sample path* or *realization* .

[a] And they obey some regularity conditions cf. Kolmogorov Extension thm.

- A finite collection of random variables is just a random vector
- T is typically called time but stochastic processes are not necessarily time series
- $\mathrm{im}(X_t)$ and T can both be either continuous or discrete

$\mathrm{im}(X_t) \setminus T$	cont.	disc.
cont.	Brownian motion (particle motion), Cox process (neuron spike trains)	Rust models
disc.	Contact process (epidemiology), Telegraph process (phase transitions)	Markov chain (noisy logic), Bernoulli process (gambling), Poisson process (queuing)

- Other: Dirichlet process, Pitman-Yor process, Random field

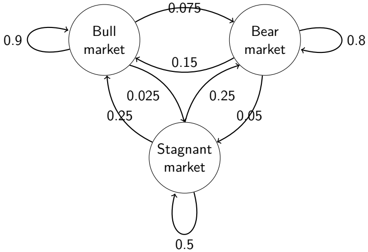
Definition

A stochastic process is *Markov* if $P(X_t \in A | \mathcal{F}_s) = P(X_t \in A | X_s)$.

- In particular, for a Markov chain

$$P\left(X_n = x_n \middle| X_{n-1} = x_{n-1}, X_{n-2} = x_{n-2}, \ldots, X_0 = x_0\right) = P\left(X_n = x_n \middle| X_{n-1} = x_{n-1}\right)$$

  i.e. only short term memory
- Intuitively a DFA with probabalistic transition function (not NFA)



	Bull	Bear	Stag.
Bull	0.9	0.075	0.025
Bear	0.15	0.8	0.05
Stag.	0.25	0.25	0.5

- Hidden Markov models (hierarchical model) great for speech recognition

Definition

A random variable N is distributed Poisson(λ) on some unit interval u if at

$$P\left(N = n \text{ events in interval}\right) = \frac{\lambda^n e^{-\lambda}}{n!}$$

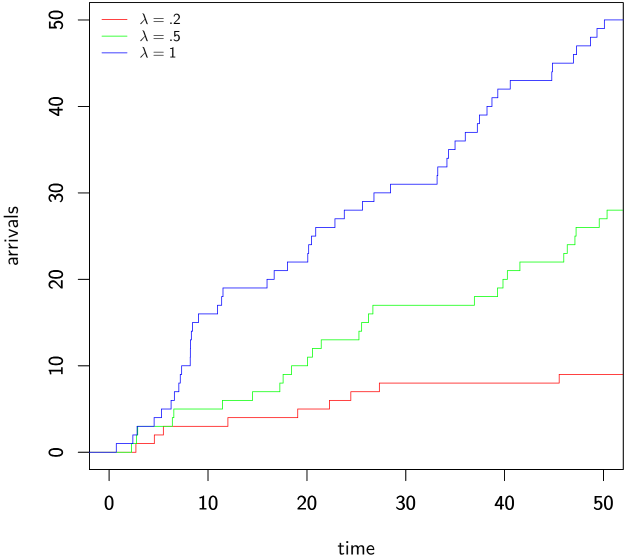
λ is called the rate parameter.

- Intuitively "time" between events is exponentially distributed but independent
- Phone calls at an exchange, arrivals bank queue, train arrivals (on a bad day!)

Example

A *Poisson process* N_t on $[0, \infty)$ with rate λ is a stochastic process where the number of events in any interval of length t is distributed Poisson(λt).

```
lambda <- 1
x1 <- cumsum(rexp(50),rate=lambda)
y1 <- cumsum(c(0,rep(1,50)))
plot(stepfun(x1,y1),xlim = c(0,50),do.points = F)
```



$\lambda = .2$
$\lambda = .5$
$\lambda = 1$
arrivals
time
0
10
20
30
40
50
0
10
20
30
40
50

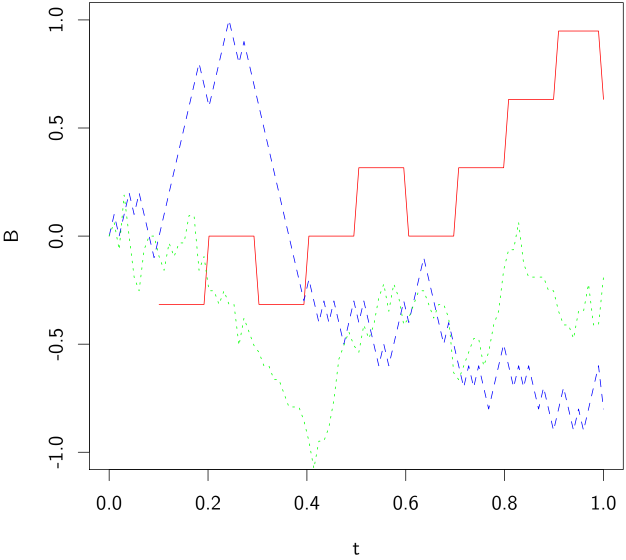
Example

A *random walk* on $\mathbb{Z}$ is a stochastic process $S_0, S_1, \ldots$ such that

$$S_n = \sum_{j=0}^{n} X_i$$

and X_i are iid Bernoulli$\left(\frac{1}{2}\right)$ on $\{1, -1\}$.

- Flip a coin and go forward or backward one unit distance in dimension
- Extensions to higher dimensions (random walk on a lattice) involve discrete uniform distribution on directions
- Precursor to Brownian motion
- Drunk man, Drunk bird


B
t
1.0
0.5
0.0
-0.5
-1.0
0.0
0.2
0.4
0.6
0.8
1.0

Brownian Motion

"Rigorous" Brownian motion

Modify S_n such that $X_i \in \{0,1\}$and suppose spatial increments are Δx and time increments Δt. Note that $E(S_n) = \frac{n}{2}$ and $\operatorname{Var}(X_i) = \frac{1}{4}$. Then

$$X(t) := X(n\Delta t) := \underbrace{S_n \Delta x}_{\text{positive dist}} - \underbrace{(n - S_n)(-\Delta x)}_{\text{negative dist}} = (2S_n - n)\,\Delta x$$

is the position of the particle at time $n\Delta t$. To use Laplace - De Moivre[1] we need

$$\operatorname{Var}(X(n\Delta t)) = (\Delta x)^2 n = \frac{(\Delta x)^2}{\Delta t} t = Dt$$

with, $D \triangleq \frac{(\Delta x)^2}{\Delta t}$. Then

$$X(n\Delta t) = \sqrt{Dt}\left[\left(\frac{S_n - \frac{n}{2}}{\sqrt{n/4}}\right)\right]$$

[1] $X_i \sim \text{Bernoulli}(p) \Rightarrow \lim_{n\to\infty} P(a \leq \sum X_i - np/\sqrt{np(1-p)}) \leq b = (2\pi)^{-1/2}\int_a^b e^{-\frac{x^2}{2}}\,dx$

and finally

$$\begin{aligned}
\lim_{\substack{n\to\infty \\ t=n\Delta t \\ ,\Delta t D=(\Delta x)^2}} P\left(a \leq \sqrt{Dt} X(t) \leq b\right) &= \lim_{\substack{n\to\infty \\ t=n\Delta t \\ ,\Delta t D=(\Delta x)^2}} P\left(\sqrt{Dt} a \leq X(t) \leq \sqrt{Dt} b\right) \\
&= \frac{1}{\sqrt{2\pi}} \int_{\frac{a}{\sqrt{Dt}}}^{\frac{b}{\sqrt{Dt}}} e^{-\frac{x^2}{2}} dx \\
&= \frac{1}{\sqrt{2\pi Dt}} \int_a^b e^{-\frac{x^2}{2Dt}} dx
\end{aligned}$$

So $X(t) \sim N(0, Dt)$

Rigorous Brownian motion

Due to Hida[2]. For $s \in \mathscr{S}_{\mathbb{R}}$ the Schwartz space of rapidly decreasing functions and its topological dual[3] $\mathscr{S}'_{\mathbb{R}}$ let

$$e^{-\frac{1}{2}\|s\|^2_{\mathsf{L}_2(\mathbb{R})}} = \int_{\mathscr{S}'_{\mathbb{R}}} e^{\langle s', s\rangle} dP\left(s'\right)$$

Defining $\Omega := \mathscr{S}'_{\mathbb{R}}$ we have $\left(\Omega, \mathcal{B}\left(\mathscr{S}'_{\mathbb{R}}\right), P\right)$ the white noise space and $L_2(\Omega) \triangleq L_2\left(\Omega, \mathcal{B}\left(\mathscr{S}'_{\mathbb{R}}\right), P\right)$. The measure P is the *white noise* measure. By taking power series of the integrand above we get a definition of $\langle \omega, f\rangle$. Then

$$B(t) \triangleq B(\omega, t) \triangleq \left\langle \omega, \mathbf{1}_{[0,t]} \right\rangle$$

[2] T. Hida. *Analysis of Brownian functionals*. Carleton Univ., Ottawa, Ont., 1975. Carleton Mathematical Lecture Notes, No. 13.

[3] Space of tempered distributions (all distributions whose Fourier transforms exist).

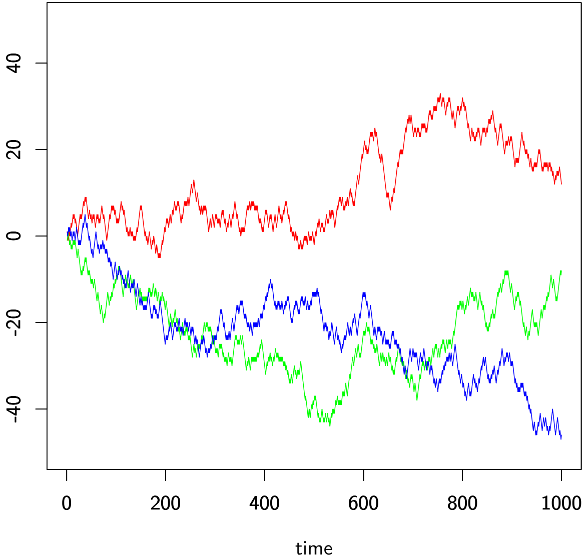
Definition

A stochastic process $B(t)$ is a Brownian motion if

1. $B(0) = 0$ almost surely, i.e. $P(B(0) = 0) = 1$
2. $B(t) - B(s) \sim N(0, t - s)$ for $t \geq s \geq 0$
3. For all $0 < t_1 < t_2 < \cdots < t_n$ it's the case that $B(t_1) \perp B(t_2) - B(t_1) \perp \cdots \perp B(t_n) - B(t_{n-1})$

Interesting facts

- $X_\omega(t) \triangleq X(\omega_0, t)$ as a sample path is a continuous function from $\mathbb{R}^+ \to \mathbb{R}$
- Brownian motion "induces" a measure on functions $\mathbb{R}^+ \to \mathbb{R}$
- Concentrated on continuous but nowhere differentiable functions (i.e. probability of "drawing" a differentiable function is 0)


40
20
0
-20
-40
0
200
400
600
800
1000
time

Stochastic Differential Equations

Two problems

- Charge $Q(t)$ in a capacitor an LRC circuit

$$LQ''(t) + RQ'(t) + \frac{1}{C}Q(t) = F(t),\; Q(0) = Q_0,\; Q'(0) = I_0 \quad (1)$$

If $F(t) = \cos(\omega t)$ and $F(0) = F_0$ the solution is simple:

$$Q(t) = c_1 e^{r_1 t} + c_2 e^{r_2 t} + A\sin(\omega t - \phi)$$

for some constants[4] $c_1, r_1, c_2, r_2, A, \phi$. But what if $F(t) = G(t) +$ "noise"?

- Consider taking noisy measurements $Z(t)$ where

$$Z(t) = Q(t) + \text{"noise"}$$

What is the best estimate of $Q(t)$ satisfying eqn 1 based on $Z(t)$? Kalman Filter.

[4] Each are constant functions of boundary values Q_0, I_0, F_0.

Baby steps

Consider the problem of finding an interpretation of

$$\frac{dx}{dt} = b(t, x) + \sigma(t, x) \cdot \text{"noise"}$$

It turns out it's reasonable to model "noise" as some stochastic process W_t and so

$$\frac{dX_t}{dt} = b(t, X_t) + \sigma(t, X_t) \cdot W_t$$

$x \to X$ a random variable because noise is stochastic. Empirical fact (experience) suggests W_t should have three properties

1. $t_1 \neq t_2 \Rightarrow W_{t_1} \perp W_{t_2}$
2. $\{W_t\}$ is stationary, i.e. the joint distribution of $\{W_{t_1+\tau}, \ldots, W_{t_k+\tau}\}$ does not depend on τ.
3. $E[W_t] = 0$ for all t.

Unfortunately property 1 not possible for continuous processes[5]. What to do? Discretize, require independent increments, take averages, redefine, and voila

$$X_k = X_0 + \int_0^t b(s, X_s)\, ds + \text{"} \int_0^t \sigma(s, X_s)\, dB_s \text{"} \tag{2}$$

But this begs the question; we still haven't defined " $\int_0^t \sigma(s, X_s)\, dB_s$". Let $0 \leq Q < T$ and start by defining $\int_Q^T \{\cdot\}\, dB_s(\omega)$ for simple processes $S_n(t, \omega) = \sum_{j=0}^{\infty} a_j(\omega)\, 1_{[j \cdot 2^{-n}, (j+1) \cdot 2^{-n})}(t)$

$$\int_Q^T S_n(s, \omega)\, dB_s \triangleq \sum_{j=0}^{\infty} a_j(\omega) \left[B_{s_{j+1}^{(n)}}(\omega) - B_{s_j^{(n)}}(\omega) \right]$$

Then extend by taking limits (which exist because Cauchy) under $\mathbf{L}_2(P)$ norm

$$\mathcal{I}[f](\omega) \triangleq \int_Q^T f(s, \omega)\, dB_s \triangleq \lim_{n \to \infty} \int_Q^T S_n(s, \omega)\, dB_s$$

[5]It is possible to represent W_t as a *generalized* process, meaning it can be constructed as a measure on the space of tempered distributions

Theorem

Some properties of the Ito integral: let $f, g \in \mathcal{V}(0, T)$ *and* $0 \leq Q < U < T$. *Then*

1. $P\left(\int_Q^T fdB = \int_Q^U fdB + \int_U^T fdB\right) = 1$
2. *For* $c \in \mathbb{R}$: $P\left(\int_Q^T (cf + g)\, dB = \int_Q^T cfdB + \int_Q^T gdB\right) = 1$
3. $E\left[\int_Q^T fdB\right] = 0$

Property 3 is related to the fact that Ito integrals are martingales.

Definition

A stochastic process X_t is a *martingale* if for $s \leq t$

$$E\left(X_t \middle| X_s\right) = X_s$$

i.e. "fair"; $E\left(X_t - X_s \middle| X_s\right) = 0$, so $E\left(X_t\right) = E\left(X_0\right)$ for all t.

Theorem (Ito formula)

Let X_t be an Ito process and $g(t,x) \in C^2([0,\infty) \times \mathbb{R})$ then

$$Y_t = g(t, X_t)$$

is again an Ito process and

$$dY_t = \frac{\partial g}{\partial t} dt + \frac{\partial g}{\partial x} dX_t + \frac{1}{2} \frac{\partial^2 g}{\partial x^2} \cdot (dX_t)^2 \tag{3}$$

where $(dX_t)^2 = (dX_t) \cdot (dX_t)$ is computed according to

$$dt \cdot dt = dt \cdot dB_t = dB_t \cdot dt = 0 \quad dB_t \cdot dB_t = dt$$

Kind of like change of variables from single variable calculus.

One example

Does

$$\int_0^T B(t)\, dB_t = \frac{1}{2}(B(t))^2$$

If this were regular calculus it would but we're not in Kansas anymore; using Ito's formula with $Y(t) = \frac{1}{2}(B(t))^2$ we get that

$$\begin{aligned} d(Y(t)) &= d\left(\frac{1}{2}(B(t))^2\right) = \frac{\partial g}{\partial t}dt + \frac{\partial g}{\partial x}dB_t + \frac{1}{2}\frac{\partial^2 g}{\partial x^2}(dB_t)^2 \\ &= B(t)\, dB_t + \frac{1}{2}(dB_t)^2 = B(t)\, dB_t + \frac{1}{2}dt \end{aligned}$$

and so

$$\int d\left(\frac{1}{2}(B(t))^2\right) = \int_0^T B(t)\, dB_t + \int_0^T \frac{1}{2}dt$$

implies

$$\int_0^T B(t)\, dB_t = \frac{1}{2}\left((B(t))^2 - T\right)$$

Two more problems

- Equity price X_t obeys SDE with known r drift, α volatility (and discount rate ρ)

$$\frac{dX_t}{dt} = rX_t + \alpha X_t \cdot \text{"noise"}$$

Know X_s up to present t - when to sell? Since noisy, *optimal stopping strategy* maximizes expected returns. Can be solved by solving a corresponding semi-elliptic second order PDE with Dirichlet boundary conditions.

- Suppose at some time t the person in problem 3 is offered the right (without obligation) to buy one unit of the risky asset at a specified price K at a specified future date $t = T$. Such a right/asset is called a *European call option*. How much should they be willing to pay for the option? Problem solved by Fischer Black and Myron Scholes - called the Black-Scholes equation for option pricing

$$\frac{\partial V}{\partial t} + \frac{1}{2}\sigma^2 S^2 \frac{\partial^2 V}{\partial S^2} + rS\frac{\partial V}{\partial S} - rV = 0$$

where V is the price of the option as a function of the price of the asset, r is the risk-free interest rate, and σ is the volatility of the stock.

References

- Probability/Measure theory
  - Schilling *Measures, Integrals, and Martingales*
  - Resnick *A Probability Path*
  - Pollard *A User's Guide to Measure Theoretic Probability*
  - Billingsley *Probability and Measure*
- Stochastic processes
  - Ross *Introduction to Probability Models*
  - Lawler *Introduction to Stochastic Processes*
- Stochastic Differential Equations
  - Klebaner *Introduction to Stochastic Calculus with Applications*
  - Shreve *Stochastic Calculus for Finance II: Continuous-Time Models*
  - Oksendal *Stochastic Differential Equations*
- SDE Numerics
  - Iacus *Simulation and Inference for Stochastic Differential Equations*

Appendix

Laplace - De Moivre

$$
\begin{aligned}
\operatorname{Var}\left(X\left(n\Delta t\right)\right) &= \operatorname{Var}\left(\left(2S_n - n\right)\Delta x\right) = \left(\Delta x\right)^2 \operatorname{Var}\left(\left(2S_n - n\right)\right) \\
&= 4\left(\Delta x\right)^2 \operatorname{Var}\left(S_n\right) = 4\left(\Delta x\right)^2 \operatorname{Var}\left(\sum_{i=1}^{n} X_i\right) \\
&= 4\left(\Delta x\right)^2 n \operatorname{Var}\left(X_i\right) = \left(\Delta x\right)^2 n = \frac{\left(\Delta x\right)^2}{\Delta t} t = Dt
\end{aligned}
$$

$$
\begin{aligned}
X(n\Delta t) &= (2S_n - n)\,\Delta x = \sqrt{n}\Delta x \left(\frac{S_n - \frac{n}{2}}{\sqrt{n/4}}\right) \\
&= \sqrt{Dt}\left(\frac{\left(\sum_{i=1}^{n} X_i\right) - \frac{n}{2}}{\sqrt{n/4}}\right) = \sqrt{Dt}n\left(\frac{\frac{1}{n}\left(\sum_{i=1}^{n} X_i\right) - \frac{1}{2}}{\sqrt{n/4}}\right) \\
&= \sqrt{Dt}\left[\left(\frac{\frac{1}{n}\left(\sum_{i=1}^{n} X_i\right) - \frac{1}{2}}{\sqrt{1/4}/\sqrt{n}}\right)\right]
\end{aligned}
$$

Differentiable nowhere

$B_t(\omega)$ has infinite total variation;

$$TV(f) := \lim_{n\to\infty} \sum_{j=1}^{m} \left| f\left(t_j^{(n)}\right) - f\left(t_{j-1}^{(n)}\right) \right|$$

over some $[Q,T]$[6]. Here's a short proof of this: first define quadratic variation

$$QV(f) := \lim_{n\to\infty} \sum_{j=1}^{m} \left| f\left(t_j^{(n)}\right) - f\left(t_{j-1}^{(n)}\right) \right|^2$$

and notice that if f is continuous then

$$\sum_{j=1}^{m} \left| f\left(t_j^{(n)}\right) - f\left(t_{j-1}^{(n)}\right) \right|^2 \leq \left(\max_{1\leq j\leq m} \left| f\left(t_j^{(n)}\right) - f\left(t_{j-1}^{(n)}\right) \right| \right) \sum_{j=1}^{m} \left| f\left(t_j^{(n)}\right) - f\left(t_{j-1}^{(n)}\right) \right|$$

and so

$$\frac{\sum_{j=1}^{m} \left| f\left(t_j^{(n)}\right) - f\left(t_{j-1}^{(n)}\right) \right|^2}{\max_{1\leq j\leq m} \left| f\left(t_j^{(n)}\right) - f\left(t_{j-1}^{(n)}\right) \right|} \leq \sum_{j=1}^{m} \left| f\left(t_j^{(n)}\right) - f\left(t_{j-1}^{(n)}\right) \right|$$

[6] Recall that $T - Q = m \cdot 2^{-n}$.

and hence any continuous f that has non-zero quadratic variation has infinite total variation[7]. So all we need to prove is that B_s has non-zero quadratic variation. First some lemmas.

Fact

If

$$\lim_{n\to\infty} Var\left[\sum_{j=1}^{m}\left(B_{t_j^{(n)}} - B_{t_{j-1}^{(n)}}\right)^2\right] = 0$$

then $\lim_{n\to\infty} QV(f) = T - Q$ *in* L^2.

Proof: Let $\Delta B_j^2 = \left(B_{t_j^{(n)}} - B_{t_{j-1}^{(n)}}\right)^2$. Then if the variance goes to 0 [8]

$$\begin{aligned}
\lim_{n\to\infty} E\left[\left(\sum_{j=1}^{m}\Delta B_j^2\right)^2\right] &= \lim_{n\to\infty}\left(E\left[\sum_{j=1}^{m}\Delta B_j^2\right]\right)^2 = \lim_{n\to\infty}\left(\sum_{j=1}^{m}E\left[\Delta B_j^2\right]\right)^2 \\
&= \lim_{n\to\infty}\left(\sum_{j=1}^{m}\left(t_j^{(n)} - t_{j-1}^{(n)}\right)\right)^2 = \lim_{n\to\infty}(T-Q)^2
\end{aligned}$$

---

[7] Since $\max_{1\le j\le m}\left|f\left(t_j^{(n)}\right) - f\left(t_{j-1}^{(n)}\right)\right| \to 0$ as $|\Pi| \to \infty$ for any continuous f and your only hope for the left side of the inequality not blowing up is if the numerator, $QV(f)$, is 0.

[8] Since$Var(X) = EX^2 - (EX)^2$.

and so

$$
\begin{aligned}
0 &= \lim_{n\to\infty}\left(E\left[\left(\sum_{j=1}^{m}\Delta B_j^2\right)^2\right]-(T-Q)^2\right)\\
&= \lim_{n\to\infty}\left(E\left[\left(\sum_{j=1}^{m}\Delta B_j^2\right)^2\right]-2\,(T-Q)^2+(T-Q)^2\right)\\
&= \lim_{n\to\infty}\left(E\left[\left(\sum_{j=1}^{m}\Delta B_j^2\right)^2\right]-2\,(T-Q)\,E\left[\left(\sum_{j=1}^{m}\Delta B_j^2\right)\right]+(T-Q)^2\right)\\
&= \lim_{n\to\infty}\left(E\left[\left(\sum_{j=1}^{m}\Delta B_j^2-(T-Q)\right)^2\right]\right)
\end{aligned}
$$

which is the definition of convergence in L^2.

Fact

On refinement of the mesh

$$\lim_{n\to\infty} Var\left[\sum_{j=1}^{m}\left(B_{t_j^{(n)}}-B_{t_{j-1}^{(n)}}\right)^2\right]=0$$

Proof:

$$
\begin{aligned}
\mathrm{Var}\left[\sum_{j=1}^{m}\left(B_{t_j^{(n)}} - B_{t_{j-1}^{(n)}}\right)^2\right] &= \sum_{j=1}^{m}\mathrm{Var}\left[\left(B_{t_j^{(n)}} - B_{t_{j-1}^{(n)}}\right)^2\right] \\
&= \sum_{j=1}^{m}\left(E\left[\left(B_{t_j^{(n)}} - B_{t_{j-1}^{(n)}}\right)^2\right] - \left(E\left[\left(B_{t_j^{(n)}} - B_{t_{j-1}^{(n)}}\right)^2\right]\right)^2\right) \\
&= \sum_{j=1}^{m}\left(E\left[\left(B_{t_j^{(n)}} - B_{t_{j-1}^{(n)}}\right)^2\right] - \left(t_j^{(n)} - t_{j-1}^{(n)}\right)^2\right) \\
&\quad \text{by Kolmogorov continuity thm.} \\
&= \sum_{j=1}^{m}\left(1\,(1+2)\left(t_j^{(n)} - t_{j-1}^{(n)}\right)^2 - \left(t_j^{(n)} - t_{j-1}^{(n)}\right)^2\right) \\
&= 2\sum_{j=1}^{m}\left(t_j^{(n)} - t_{j-1}^{(n)}\right)^2
\end{aligned}
$$

which goes to 0 as the mesh is refined.

Theorem

For $f = B_t$ it's the case that $\lim_{n\to\infty} QV(f) = T - Q$ almost surely.

Proof: Let

$$X_i^{(n)} = \Delta B_j^2 - \left(t_j^{(n)} - t_{j-1}^{(n)}\right)$$

and

$$Y_n := \sum_{j=1}^{m} X_i^{(n)} = \sum_{j=1}^{m}\left(\Delta B_j^2 - \left(t_j^{(n)} - t_{j-1}^{(n)}\right)\right) = \sum_{j=1}^{m}\Delta B_j^2 - (T - Q)$$

Then

$$\begin{aligned} EY_n &= E\left[\sum_{j=1}^{m}\left(B_{t_j^{(n)}} - B_{t_{j-1}^{(n)}}\right)^2\right] - E(T-Q) \\ &= 0 \end{aligned}$$

and

$$EY_n^2 = E\left(\sum_{j=1}^{m}\left(X_i^{(n)}\right)^2 + \sum_{i<j} X_i^{(n)}X_j^{(n)}\right) = \sum_{j=1}^{m} E\left[\left(X_i^{(n)}\right)^2\right] + \sum_{i<j} E\left[X_i^{(n)}X_j^{(n)}\right]$$

but $E\left[X_i^{(n)}X_j^{(n)}\right] = 0$ so

$$EY_n^2 = \sum_{j=1}^{m} E\left[\left(X_i^{(n)}\right)^2\right]$$

and so by Chebyshev's inequality[9]

$$\begin{aligned}
P\left(|Y_n| \geq \epsilon\right) &\leq \frac{E\left[(Y_n)^2\right]}{\epsilon^2} \\
&= \frac{1}{\epsilon^2}\sum_{j=1}^{m} E\left[\left(X_i^{(n)}\right)^2\right] \\
&= \frac{1}{\epsilon^2}\sum_{j=1}^{m}\left(t_j^{(n)} - t_{j-1}^{(n)}\right)^2 \\
&\leq \frac{1}{\epsilon^2}\frac{1}{2^n}\sum_{j=1}^{m}\left(t_j^{(n)} - t_{j-1}^{(n)}\right) \\
&= \frac{T-Q}{2^n\epsilon^2}
\end{aligned}$$

[9] $P\left(|X-\mu| \geq \epsilon\right) \leq \frac{E\left[(X-\mu)^2\right]}{\epsilon^2}$

and finally using Borel-Cantelli[10] with

$$\sum_{n=1}^{\infty} P\left(|Y_n| \geq \epsilon\right) \leq \sum_{n=1}^{\infty} \frac{T-Q}{2^n \epsilon^2} = \frac{T-Q}{\epsilon^2}$$

which implies almost sure convergence[11] of $Y_n \to 0$.

[10] If $\sum_{n=1}^{\infty} P\left(E_n\right) < \infty$ for some sequence of events E_n then $P\left(\limsup_{n\to\infty} E_n\right) = 0$.

[11] $P\left(\liminf_{n\to\infty} |X_n - X| < \epsilon\right) = 1$ for all ϵ. Naturally this is to equivalent $P\left(\liminf_{n\to\infty} |X_n - X| > \epsilon\right) = 0$ for all ϵ. Why? lim inf is the set of points ω that is ultimately in all of the sets and lim sup is the set of points ω appear infinitely often. So if the set of ω for which $|Y_n| \geq \epsilon$ occur infinitely often has measure 0 then set of ω for which $|Y_n| \leq \epsilon$ eventually always is almost all of them (otherwise $|Y_n| \geq \epsilon$ would keep happening once in a while).

Motivation

$$\frac{dx}{dt} = b(t,x) + \sigma(t,x) \cdot \text{"noise"}$$

becomes

$$X_{k+1} - X_k = b(t_k, X_k)\,\Delta t_k + \sigma(t_k, X_k)\, W_k \Delta t_k$$

where $X_k := X_{t_k}$. Restated the question is: does there exist some V_t such that for $\Delta V_k := V_{k+1} - V_k := V_{t_{k+1}} - V_{t_k}$

$$\begin{aligned} X_{k+1} - X_k &= b(t_k, X_k)\,\Delta t_k + \sigma(t_k, X_k)(V_{k+1} - V_k) \\ &= b(t_k, X_k)\,\Delta t_k + \sigma(t_k, X_k)\,\Delta V_k \end{aligned}$$

Assumptions 1,2,3 above suggest that stationary, independent, and mean 0 **increments**. Why? Because what appears in the discretized model are the increments. Turns out the only such process with continuous paths is Brownian motion B_t. Thus putting $V_t = B_t$ and taking sums we get

$$\sum_{j=0}^{k-1} (X_{k+1} - X_k) = X_k - X_0 = \sum_{j=0}^{k-1} \left(b(t_j, X_j)\,\Delta t_j + \sigma(t_j, X_j)\,\Delta B_j \right)$$

Martingale

First for $I_n(s,\omega)$ with $\phi_n(\omega,t) = \sum_j a_j^{(n)}(\omega)\mathbf{1}_{\left[t_j^{(n)},t_{j+1}^{(n)}\right)}(t)$:

$$
\begin{aligned}
E\left(I_n(\omega,s)\middle|\mathcal{F}_t\right) &= E\left(\int_0^t \phi_n dB + \int_s^t \phi_n dB \middle| \mathcal{F}_t\right) \\
&= \int_0^t \phi_n dB + E\left(\sum_{t\le t_j^{(n)}\le t_{j+1}^{(n)}\le s} a_j^{(n)}\Delta B_j \middle| \mathcal{F}_t\right) \\
&= \int_0^t \phi_n dB + \sum_j E\left(E\left(a_j^{(n)}\Delta B_j \middle| \mathcal{F}_{t_j^{(n)}}\right)\middle| \mathcal{F}_t\right) \\
&= \int_0^t \phi_n dB + \sum_j E\left(a_j^{(n)} \cancel{E\left(\Delta B_j \middle| \mathcal{F}_{t_j^{(n)}}\right)}\middle| \mathcal{F}_t\right) \\
&= I_n(\omega,t)
\end{aligned}
$$

Then by convergence of a.s convergence of $I_n(\omega, t) \to I(\omega, t)$ we get that $I(\omega, t)$ is a martingale.