



## CHAPTER 6 Simulation

### Related Readings ...

- 
 Loss Models: from Data to Decisions (2<sup>nd</sup> edition)  
 Chapter 17 (except Section 17.2.4)
- 
 Derivative Markets (2<sup>nd</sup> edition)  
 Section 18.6, Chapter 19



### Learning Objectives

Inverse transform, applications in actuarial science and derivatives pricing, variance reduction, jump-diffusion model for stock returns

In this chapter, we study simulation (模擬) and its use in actuarial science and derivatives pricing.

### 1 Simulating a Random Variable

Simulating a random variable means the activity of obtaining an observation of a random variable from a desired distribution. For example, to simulate a  $B(1, 0.5)$  random variable, we can flip a fair coin. If the coin lands on head, then the observation of the random variable is 1, and is 0 otherwise.

Obviously, coin tossing is too time consuming and cannot be done by a computer. Most computers, however, have built in random number generator (RNG) (亂數產生程式). For example, in EXCEL, the function `RAND()` is a RNG.

#### Random Number Generator

A random number generator is a program to generate an observation of a continuous  $U(0, 1)$  random variable.

Of course, everything in a computer is deterministic and the result of an RNG is not truly random (we call this pseudo random). But we shall treat the RNG as a black box that can generate sequence of independent  $U(0, 1)$  random variables.

Given that we have the ability to generate  $U(0, 1)$  using an RNG, how can we simulate random variables with an arbitrary probability distribution? The following is the one of the various methods.

### The Inverse Transformation Method

Let  $F$  be a distribution function and  $U \sim U(0, 1)$ . Then

$$X = F^{-1}(U)$$

follows  $F$ .

What is the meaning of  $F^{-1}$ ?

- (1) If  $F$  is a continuous distribution, then it is strictly increasing, and hence there is only one  $x$  such that  $F(x) = u$ . That unique  $x$  is  $F^{-1}(u)$ .
- (2) If  $F$  is a discrete distribution, then it is a right-continuous step function and there will be an interval  $(a, b]$  such that  $F(x) = u$  for  $x \in (a, b]$ . In this case,  $F^{-1}(u) = a$ .
- (3) If  $F$  is a mixed distribution, then one of (1) and (2) happens depending on the value of  $u$ . The value of  $F^{-1}(u)$  can be determined as in (1) or (2).

Note that since  $F$  is non-decreasing,  $F^{-1}$  is also non-decreasing, and hence a greater value of  $U$  gives rise to a greater value of  $X$ .

#### Example 1.1

By using the inverse transformation method, generate 5,000  $\text{Exp}(\theta = 2)$  random variables. Calculate the sample mean and sample variance of the 5,000 simulated results and compare with the theoretical mean and variance. Repeat for the  $\text{Pareto}(\alpha = 4.5, \theta = 2)$  distribution.

#### Example 1.2

Simulate values from a  $B(4, 0.6)$  distribution using the random numbers 0.3, 0.6875 and 0.95. Use the convention that low random numbers correspond to low values of the random variables.

$k$	0	1	2	3	4
$p_k$	0.0256	0.1536	0.3456	0.3456	0.1296
$F(k)$	0.0256				1

### Example 1.3

Suppose that

$$F(x) = \begin{cases} 0.5x & \text{for } 0 \leq x < 1 \\ 0.5 + 0.25x & \text{for } 1 \leq x < 2 \end{cases}$$

Determine the simulated values of  $x$  resulting from the uniform numbers 0.3, 0.6 and 0.9.

### Monte Carlo Simulation

One of the main uses of simulation is to obtain estimates of expectation that is hard to calculate theoretically. Suppose we want to evaluate

$$E[g(Y)] = \int_{-\infty}^{\infty} g(y)f(y)dy.$$

If  $g$  is complicated (e.g.  $Y$  is the price of an asset and  $g(Y)$  is the payoff of an exotic option), then sometimes it would be impossible to calculate  $E[g(Y)]$  using integration. One way to estimate  $E[g(Y)]$  is to use Monte Carlo simulation.

#### Monte Carlo Simulation Methodology

*Step 1:* Simulate  $r$  independent and identically distributed random variables  $Y_1, Y_2, \dots, Y_r$ .

*Step 2:* Calculate the corresponding  $g(Y_1), g(Y_2), \dots, g(Y_r)$ .

*Step 3:*  $\bar{g}(r) = \frac{g(Y_1) + g(Y_2) + \dots + g(Y_r)}{r}$  gives an estimate of  $E[g(Y)]$ ,

Based on the above, we can simulate anything functions of expectations. For example:

Quantity	$g(y)$
mean	$y$
$k^{\text{th}}$ moment	$y^k$
cdf $F_Y(t)$	$I(y \leq t)$
call option	$(y - K)_+$

Even quantiles and percentiles of a random variable can be simulated. For example, we can use the simulated 25% smoothed empirical estimate to estimate the first quantile.

Why does the approach above work?

First of all, before we really use the RNG to get the values of  $Y_i$ 's,  $\bar{g}(r)$  is a random variable. We call  $\bar{g}(r)$  an **estimator** of  $E[g(Y)]$ . Suppose that  $g(Y)$  has mean  $\mu$  and variance  $\sigma^2$ . Both  $\mu$  and  $\sigma^2$  are unknown and  $\mu$  is the quantity that we want to estimate. Observe that

### Moments of an Estimator

$$(1) E[\bar{g}(r)] = \frac{1}{r}(E[g(Y_1)] + \dots + E[g(Y_r)]) = \frac{1}{r}(rE[g(Y)]) = \mu$$

$$(2) \text{Var}[\bar{g}(r)] = \frac{1}{r^2}(\text{Var}[g(Y_1)] + \dots + \text{Var}[g(Y_r)]) = \frac{1}{r^2}(r\text{Var}[g(Y)]) = \frac{\sigma^2}{r}$$

Equation (1) above states that the mean of  $\bar{g}(r)$  is an **unbiased** estimator, while equation (2) states that the variance of the estimator decreases with  $r$ . Finally, by the weak law of large numbers,  $\bar{g}(r) \rightarrow \mu$  as  $r$  increases to infinity. Both this and (2) states that the accuracy of  $\bar{g}(r)$  improves with  $r$ . As we simulate more and more number of times, we can get a more and more accurate estimate of  $\mu$ .

### Number of Runs (Advanced Topic)

For the estimate to be close to  $\mu$ , we need a large  $r$ . But how large should  $r$  be? There are numerous ways to determine the number of runs needed.

If our criterion for accuracy is that the variance of the estimate is less than a tolerance level  $\varepsilon$ , then we would choose  $r$ , so that  $\sigma^2 / r$  is less than  $\varepsilon$ . However,  $\sigma^2$  is unknown. One way suggested in Klugman et al. to get around this is to estimate it using sample variance  $s^2(r)$  where

$$s^2(r) = \frac{1}{r-1} \sum_{k=1}^r [g(Y_k) - \bar{g}(r)]^2$$

and continue to simulate until  $\frac{s^2(r)}{r} \leq \varepsilon$ .

There are other criteria. The next two examples illustrate a tolerance level based on relative error:

### Example 1.4 [00 Exam C Nov #11]

You are planning a simulation to estimate the mean of a non-negative random variable. It is known that the population standard deviation is 20% larger than the population mean.

Use the central limit theorem to estimate the smallest number of trials needed so that you will be at least 95% confident that the simulated mean is within 5% of the population mean.

- (A) 944      (B) 1299      (C) 1559      (D) 1844      (E) 2213      (Ans: E)

### Example 1.5

Outline how you can use simulation to estimate  $F_X(1000)$  of a Pareto distribution with  $\alpha = 3$  and  $\theta = 1000$ . The simulation should be carried out until you are 95% confident that the answer is within  $\pm 1\%$  of the true value. Repeat for the mean and the first quantile of the distribution.

**Part I:  $F_X(1000)$ :** We let  $p = F_X(1000)$  to be the number that we want to obtain.

Simulate a small number (say, 30) of Pareto(3, 1000) random variables (call them  $X_i$ ).

– The estimated value of  $p = E[I(X \leq 1000)]$  is  $\bar{g}(r) = \frac{1}{r} \sum_{i=1}^r I(X_i \leq 1000)$ .

– The variance of the estimate is  $\text{Var}[I(X \leq 1000)] / r = p(1-p) / r$ .

$P(|\bar{g}(r) - p| \leq 0.01p) \geq 0.95$  is equivalent to  $P\left(\frac{|\bar{g}(r) - p|}{\sqrt{p(1-p)/r}} \leq \frac{0.01p}{\sqrt{p(1-p)/r}}\right) \geq 0.95$ .

For sufficiently large  $r$  (say,  $r \geq 30$ ), CLT says that  $\frac{0.01p}{\sqrt{p(1-p)/r}} > 1.96$ . But since  $p$  is unknown,

we continue to simulate and calculate  $\bar{g}(r)$  until  $\frac{0.01\bar{g}(r)}{\sqrt{\bar{g}(r)(1-\bar{g}(r))/r}} > 1.96$ , or equivalently

$r > \frac{196^2(1-\bar{g}(r))}{\bar{g}(r)}$ . By using a rough estimate of  $\bar{g}(r)$  based on the first 30 trials, we can get a rough idea of the order of magnitude of  $r$ .

**Part II: Population mean:** Let  $\mu = E(X)$ .

– The estimated value of  $\mu$  is  $\bar{g}(r) = \frac{1}{r} \sum_{i=1}^r X_i$ .

– The variance of the estimate is  $\sigma^2 / r$  where  $\sigma^2$  is the population variance.

We shall continue to simulate until  $\frac{0.01\mu}{\sigma\sqrt{r}} > 1.96$ . Plugging in the relevant estimates, we get the

criterion “stop until  $r > \frac{196^2 s^2(r)}{\bar{g}^2(r)}$ ”.

**Part III: First quantile  $Q_1$ :** Let  $X_{(1)}, X_{(2)}, \dots, X_{(r)}$  be the order statistics from the simulated sample, and recall in FIN4220 that the smoothed empirical estimate of the  $g$ -percentile is  $\hat{\pi}_g = (1-h)X_{(j)} + hX_{(j+1)}$  where  $j = \lfloor (r+1)g \rfloor$ ,  $h = (r+1)g - j$ . In our case  $g = 25\%$ .

Let  $a_+ = \lfloor rg + 0.5 - 1.96\sqrt{rg(1-g)} \rfloor$  and  $a_- = \lceil rg + 0.5 + 1.96\sqrt{rg(1-g)} \rceil$ . By CLT,  $P(X_{(a_-)} \leq Q_1 \leq X_{(a_+)}) \approx 0.95$  and hence the simulation stops when  $\max\left(\frac{\hat{\pi}_g - X_{(a_-)}}{\hat{\pi}_g}, \frac{X_{(a_+)} - \hat{\pi}_g}{\hat{\pi}_g}\right) \leq 0.01$ .

## 2 Actuarial Application of Simulation

The following examples show how simulation can be applied in solving actuarial problems.

### **?** Example 2.1 [00 Exam M May #32]

Insurance for a city's snow removal costs covers four winter months.

- (i) There is a deductible of 10,000 per month.
- (ii) The insurer assumes that the city's monthly costs are independent and normally distributed with mean 15,000 and standard deviation 2,000.
- (iii) To simulate four months of claim costs, the insurer uses the Inverse Transform Method (where small random numbers correspond to low costs).
- (iv) The four numbers drawn from the uniform distribution on  $[0, 1]$  are:

0.5398      0.1151      0.0013      0.7881

Calculate the insurer's simulated claim cost.

- (A) 13,400    (B) 14,400    (C) 17,800    (D) 20,000    (E) 26,600    (Ans: B)

### **?** Example 2.2 [04 Exam M Nov #5]

You are simulating the future lifetimes of newborns in a population.

- (i) For any given newborn, mortality follows De Moivre's law with maximum lifetime  $\Omega$ .
- (ii)  $\Omega$  has distribution function  $F(\omega) = (\omega/80)^2$ ,  $0 \leq \omega \leq 80$ .
- (iii) You are using the inverse transform method, with small random numbers corresponding to small values of  $\Omega$  or short future lifetimes.
- (iv) Your first random numbers from  $[0, 1]$  for simulating  $\Omega$  and the future lifetime are 0.4 and 0.7 respectively.

Calculate your first simulated value of the future lifetime.

- (A) 22      (B) 35      (C) 46      (D) 52      (E) 56      (Ans: B)

**Example 2.3 [01 Exam M May #11]**

You are using the inverse transform method to simulate  $Z$ , the present value random variable for a special two-year term insurance on  $(70)$ . You are given:

- (i)  $(70)$  is subject to only two causes of death, with

$k$	${}_k q_{70}^{(1)}$	${}_k q_{70}^{(2)}$
0	0.1	0.1
1	0.1	0.5

- (ii) Death benefits, payable at the end of the year of death, are:

During year	Benefit for Cause 1	Benefit for Cause 2
1	1000	1100
2	1100	1200

- (iii)  $i = 0.06$   
 (iv) For this trial your random number, from the uniform distribution on  $[0, 1]$ , is 0.35.  
 (v) High random numbers correspond to high values of  $Z$ .

Calculate the simulated value of  $Z$  for this trial.

- (A) 943      (B) 979      (C) 1000      (D) 1038      (E) 1068      (Ans: B)

### 3 Simulation of European Derivatives

Recall in FIN3250 that under risk neutral pricing, for derivatives whose payoff is paid at maturity  $T$ ,

$$V = e^{-rT} E^*[\text{payoff at } T].$$

To estimate  $V$ , we only need to use simulation to estimate the risk-neutral expectation

$$E^*[\text{terminal payoff at } T].$$

Under the binomial model, this can be done easily:

#### Example 3.1

Consider a non-dividend-paying stock with  $S(0) = \$41$  and  $\sigma = 30\%$ .

- By construct a binomial tree with 3 time steps, calculate the price of a \$40-strike European call on the stock maturing after 1 year. Use  $r = 8\%$ .
- Verify your result by using 1000 simulations to calculate the risk-neutral expectation.

What if we are dealing with the Black-Scholes framework? In this case, recall in FIN3250 that under the risk-neutral measure,

$$S(T) = S(0) \exp\left[\left(r - \delta - \frac{\sigma^2}{2}\right)T + \sigma Z(T)\right].$$

where  $Z(T) \sim N(0, T)$ . To simulate the price of a European derivative with maturity  $T$ ,

*Step 1:* Simulate an  $N(0, 1)$  distributed random variable  $Z$ .

*Step 2:* Multiple  $Z$  by  $\sqrt{T}$  to get the  $N(0, T)$  distributed random variable  $Z(T)$ .

*Step 3:* Calculate  $S(T)$  by using the formula above.

*Step 4:* Calculate the terminal payoff.

*Step 5:* Repeat Steps 1 to 4 above to get  $r$  terminal payoffs, take average and multiply by  $e^{-rT}$ .

### Example 3.1 (continue)

Repeat Example 3.1 under the Black-Scholes option pricing model.

#### A Remark about Simulating an $N(0, 1)$

McDonald presented a method to simulate an (approximate)  $N(0, 1)$  on p.624. Consider

$$\tilde{Z} = \sum_{i=1}^{12} U_i - 6$$

where each  $U_i \sim U(0, 1)$ . The mean and variance of  $\tilde{Z}$  are 0 and 1. McDonald claims that by central limit theorem,  $\tilde{Z}$  is approximately normal. The advantage of this method is that it only requires addition and does not require the calculation of

$$Z = \Phi^{-1}(U)$$

which requires time-consuming numerical approximation.

#### Accuracy of Monte Carlo Simulation

We want to know the number of runs to obtain an accurate price. The simulated price is

$$\bar{V}(r) = \frac{1}{r} \sum_{i=1}^r V(S_i)$$

where  $V(S_i)$  is the simulated derivative price in the  $i$ th trial. If  $\sigma_V$  is the standard deviation of the random variable  $V(S)$ , then the standard deviation of the price resulted from  $r$  simulations is

$$\sigma(\bar{V}_r) = \frac{1}{\sqrt{r}} \sigma_V.$$

A 95% confidence interval about the true price  $V$  is therefore given by

$$\bar{V}(r) - \frac{1.96\sigma_V}{\sqrt{r}} < V < \bar{V}(r) + \frac{1.96\sigma_V}{\sqrt{r}}.$$

As usually,  $\sigma_V$  is actually unknown and it is estimated by the sample standard deviation. Since  $r$  is typically very large (usually  $> 10,000$ ), a  $t$ -distribution is not needed to replace 1.96.

The CI above shows that the uncertainty about  $V$  is inversely proportional to the square root of the number of runs. To double the accuracy, we must quadruple the number of runs. To increase the accuracy by a factor of 10 (that is, one more decimal place), the number of necessary runs must increase by a factor of 100!

### Extension 1: Pricing a Path-Dependent European Option

A path-dependent European option has a terminal payoff that does not only depend on the terminal stock price  $S(T)$ , but also on some immediate values of the stock before expiration.

For example, for an arithmetic average-strike European call with 1 year maturity and observation period of 4 months, the terminal payoff is

$$\left( \frac{S(1/3) + S(2/3) + S(1)}{3} - K \right)_+.$$

To simulate a single terminal payoff, we need to simulate  $S(1/3)$ ,  $S(2/3)$  and  $S(1)$ . In the Black-Scholes model, this can be achieved by

$$S(t_2) = S(t_1) \exp\left[\left(r - \delta - \frac{\sigma^2}{2}\right)(t_2 - t_1) + \sigma(Z(t_2) - Z(t_1))\right]$$

and that  $Z(t_2) - Z(t_1) \sim N(0, t_2 - t_1)$  and is independent of all events happened on or before  $t_1$ .

To get a single payoff, we may proceed as follows:

*Step 1:* Simulate  $S(1/3)$  using a  $N(0, 1)$  and the method introduced previously.

*Step 2:* Simulate another  $N(0, 1)$  and multiply by  $\sqrt{1/3}$  to get  $Z(2/3) - Z(1/3)$ .

*Step 3:* Get  $S(\frac{2}{3}) = S(\frac{1}{3}) \exp\left[\left(r - \delta - \frac{\sigma^2}{2}\right)\frac{1}{3} + \sigma\left(Z(\frac{2}{3}) - Z(\frac{1}{3})\right)\right]$ .

*Step 4:* Simulate another  $N(0, 1)$  and multiply by  $\sqrt{1/3}$  to get  $Z(1) - Z(2/3)$ .

*Step 5:* Get  $S(1) = S(\frac{2}{3}) \exp\left[\left(r - \delta - \frac{\sigma^2}{2}\right)\frac{1}{3} + \sigma\left(Z(1) - Z(\frac{2}{3})\right)\right]$ .

*Step 6:* Get the terminal payoff.

### Extension 2: Pricing an Exchange Option (Advanced Topic)

Recall in Chapter 6 of FIN3250 that for exchange options, the SDEs of the underlying asset  $S$  and strike asset  $K$  (with return correlation  $\rho$ ) are

$$\begin{aligned} dS(t) &= (r - \delta_S)S(t)dt + \sigma_S S(t)dZ_1(t) \\ dK(t) &= (r - \delta_K)K(t)dt + \sigma_K K(t)dZ_3(t) \end{aligned}$$

where  $\{Z_1(t): t \geq 0\}$  and  $\{Z_3(t): t \geq 0\}$  are correlated standard Brownian motions with correlation coefficient  $\rho$  such that

$$\text{Cov}[Z_1(t), Z_3(t)] = \rho t.$$

An alternative way to write the SDE of  $K$  is

$$dK(t) = (r - \delta_K)K(t)dt + \sigma_K K(t) [\rho dZ_1(t) + \sqrt{1 - \rho^2} dZ_2(t)]$$

where  $\{Z_2(t): t \geq 0\}$  is a standard Brownian motion that is independent with  $\{Z_1(t): t \geq 0\}$ .

The representation

$$dZ_3(t) = \rho dZ_1(t) + \sqrt{1 - \rho^2} dZ_2(t)$$

or equivalently

$$Z_3(t) = \rho Z_1(t) + \sqrt{1 - \rho^2} Z_2(t)$$

is called the **Cholesky decomposition**. By using this decomposition, we can simulate a pair of correlated normal random variables  $(Z_1(t), Z_3(t))$  for  $t$  being a fixed number:

*Step 1:* Simulate  $Z_1(t)$  using a  $N(0, 1)$ .

*Step 2:* Simulate  $Z_2(t)$  (independent of  $Z_1(t)$ ) using another  $N(0, 1)$ .

*Step 3:* Get  $Z_3(t) = \rho Z_1(t) + \sqrt{1 - \rho^2} Z_2(t)$ .

To simulate a terminal payoff of an exchange option,

*Step 1:* Simulate the correlated normal variables  $(Z_1(T), Z_3(T))$  using Cholesky decomposition.

*Step 2:* Obtain the terminal asset prices

$$\begin{cases} S(T) = S(0) \exp[(r - \delta_S - \frac{\sigma_S^2}{2})T + \sigma_S Z_1(T)] \\ K(T) = K(0) \exp[(r - \delta_K - \frac{\sigma_K^2}{2})T + \sigma_K Z_3(T)] \end{cases}$$

*Step 3:* Calculate  $[S(T) - K(T)]_+$ .

To price basket or rainbow options, we need sets of  $n$  correlated normal variables.

We can simulate  $n$  correlated **standard** normal random variables  $(Z_1, Z_2, \dots, Z_n)$  with a positive-definite correlation matrix  $[\rho_{ij}]$  where  $\rho_{ii} = 1$ . The Cholesky decomposition is of the **lower triangular form**

$$Z_k = \sum_{i=1}^k a_{ki} \varepsilon_i$$

where all  $\varepsilon_i$  are independent  $N(0, 1)$  random variables. To obtain the values of  $a_{ki}$ 's, observe that

(1) variance constraint:  $\text{Var}(Z_k) = \sum_{i=1}^k a_{ki}^2 = 1$

(2) covariance constraint: For  $j < k$ ,  $\text{Cov}(Z_j, Z_k) = \sum_{i=1}^j a_{ji} a_{ki} = \rho_{jk}$

The set of equations can be solved starting from the top of the triangle and proceed down the row in the order  $a_{11} (= 1)$ ,  $a_{21}$ ,  $a_{22}$ ,  $a_{31}$ ,  $a_{32}$ ,  $\dots$ .

If a set of  $n$  correlated normal variables with standard deviations not necessarily equal to 1 is required, we may first simulate a set of normal variables with the correct correlations, then scale up each normal variable by the correct standard deviation. If the means of the correlated normal variables are also not 0, we can also add the means to a set of correlated normal variables with correct correlations and standard deviations.

### Variance Reduction (Extremely Advanced Topic)

If the simulation is carried out as described so far, a very large number of runs is usually needed for reasonable accuracy. In this subsection we examine a number of variance reduction procedures that can lead to dramatic decrease in number of necessary runs.

#### (1) Antithetic (對偶) variable

*Step 1:* Simulate the price  $f_1$  of the derivative using normal random numbers.

*Step 2:* Add a minus sign to all normal random numbers used in Step 1 and recalculate another simulated price  $f_2$ .

*Step 3:* Calculate  $\bar{f} = \frac{f_1 + f_2}{2}$

*Step 4:* Repeat Steps 1 to 3 using other sets of normal random numbers and use the average of  $\bar{f}$ 's as the estimate of the true value  $f$ .

The idea behind Step 3 is that

$$\text{Var}(\bar{f}) = \frac{1}{4}[\text{Var}(f_1) + \text{Var}(f_2) + 2\text{Cov}(f_1, f_2)] = \frac{1}{2}[\text{Var}(f) + \text{Cov}(f_1, f_2)]$$

and by using antithetic random numbers, the covariance is (hopefully) negative.

#### (2) Control variate (控制變量)

This is best illustrated with an example. Define the followings:

$A_{\text{sim}}$ : simulated price of an arithmetic Asian average price call using Monte Carlo simulation

$G_{\text{sim}}$ : simulated price of the corresponding geometric Asian average price call, using the **same** set of random numbers

$A$ : true price of the arithmetic call (we want to know  $A$ )

$G$ : true price of the geometric call (we have closed-form formula for  $G$ )

Heuristically, the errors in the simulated arithmetic and geometric prices are correlated: when  $G_{\text{sim}}$  is high, the stock price has high return and  $A_{\text{sim}}$  is also high. Thus

$$A - A_{\text{sim}} \approx G - G_{\text{sim}}$$

and hence

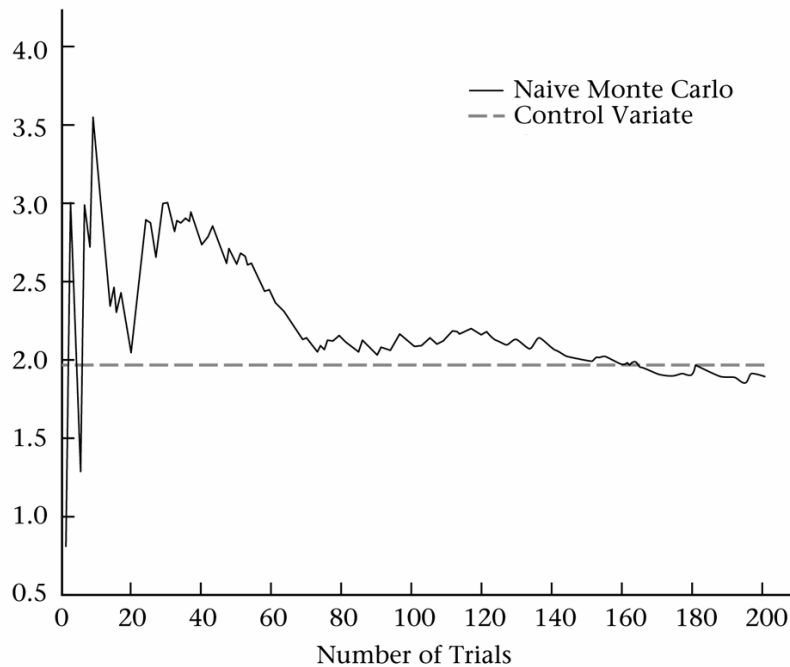
$$A^* = A_{\text{sim}} + (G - G_{\text{sim}})$$

can be used as an estimate of the true value  $A$ . The idea behind is the following two equations:

$$\begin{aligned} E(A^*) &= E(A_{\text{sim}}) + E[G] - E[G_{\text{sim}}] = A + E[G] - E[G] = A \\ \text{Var}(A^*) &= \text{Var}(A_{\text{sim}}) + \text{Var}(G - G_{\text{sim}}) + 2\text{Cov}(A_{\text{sim}}, G - G_{\text{sim}}) \\ &= \text{Var}(A_{\text{sim}}) + \text{Var}(G_{\text{sim}}) - 2\text{Cov}(A_{\text{sim}}, G_{\text{sim}}) \end{aligned}$$

As long as  $A_{\text{sim}}$  and  $G_{\text{sim}}$  are highly correlated, we have  $\text{Var}(A^*) < \text{Var}(A_{\text{sim}})$ :

Price of Arithmetic Average Price Call



Comparison of “naive” Monte Carlo estimate of arithmetic average option price with control variate method. Graph depicts first 200 simulations for an option with  $S = \$40$ ,  $K = \$40$ ,  $\sigma = 0.3$ ,  $r = 0.08$ ,  $T = 0.25$ ,  $\delta = 0$ , and the final price computed with three averages.

Actually, for  $A^*$  to be an unbiased estimator of  $A$ , we can use

$$A^* = A_{\text{sim}} + \beta(G - G_{\text{sim}})$$

for any  $\beta$  because  $E(G) = E(G_{\text{sim}})$ . The formula for  $\text{Var}(A^*)$  becomes

$$\text{Var}(A^*) = \text{Var}(A_{\text{sim}}) + \beta^2 \text{Var}(G_{\text{sim}}) - 2\beta \text{Cov}(A_{\text{sim}}, G_{\text{sim}}),$$

which shows that  $\text{Var}(A^*)$  is minimized when

$$\beta = \text{Cov}(A_{\text{sim}}, G_{\text{sim}}) / \text{Var}(G_{\text{sim}}).$$

For this value of  $\beta$ , the variance of  $A^*$  is found to be

$$\text{Var}(A^*) = \text{Var}(A_{\text{sim}}) - \frac{[\text{Cov}(A_{\text{sim}}, G_{\text{sim}})]^2}{\text{Var}(G_{\text{sim}})}$$

so that that variance would be reduced as long as  $A_{\text{sim}}$  and  $G_{\text{sim}}$  are correlated. In practice,  $\text{Cov}(A_{\text{sim}}, G_{\text{sim}})$  and  $\text{Var}(G_{\text{sim}})$  can be estimated from Monte Carlo simulations:

$$\text{Cov}(A_{\text{sim}}, G_{\text{sim}}) \approx \frac{1}{r-1} \sum_{i=1}^r (A_i - \bar{A})(G_i - \bar{G})$$

Then the value of  $\beta$  can be chosen.

**(3) Stratified sampling** (分層隨機抽樣)

Stratified sampling of a distribution (say, normal) using random numbers involves dividing the desired distribution into  $n$  intervals and sampling from each interval according to its probability.

Suppose we want to generate  $n$  independent  $N(0, 1)$  from  $U(0, 1)$ , then

*Step 1:* Generate  $n$   $U(0, 1)$  random numbers, denoted by  $U_1, U_2, U_3, \dots, U_n$ .

(Old *Step 2:* Calculate  $Z_i = N^{-1}(U_i)$ . But it may happen that all simulated  $Z_i > -1.645$  if it happens unluckily that all  $U_i > 0.1$ . Thus the frequency distribution of the  $Z_i$ 's may not look like a normal distribution when  $n$  is only moderately large.)

*Step 2:* Note that the  $U_i$ 's may not be evenly distributed on  $(0, 1)$ . So we do a transformation:

$$V_i = \frac{1}{n}(U_i + i - 1) \sim U\left(\frac{i-1}{n}, \frac{i}{n}\right).$$

Notice that now the  $V_i$ 's fill in  $(0, 1)$  much evenly. They are only random distributed in each subinterval.

*Step 3:* Calculate  $Z_i = N^{-1}(V_i)$ . The frequency distribution of  $Z_i$ 's looks like a normal distribution even if  $n$  is quite small.

Stratified sampling works when the payoff of a derivative depends on only one random variable (as in the case of European-type option). A generalization of this technique when the payoff depends on more than one random variable is **Latin hypercube sampling**. [http://en.wikipedia.org/wiki/Latin\\_hypercube\\_sampling](http://en.wikipedia.org/wiki/Latin_hypercube_sampling) provides some basic ideas of why this is better than a random sampling.

The vigorous proof of why variance is reduced is based on the following two formulas:

(a)  $E\left[\frac{1}{r} \sum_{i=1}^r g(V_i)\right] = \int_0^1 g(u) du$  so that the stratified sampling estimator is unbiased.

(b)  $\text{Var}\left[\frac{1}{r} \sum_{i=1}^r g(V_i)\right] \leq \text{Var}\left[\frac{1}{r} \sum_{i=1}^r g(U_i)\right]$  so that the variance is reduced.

The first formula can be proved directly, while the proof of the second formula requires a simple conditioning argument.

**(4) Other methods**

This includes importance sampling and low discrepancy (or quasi random) sequences (超均勻分佈列). McDonald did not give much detail about these two methods. He only referred readers to a paper by Boyle et al. (1997). The theories behind importance sampling and low discrepancy sequence are well beyond the scope of this course.

#### 4 Simulation of American Derivatives

A number of researchers have adopted special approximations to value American options using Monte Carlo simulation. McDonald discusses two approaches.

##### (1) The Longstaff-Schwartz Least-Square approach

First, recall a basic fact about least-square regression:

For a linear regression of a dependent variable  $V$  on an independent variable  $S$  of the form

$$V = a + bS + cS^2,$$

the least-square estimate of  $(a, b, c)$  is given by the solution of the linear system

$$\begin{bmatrix} n & \sum S_i & \sum S_i^2 \\ \sum S_i & \sum S_i^2 & \sum S_i^3 \\ \sum S_i^2 & \sum S_i^3 & \sum S_i^4 \end{bmatrix} \begin{bmatrix} a \\ b \\ c \end{bmatrix} = \begin{bmatrix} \sum V_i \\ \sum V_i S_i \\ \sum V_i S_i^2 \end{bmatrix}.$$

To value an American option, it is necessary to choose between exercising and continuing at each early exercise point. In the Longstaff-Schwartz approach,

- (a) the value of exercising is the payoff that one can get from early exercise;
- (b) the value of continuing (that is, not exercising the option) is obtained from least-square regression.

To illustrate, consider a special 3-year American put on a non-dividend-paying stock with  $S_0 = 1$ ,  $K = 1.1$ , and  $r = 6\%$ . The option can only be exercised at each year end. By simulation, we obtain the following 8 stock price paths:

Path	$t = 1$	$t = 2$	$t = 3$
1	1.09	1.08	1.34
2	1.16	1.26	1.54
3	1.22	1.07	1.03
4	0.93	0.97	0.92
5	1.11	1.56	1.52
6	0.76	0.77	0.90
7	0.92	0.84	1.01
8	0.88	1.22	1.34

We work backward through time to determine when will the option be exercised for each path by comparing (a) and (b).

At  $t = 3$ :

Criterion: For each path, if the option was not exercised before, then

$$\text{payoff} = (1.1 - S_3)_+$$

Exercising is optimal only for paths 3, 4, 6 and 7 (if it has not happened before).

At  $t = 2$ :

Criteria: For each path, if the option was not exercised at time 1, then

$$\text{Payoff from immediate exercise} = (1.1 - S_2)_+$$

Continuing value = price of the option at  $t = 2$  if it is not exercised, *unknown* at  $t = 2$

Notice that for paths 2, 5 and 8, it is not optimal to exercise because the payoff from immediate exercise is negative. So we only need to consider paths 1, 3, 4, 6, 7.

We assume that the continuing value  $V$  is equal to  $a + bS + cS^2$  and that  $V$  can be approximated by the present value of the payoff at time 3.

Data for regression:

Path 1:  $V = 0$ ,  $S = 1.08$

Path 3:  $V = (1.1 - 1.03)\exp(-0.06)$ ,  $S = 1.07$

Path 4:  $V = (1.1 - 0.92)\exp(-0.06)$ ,  $S = 0.97$

Path 6:  $V = (1.1 - 0.90)\exp(-0.06)$ ,  $S = 0.77$

Path 7:  $V = (1.1 - 1.01)\exp(-0.06)$ ,  $S = 0.94$

This gives  $a = -1.070$ ,  $b = 2.983$  and  $c = -1.813$ , and the forecasted value of  $V$ :

Path	forecasted value of $V$	early exercise
1	0.0369	0.02
3	0.0461	0.03
4	0.1176	0.13
6	0.1520	0.33
7	0.1565	0.26

Thus it is optimal to early exercise at  $t = 2$  (if it has not happened at  $t = 1$ ) for paths 4, 6 and 7.

At  $t = 1$ :

Criteria: For each path,

$$\text{Payoff from immediate exercise} = (1.1 - S_1)_+$$

Continuing value = price of the option at  $t = 1$  if it is not exercised, *unknown* at  $t = 1$

Notice that for paths 2, 3 and 5, it is not optimal to exercise because the payoff from immediate exercise is negative. So we only need to consider paths 1, 4, 6, 7, 8.

We assume that the continuing value  $V$  is equal to  $a + bS + cS^2$  and that  $V$  can be approximated by the greater of 0 and the present values of exercising at  $t = 2$  and at  $t = 3$ .

Data for regression:

Path 1:  $V = 0, S = 1.09$  (since it is not optimal to exercise at  $t = 2, 3$ )

Path 4:  $V = (1.1 - 0.97)\exp(-0.06), S = 0.93$  (since it is optimal to exercise at  $t = 2$ )

Path 6:  $V = (1.1 - 0.77)\exp(-0.06), S = 0.76$  (since it is optimal to exercise at  $t = 2$ )

Path 7:  $V = (1.1 - 0.84)\exp(-0.06), S = 0.92$  (since it is optimal to exercise at  $t = 2$ )

Path 8:  $V = 0, S = 0.88$  (since it is not optimal to exercise at  $t = 2, 3$ )

This gives  $a = 2.038, b = -3.335$  and  $c = 1.356$ , and the forecasted value of  $V$ :

Path	forecasted value of $V$	early exercise
1	0.0139	0.01
4	0.1092	0.17
6	0.2866	0.34
7	0.1175	0.18
8	0.1533	0.22

Thus it is only optimal to early exercise at  $t = 1$  for paths 4, 6 and 7 and 8.

To summarize:

Path	$t = 1$	$t = 2$	$t = 3$	When to exercise?
1	1.09	1.08	1.34	never
2	1.16	1.26	1.54	never
3	1.22	1.07	<b>1.03</b>	3
4	<b>0.93</b>	<b>0.97</b>	<b>0.92</b>	1
5	1.11	1.56	1.52	never
6	<b>0.76</b>	<b>0.77</b>	<b>0.90</b>	1
7	<b>0.92</b>	<b>0.84</b>	<b>1.01</b>	1
8	<b>0.88</b>	1.22	1.34	1

After noting when the option should be exercised for each path, we determine the payoff for each path and calculate the discounted payoff, and finally take average:

$$\text{Simulated price} = \frac{1}{8}[0.07e^{-0.18} + 0.17e^{-0.06} + 0.34e^{-0.06} + 0.18e^{-0.06} + 0.22e^{-0.06}] = 0.1144$$

The difficulties of this approach:

- The functional form of the continuing value is hard to determine: if a cubic function is used instead, the result would be different.
- If the option can be exercised any time, then one must use a very fine grid of time intervals and determine at which time point would exercise be optimal.

## (2) The Broadie and Glasserman approach

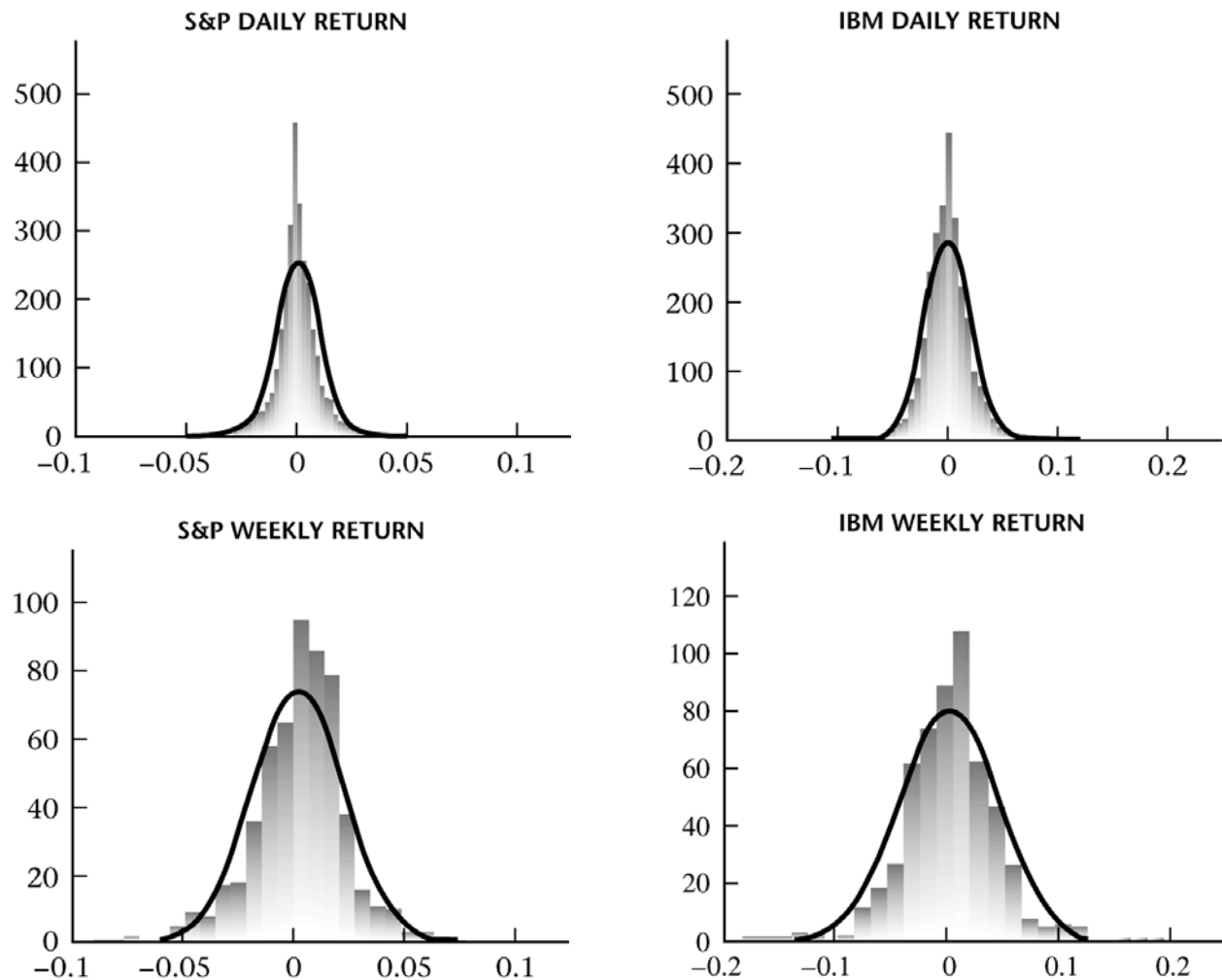
This is a very computationally intensive method and McDonald only gives a surface treatment of this topic. It is unlikely that SoA will set question on this.

## 5 The Jump-Diffusion Model

Recall that under the Black-Scholes framework, we assume that the mean and variance of the return is constant through time, and that stock price follows a geometric Brownian motion. This implies that returns are normally distributed.

### How Good is the Black-Scholes Model?

The following graphs show the histogram of the returns of S&P 500 index and IBM from 1991/1/2 to 2001/10/24 and the normal distribution with mean and standard deviation equal to the sample mean and sample standard deviation of the returns.



A simple method to see if the returns are normal distributed is to calculate the skewness and kurtosis of the returns. Skewness and kurtosis are used to measure the lack of symmetry and the thickness of the tail of a distribution. They are defined as:

$$\text{Skew}(X) = \frac{1}{\sigma^3} E[(X - \mu)^3] = \frac{\mu_3}{\sigma^3}, \quad \text{Kurt}(X) = \frac{1}{\sigma^4} E[(X - \mu)^4] = \frac{\mu_4}{\sigma^4}$$

For a (not necessarily standard) normal distribution, skewness = 0 and kurtosis = 3. These two results can be proved by noting that  $\text{Skew}(X)$  and  $\text{Kurt}(X)$  are of the form  $E[(\frac{X-\mu}{\sigma})^k]$ ,  $k = 3, 4$ .

Since the MGF of  $\frac{X-\mu}{\sigma}$  is  $\exp(0.5t^2)$ ,  $\text{Skew}(X)$  and  $\text{Kurt}(X)$  can be obtained from a Taylor series expansion (quicker) or repeated differentiations (slower).

We have the following classifications:

$$\text{Skewness} \begin{cases} > 0 & \text{positively / right skewed} \\ = 0 & \text{symmetric} \\ < 0 & \text{negatively / left skewed} \end{cases} \quad \text{Kurtosis} \begin{cases} > 3 & \text{leptokurtic} \\ = 3 & \text{mesokurtic} \\ < 3 & \text{platykurtic} \end{cases}$$

it is evident from the histogram that the four return series are left skewed and **leptokurtic** (i.e., having a sharper peak and fatter tail than a normal distribution) and thus the returns are not normally distributed. A rigorous statistical test can be based on the Jarque-Bera statistic

$$J = \frac{n}{6} (\text{Skew}(X))^2 + \frac{(\text{Kurt}(X) - 3)^2}{4}$$

which follows  $\chi^2$  if the null hypothesis of normality holds.

Still another method is to draw a **normal probability plot**. We shall illustrate its construction with a data set.

*Step 1:* Arrange the data in ascending order.

*Step 2:* For the  $i$ th order statistic, assign the quantile  $q = \frac{i-0.5}{n}$  to it. (In statistics language, the  $q$ -quantile of a distribution  $F$  is the value  $z$  such that  $F(z) = q$ .)

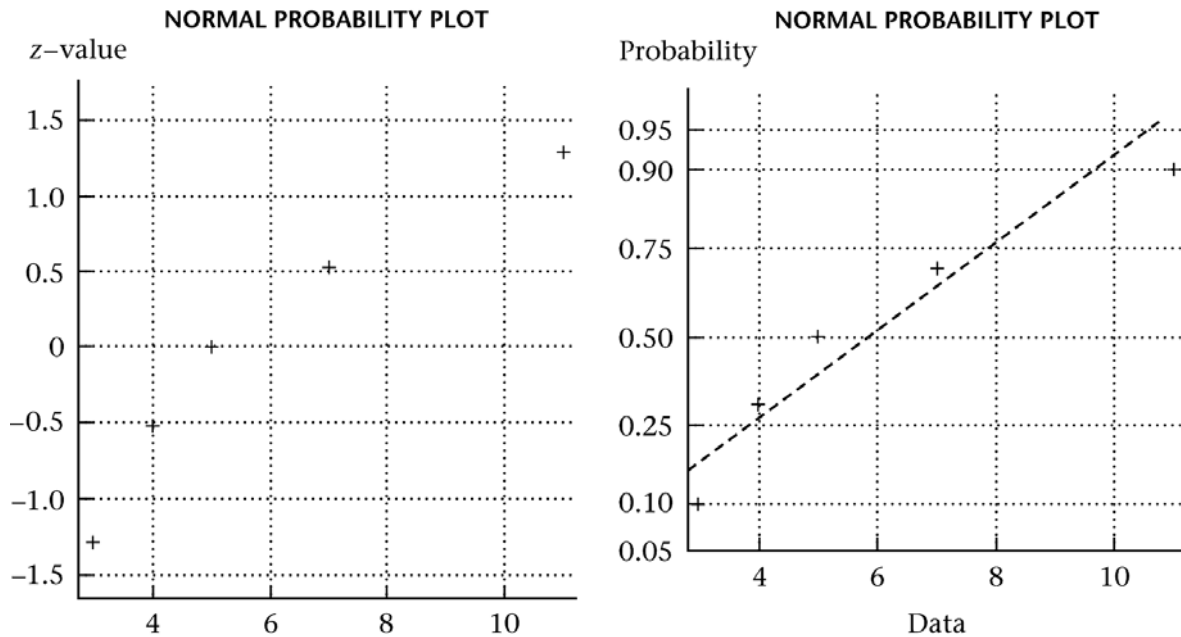
*Step 3:* Convert the quantiles  $q$  in Step 3 into the  $z$  values such that  $N(z) = q$ .

*Step 4:* Plot the data and the associated  $z$ . If the underlying distribution is normal, the data should fall on a straight line.

### Example 5.1

Consider the five data points: {5, 4, 7, 11, 3}. Plot the normal probability plot of the data set.

Data	Quantile assigned	$z$ value
3	$0.5/5 = 10\%$	-1.282
4	$1.5/5 = 30\%$	-0.524
5	$2.5/5 = 50\%$	0
7	70%	+0.524
11	90%	+1.282



The normal probability plot for the four return series considered are given on p.610 of McDonald. It is again evident that the returns are not normally distributed.

Various alternative continuous-time models have been proposed to fit the stock price process observed in reality. While most of these models are complicated Levy processes, the following three are easier to explain than others:

- (1) Mixture of two normal distributions for stock returns
- (2) Two-state log-normal regime-switching model (RSLN2) for stock price
- (3) Geometric Brownian motion augmented with jumps arriving according to a Poisson process

In Exam C we study (3).

### The Poisson Jump Diffusion Model with Lognormal Jumps

The Poisson Jump Diffusion model with lognormal jumps was advocated by Merton (1976).

The jump in this model has a **multiplicative effect** on the stock price. Let  $Y(t)$  be the size of the jump at  $t$ . If  $S(t-)$  is the pre-jump stock price, then the post-jump price is  $S(t-)Y(t)$ . If there is no jump at  $t$ , then  $Y(t) = 1$ .

Suppose the stock has no jumps, then we model its price at  $t$  in the real world by

$$S(t) = S(0) \exp[(\alpha - \delta - 0.5\sigma^2)t + \sigma Z(t)].$$

Consider an otherwise identical stock that has already jumped  $N(t)$  times by time  $t$ . Then the price at time  $t$  is

$$S(t) = S(0) \exp[(\alpha - \delta - 0.5\sigma^2)t + \sigma Z(t)] \prod_{i=1}^{N(t)} Y_i.$$

Notice that the exact time of the jumps does not matter as long as they are less than  $t$ .

The assumptions in Merton (1976) are

- (a)  $\{N(t): t \geq 0\}$  is a Poisson process,
- (b) Jump sizes  $Y_i$  are independent and are lognormal distributed,
- (c) Jumps do not affect the risk premium of the asset.

Assumption (b) means that to simulate a lognormal  $Y_i$ , we can use

$$Y_i = \exp[\alpha_J - 0.5\sigma_J^2 + \sigma_J W_i]$$

where  $W_i \sim N(0, 1)$ . The interpretation of  $\alpha_J$  and  $\sigma_J$  are that

- (1)  $E(Y_i) = \exp(\alpha_J)$  so that the expected change immediately after a jump is

$$k = E\left[\frac{Y_i S - S}{S}\right] = E(Y_i) - 1 = e^{\alpha_J} - 1$$

- (2) The standard deviation of the log of the jump ( $\ln Y_i$ ) is  $\sigma_J$ .

Under this assumption,  $\prod_{i=1}^{N(t)} Y_i = \exp\left[(\alpha_J - 0.5\sigma_J^2)N(t) + \sigma_J \sum_{i=1}^{N(t)} W_i\right]$ .

### Example 5.2

- (a) Compute the mean of  $\prod_{i=1}^{N(t)} Y_i$ .
- (b) Hence, compute the mean of  $S(t)$  under Merton's model.
- (c) Find the expected rate of return of the stock under Merton's model.

*Solution:* Notice that the  $E[Y_1 Y_2 \dots Y_n] = \exp(n\alpha_J)$  and hence  $E\left[\prod_{i=1}^{N(t)} Y_i \mid N(t)\right] = \exp[N(t)\alpha_J]$ . Since

the MGF of a  $Po(\lambda)$  is  $M_X(t) = \exp[\lambda(e^t - 1)]$ ,

$$E\left(\prod_{i=1}^{N(t)} Y_i\right) = E\left(E\left[\prod_{i=1}^{N(t)} Y_i \mid N(t)\right]\right) = E \exp[N(t)\alpha_J] = \exp[\lambda t(e^{\alpha_J} - 1)].$$

Finally,

$$E[S(t)] = \exp[(\alpha - \delta)t] E\left[\prod_{i=1}^{N(t)} Y_i\right] = \exp[(\alpha - \delta)t + \lambda t(e^{\alpha_J} - 1)] = \exp[(\alpha - \delta)t + \lambda k t],$$

which shows that the expected rate of return is  $\alpha - \delta + \lambda k$ .

The importance of Example 5.2 is that

Incorporating multiplicative jumps changes the expected rate of return from  $\alpha - \delta$  to  $\alpha - \delta + \lambda k$ .

To simulate the value of  $S(t)$  (for a fixed  $t$ ) in the real world under Merton's model,

*Step 1:* Choose  $\alpha$  such that the  $\alpha - \delta + \lambda k$  matches the desired rate of return.

*Step 2:* Simulate  $S(t)$  using the  $\alpha$  in Step 1, assuming that there is no jumps.

*Step 3:* Simulate  $N(t) \sim \text{Po}(\lambda t)$  using inverse transform.

*Step 4:* For  $N(t) = n$ ,  $\prod_{i=1}^{N(t)} Y_i = \exp\left[(\alpha_j - 0.5\sigma_j^2)n + \sigma_j \sum_{i=1}^n W_i\right]$ .

Since  $\sum_{i=1}^n W_i \sim N(0, n)$ , we can simulate  $\sum_{i=1}^n W_i$  by using  $\sqrt{n}Z$  where  $Z \sim N(0, 1)$ .

*Step 5:* Multiply  $S(t)$  and  $\prod_{i=1}^{N(t)} Y_i$  obtained in Steps 2 and 4 together to get the post-jump price.

If you are asked to simulate a path of  $\{S(t), t \geq 0\}$ , then you may first simulate  $S(t_1)$  using the procedure above, then note that

$$S(t_1 + h) = S(t_1) \exp[(\alpha - \delta - 0.5\sigma^2)h + \sigma(Z(t_1 + h) - Z(t_1))] \prod_{i=1}^{N(t+h)-N(t)} Y_i.$$

To simulate  $S(t_1 + h)$ , you only know to simulate  $Z(t_1 + h) - Z(t_1) \sim N(0, h)$ ,  $N(t + h) - N(t) \sim \text{Po}(\lambda h)$ , and another standard normal random variable to form the sum in

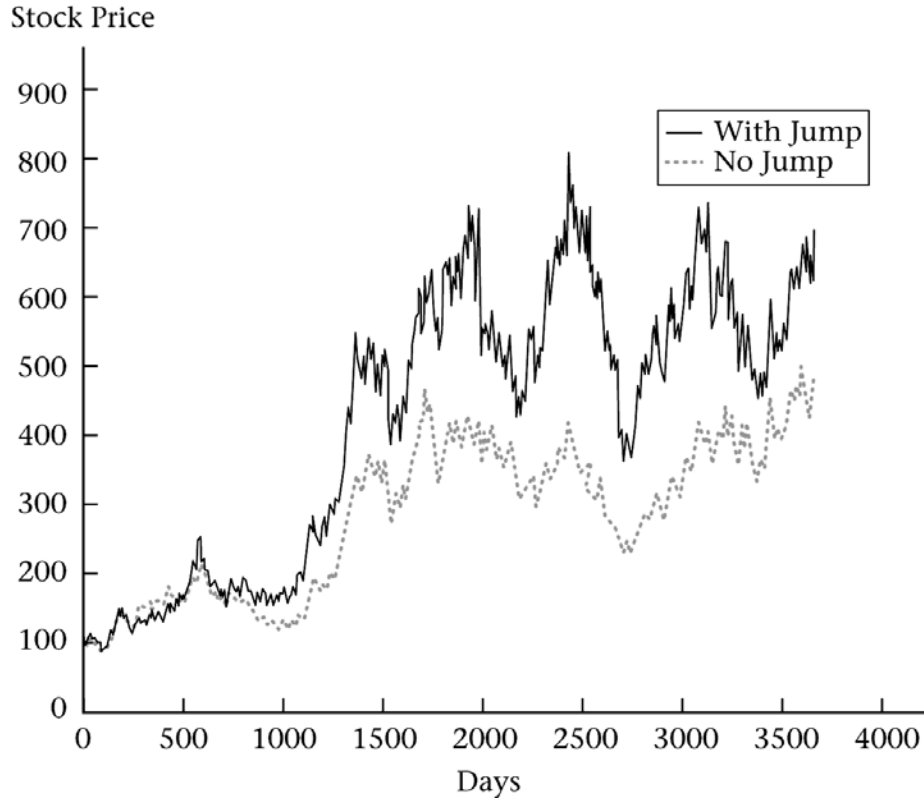
$$\prod_{i=1}^{N(t+h)-N(t)} Y_i = \exp\left[(\alpha_j - 0.5\sigma_j^2)n + \sigma_j \sum_{i=1}^n W_i\right]$$

for  $N(t + h) - N(t) = n$ .

If we are simulating the stock price under the risk-neutral measure, then owing to assumption (3) in Merton's model, we should set the value of  $\alpha$  such that

$$r = \alpha - \delta + \lambda k = \alpha - \delta + \lambda[\exp(\alpha_j) - 1].$$

The following is a simulation of the stock price using  $\alpha = 8\%$ ,  $\delta = 0$ ,  $\sigma = 30\%$ ,  $\lambda = 3$ ,  $\alpha_j = -2\%$  and  $\sigma_j = 5\%$ :



The sample kurtosis of the stock returns with and without jump are 7.40 and 2.93 ( $\approx 3$ ). This shows that incorporating multiplicative jumps can increase kurtosis a lot.

### Direction of Jumps

In Merton's model, we can control the expected size of the jump by adjusting  $\sigma_J$ . If we increasing  $\sigma_J$  we get larger ups and downs. Some researchers propose that a better model should have two independent Poisson jumps, one controlling up jumps and one controlling down jumps. The lognormal moves associated with each can have different mean and standard deviations so that we can control skewness.