Topics in Market Microstructure Models, Summer 2015

Lecture 2: Crash course on stochastic calculus and stochastic control theory

Tai-Ho Wang (王 太 和)

Baruch COLLEGE
The City University of New York

and

北京大學

Outline of Lecture 2

- Brownian motion
- Stochastic integral
- Ito's formula
- Feymann-Kac formula
- Stochastic control theory
  - Dynamic programming principle
  - Hamilton-Jacobi-Bellman (HJB) equation

Definition of Brownian motion

Let $(\Omega, \mathcal{F}_t, \mathbb{P})$ be a filtered probability space.

A stochastic process B_t adapted to $\mathcal{F}_t$ is called a (standard) Brownian motion or a Wiener process if it satisfies the following conditions

- $\mathbb{P}[\omega : B_0(\omega) = 0] = 1$, i.e., the process starts at zero almost surely.

- For any $0 \leq s < t$, the random variable $B_t - B_s$ is normally distributed with mean 0 and variance $t - s$, i.e., for any $a < b$,
  $\mathbb{P}[a \leq B_t - B_s \leq b] = \frac{1}{\sqrt{2\pi(t-s)}} \int_a^b e^{-\frac{x^2}{2(t-s)}}\, dx.$
- B_t has independent increment, i.e., for any $0 \leq t_1 < t_2 < \cdots < t_n$, the random variables
  $B_{t_1},\ B_{t_2} - B_{t_1},\ \cdots,\ B_{t_n} - B_{t_{n-1}}$
  are independent.
- Almost all sample paths of B_t are continuous functions, i.e.,
  $\mathbb{P}[\omega : B_t(\omega) \text{ is continuous }] = 1$

Remark

- A Brownian motion is sometimes defined as a stochastic process satisfying only the first 3 conditions in the definition. Such a process always has continuous modification by applying Kolmogorov's continuity criterion.
- The standard Brownian motion starts at 0. A Brownian motion starts at $x \neq 0$ is obtain by shifting $x + B_t$.

Kolmogorov's continuity criterion

Theorem

A process X, for which there exist three constants α, β, $C > 0$ such that
$\mathbb{E}|X_{t+h} - X_t|^\alpha \leq Ch^{1+\beta}$
for every t and h, has a modification which is almost surely continuous.

For Brownian motion B_t, since the random variable $B_{t+h} - B_t$ is centered Gaussian with variance h, we have

$\mathbb{E}\left[(B_{t+h} - B_t)^4\right] = 3h^2.$

Therefore, by taking $\alpha = 4$, $\beta = 1$, and $C = 3$, the Kolmogorov's continuity criterion applies.

Historical note

Quotes from Wikipedia:

"Brownian motion or pedesis (from Greek: πήδησις /pɛ̌ːdɛːsis/ "leaping") is the random motion of particles suspended in a fluid (a liquid or a gas) resulting from their collision with the quick atoms or molecules in the gas or liquid."

"This transport phenomenon is named after the botanist Robert Brown. In 1827, while looking through a microscope at particles found in pollen grains in water, he noted that the particles moved through the water but was not able to determine the mechanisms that caused this motion."

"The first person to describe the mathematics behind Brownian motion was Thorvald N. Thiele in a paper on the method of least squares published in 1880. This was followed independently by Louis Bachelier in 1900 in his PhD thesis "The theory of speculation", in which he presented a stochastic analysis of the stock and option markets. Albert Einstein (in one of his 1905 papers) and Marian Smoluchowski (1906) brought the solution of the problem to the attention of physicists, and presented it as a way to indirectly confirm the existence of atoms and molecules. Their equations describing Brownian motion were subsequently verified by the experimental work of Jean Baptiste Perrin in 1908."

Properties of Brownian motion

The following properties hold for Brownian motion B_t.

- *Time-homogeneity* For any $s > 0$, the process $B_{t+s} - B_s$, $t \geq 0$ is also a Brownian motion and is independent of the σ-algrebra $\sigma(B_u, u \leq s)$.
- *Symmetry* The process $-B_t$, $t \geq 0$, is a Brownian motion.
- *Self-similarity* For every $c > 0$, the process cB_{t/c^2}, $t \geq 0$, is a Brownian motion.
- *Time inversion* The process X defined by $X_0 = 0$, $X_t = tB_{1/t}$ for $t > 0$, is a Brownian motion.

Distributional properties of Brownian motion

Brownian motion is a Gaussian process, it is fully characterized by the mean and the covariance functions.

- $\mathbb{E}[B_t] = 0$ for all t
- $\operatorname{cov}(B_t, B_s) = \min\{s, t\}$

To calculate the covariance, without loss of generality, we assume $s < t$.

Review: Gaussian process

A stochastic process X_t is called a *Gaussian process* if all its finite dimensional distributions are multivariate normally distributed. Thus, a Gaussian process is fully characterized by its mean function $\mu(t) = \mathbb{E}[X_t]$ and covariance function $\gamma(t, s) = \operatorname{cov}(X_t, X_s)$.

Commonly encountered Gaussian processes

- Brownian motion
- Brownian motion with deterministic drift
- Brownian bridge
- Ornstein-Uhlenbeck process
- fractional Brownian motion

Variation of a function

Let $f : [0, T] \to \mathbb{R}$. Let $\Pi_n = \{0 = t_0 < t_1 < \cdots < t_n = T\}$ be a sequence of partitions of the finite interval $[0, T]$ and denote $\|\Pi_n\| = \max_n\{t_i - t_{i-1}\}$. The variation $V_f(T)$ of f in $[0, T]$ is defined as

$$V_f(T) = \lim_{\|\Pi_n\| \to 0} \sum_{i=1}^{n} |f(t_i) - f(t_{i-1})|$$

provided the limit exists.

Remark

An important property for functions of finite variation is that it can be uniquely written as the sum of an increasing function and a decreasing function.

Quadratic variation and covariation

Let $f, g : [0, T] \to \mathbb{R}$. Let $\Pi_n = \{0 = t_0 < t_1 < \cdots < t_n = T\}$ be a sequence of partitions of the finite interval $[0, T]$ and denote $\|\Pi_n\| = \max_n \{t_i - t_{i-1}\}$.

Quadratic variation

The quadratic variation of f, denoted by $[f](T)$, in $[0, T]$ is defined as

$$[f](T) = \lim_{\|\Pi_n\| \to 0} \sum_{i=1}^{n} |f(t_i) - f(t_{i-1})|^2$$

provided the limit exists.

Quadratic covariation

The quadratic covariation of f and g, denoted by $[f, g](T)$, in $[0, T]$ is defined as

$$[f, g](T) = \lim_{\|\Pi_n\| \to 0} \sum_{i=1}^{n} [f(t_i) - f(t_{i-1})][g(t_i) - g(t_{i-1})]$$

provided the limit exists.

Remark

One can define even higher order variations, say, cubic variation. However, if the variation at some order is finite, all the higher order variations vanish. For example, the cubic variation of Brownian motion vanishes since it has finite quadratic variation.

Quadratic variation of Brownian motion

Let $\Delta_n = \{a = t_0 < t_1 < \cdots < t_n = b\}$ be a partition of a finite interval $[a, b]$. Then

$$\lim_{n \to \infty} \sum_{i=1}^{n} (B_{t_i} - B_{t_{i-1}})^2 = b - a \quad \text{ in } L^2(\Omega)$$

as $\|\Delta_n\| = \max\limits_{1 \le i \le n} (t_i - t_{i-1})$ goes to 0.

In other words, the quadratic variation $[B]_t$ of Brownian motion B_t in $[0, t]$ is $[B]_t = t$.

Remark

Almost sure convergence is guaranteed if the sequence $\{\Delta_n\}$ satisfies the condition

$$\Delta_1 \subset \Delta_2 \subset \cdots \subset \Delta_n \subset \cdots.$$

Almost sure convergence is also guaranteed when $\{\Delta_n\}$ satisfies $\sum_{n=1}^{\infty} \|\Delta_n\| < \infty$.

Technical note: Convergence in L^p

A sequence of random variables X_n defined on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$ is called *covergent to X in L^p* if

$$\lim_{n\to\infty} \|X_n - X\|_p = 0 \quad \text{or equivalently} \quad \lim_{n\to\infty} \mathbb{E}\left[|X_n - X|^p\right] = 0$$

where $\|X\|_p = \sqrt[p]{\mathbb{E}\left[|X|^p\right]}$ is the L^p-norm of the random variable X.

Remark

Recall that we have the relationship among different types of convergence

$$\begin{array}{ccccc} \text{in } L^p & \Rightarrow & \text{in probability} & \Rightarrow & \text{in distribution or weakly} \\ & & \Uparrow & & \\ & & \text{almost surely} & & \end{array}$$

Local properties of Brownian paths

Let B_t be a Brownian motion.

- B_t is locally Hölder continuous of order α for every $\alpha < \frac{1}{2}$.
- The Brownian paths are almost surely of infinite variation on any interval.
- The Brownian paths are almost surely nowhere locally Hölder continuous of order α for $\alpha > \frac{1}{2}$.
- Levy's modulus of continuity. Let $h(t) = \sqrt{2t\log(1/t)}$. Then

$$\mathbb{P}\left[\overline{\lim_{\epsilon\to 0}}\left(\sup_{0\le t_1<t_2\le 1, t_2-t_1\le\epsilon} \frac{|B_{t_2} - B_{t_1}|}{h(\epsilon)}\right) = 1\right] = 1$$

Technical note: Lipschitz and Hölder continuity

Definition (Lipschitz continuous)

A function f is called Lipschitz or Lipschitz continuous on the interval $[a, b]$ if there exists a constant L such that

$$|f(t) - f(s)| \le L|t - s|$$

for all $t, s \in [a, b]$.

Definition (Hölder continuous)

A function f is called Hölder continuous of order α on the interval $[a, b]$ if there exists a constant K such that

$$|f(t) - f(s)| \le K|t - s|^{\alpha}$$

for all $t, s \in [a, b]$.

The Lévy-Ciesielski construction of Brownian motion

Let $\{\psi_i\}$ be a complete orthonormal basis for $L^2[0,1]$ and ξ_i, $i = 1, 2, \cdots$, an iid sequence of standard normal random variables defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Define $\phi_i(t) = \int_0^t \psi_i(s)ds$, for $t \in [0,1]$. Then the stochastic process W defined by

$$W_t = \sum_{i=1}^{\infty} \xi_i \phi_i(t)$$

is a Brownian motion.

Paley-Wiener expansion of Brownian motion

The Paley-Wiener representation of a Brownian path in terms of a random Fourier series.

Let ξ_n be an iid sequence of standard normal variables. Then

$$B_t = \xi_0 t + \sqrt{2} \sum_{n=1}^{\infty} \xi_n \frac{\sin(n\pi t)}{n\pi}$$

and

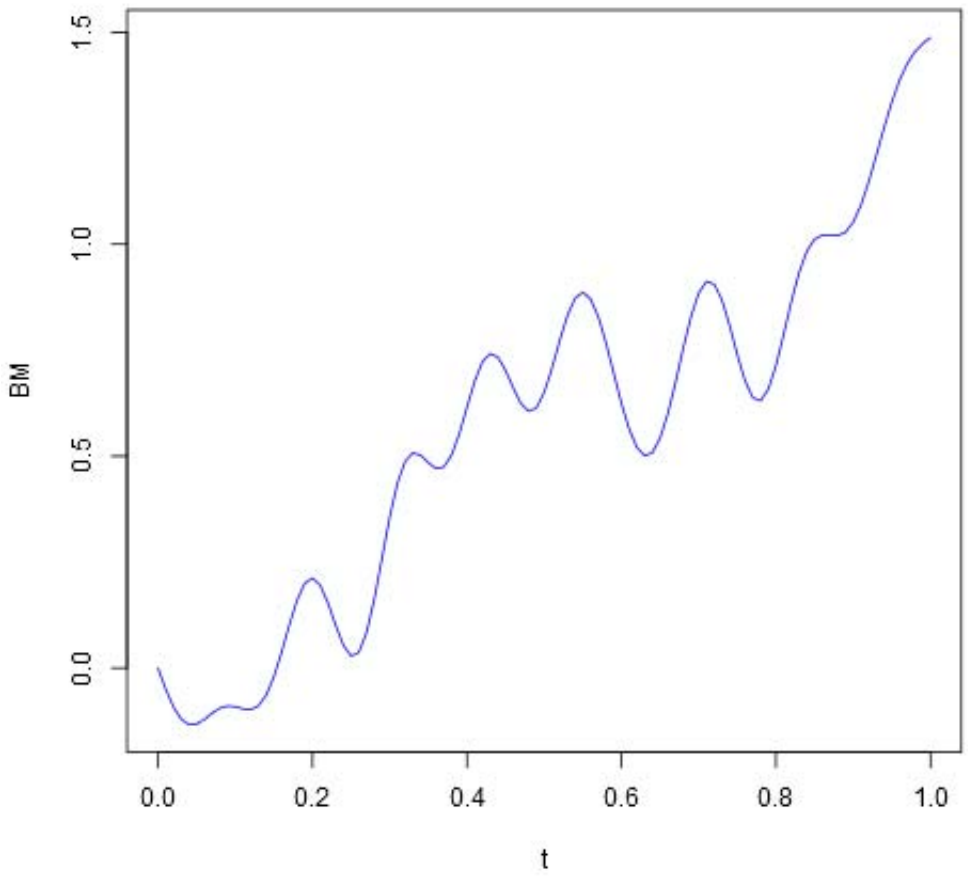
$$B_t = \sqrt{2} \sum_{n=1}^{\infty} \xi_n \frac{\sin\left(\left(n - \frac{1}{2}\right)\pi t\right)}{\left(n - \frac{1}{2}\right)\pi}$$

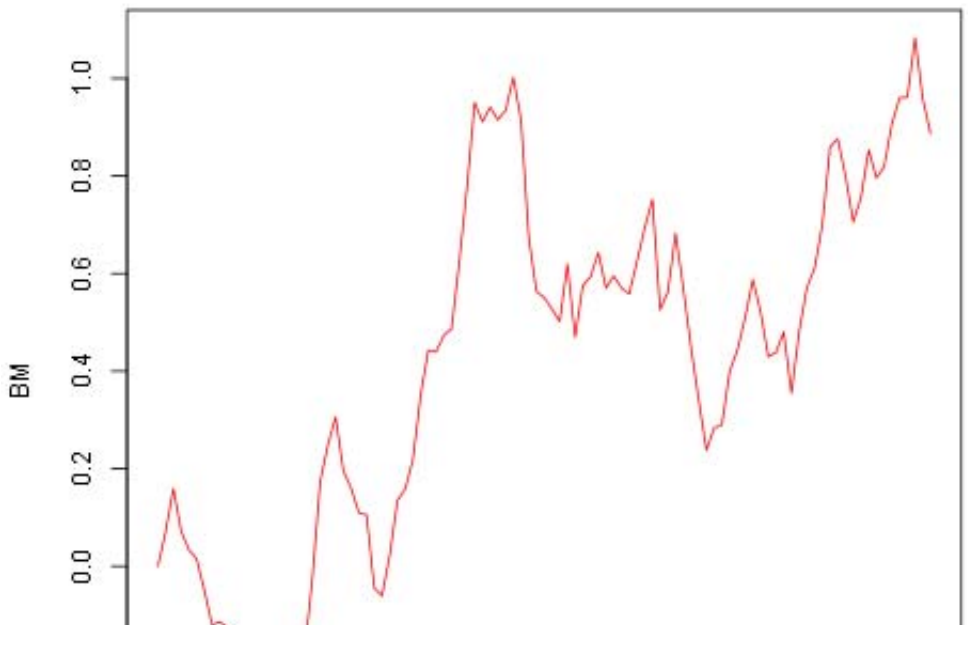
represent a Brownian motion on [0,1].

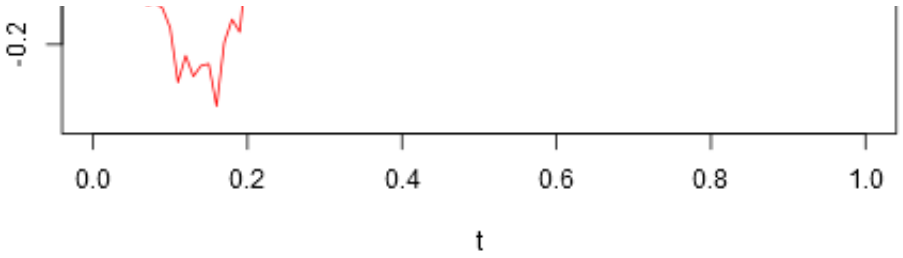
In [1]:
```
%load_ext rpy2.ipython
```

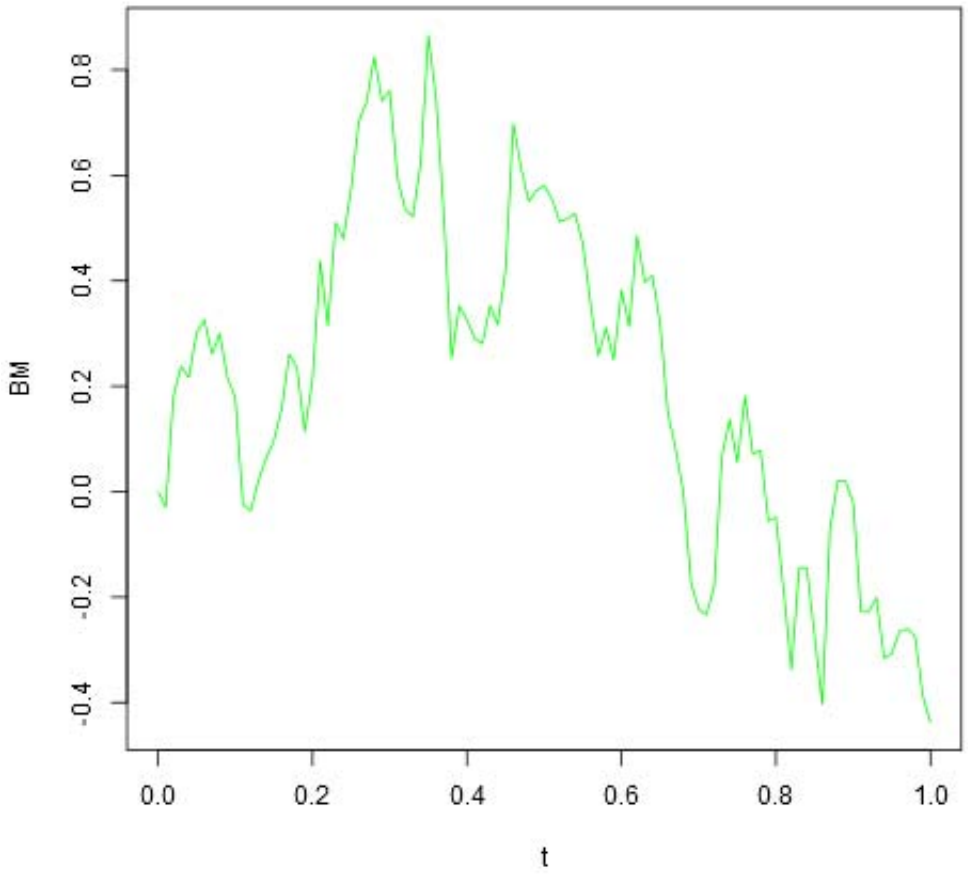
In [2]:
```
%%R
# the following function plots the Brownian motion path by the Paley-Wiener expansion
plotBM <- function(N,col='blue')
{
  xi0 <- rnorm(1)
  xi <- rnorm(N)
  W <- function(N) {
    if (N==0)
      res <- function(t) xi0*t
    else {
      coeff <- xi/(1:N)/pi
      res <- function(t) xi0*t + sqrt(2)*sum(coeff*sin(seq(1:N)*pi*t))
    }
    res
  }

  vW <- function(t) sapply(t,function(x) W(N)(x))
  curve(vW,from=0,to=1,col=col,xlab='t',ylab='BM')
```

```
}

plotBM(20)
plotBM(200,col='red')
plotBM(20000,col='green')
```









Donsker's invariance principle

- Donsker's invariance principle is also referred to as the Donsker's theorem.

Suppose $\{X_i\}_{i=1}^{\infty}$ is an iid sequence of random variables with mean 0 and and variance 1. Let $S_n = \sum_{i=1}^{n} X_i$. Define the function $\mathfrak{S}_n$ of t by

$$\mathfrak{S}_n(t) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left[S_{i-1} + n \left(t - \frac{i-1}{n} \right) X_i \right] \mathbf{1}_{\left(\frac{i-1}{n}, \frac{i}{n}\right]}(t).$$

In fact, $\mathfrak{S}_n$ is simply the linear interpolation of the scaled random walk $\left\{ \frac{S_1}{\sqrt{n}}, \frac{S_2}{\sqrt{n}}, \cdots, \frac{S_n}{\sqrt{n}} \right\}$.

Then, $\mathfrak{S}_n \Longrightarrow W$ as $n \to \infty$, where W denotes a Brownian motion.

In other words, as $n \to \infty$, the linearly interpolated scaled random walk $\mathfrak{S}_n$ converges weakly to a Brownian motion.

Technical note: Weak convergence or convergence in distribution

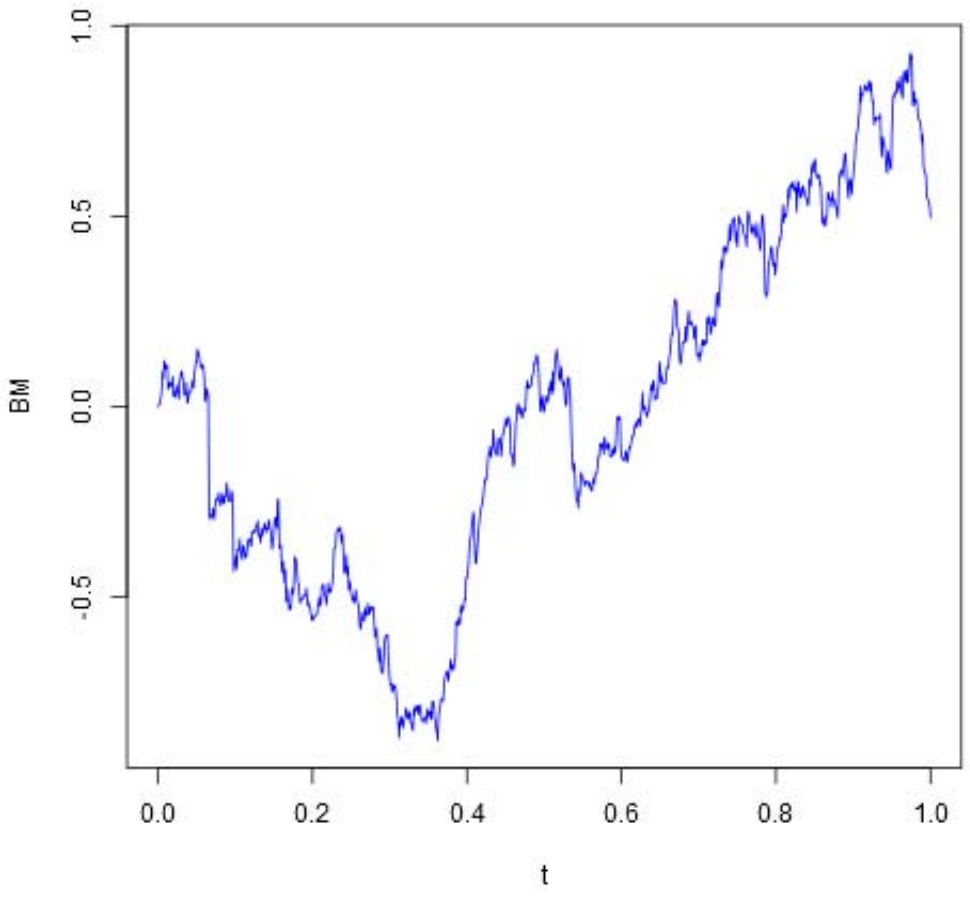
A sequence of random variables X_n is called convergent weakly or convergent in distribution to X if it satisfies one of the following equivalent conditions.

- $\lim_{n\to\infty} \phi_{X_n}(u) = \phi_X(u)$ for every u, where ϕ is the characteristic function.
- $\lim_{n\to\infty} \mathbb{E}\left[f(X_n)\right] = \mathbb{E}\left[f(X)\right]$ for all bounded continuous function f
- $\lim_{n\to\infty} F_n(x) = F(x)$ for every x at which F is continuous. F_n and F are cdfs of X_n and X respectively.

Remark

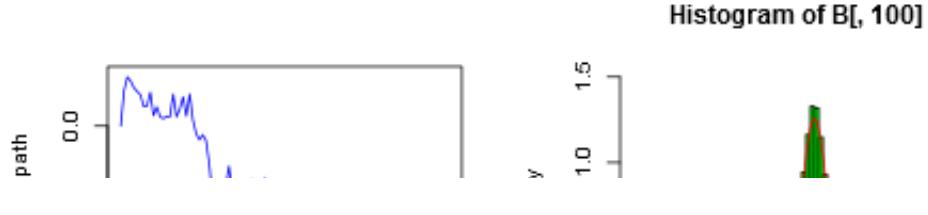
As opposed to the definition of convergence a.s., in L^p, and in probability, in defining weak convergence, the random variables X_n's and X need not to be defined on the same probability space.

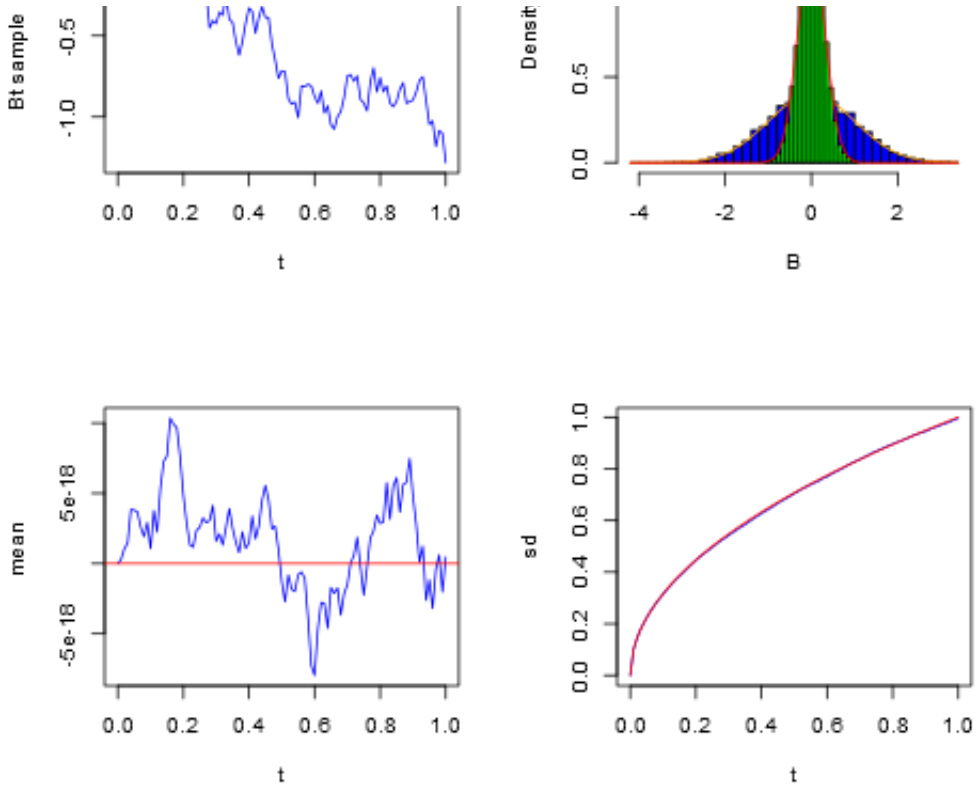
In [3]:
```
%%R
# Simulate BM by using the Donsker's invariance principle
N <- 1000
nu <- 3
X <- rt(N,df=nu)/sqrt(nu/(nu-2))
BM <- c(0,cumsum(X)/sqrt(N))
t <- c(0,(1:N)/N)
plot(t,BM,type='l',col='blue')
```



Simulation of Brownian motion

In [4]:
```
%%R
# the code demonstrates how to simulate Brownian paths

NSim <- 1e4   # number of samples at each step
NSteps <- 100   # number of steps
Tfin <- 1    # terminal time
B <- matrix(0,NSim,NSteps+1)  # initialize the Brownians

# simulation step
for (i in 1:NSteps){
  dB <- rnorm(NSim)
  dB <- dB - mean(dB)  # now dB has mean 0
  dB <- dB/sd(dB)  # now dB has variance 1
  dB <- sqrt(Tfin/NSteps)*dB
  B[,i+1] <- B[,i] + dB
}

t <- (0:NSteps)/NSteps*Tfin
par(mfrow=c(2,2))

# plot a Browian path
plot(t,B[5000,],type='l',col='blue',ylab='Bt sample path')

# histograms superimposed with normal distributions at different times
hist(B[,100],prob=T,breaks=50,col='blue',ylim=c(0,1.5),xlab='B')
curve(dnorm,col='orange',add=T)
hist(B[,10],prob=T,breaks=25,col='green',add=T)
curve(dnorm(x,sd=1/sqrt(10)),add=T,col='red')

# evolution of mean
plot(t,apply(B,2,FUN=mean),type='l',col='blue',xlab='t',ylab='mean')
abline(h=0,col='red')

# evolution of standard deviation
plot(t,apply(B,2,FUN=sd),type='l',col='blue',xlab='t',ylab='sd')
curve(sqrt(x),add=T,col='red')
```





Brownian motion with drift

Let $(\Omega, \mathcal{F}_t, \mathbb{P})$ be a filtered probability space and B_t a Brownian motion on Ω. A stochastic process X of the form

$$X_t = x + B_t + \int_0^t \mu_s \, ds \quad \iff \quad dX_t = dB_t + \mu_t \, dt$$

is called a Brownian motion with drift μ_t , where μ is adapted to the filtration $\mathcal{F}_t$.

Remark

- X_t is a Gaussian process if μ_t is deterministic. Apparently, the mean function is $\mathbb{E}\left[X_t\right] = x + \int_0^t \mu_s \, ds$ and the covariance function $\gamma(t, s) = \operatorname{cov}(X_t, X_s) = \min\{t, s\}$.
- We can always transform a Brownian motion with drift into a standard Brownian motion by change of the underlying probability measure so long as the drift μ_t satisfies certain conditions, say, bounded.

Markov process

A stochastic process X_t defined on a filtered probability space $(\Omega, \mathcal{F}_t, \mathbb{P})$ is a *Markov process* if, for $s > t$,

$$\mathbb{P}[X_s \in A | \mathcal{F}_t] = \mathbb{P}[X_s \in A | X_t] \quad \forall \text{ Borel measurable set } A.$$

In words, to determine the probability of the process X in the set A at time s given the information up to time t, it suffices to condition on the position of the process at time t. The Markov property can also be written in terms of conditional expectations as

$$\mathbb{E}\left[f(X_s) | \mathcal{F}_t\right] = \mathbb{E}\left[f(X_s) | X_t\right]$$

for all bounded measurable function f.

Typical examples of Markov processes

- Random walk (in discrete time)
- Brownian motion
- Brownian motion with drift
- Poisson process

Technical note: Poisson process

N_t is a called a *Poisson process* with intensity λ if

- $N_0 = 0$ almost surely
- Given $0 < t_1 < t_2 < \cdots < t_n$, the increments N_{t_1}, $N_{t_2} - N_{t_1}, \cdots, N_{t_n} - N_{t_{n-1}}$ are independent
- For $i = 1, 2, \cdots, n$, the random variable of increment $N_{t_i} - N_{t_{i-1}}$ is Poisson distributed with parameter $\lambda(t_i - t_{i-1})$.

λ is regarded as the expected number of jumps per unit time because $\mathbb{E}\left[N_{t_i} - N_{t_i - 1}\right] = \lambda(t_i - t_{i-1})$.

Remark

- We will assume the sample paths of Poisson process is right continuous.
- $\mathrm{Var}(N_t - N_s) = \lambda(t - s)$
- $\mathbb{E}[e^{i\xi N_t}] = e^{\lambda t(e^{i\xi} - 1)}$

Transition probability and transition density

A Markov process is fully determined by an initial distribution and the transition probability (or density if there exists any). The transition probability $p(s, A|t, x)$ of a Markov process is defined by

$$p(s, A|t, x) = \mathbb{P}\left[X_s \in A | X_t = x\right]$$

where $s > t$ and $A \subset \mathbb{R}$ is measurable. If the transition probability p is absolutely continuous with respect to the Lesbegue measure, we refer to the density as the *transition density* and denote it by, abusing the use of notation, $p(s, y|t, x)$. Intuitively,

$$p(s, y|t, x) = \mathbb{P}\left[X_s = y | X_t = x\right].$$

Ideally, we would expect

$$p(s, \mathbb{R}|t, x) = \int_{\mathbb{R}} p(s, y|t, x) dy = 1.$$

Chapman-Kolmogorov equation

The transition density p of a Markov process satisfies the *Chapman-Kolmogorov* equation.

For and $t < u < s$, we have

$$\int p(s, y|u, z)p(u, z|t, x)dz = p(s, y|t, x).$$

Intuitively, we can reconcile the Chapman-Kolmogorov equation as

$$\sum_z \mathbb{P}\left[X_s = y | X_u = z\right] \mathbb{P}\left[X_u = z | X_t = x\right] = \mathbb{P}\left[X_s = y | X_t = x\right]$$

Infinitesimal generator

The infinitesimal generator L of a Markov process X is defined by

$$Lf(x) = \lim_{h \to 0^+} \frac{1}{h}\{\mathbb{E}\left[f(X_h)\right] - f(x)\} = \frac{d}{dt}\mathbb{E}\left[f(X_t)\right]$$

Examples

- The infinitesimal generator of Brownian motion is the Laplacian, i.e., $\frac{1}{2}\frac{d^2}{dx^2}$. In general, the infinitesimal generator of diffusion process is a second order differential operator.
- The infinitesimal generator L for the Poisson process N_t with intensity λ is the difference operator $Lf(x) = \lambda[f(x+1) - f(x)]$.

Infinitesimal generator for Brownian motion

For Brownian motion, Thus,

Infinitesimal generator for Poisson process

For Poisson process, note that Thus,

Wiener integral

- Let f be a (deterministic) step function defined by $f = \sum_{i=1}^n a_i \mathbf{1}_{[t_{i-1}, t_i)}$, where $t_0 = a$ and $t_n = b$, $a_i \in \mathbb{R}$. The *Wiener integral* $I(f)$ of f is defined by

  $$I(f) = \int_a^b f(t)dB_t = \sum_{i=1}^n a_i \Delta B_{t_i}, \quad \Delta B_{t_i} = B_{t_i} - B_{t_{i-1}}.$$

- Let $f \in L^2[a, b]$ and f_n be a sequence of step functions such that $f_n \to f$ in $L^2[a, b]$. The Wiener integral $I(f)$ of f is defined by

  $$I(f) = \int_a^b f(t)dB_t = \lim_{n \to \infty} \int_a^b f_n(t)dB_t, \quad \text{in } L^2(\Omega).$$

A technical issue here: Is $I(f)$ well-defined?

Wiener integral is normally distributed

Theorem

For each $f \in L^2[a,b]$, the Wiener integral $\int_a^b f(t)dB_t$ is a Gaussian random variable with mean 0 and variance $\|f\|_2^2 = \int_a^b f^2(t)dt$. In short,

$$\int_a^b f(s)dB_s \sim N(0, \|f\|_2^2).$$

- In particular, recall that if the integrand f is the step function $f(t) = \sum_{i=1}^n a_i \mathbf{1}_{[t_{i-1},t_i)}(t)$, apparently the Wiener integral $I(f)$ of f is normally distributed

  $$I(f) = \sum_{i=1}^n a_i \Delta B_i \sim N\left(0, \sum_{i=1}^n a_i^2 \Delta t_i\right)$$

  since the ΔB_i's are independent normal random variables.
- The proof for general $f \in L^2[a,b]$ is based on limiting process.

Example

$\int_0^t s dB_s \sim N\left(0, \frac{t^3}{3}\right)$

Corollary

If $f, g \in L^2[a,b]$, then

$$\mathbb{E}[I(f)I(g)] = \mathbb{E}\left[\int_a^b f(s)dB_s \int_a^b g(s)dB_s\right] = \int_a^b f(t)g(t)dt.$$

Thus, the Wiener integral $I : L^2[a,b] \to L^2(\Omega)$ is an isometry. In particular, if f and g are orthogonal, i.e., $\int_a^b fg = 0$, then the Gaussian random variables $I(f)$ and $I(g)$ are independent.

Properties of Wiener integral

Let f and g be deterministic $L^2[a,b]$ functions, α and β are constants. Then

- $\int_a^b [\alpha f(t) + \beta g(t)]\, dB_t = \alpha \int_a^b f(t)dB_t + \beta \int_a^b g(t)dB_t$.
- $\int_a^b f(t)dB_t = \int_a^c f(t)dB_t + \int_c^b f(t)dB_t$, for $c \in [a,b]$.

Integration by parts formula

Let f be a continuous function of bounded variation. Then almost surely

$$\int_a^b f(t)dB_t = f(t)B_t|_a^b - \int_a^b B_t df(t).$$

- Note that the integral on the left hand side is in the sense of Wiener, whereas on the right hand side is in the sense of Riemann-Stietjes.

Example

Determine the distribution of the random variable $\int_0^1 B_t dt$.

Wiener integral defines a continuous martingale

Let $f \in L^2[a,b]$. Then the stochastic process M_t obtained through Wiener integral

$$M_t = \int_a^t f(s) dB_s, \quad a \leq t \leq b,$$

is a martingale with respect to $\mathcal{F}_s = \sigma(B_s; s \leq t)$.

- Don't be confused with a continuous time martingale and a continuous martinagle.
- Since Wiener integral defines a Gaussian process, continuity of the process can be obtained by applying Kolmogorov's continuity criterion.

Technical note: martingale conditions

To show if the process M_t is a martingale, we need to verify the three defining conditions

- M_t is adapted
- M_t is integrable for every t
- For every $s < t$, $\mathbb{E}\left[M_t | \mathcal{F}_s\right] = M_s$ almost surely

An illustrative example for Ito integral

Let's start with defining the simple integral as

$$\int_0^t B_s dB_s.$$

As in the theory of Riemann-Stieltjes integral, we shall start with partitioning the interval $[0,t]$ into, say, n subintervals. Within each subinterval, we pick a point and evaluate the integrand at that point, multiply that value by the increment of the integrator in that subinterval. Then we sum up the results from each subinterval and take limit as the mesh of the partition approaches zero. Possible choices for selecting points from each subinterval may be, denoting $\Delta B_{t_k} = B_{t_k} - B_{t_{k-1}}$:

- The right point rule:
  $R_n = \sum_{k=1}^n B_{t_k} \Delta B_{t_k}$
- The left point rule:
  $L_n = \sum_{k=1}^n B_{t_{k-1}} \Delta B_{t_k}$
- The midpoint rule:
  $M_n = \sum_{k=1}^n B_{t_*} \Delta B_{t_k}, \quad \text{where } t_* = \dfrac{t_k + t_{k-1}}{2}$

Which rule rules?

Question: Which rule yields convergent integral? in what sense? We knew that it can't be pathwise because the integrator, in this case the Brownian motion, is not of finite variation (because it has nonzero second variation) almost surely.
Note that the following identities hold. Hence, Notice that the first term in both expressions is independent of partitions and the second term, as we have seen in previous lecture, will converge to the quadratic variation of Brownian motion in L^2 as the mesh approaches zero! Consequently,
So we learnt from this simple example that

- The right point rule and the left end point rule yield different "integrals".
- The difference between the "right integral" and the "left integral" is exactly the quadratic variation.
- The convergence is in L^2 sense.

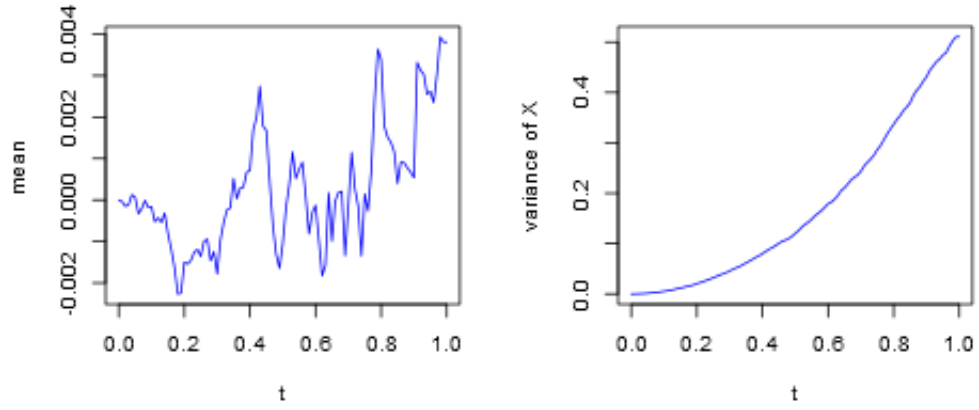
Remark

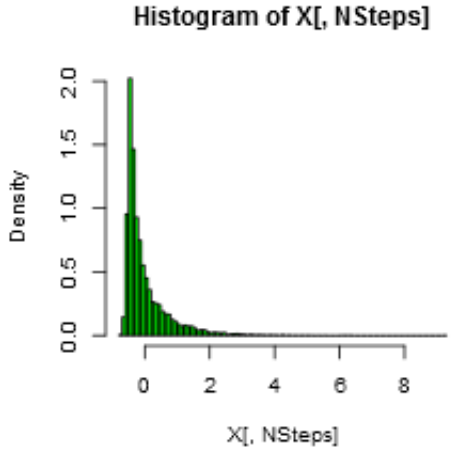
- We need to stick with one specific rule in order to have convergence.
- Ito picked the left end point rule because of adaptivity and martingality.
- L_n is a martingale whereas R_n isn't.
- The midpoint rule leads to the Stratonovich integral.

Simulation of stochastic integral: left endpoint rule

In [5]:
```
%%R
# Simulate the stochastic integral int f(B) dB from 0 to t, for t in [0,1]
# Demonstrate the effect of Ito correction/term

NSim <- 1e4   # number of samples at each step
NSteps <- 100   # number of steps
Tfin <- 1    # terminal time
dt <- Tfin/NSteps
f <- function(x) x
B <- matrix(0,NSim,NSteps+1)  # initialize the Brownians
dB <- matrix(0,NSim,NSteps+1)

# Simulate Browian paths
for (i in 1:NSteps){
  db <- rnorm(NSim)
  db <- db - mean(db)  # now db has mean 0
  db <- db/sd(db)  # now db has variance 1
  dB[,i] <- sqrt(dt)*db
  B[,i+1] <- B[,i] + dB[,i]
}
```

```
# Calculate the (discretized) stochastic integral using left endpoints
dX <- f(B)*dB
X <- t(apply(dX,1,FUN=cumsum))
t <- (0:NSteps)*dt

# plots
par(mfrow=c(2,2))

# evolution of mean
plot(t,colMeans(X),type='l',col='blue',ylab='mean')

# evolution of variance
plot(t,apply(X,2,FUN=var),type='l',col='blue',ylab='variance of X')

# histogram at terminal time
hist(X[,NSteps],prob=T,col='green',breaks=100)
```



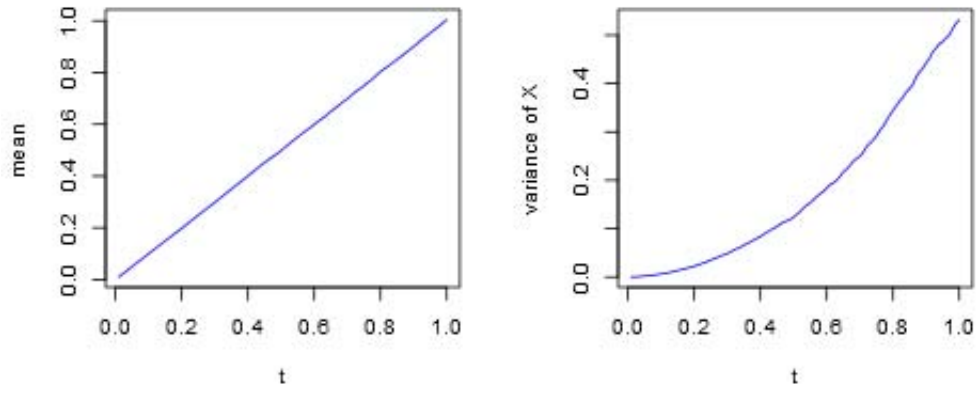


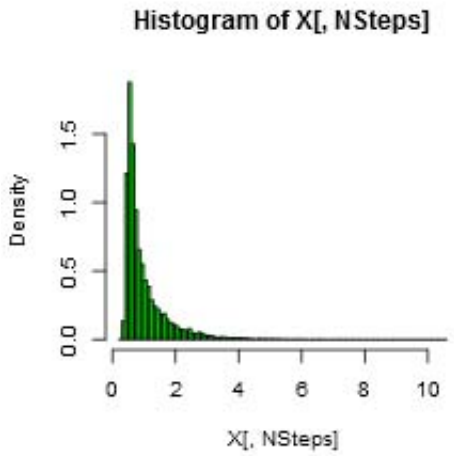
Simulation of stochastic integral: right endpoint rule

In [6]:

```
%%R
```

```
# What happens if we use right endpoints?

# Calculate the (discretized) stochastic integral using right endpoints
B <- B[,-1]
dB <- dB[,1:NSteps]
dX <- f(B)*dB
X <- t(apply(dX,1,FUN=cumsum))
t <- (1:NSteps)/NSteps*Tfin

# plots
par(mfrow=c(2,2))

# evolution of mean
plot(t,colMeans(X),type='l',col='blue',ylab='mean')

# evolution of variance
plot(t,apply(X,2,FUN=var),type='l',col='blue',ylab='variance of X')

# histogram at terminal time
hist(X[,NSteps],prob=T,col='green',breaks=100)
```





伊藤 清 先生 (Itô, Kiyoshi sensei)

Courtesy: Photo from Wikipedia

Ito integral of simple processes

Definition (simple process)

A process φ_t is called *simple* if it is of the form

$$\varphi_t(\omega) = \sum_{k=1}^{n} \xi_{k-1}(\omega)\mathbf{1}_{[t_{k-1},t_k)}(t),$$

where $\xi_k \in \mathcal{F}_{t_k}$ for $k = 0, \cdots, n$.

- Basically, a simple process is simply a step function with random coefficients that are measurable with respect to the left endpoints.
- Simple process is defined as such for mimicking a) the step functions in the Wiener integral and b) the left endpoint rule in the Riemann integral.
- The left endpoint measurability is key to the martingality of Ito integral, as we expected.

Definition (Ito integral of a simple process)

The stochastic integral of a simple process φ_t with respect to Brownian motion B_t over $[0, T]$ is defined by

$$\int_0^T \varphi_t dB_t = \sum_{k=1}^{n} \xi_{k-1}\Delta B_{t_k}, \qquad \text{where } \Delta B_{t_k} = B_{t_k} - B_{t_{k-1}}.$$

Integrand for Ito integral

Definition

We will use $L^2_{ad}(\Omega \times [a,b])$ to denote the space of all stochastic processes $\varphi_t(\omega)$, $a \leq t \leq b$, satisfying

- φ_t is adapted to the filtration $\mathcal{F}_t$.
- $\int_a^b \mathbb{E}|\varphi_t|^2 dt < \infty$.

In other words, φ_t is adatped and in $L^2(\Omega \times [a,b])$.
Note that, for notational simplicity, we usually omit the reference to the sample space Ω and deonte the space as $L^2_{ad}[a,b]$.

Lemma

Any L^2_{ad} process is the L^2 limit of a sequence of simple processes. Precisely, let $\varphi_t \in L^2_{\text{ad}}([a,b])$. Then there exists a sequence of simple processes $\{\varphi_t^{(n)}\}$ in $L^2_{\text{ad}}([a,b])$ such that

$$\lim_{n\to\infty} \mathbb{E}\int_a^b |\varphi_t^{(n)} - \varphi_t|^2 dt = 0.$$

In other words, $\varphi_t^{(n)} \to \varphi_t$ in $L^2_{\text{ad}}([a,b])$.

Properties of Ito integral

Let $\varphi \in L^2_{ad}$ and, for $t \in [0,T]$,

$$X_t = \int_0^t \varphi_s dB_s$$

be the stochastic integral of φ_t with respect to Brownian motion B_t up to time t. Then X_t has the following properties.

- *Adaptivity*: $X_t \in \mathcal{F}_t$ for each t.
- *Martingality*: X_t is a martingale.
- *Ito isometry*:
  $$\mathbb{E}[X_t^2] = \mathbb{E}\int_0^t \varphi_s^2 ds \iff \mathbb{E}\left[\int_0^t \varphi_s dB_s\right]^2 = \int_0^t \mathbb{E}\left[\varphi_s^2\right] ds.$$
- *Quadratic variation*: $[X]_t = \int_0^t \varphi_s^2 ds \iff d[X]_t = \varphi_t^2 dt$.
- *Continuity*: X_t is continuous in t a.s.
- *Linearity*: Let ψ_t be another adapted process with $\mathbb{E}\int_0^T \psi_t^2 dt < \infty$ and $\alpha, \beta \in \mathbb{R}$. Then
  $$\int_0^t (\alpha\varphi_s + \beta\psi_s) dB_s = \alpha\int_0^t \varphi_s dB_s + \beta\int_0^t \psi_s dB_s$$

Ito processes

Definition

An adapted process X_t is called an *Ito process* if it is of the form

$$X_t = x + \int_0^t \sigma_s dB_s + \int_0^t b_s ds,$$

where σ_t and b_s are adapted processes. We also write it in differential form as

$$dX_t = \sigma_t dB_t + b_t dt, \quad \text{with initial condition } X_0 = x.$$

The coefficient b_t is referred to as the *drift* (term) and σ_t as the *diffusion* (term) of the Ito process X_t.

Lemma

The quadratic variation of the Ito process X_t is

$$[X]_t = \int_0^t \sigma_s^2 ds \quad \longleftrightsquigarrow \quad d[X]_t = \sigma_t^2 dt.$$

Note that

- though the quadratic variation of Brownian motion is deterministic (recall $[B]_t = t$ a.s.), in general the quadratic variation of an Ito process is stochastic;
- the proof is very similar to that of Brownian and is left as an exercise.

Ito integral with respect to Ito processes

Let X_t be an Ito process with drift b_t and diffusion σ_t, i.e., X_t is defined by

$$X_t = x + \int_0^t \sigma_s dB_s + \int_0^t b_s ds \quad \longleftrightsquigarrow \quad dX_t = \sigma_t dB_t + b_t dt,\ X_0 = x.$$

and φ_t be an adapted process. We define the stochastic integral of φ_t with respect to X_t as

$$\int_0^t \varphi_s dX_s = \int_0^t \varphi_s \sigma_s dB_s + \int_0^t \varphi_s b_s ds$$

provided the integrals on the right hand side are defined.

Ito's formula for Brownian motion

Theorem

Let $f(t,x)$ be a function with continuous partial derivatives f_t, f_x, and f_{xx}. Let $W_t = W_0 + B_t$ be a Brownian motion starting at W_0 (nonrandom). Then, for every $T \geq 0$,

$$f(T, W_T) - f(0, W_0) = \int_0^T f_x(t, W_t) dW_t + \int_0^T \left[f_t(t, W_t) + \frac{1}{2} f_{xx}(t, W_t) \right] dt.$$

Or equivalently in differential form

$$df(t, W_t) = f_x(t, W_t) dW_t + \left[f_t(t, W_t) + \frac{1}{2} f_{xx}(t, W_t) \right] dt.$$

Remark

The idea is that we Taylor expand $f(t, W_t)$ to second order then formly apply the following rule:

$$(dB_t)^2 \rightsquigarrow dt, \quad (dt)^2 \rightsquigarrow 0, \quad dB_t dt \rightsquigarrow 0.$$

Ito's formula for Ito process

Let $f(t,x)$ be a function with continuous partial derivatives f_t, f_x, and f_{xx}. Then for every $T \geq 0$, Or in differential form

Remark

- Ito's formula natually decomposes $f(t, X_t)$ into a drift/finite variation part plus a diffusion/martingale part; reminiscent of the Doob decomposition. Processes consist of a finite variation part and a martingale part are also referred to as *semimartingales*.
- Note that the second order differential operator $\frac{\sigma_t^2}{2}\partial_x^2 + b_t \partial_x$ in the drift part is the infinitesimal generator of the process X_t.

Review: Fundamental theorem of calculus and Taylor's theorem

Fundamental theorem of calculus

Let f be a continuously differentiable function. Then
$f(y) - f(x) = \int_x^y f'(\xi) d\xi.$

Taylor's theorem

For a second differentiable function f, there exists some ξ between y and x such that
Notice that the equation is exact, however in general it is not possible to specify what ξ is.

Applications of Ito's formula I: Evaluating stochastic integral

In classical calculus, we barely evaluate an integral from the definitioin *per se*, i.e., partition the integrating interval, form Riemann sum, then take limit as the mesh of the partition approaches zero. Instead, we evaluate an integral by applying the Fundamental Theorem of Calculus. Though in stochastic calculus the Fundamental Theorem of Calculus does not really exist, we evaluate stochastic integrals by applying Ito's formula.

Theorem

$$\int_a^b f(B_t) dB_t = F(B_t)|_{t=a}^b - \frac{1}{2}\int_a^b f'(B_t) dt,$$

where F is an antiderivative of f, i.e., $F' = f$.

Theorem

$$\int_a^b f(t, B_t) dB_t = F(t, B_t)|_{t=a}^b - \int_a^b \left[F_t(t, B_t) + \frac{1}{2} f_x(t, B_t)\right] dt,$$

where $F_x = f$, i.e., F is an antiderivative of f with respect to x.

- The idea is, we find an antiderivative of f (with respect to x), say, F; apply Ito's formula to F, then rearrange terms.
- However, the price we pay is that in general the last (Riemann) integral on the right hand side usually has no simple analytical expression.

Examples of stochastic integral evaluation

- Evaluate the stochastic integral $\int_0^t B_s dB_s$.
  Note that in this case $f(x) = x$. Hence an antidrivative of f is $F(x) = \frac{x^2}{2}$.
  Apply Ito's formula to F we have
- Evaluate the stochastic integral $\int_0^t s e^{B_s} dB_s$.
  Note that in this case $f(t, x) = te^x$. Hence an antidrivative F of f is $F(t, x) = te^x$.
  Apply Ito's formula to F we have

Applications of Ito's formula II: Solving SDEs

A stochastic differenital equation (SDE) is an differential equation with random noise of the form

$$dX_t = \mu(X_t, t)dt + \sigma(X_t, t)dB_t$$

In cases, we can solve SDEs by applying Ito's formula to certain function of X_t. We demonstrate the technique by solving the following two very important examples.

- *Geometric Brownian motion/Black-Scholes model*

  $dX_t = \mu X_t dt + \sigma X_t dB_t$, where μ and σ are constants.

  To solve it, we apply Ito's formula to $\log X_t$.
- *Ornstein-Uhlenbeck process/Vasicek model*

  $dX_t = \lambda(m - X_t)dt + \sigma dB_t$, where m, λ, and σ are constants.

  To solve it, we apply Ito's formula to $e^{\lambda t} X_t$.

Remark

Geometric Brownian motion and Ornstein-Uhlenbeck process are special cases of the so called *linear SDEs* which has the general form

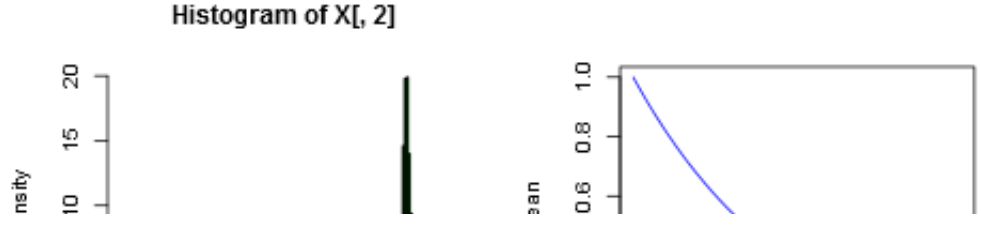
$$dX_t = (\mu_1 X_t + \mu_0)dt + (\sigma_1 X_t + \sigma_0)dB_t.$$

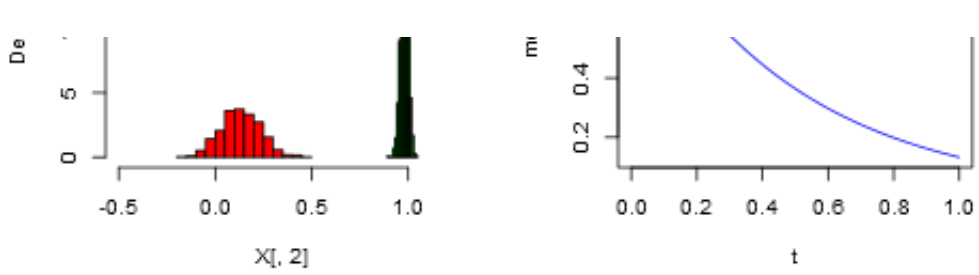
Such SDEs have "closed form" solutions.

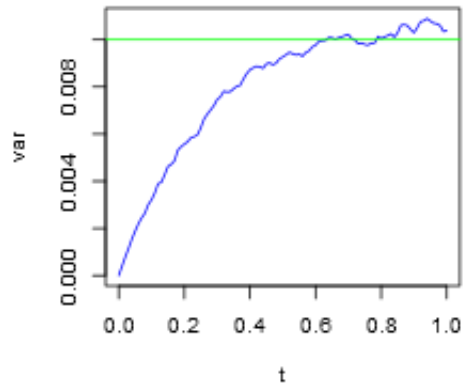
Simulation of the Ornstein-Uhlenbeck process

In [7]:
```
%%R
```

```
# The code simulates the OU process dX = lambda (m - X) dt + sigma dB
# by Euler-Maruyama scheme

# Parameters of the OU process
lambda <- 2
m <- 0
sigma <- 0.2
x0 <- 1
Tfin <- 1

# number of paths and number of time steps
NSim <- 1e3
NSteps <- 100

# initialize X
X <- matrix(0,NSim,NSteps+1)
X[,1] <- x0

# Euler-Maruyama scheme
dt <- Tfin/NSteps
for (i in 1:NSteps){
  db <- rnorm(NSim)
  db <- db - mean(db)  # now db has mean 0
  db <- db/sd(db)  # now db has variance 1

  X[,i+1] <- X[,i] + lambda*(m - X[,i])*dt + sigma*sqrt(dt)*db
}

par(mfrow=c(2,2))
# histograms at different times
hist(X[,2],prob=T,col='green',breaks=20,xlim=c(-.5,1.2))
hist(X[,NSteps+1],prob=T,col='red',breaks=20,add=T)

# time evolution of mean
t <- (0:NSteps)/NSteps*Tfin
plot(t,colMeans(X),type='l',col='blue',ylab='mean')

# time evolution of variance
t <- (0:NSteps)/NSteps*Tfin
plot(t,apply(X,2,FUN=var),type='l',col='blue',ylab='var')
abline(h=sigma^2/2/lambda,col='green')
```







Stochastic differential equation

Let $(\Omega, \mathcal{F}_t, \mathbb{P})$ be a filtered probability space and W_t a Brownian motion defined on it. A stochastic differential equation (SDE) driven by the Brownian motion W_t is an equation of the form

$$dX_t = \mu(X_t, t)dt + \sigma(X_t, t)dW_t, \quad X_0 = x$$

or in integral form

$$X_t = x_0 + \int_0^t \mu(X_s, s)ds + \int_0^t \sigma(X_s, s)dW_s.$$

As usual, $\mu(x, t)$ is referred to as the drift part and $\sigma(x, t)$ the diffusion part.

Connection to partial differential equation

Stochastic differential equation provides a way to numerically solve second order parabolic partial differential equations by Monte Carlo simulation. The key point is a stochastic representation of the solution to partial differential equations which we develop in the following.

Let X_t be the diffusion process driven by

$$dX_t = \mu(X_t, t)dt + \sigma(X_t, t)dW_t.$$

We shall suppress the dependence on x, t of μ and σ for notational simplicity.

Theorem

Let $u = u(x, t)$ be the solution to the terminal value problem

Then u has the representation

$$u(x,t) = \mathbb{E}_{t,x}\left[h(X_T)\right],$$

where $\mathbb{E}_{t,x}[\cdot]$ denotes the conditional expectation $\mathbb{E}[\cdot|X_t = x]$.

Proof

Applying Ito's formula to $u(X_t, t)$ yields

since u satisfies the PDE

$$u_t + \frac{\sigma^2}{2}u_{xx} + \mu u_x = 0.$$

Thus taking conditional expectation $\mathbb{E}_{t,x}[\cdot]$ on both sides and taking into account the terminal condition $u(x,T) = h(x)$ we end up with

$$u(x,t) = \mathbb{E}_{t,x}[h(X_T)].$$

The Feynman-Kac formula

Let $u = u(x,t)$ be the solution to the terminal value problem

Then u has the representation

$$u(x,t) = \mathbb{E}_{t,x}\left[e^{-\int_t^T V(X_s,s)ds}\, h(X_T)\right],$$

where $\mathbb{E}_{t,x}[\cdot]$ denotes the conditional expectation $\mathbb{E}[\cdot|X_t = x]$. This is the celebrated *Feynman-Kac formula*.

Proof of Feynman-Kac formula

Applying Ito's formula to $u(X_t,t)e^{-\int_0^t V(X_s,s)ds}$ yields

since u satisfies the PDE

$$u_t + \frac{\sigma^2}{2}u_{xx} + \mu u_x = V(x,t)u.$$

In integral form

therefore, by dividing on both sides the term $e^{-\int_0^t V(X_s,s)ds}$ we have

Richard Phillips Feynman

Courtesy: Photo from Wikipedia

Marek Kac

Courtesy: Photo from Wikipedia
Quoted from Wikipedia:

- His question, "Can one hear the shape of a drum?" set off research into spectral theory, with the idea of understanding the extent to which the spectrum allows one to read back the geometry.
- When Kac and Richard Feynman were both on the Cornell faculty he went to a lecture of Feynman's and saw that the two of them were working on the same thing from different directions. The Feynman-Kac formula resulted, which proves rigorously the real case of Feynman's path integrals. The complex case, which occurs when a particle's spin is included, is still unproven.

Adding nonhomogeneous term

Let $u = u(x,t)$ be the solution to the terminal value problem

Then u has the representation

$$u(x,t) = \mathbb{E}_{t,x}\left[h(X_T) - \int_t^T f(X_\tau,\tau)d\tau\right],$$

where $\mathbb{E}_{t,x}[\cdot]$ denotes the conditional expectation $\mathbb{E}[\cdot|X_t = x]$.

Backward second order parabolic PDEs

Finally, we have the stochastic representation for any backward second order parabolic linear PDE with terminal condition as follows.

Theorem

Let $u = u(x,t)$ be the solution to the terminal value problem

Then u has the representation

$$u(x,t) = \mathbb{E}_{t,x}\left[e^{-\int_t^T V(X_s,s)ds}h(X_T) - \int_t^T e^{-\int_t^\tau V(X_s,s)ds}f(X_\tau,\tau)d\tau\right],$$

where $\mathbb{E}_{t,x}[\cdot] = \mathbb{E}[\cdot|X_t = x]$ is the conditional expectation.

Optimal control theory

Quotes from Wikipedia:

"Optimal control theory, an extension of the calculus of variations, is a mathematical optimization method for deriving control policies. The method is largely due to the work of Lev Pontryagin and his collaborators in the Soviet Union and Richard Bellman in the United States."

"Optimal control deals with the problem of finding a control law for a given system such that a certain optimality criterion is achieved. A control problem includes a cost functional that is a function of state and control variables. An optimal control is a set of differential equations describing the paths of the control variables that minimize the cost functional. The optimal control can be derived using Pontryagin's maximum principle (a necessary condition also known as Pontryagin's minimum principle or simply Pontryagin's Principle), or by solving the Hamilton–Jacobi–Bellman equation (a sufficient condition)."

Richard Ernest Bellman

Courtesy: Photos from Wikipedia

Лев Семёнович Понтря́гин

Courtesy: Photo from Wikipedia

Stochastic control problem

A stochastic control problem is a control problem which aims to minimize certain expected costs among all admissible controls.
Specifically, consider

$$\min_{v \in \mathcal{G}[0,T]} \mathbb{E}\left[g(X_T^{(v)}) + \int_0^T h(s, X_s^{(v)}, v_s) ds\right]$$

where

- g is referred to as the *terminal cost* and h is the *running cost*

- the state variable $X_t^{(v)}$ is driven by the controlled SDE

  $dX_t^{(v)} = \mu(t, X_t^{(v)}, v_t)dt + \sigma(t, X_t^{(v)}, v_t)dW_t.$
- $\mathcal{G}[0, T]$ is the collection of admissible controls in the time interval $[0, T]$

Bellman's principle of optimality

"An optimal policy has the property that whatever the initial state and initial decision are, the remaining decisions must constitute an optimal policy with regard to the state resulting from the first decision."

(See Bellman, 1957, Chap. III.3.)

Value function

- For a given admissible control $v \in \mathcal{G}[t, T]$, define the expected objective functional $J^{(v)}$ as

  $$J^{(v)}(t, x) = \mathbb{E}\left[g(X_T^{(v)}) + \int_t^T h(s, X_s^{(v)}, v_s)ds \middle| X_t = x\right].$$
- The value function $J(t, x)$ for a stochastic control problem is defined as

  $J(t, x) = \min_{v \in \mathcal{G}[t,T]} J^{(v)}(t, x).$
- The value function J at (t, x) is the optimal value of the control problem conditioned on the process starting at (t, x) and applying the optimal control thereafter.

Bellman's principle again

The Bellman's principle can be rephrased in terms of value function as follow. For any $0 < \epsilon < T - t$,

$$J(t, x) = \min_{v \in \mathcal{G}[t,t+\epsilon]} \mathbb{E}\left[\int_t^{t+\epsilon} h(s, X_s^{(v)}, v_s)ds + J(t + \epsilon, X_{t+\epsilon}^{(v)}) \middle| \mathcal{F}_t\right].$$

The Hamilton-Jacobi-Bellman (HJB) equation

The value function J satisfies the terminal value problem

$$\partial_t J(t, x) + \min_{v \in \mathcal{G}[t]}\{\mathcal{L}^{(v)} J(t, x) + h(t, x, v)\} = 0, \text{ for } t < T$$

with terminal condition

$$J(T, x) = g(x),$$

where $\mathcal{L}^{(v)} := \frac{\sigma^2}{2}\partial_x^2 + \mu\partial_x$ is the associated infinitesimal generator of the controlled process

$$dX_t^{(v)} = \mu(t, X_t^{(v)}, v_t)dt + \sigma(t, X_t^{(v)}, v_t)dW_t.$$

Note

- The HJB equation is bascially an infinitesimal version of the Bellman's principle.
- The optimal policy (control) is given implicitly in terms of the value function J.