Probabilities & Data
Week 10: Brownian Motion/Stocks

DR NICK FEWSTER-YOUNG

Topics and Concepts

- We look at simulating Normally distributed random walks.
- We define Brownian motion in terms of the normal distribution of the increments, the independence of the increments, the value at 0, and its continuity.
- The joint density function for the value of Brownian motion at several times is a multivariate normal distribution.
- We also look at simulating Brownian motion and applications to stock market pricing.
- Calculating probabilities of loss in markets over a time period.
- We also look at simulating insurance claims.

Motivation

- **Brownian motion** is the physical phenomenon named after the English botanist Robert Brown who discovered it in 1827. Brownian motion is the zig-zagging motion exhibited by a small particle, such as a grain of pollen, immersed in a liquid or a gas. Albert Einstein gave the first explanation of this phenomenon in 1905. He explained Brownian motion by assuming the immersed particle was constantly buffeted by the molecules of the surrounding medium. Since then the abstracted process has been used for modelling the stock market and in quantum mechanics.

Random Walks

- Remember that sums of independent Normally distributed random variables are still Normally Distributed.
- The overall mean being sum of the individual means and the overall variance is the sum of the individual variances.
- A **Random Walk** is known as a stochastic or random process that describes a path that consists of a succession of random steps on some real space.

Example

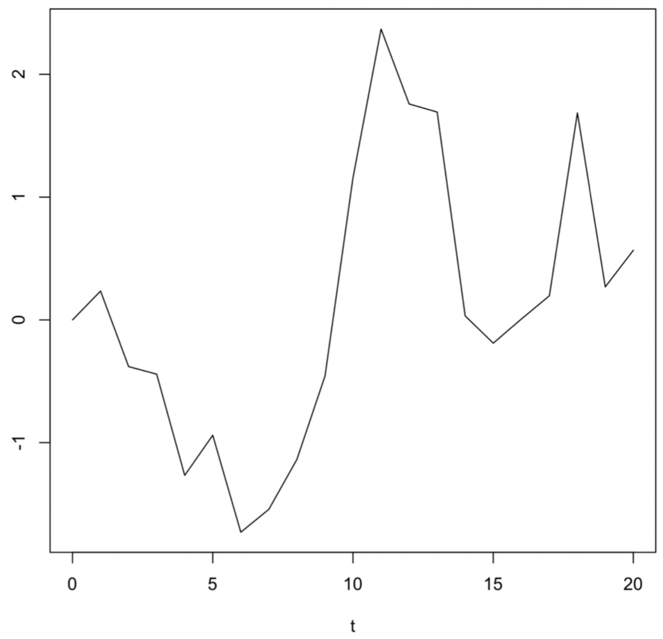
Example. Consider a drunk man walking on a straight line. He starts off at point 0. For each step that he takes, the length X_i of his ith stride is distributed as normal N(0,σ). We assume that each stride is independent of each other and a negative value means a step backwards.

- $Let\ \ S_t = \sum_{i=1}^{t} X_i$. From Week 8 lectures, we have that $S_t \sim N(0, \sigma_S)$ where $\sigma_S^2 = \sum \sigma^2 = \mathrm{t}\sigma^2$. Thus $S_t \sim N\left(0, \sigma\sqrt{t}\ \right)$.

- Let's simulate a drunk man. To simulate him, we start by plotting the position of the man on the line as a function of time when $X_i \sim N(0,1)$. Assume he takes 20 steps.

Example - Simulation

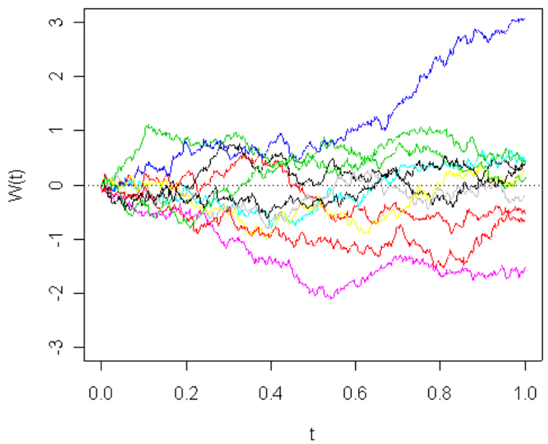
```
> t <- seq(1,20)
> x <- rnorm(20,0,1)
 > s <- cumsum(x)
> t <- c(0,t)
> s <- c(0,s)
> plot(t,s,"l")
```



The c(0,) commands add a 0 in front of both the t and the s vectors so that my plot starts from (0,0). Here is the plot.

Brownian Motion

- Is just a category of Random walks where each step is independent and identically distributed.
- In the general case, the size of the time intervals can be very small, in the ideal case, the time is continuous instead of discrete but of course, if we want to do plots, we have to discretise the time interval. Also the individual components need not be independent standard normal, they can be, $X_i \sim N(\mu, \sigma)$.



Formal Definition of Brownian Process

Definition. The Standard Process is a stochastic process B(t), for t ≥ 0, with the following properties:

1) Every increment B(t)−B(s) over an interval of length t−s is Normally distributed with mean 0 and variance t − s, that is B(t) − B(s) ~ N (0, t−s).
2) For every pair of disjoint time intervals $[t_1, t_2]$ and $[t_3, t_4]$ with $t_1 < t_2 < t_3 < t_4$ the increments $B(t_4) - B(t_3)$ and $B(t_2) - B(t_1)$. are independent random variables with distributions, and similarly for n disjoint time intervals where n is an arbitrary positive integer.
3) B(0)=0.
4) B(t) is continuous for all t.

- **Property 2:** Can we written as

$$P(B(t) \geq x \mid B(t_0) = x_0, B(t_1) = x_1, \cdots, B(t_n) = x_n) = P(B(t) \geq x \mid B(t_n) = x_n)$$

Brownian Motion

Standard Brownian motion is denoted by B_t or W_t in the literature. If we slice the trajectory of the Brownian motion at time t, then we denote it by $B_t \sim N(0, \sqrt{t}\,)$.

Suppose we want to plot a Brownian motion from the start, $t = 0$ for 20 steps. Let the interval length be 1. That is

$$B_t = \textstyle\sum_{i=1}^{20} Z_i$$

where $Z_i \sim N(0, 1)$. However, if we wanted to have a more continuous and finer looking plot then we would discretise the time interval further. We could do this by subdividing the interval into 10 subintervals, then we can use

$$B_t = \sum_{i=1}^{200} Z_i$$

where $Z_i \sim N(0, \sqrt{0.1})$.

Let's simulate some Brownian motion now.

Steps to Simulate

Simulate a sample path of the process as follows.

1. Divide the interval [0,T] into a grid

$$0 = t_0 < t_1 < t_2 < \cdots < t_n = T.$$

Set i = 1 and $B(0) = B(t_0) = 0$ and iterate the following algorithm.

2. Generate a new random number z from the standard normal distribution.

3. Set i to i + 1.

4. Set $\mathrm{B}(t_i) = B(t_{i-1}) + z\sqrt{t_i - t_{i-1}}$

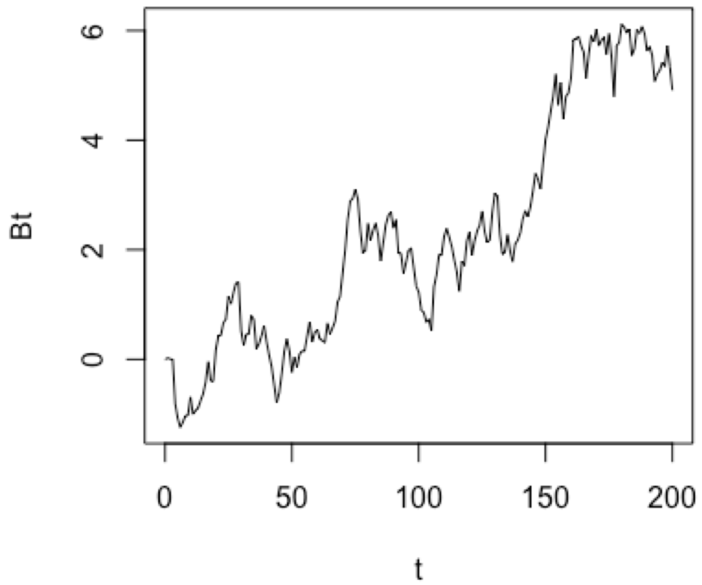
5. If i < N, iterate from step 1.

This method of approximation is valid only on the points of the grid. In between any two points t_i *and* t_{i-1} , the Brownian process is approximated by linear interpolation.

In R – Brownian Simulation

By using **R** we can simulate:

```
> t <- seq(1,200)
> x <- rnorm(200,0,sqrt(0.1))
 > t <- c(0,t)
> x <- c(0,x)
> Bt <- cumsum(x)
> plot(t,Bt,"l")
```



Here is the plot. TRY THIS YOURSELF!

Stock Prices

From the previous plot, it looks like it could be used to model daily changes in the stock prices, that is, the daily change could be Normally distributed, and so we can get both positive and negative changes, which appears to be what is observed in the markets.

However, there is a flaw in this model in that a very large negative change (possible under the normal distribution) will push the stock price into negative territory, which is realistically not at all possible.

One way to get around this flaw is to look at the log-returns. This is due to the work by Paul Samuelson (around 1964) and Fisher Black and Myron Scholes (1972), and also by Robert Merton.

Stock Prices

Let S_t be the stock price at time t and S_0 be the price at time 0. Then the log-return is

$$\ln(S_t) - \ln(S_0).$$

If looking at the growth of the stock price from initial time zero.

If we are concerned with day to day changes then

$$\ln(S_t) - \ln(S_{t-1}).$$

However, we can write $\ln(S_t) - \ln(S_0)$ as the following:

$$\ln(S_t) - \ln(S_{t-1}) + \ln(S_{t-1}) \cdots + \ln(S_1) - \ln(S_0).$$

Stock Prices (cont.)

So all we need is to assume that each

$$\ln(S_i) - \ln(S_{i-1}) \text{ for } i = 1,2,\cdots,t$$

is independent and identically distributed, and in the Samuelson, Black- Scholes-Merton models, they are Normally distributed with $N(\mu, \sigma)$. In finance terminology, μ is called the drift of the stock and σ the volatility, but in probability, μ is simply the mean and σ is the standard deviation and μ is just the mean rate of return.

Hence, the model can be written as

$$X_i = \ln\left(\frac{S_i}{S_{i-1}}\right) \sim N(\mu, \sigma).$$

$$\ln\left(\frac{S_t}{S_0}\right) = \sum_i^t X_i = \mu t + \sigma \sum_i^t Z_i = \mu t + \sigma B_t$$

Brownian Motion

We could rearrange to obtain an expression for the Stock Price in terms of Brownian Motion.

$$S_t = S_0 e^{\mu t + \sigma B_t}$$

Note that the right hand side is just a combination of constants and B_t is the Brownian Motion.

Let’s now simulate some stock prices.

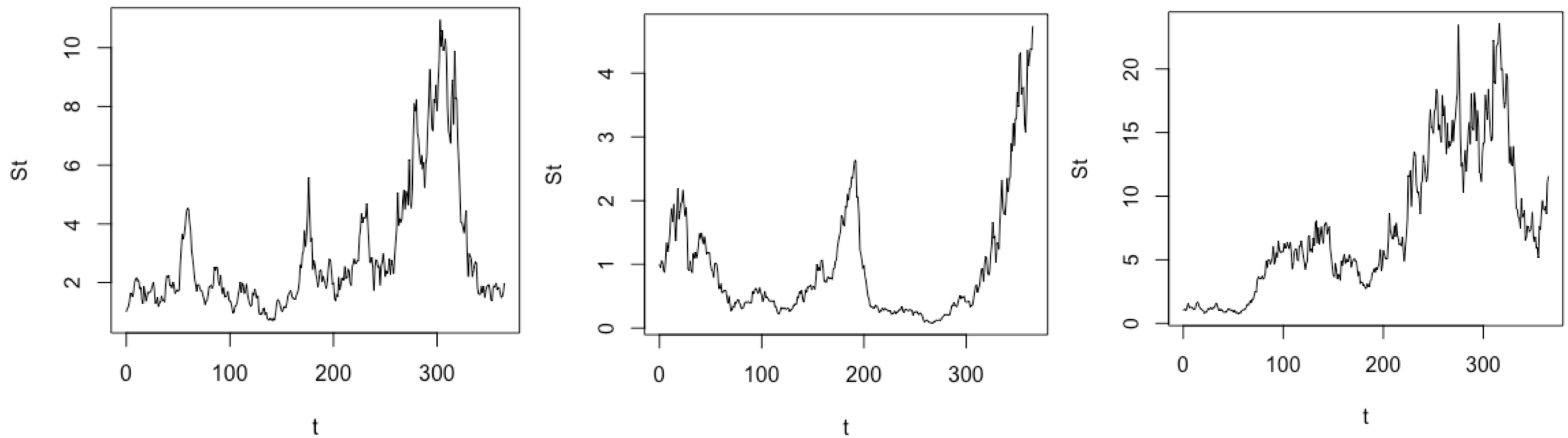
Example (Stock)

Example. Suppose the daily log returns of a stock is distributed as normal with annual drift 0.1 and annual volatility 50, then we can simulate a possible daily stock price movement (with S(0) = 1) as follows after first adjusting the drift and volatility to their respective daily figures.

```
> N <- 365
> mu <- 0.1/N;  sigma <- 50/N
> t <- seq(0,N)
> logreturns <- cumsum(rnorm(N,mu,sigma))
> S0 <- 1
> logreturns <- c(0,logreturns)
> St <- S0*exp(logreturns)
> plot(t,St,"l")
```

This is what I get for my simulated stock prices. We can see that the general upward trend is exponential.

Different Stock Price Paths



Random Poisson Motion

A Random Poisson motion is a distributed random variable which is made up of N, X_i independently, and identically distributed according to a particular distribution which are distributed as a Poisson Distribution. It takes the following form:

$$S_N := \sum_{i=1}^{N} X_i .$$

In previous weeks and for Brownian motion, we saw the sum of random variables from the same distribution results in being closely related or the same type of distribution!

However, what does the sum of Poisson distribution result in?

Examples

1. Suppose X_i are independently and identically distributed Bernoulli variables, $Bern(p)$, then the sum, $S_n = \sum_i^n X_i$ is Binomial, $Bin(n, p)$.
2. Suppose X_i are independently and identically distributed Normal variables, $N(0,1)$, then the sum, $S_n = \sum_i^n X_i$ is Normal, $\mathrm{N}(0, \sqrt{n})$.
3. Suppose X_i are independently and identically distributed Poisson variables, $Poisson(\lambda)$, then the sum, $S_n = \sum_i^n X_i$ is Gamma, $\mathrm{Gamma}(n, \lambda)$.

Applications

If we have S_N where the number of terms in the sum is random and distributed as **Poisson**, we can deduce the distribution of S_N given $N = n$.

In fact, S_N given $N = n$ is simply the sum of the first n of the X_i for each fixed n.

- The compound Poisson distributed random variable S_N is useful for modelling total amount of customer spent and insurance claims.

Example

Example. The number of customers N that arrive at a shop within a 1 hour period is distributed as Poisson with λ = 6. Suppose the amount that the ith customer spends is independent of each other and distributed as Exponential Exp(0.01), that is, the mean amount spent is $\frac{1}{0.01} = 100,$ then the distribution of the total amount of the customer spent given that 5 customers have arrived within the hour is $S_N \mid N = 5$ and therefore

$$S_5 = \sum_{i}^{5} X_i$$

which is distributed by a Gamma distribution, Gamma(5,0.01).

- Use **R** to simulate number of customers entering the ship and the total amount spent.

Example

We can also use it to model insurance claim amounts. If N is distributed as Poisson with λ = 1000, this means that the mean number of claims for a particular type of policy coming in within any particular year is 1000.

The claim amount for the ith claim can be modelled by an Exponential random variable $Exponential(\lambda_e)$ where perhaps $\lambda_e = 0.002$.

So the average claim amount is $\frac{1}{\lambda_e} = 2000$. If $N = 900$ for that particular $\lambda_e = 0.002$ year, then

$$S_N \mid N = 900$$

is distributed as Gamma(900, 0.002).

We can simulate 1 year of claims as follows using **R**.

Simulation of Example

```
> N <- rpois(1,1000); N
[1] 1014
> X <- rgamma(N,1,0.002)
> mean(X)
[1] 519.5337
> X
[1] 373.3899968 471.7026487 1135.7789690 1202.6279642 85.2289290 [6] 1212.9139223
492.0690056 ........ (note I am not pasting all here since there are 1014)
> sum(X)
[1] 526807.1
```

Thus for my simulation run, I have 1014 claims, the mean of which is $519.5337, which is close to the actual theoretical value of $500. You can see the size of the first 7 claims. The total amount of the 1014 claims is $526,807.10.

Example (Insurance)

Suppose the number of claims for the insurance company distributed as Poisson(10), and assume that each claim amount X_i is distributed as $X_i = e^{Y_i}$ where $Y_i \sim N(4{,}2)$, then the total claim amount is actually

$$\sum_i^N X_i = \sum_i^N e^{Y_i}.$$

(For your information only, note that e^{Y_i} is called a log-normal distribution, and it is also used to get positive values for claim amounts.)

So now we need to simulate a value of N first, get it, then simulate N values of X_i to get the individual claim amounts for a particular year.

Simulate Claims

```
> n<-rpois(1,10)

> n
[1] 12
> y<-rnorm(n,4,2)

> X<-exp(y)

> X
[1] 3.319900 18.980818 1683.977213 244.994616 41.007877 21.275428 259.798199

[7] 67.391715 364.688535 3.796879 37.969903 16.764678

> sum(X)

[1] 2763.966
```

Simulation (cont.)

So here in my simulation run for the first year, I have 12 claims. Let us do it again and this is what I get.

```
> n<-rpois(1,10)

> n
[1] 7
> y<-rnorm(n,4,2)

> X<-exp(y)

> X
[1] 568.15834 275.47987 36.84227 8.22904 802.82309 44.55595 91.96100

> sum(X)
[1] 1828.05
```

Of course, the insurance company would need to do it over and over again in order to get a sensible average of the total amount of annual claims and its standard deviation and so on.

That’s Brownian Motion

- That is it for Week 10!

- ## Next week:

- Logistic Regression ☺

We will investigate defaulting on loans and probability of this happening and the odds.