Monte Carlo Techniques

We know from earlier that the SDE

$$\frac{dS_t}{S_t} = rdt + \sigma dW_t$$

with constant r and σ has the solution

$$S_T = S_0 \exp\left\{\left(r - \tfrac{1}{2}\sigma^2\right)T + \sigma\phi\sqrt{T}\right\},$$

for some time horizon T; with $\phi \sim N(0,1)$; $W_t \sim N(0,t)$ and can be written $\phi\sqrt{T}$. It is often more convenient to express in time stepping form

$$S_{t+\delta t} = S_t \exp\left\{\left(r - \tfrac{1}{2}\sigma^2\right)\delta t + \sigma\phi\sqrt{\delta t}\right\}.$$

In general a closed form solution of an arbitrary SDE is difficult if e.g.

1. $r = r(t)$ and $\sigma = \sigma(S,t)$, i.e. the parameters are no longer constant
2. the SDE is complicated.

The need for Monte Carlo requires numerical integration of stochastic differential equations. Previously we considered the Forward **Euler-Maruyama** method. Why did this work?

Consider

$$dX_t = a(X_t,t)\,dt + b(X_t,t)\,dW_t \quad ; X_{t_0} = x_0 \tag{1}$$

The simplest scheme for solving (1) is using the E-M method. That is

$$\int_{t_n}^{t_{n+1}} dX_s = \int_{t_n}^{t_{n+1}} a(X_s,s)\,ds + \int_{t_n}^{t_{n+1}} b(X_s,s)\,dW_s$$

$$X_{n+1} = X_n + \int_{t_n}^{t_{n+1}} a(X_s,s)\,ds + \int_{t_n}^{t_{n+1}} b(X_s,s)\,dW_s$$

Using the left hand integration rule:

$$\int_{t_n}^{t_{n+1}} a(s,X_s)\,ds \approx a(t_n,X_n)\int_{t_n}^{t_{n+1}} ds = a(t_n,X_n)\,\delta t$$

$$\int_{t_n}^{t_{n+1}} b(s,X_s)\,ds \approx b(t_n,X_n)\int_{t_n}^{t_{n+1}} dW_s = b(t_n,X_n)\,\Delta W_n$$

$$X_{n+1} = X_n + a(t_n,X_n)\,\delta t + b(t_n,X_n)\,\Delta W_n$$

where $\Delta W_n = (W_{n+1} - W_n)$.

The Forward **Euler-Maruyama** method for GBM gives

$$\frac{\delta S_t}{S_t} = \frac{S_{t+\delta t} - S_t}{S_t} \sim r\delta t + \sigma\phi\sqrt{\delta t}$$

i.e

$$S_{t+\delta t} \sim S_t\left(1 + r\delta t + \sigma\phi\sqrt{\delta t}\right).$$

Now do a Taylor series expansion of the exact solution, i.e.

$$e^{\left(r-\frac{1}{2}\sigma^2\right)\delta t+\sigma\phi\sqrt{\delta t}} \sim 1+\left(r-\frac{1}{2}\sigma^2\right)\delta t+\sigma\phi\sqrt{\delta t}+\frac{1}{2}\sigma^2\phi^2\delta t$$

so we have

$$S_{t+\delta t} \sim S_t\left(1+r\delta t+\sigma\phi\sqrt{\delta t}+\frac{1}{2}\sigma^2\left(\phi^2-1\right)\delta t+..\right)$$

which differs from the Euler method at $O\left(\delta t\right)$ by the term $\frac{1}{2}\sigma^2\left(\phi^2-1\right)\delta t$. The term

$$\boxed{\frac{1}{2}\left(\phi^2-1\right)\delta t,}$$

is called the *Milstein correction.*

This unusual looking effect only arises for SDEs. To see this we do something similar to an earlier section where we considered the case of zero volatility, in the absence of randomness so set $\sigma=0$

$$\frac{dS_t}{S_t}=rdt$$

which has exact solution

$$S_{t+\delta t}=S_t e^{r\delta t}.$$

The Euler approximation is

$$\frac{\delta S_t}{S_t}=\frac{S_{t+\delta t}-S_t}{S_t}\sim r\delta t$$

or written explicitly

$$S_{t+\delta t}=S_t\left(1+r\delta t+...\right)$$

Expanding the exact solution in a series about t we find that

$$\begin{aligned} S_{t+\delta t} &= S_t e^{r\delta t} \\ &= S_t\left(1+r\delta t+O\left(\delta t^2\right)\right) \end{aligned}$$

so both the exact solution and Euler approximation agree to $O\left(\delta t^2\right)$.

The Milstein correction can be thought of as being a stochastic effect (a result of Itô's lemma in a sense).

Milstein Integration

We approximate the solution of the SDE

$$dG_t=A\left(G_t,t\right)dt+B\left(G_t,t\right)dW_t$$

which is compact form for

$$G_{t+\delta t}=G_t+\int_t^{t+\delta t}A\left(G_s,s\right)ds+\int_t^{t+\delta t}B\left(G_s,s\right)dW_s,$$

by

$$\boxed{G_{t+\delta t}\sim G_t+A\left(G_t,t\right)\delta t+B\left(G_t,t\right)\sqrt{\delta t}\phi+B\left(G_t,t\right)\frac{\partial}{\partial G_t}B\left(G_t,t\right)\cdot\frac{1}{2}\left(\phi^2-1\right)\delta t.}$$

Note: We use the same value of the random number $\phi\sim N\left(0,1\right)$ in both of the expressions

$$B\left(G_t,t\right)\sqrt{\delta t}\phi$$

and

$$B\left(G_t,t\right)\frac{\partial}{\partial G_t}B\left(G_t,t\right)\cdot\tfrac{1}{2}\left(\phi^2-1\right)\delta t.$$

The error of the Milstein scheme is $O\left(\delta t\right)$ which makes it better than the Euler-Maruyama method which is $O\left(\delta t^{1/2}\right)$. The Milstein makes use of Itô's lemma to increase the accuracy of the approximation by adding the second order term.

Some texts express the scheme in difference form. So a SDE written

$$dY_t=A\left(Y_t,t\right)dt+B\left(Y_t,t\right)dW_t$$

can be discretized as

$$Y_{i+1}=Y_i+A\Delta t+B\Delta W_t+\tfrac{1}{2}B\frac{\partial B}{\partial Y_i}\left(\left(\Delta W_t\right)^2-\Delta t\right)$$

Applying Milstein to the earlier example of GBM

$$dS_t=rS_tdt+\sigma S_tdW_t$$

where

$$\begin{aligned}A\left(S_t,t\right)&=rS_t\\B\left(S_t,t\right)&=\sigma S_t\end{aligned}$$

gives

$$\begin{aligned}S_{t+\delta t}&\sim S_t+rS_t\delta t+\sigma S_t\sqrt{\delta t}\phi+\sigma S_t\frac{\partial}{\partial S_t}\sigma S_t\cdot\tfrac{1}{2}\sigma^2\left(\phi^2-1\right)\delta t\\&=S_t\left(1+r\delta t+\sigma\phi\sqrt{\delta t}+\tfrac{1}{2}\sigma^2\left(\phi^2-1\right)\delta t\right)\end{aligned}$$

As another example, the CIR model for the spot rate is

$$dr_t=\left(\eta-\gamma r_t\right)dt+\sqrt{\alpha r_t}dW_t.$$

So identifying

$$\begin{aligned}A\left(r_t,t\right)&=\eta-\gamma r_t\\B\left(r_t,t\right)&=\sqrt{\alpha r_t}\end{aligned}$$

and substituting into the Milstein scheme gives

$$\begin{aligned}r_{t+\delta t}&\sim r_t+\left(\eta-\gamma r_t\right)\delta t+\sqrt{\alpha r_t\delta t}\phi+\sqrt{\alpha r_t}\frac{\partial}{\partial r_t}\sqrt{\alpha r_t}\cdot\tfrac{1}{2}\left(\phi^2-1\right)\delta t\\&=r_t+\left(\eta-\gamma r_t\right)\delta t+\sqrt{\alpha r_t\delta t}\phi+\tfrac{1}{4}\alpha\left(\phi^2-1\right)\delta t.\end{aligned}$$

Derivation of Milstein

Recall if

$$dG_t=A\left(G_t,t\right)dt+B\left(G_t,t\right)dW_t,$$

where A,B only depend on G, not t directly and $F=F\left(G_t\right)$ then Itô gives

$$dF=\left(A\frac{dF}{dG}+\tfrac{1}{2}B^2\frac{d^2F}{dG^2}\right)dt+B\frac{dF}{dG}dW_t$$

meaning: $dG_t=A(G_t)dt+B(G_t)dW_t$

One dimensional formula, so use Itô III

<= Itô III

Now consider a GBM

$$dS_t = \mu(S_t)dt + \sigma(S_t)dW_t$$

~~$dS_t = \mu S_t dt + \sigma S_t dW_t$~~

The Milstein Scheme works for SDEs for which the coefficients $\mu(S_t)$ and $\sigma(S_t)$ depend only on S, and do not depend on t directly. Here we assume that the stock price S_t is driven by the SDE:

$$dS_t = \mu(S_t)dt + \sigma(S_t)dW_t$$
$$= \mu_t dt + \sigma_t dW_t$$

Put

$$\mu_t = \mu(S_t); \quad \sigma_t = \sigma(S_t)$$

to give

$$dS_t = \mu_t dt + \sigma_t dW_t$$

~~$dS_t = \mu S_t dt + \sigma S_t dW_t;$~~

$\therefore$ In integral form:

$$\int_t^{t+\delta t} dS_s = \int_t^{t+\delta t} \mu_s ds + \int_t^{t+\delta t} \sigma_s dW_s$$
$$S_{t+\delta t} = S_t + \int_t^{t+\delta t} \mu_s ds + \int_t^{t+\delta t} \sigma_s dW_s$$

which in integral form is

$$S_{t+\delta t} = S_t + \int_t^{t+\delta t} \mu_s ds + \int_t^{t+\delta t} \sigma_s dW_s, \tag{2}$$

We want to improve the accuracy of discreatization by considering expansions of coefficients μ_t, σ_t using Itô. Here we note the coefficients are functions of S and do not depend directly on t. To minimize the amount of working, primed variables $' \equiv \frac{d}{dS}$ are used to denote differentiation w.r.t. S. Then by Itô

So: $\mu_t = \mu(S_t)$ ← same as $F = F(G_t)$ above and $dS_t = \mu_t dt + \sigma_t dW_t$

Use Itô III we have:

$$d\mu_t = \left(\mu_t \cdot \frac{d\mu_t}{dS_t} \cdot dt + \frac{1}{2}\sigma_t^2 \cdot \frac{d^2\mu_t}{dS_t^2}\right)dt + \sigma_t \cdot \frac{d\mu_t}{dS_t} dW_t$$
$$\therefore d\mu_t = \left(\mu_t \mu_t' + \frac{1}{2}\sigma_t^2 \mu_t''\right)dt + (\sigma_t \mu_t')dW_t$$

$$\begin{aligned} d\mu_t &= \left(\mu_t \mu_t' + \tfrac{1}{2}\sigma_t^2 \mu_t''\right) dt + \left(\sigma_t \mu_t'\right) dW_t \\ d\sigma_t &= \left(\mu_t \sigma_t' + \tfrac{1}{2}\sigma_t^2 \sigma_t''\right) dt + \left(\sigma_t \sigma_t'\right) dW_t \end{aligned}$$

The integral form of the two SDEs above at time s such that $t < s < t + dt$

$$\begin{aligned} \mu_s &= \mu_t + \int_t^s \left(\mu_u \mu_u' + \tfrac{1}{2}\sigma_u^2 \mu_u''\right) du + \int_t^s \left(\sigma_u \mu_u'\right) dW_u \\ \sigma_s &= \sigma_t + \int_t^s \left(\mu_u \sigma_u' + \tfrac{1}{2}\sigma_u^2 \sigma_u''\right) du + \int_t^s \left(\sigma_u \sigma_u'\right) dW_u. \end{aligned}$$

Substituting for μ_s and σ_s in (2) gives

$$\begin{aligned} S_{t+\delta t} = S_t &+ \int_t^{t+\delta t} \left(\mu_t + \int_t^s \left(\mu_u \mu_u' + \tfrac{1}{2}\sigma_u^2 \mu_u''\right) du + \int_t^s \left(\sigma_u \mu_u'\right) dW_u\right) ds + \\ &\int_t^{t+\delta t} \left(\sigma_t + \int_t^s \left(\mu_u \sigma_u' + \tfrac{1}{2}\sigma_u^2 \sigma_u''\right) du + \int_t^s \left(\sigma_u \sigma_u'\right) dW_u\right) dW_s. \end{aligned}$$

ignored

Now look at the orders

$$\begin{aligned} ds du &= O\left(dt^2\right) && \text{(a)} \\ ds dW_u &= O\left(dt^{3/2}\right) = du\, dW_s && \text{(b)} \\ dW_u dW_s &= O\left(dt\right) && \text{(c)} \end{aligned}$$

therefore we can ignore double integrals of type (a), (b). This gives

$$\begin{aligned} S_{t+\delta t} &\approx S_t + \int_t^{t+\delta t} \mu_t ds + \int_t^{t+\delta t} \sigma_t dW_s + \int_t^{t+\delta t} \int_t^s \left(\sigma_u \sigma_u'\right) dW_u dW_s \\ &= S_t + \mu_t \delta t + \sigma_t \underbrace{\Delta W_t}_{= W_{t+\delta t} - W_t} + \int_t^{t+\delta t} \int_t^s \left(\sigma_u \sigma_u'\right) dW_u dW_s \end{aligned} \tag{3}$$

Now focus on approximating the double integral

$$\begin{aligned} \int_t^{t+\delta t} \int_t^s \left(\sigma_u \sigma_u'\right) dW_u dW_s &= \sigma_t \sigma_t' \int_t^{t+\delta t} \left(W_s - W_t\right) dW_s \\ &= \sigma_t \sigma_t' \left(\int_t^{t+\delta t} W_s dW_s - W_t dW_s\right) \end{aligned} \tag{4}$$

We know from earlier work using the stochastic integral formula that

$$\int_t^{t+\delta t} W_s dW_s = \frac{1}{2}W_{t+\delta t}^2 - \frac{1}{2}W_t^2 - \frac{1}{2}\delta t$$

and

$$\begin{aligned}\left(\int_t^{t+\delta t} W_t dW_s\right) &= W_t \int_t^{t+\delta t} dW_s \\ &= W_t \left(W_{t+\delta t} - W_t\right) = W_t W_{t+\delta t} - W_t^2\end{aligned}$$

Putting these in the integral term of expression (4)

$$\begin{aligned}\int_t^{t+\delta t} W_s dW_s - W_t dW_s &= \frac{1}{2}W_{t+\delta t}^2 - \frac{1}{2}W_t^2 - \frac{1}{2}\delta t - W_t W_{t+\delta t} + W_t^2 \\ &= \frac{1}{2}\left(W_{t+\delta t}^2 + W_t^2 - 2W_t W_{t+\delta t} - \delta t\right) \\ &= \frac{1}{2}\left(W_{t+\delta t} - W_t\right)^2 - \frac{1}{2}\delta t \\ &= \frac{1}{2}\left(\Delta W_t\right)^2 - \frac{1}{2}\delta t\end{aligned}$$

So (4) becomes

$$\sigma_t \sigma_t' \left(\int_t^{t+\delta t} W_s dW_s - W_t dW_s\right) = \sigma_t \sigma_t' \times \frac{1}{2}\left(\phi^2 - 1\right)\delta t$$

and we are able to write the earlier expression (3) as

$$\begin{aligned}S_{t+\delta t} &\approx S_t + \int_t^{t+\delta t} \mu_t ds + \int_t^{t+\delta t} \sigma_t dW_s + \int_t^{t+\delta t}\int_t^s \left(\sigma_u \sigma_u'\right) dW_u dW_s \\ &= S_t + \mu_t \delta t + \sigma_t \Delta W_t + \sigma_t \sigma_t' \times \frac{1}{2}\left(\phi^2 - 1\right)\delta t \\ &= S_t\left(1 + \mu\delta t + \sigma\phi\sqrt{\delta t} + \sigma^2 \times \frac{1}{2}\left(\phi^2 - 1\right)\delta t\right).\end{aligned}$$

To conclude, a SDE for the process Y_t

$$dY_t = A\left(Y_t, t\right)dt + B\left(Y_t, t\right)dW_t$$

can be discretized using Milstein as

$$\boxed{Y_{i+1} = Y_i + A\delta t + B\phi\sqrt{\delta t} + \frac{1}{2}B\frac{\partial B}{\partial Y_i}\left(\phi^2 - 1\right)\delta t,}$$

where $\frac{1}{2}\left(\phi^2 - 1\right)\delta t$ is the **Milstein correction term.** The same random number $\phi \sim N\left(0, 1\right)$ is used per time-step.

Let us remind ourselves of some basic statistics terminology:

An *estimator* is a rule for calculating an estimate for a given estimate (e.g. some statistical parameter) based on observed data. As a simple example consider a fixed set of n i.i.d observations $\{x_i\}_{1\leq i\leq n}$ from a given distribution. Then the sample mean

$$\widehat{\mu} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

is an **estimator** for the population mean μ.

In general an estimator $\widehat{\theta}$ is an unbiased estimator of θ if

$$\mathbb{E}\left[\widehat{\theta}\right] = \theta.$$

In other words in the average we get to the correct value for our estimate. So in the earlier example Now define the *bias* of an estimator $\widehat{\theta}$ as

$$B\left(\widehat{\theta}\right) \equiv \mathbb{E}\left[\widehat{\theta}\right] - \theta.$$

We can write

$$\begin{aligned} \widehat{\theta} - \theta &= \left(\widehat{\theta} - \mathbb{E}\left[\widehat{\theta}\right]\right) + \left(\mathbb{E}\left[\widehat{\theta}\right] - \theta\right) \\ &= \left(\widehat{\theta} - \mathbb{E}\left[\widehat{\theta}\right]\right) + B\left(\widehat{\theta}\right) \end{aligned}$$

hence an estimator for which $B = 0$ is an unbiased estimator.

So in the example of the mean above, in the average we want to get the correct value for our estimate so that

$$\mathbb{E}\left[\widehat{\mu}\right] = \mu,$$

i.e. the estimate $\widehat{\mu}$ is said to be unbiased if its expected value $\mathbb{E}\left[\widehat{\mu}\right]$ is equal to its theoretical value, μ. The bias which was defined as the difference of the expecetd value and the true value becomes

$$B\left(\widehat{\mu}\right) \equiv \mathbb{E}\left[\widehat{\mu}\right] - \mu.$$

The *mean square error,* $\mathbf{MSE}\left(\widehat{\theta}\right)$ is given by

$$\begin{aligned} \mathbf{MSE}\left(\widehat{\theta}\right) &= \mathbb{E}\left[\left(\widehat{\theta} - \theta\right)^2\right] \\ &= 2^{\text{nd}} \text{ moment} \end{aligned}$$

and the *root mean square* $\mathbf{RMSE}\left(\widehat{\theta}\right)$ is simply the square root of the above

$$\begin{aligned} \mathbf{RMSE}\left(\widehat{\theta}\right) &= \sqrt{\mathbf{MSE}\left(\widehat{\theta}\right)} \\ &= \text{standard deviation} \end{aligned}$$

There are two properties of a *good* estimator $\widehat{\theta}$

1. $\widehat{\theta}$ is unbiased. That is

$$\mathbb{E}\left[\widehat{\theta}\right] = \theta$$

2. $\widehat{\theta}$ is consistent. That is

$$\widehat{\theta} \longrightarrow \theta \text{ with probability 1.}$$

This follows from the Strong Law of Large Numbers.

Monte-Carlo methods are centred on evaluating definite integrals as expectations (or averages). Before studying this in greater detail, we consider the simple problem of estimating expectations of functions of uniformly distributed random numbers.

Motivating Example: Estimate $\theta = \mathbb{E}\left[e^{U^2}\right]$, where $U \sim U(0,1)$.

We note that $\mathbb{E}\left[e^{U^2}\right]$ can be expressed in integral form, i.e.

$$\mathbb{E}\left[e^{U^2}\right] = \int_0^1 e^{x^2} p(x)\, dx$$

where $p(x)$ is the density function of a $U(0,1)$

$$p(x) = \begin{cases} 1 & 0 < x < 1 \\ 0 & \text{otherwise} \end{cases}$$

hence

$$\mathbb{E}\left[e^{U^2}\right] = \int_0^1 e^{x^2} dx.$$

This integral does not have an analytical solution. The theme of this section is to consider solving numerically, using simulations. We use the Monte Carlo simulation procedures:

1. Generate a sequence $U_1, U_2, ..., U_n \sim U(0,1)$ where U_i are i.i.d (independent and identically distributed)
2. Compute $Y_i = e^{U_i^2}$ $(i = 1, ..., n)$
3. Estimate θ by

$$\begin{aligned} \widehat{\theta}_n &\equiv \frac{1}{n}\sum_{i=1}^{n} Y_i \\ &= \frac{1}{n}\sum_{i=1}^{n} e^{U_i^2} \end{aligned}$$

i.e. use the sample mean of the $e^{U_i^2}$ terms.

Why is this a good procedure? That is, why is $\widehat{\theta}_n$ a good estimator of θ?

Recall there are two properties that need to be a satisfied

1. $\widehat{\theta}_n$ is unbiased. So we need to show $\mathbb{E}\left[\widehat{\theta}_n\right] = \theta$.
   We know

$$\begin{aligned}
\mathbb{E}\left[\widehat{\theta}_n\right] &= \mathbb{E}\left[\frac{\sum_{i=1}^{n} e^{U_i^2}}{n}\right] \\
&= \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}\left[e^{U_i^2}\right] \\
&= \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}\left[e^{U^2}\right] = \frac{1}{n} \times n\mathbb{E}\left[e^{U^2}\right] \\
&= \mathbb{E}\left[e^{U^2}\right] = \theta.
\end{aligned}$$

2. $\widehat{\theta}_n$ is consistent, i.e. $\widehat{\theta}_n \to \theta$ with probability 1 as $n \to \infty$.

$$\begin{aligned}
\lim_{n\to\infty}\widehat{\theta}_n &= \lim_{n\to\infty}\frac{\sum_{i=1}^{n} e^{U_i^2}}{n} \\
&= \lim_{n\to\infty}\frac{\sum_{i=1}^{n} Y_i}{n}, \text{ where } Y_i \equiv e^{U_i^2} \\
&= \mathbb{E}[Y] \text{ by the Strong Law of large numbers} \\
&= \mathbb{E}\left[e^{U^2}\right] \text{ since } Y \equiv e^{U^2} \\
&= \theta.
\end{aligned}$$

Monte Carlo Integration

When a closed form solution for evaluating an integral is not available, numerical techniques are used. The purpose of Monte Carlo schemes is to use simulation methods to approximate integrals in the form of expectations.

Suppose $f(\cdot)$ is some function such that $f : [0,1] \to \mathbb{R}$. The basic problem is to evaluate the integral

$$I = \int_0^1 f(x)\, dx$$

i.e. diagram

Consider e.g. the earlier problem $f(x) = e^{x^2}$, for which an analytical solution cannotbe obtained.

Note that if $U \sim U(0,1)$ then

$$\mathbb{E}[f(U)] = \int_0^1 f(u)\, p(u)\, du$$

where the density $p(u)$ of a uniformly distributed random variable $U(0,1)$ is given earlier. Hence

$$\begin{aligned} \mathbb{E}[f(U)] &= \int_0^1 f(u)\, p(u)\, du \\ &= I. \end{aligned}$$

So the problem of estimating I becomes equivalent to the exercise of estimating $\mathbb{E}[f(U)]$ where $U \sim U(0,1)$.

Very often we will be concerned with an arbitrary domain, other than $[0,1]$. This simply means that the initial part of the problem will involve seeking a transformation that converts $[a,b]$ to the domain $[0,1]$. We consider two fundamental cases.

1. Let $f(\cdot)$ be a function s.t. $f : [a,b] \to \mathbb{R}$ where $-\infty < a < b < \infty$. The problem is to evaluate the integral
$$I = \int_a^b f(x)\, dx.$$
In this case consider the following substitution
$$y = \frac{x-a}{b-a}$$
which gives $dy = dx/(b-a)$. This gives
$$\begin{aligned} I &= (b-a)\int_0^1 f(y \times (b-a) + a)\, dy \\ &= (b-a)\,\mathbb{E}[f(U \times (b-a) + a)] \end{aligned}$$
where $U \sim U(0,1)$. Hence I has been expressed as the product of a constant and expected value of a function of a $U(0,1)$ random number; the latter can be estimated by simulation.

2. Let $g(\cdot)$ be some function s.t. $g : [0,\infty) \to \mathbb{R}$ where $-\infty < a < b < \infty$. The problem is to evaluate the integral
$$I = \int_0^\infty g(x)\, dx,$$

provided $I < \infty$. So this is the area under the curve $g(x)$ between 0 and ∞. In this case use the following substitution

$$y = \frac{1}{1+x}$$

which is equivalent to $x = -1 + \frac{1}{y}$. This gives

$$\begin{aligned} dy &= -dx/(1+x)^2 \\ &= -y^2 dx. \end{aligned}$$

The resulting problem is

$$\begin{aligned} I &= \int_0^1 \frac{g\left(\frac{1}{y}-1\right)}{y^2} dy \\ &= \mathbb{E}\left[\frac{g\left(-1+\frac{1}{U}\right)}{U^2}\right] \end{aligned}$$

where $U \sim U(0,1)$. Hence I has again been expressed as the expected value of a function of a $U(0,1)$ random number; to be estimated by simulation.

Consider the integral over the unit interval

$$\overline{f} = \int_0^1 f(x)\, dx = \mathbb{E}_{U(0,1)}[f]$$

where $\mathbb{E}_{U(0,1)}[\cdot]$ is an expectation with respect to the uniformly distributed random variable on $(0,1)$.

Similarly on a unit cube $\mathbf{I}^3$ the problem becomes

$$\int_{\mathbf{I}^3} f(x)\, dx = \mathbb{E}_{\mathbf{Unif}_{[\mathbf{I}^3]}}[f],$$

which we denote as $\overline{f}$.

We could write $\theta = \int_0^1 f(x)\, dx$, i.e. θ being an estimator for the integral.

To approximate the integral we can take a sequence $\{x_n\}_{1 \le n \le N}$ of random numbers drawn from a uniform distribution on $\mathbf{I}^3$ and compute

$$J_N(f) = \frac{1}{N}\sum_{n=1}^{N} f(x_i)$$

Monte Carlo Estimation

Consider a *random vector* $\mathbf{Y} = (Y_1, ..., Y_n)^{\mathrm{T}} \in \mathbb{R}^n$ and a function $h(\cdot)$ s.t.

$$h : \mathbb{R}^n \longrightarrow \mathbb{R}$$

The aim is to estimate $\theta = \mathbb{E}[h(\mathbf{Y})]$. Clearly we want $\mathbb{E}[h(\mathbf{Y})] < \infty$.

$\mathbf{Y}$ could represent the values of a stochastic processes at different points in time. As a particular example suppose Y_i is a stock price at time i, with $h(\cdot)$ defined by

$$h(\mathbf{Y}) = \frac{1}{n}\sum_{i=1}^{n} Y_i$$

So then θ is the expected average value of the stock price.

The MC algorithm for the estimation of θ can be written

$$\begin{array}{c} \textbf{for } i = 1 \textbf{ to } n \\ \textbf{simulate } Y_i \\ \textbf{set } h_i = h\left(Y_i\right) \\ \textbf{set } \widehat{\theta}_n = \dfrac{h_1 + h_2 + ... + h_n}{n} \end{array}$$

A note on computation: If n is large, a sensible step would be to keep track of $\sum_i h_i$ within the for loop so as not to store each value of h_i.

So why is $\widehat{\theta}$ a good estimator? Because of the two reasons considered earlier.

1. $\widehat{\theta}$ **is unbiased**. That is

$$\mathbb{E}\left[\widehat{\theta}\right] = \frac{1}{n}\mathbb{E}\left[\sum_i h_i\right] = \frac{1}{n}\mathbb{E}\left[\sum_i h\left(Y_i\right)\right] = \frac{1}{n} n\theta = \theta.$$

2. $\widehat{\theta}$ **is consistent.** That is

$$\widehat{\theta}_n \to \theta \text{ with probabilty 1 as } n \to \infty.$$

This follows from the Strong Law of Large Numbers (SLLN).

Example: Describe a Monte Carlo algorithm for estimating

$$\theta = \int_0^\infty e^{-x^3} dx.$$

To estimate θ requires a change of variable (and limits of integration) since we are working with $\mathbf{Unif}_{[0,1]}$. Consider the transformation

$$\begin{array}{rcl} x & = & \dfrac{1-y}{y} \\ dx & = & -\dfrac{1}{y^2} dy \\ \displaystyle\int_0^\infty f\left(x\right) dx & \longrightarrow & \displaystyle\int_1^0 F\left(y\right) dy \end{array}$$

So

$$\begin{array}{rcl} \theta & = & \displaystyle\int_0^\infty e^{-x^3} dx = \int_1^0 e^{-(-1+1/y)^3} \left(-\tfrac{1}{y^2}\right) dy \\ & = & \displaystyle\int_0^1 \tfrac{1}{y^2} e^{-(-1+1/y)^3} dy \end{array}$$

where the final integral is an expectation, written

$$\mathbb{E}\left[\tfrac{1}{U^2} e^{-(-1+1/U)^3}\right]$$

of a random variable $U \sim U\left(0,1\right)$.

The Monte Carlo algorithm now becomes

1. Simulate $\{U_i\}_{i=1,.....,N} \sim U(0,1)$

2. Calculate $X_i = \frac{1}{U_i^2} e^{-(-1+1/U_i)^3}$ for each $i = 1, ..., N$

3. Put $\widehat{\theta} = \frac{1}{N} \sum_{n=1}^{N} X_i$

The integral approximation $J_N(f)$ defined earlier is an unbiased estimator, i.e.

$$\mathbb{E}_{U(0,1)}[J_N(f)] = \mathbb{E}_{U(0,1)}[f] = \int_0^1 f(x)\, dx$$

$J_N(f)$ converges, i.e.

$$\lim_{N \longrightarrow \infty} J_N(f) = \mathbb{E}_{U(0,1)}[f].$$

Now define the error $\varepsilon_N(f)$ as

$$\begin{aligned} \varepsilon_N(f) &= \mathbb{E}_{U(0,1)}[f] - J_N(f) \\ &= \int_{\mathbf{I}^3} f(x)\, dx - J_N(f). \end{aligned}$$

Since the distribution is $U(0,1)$,

$$\mathbb{E}[\varepsilon_N(f)] = 0.$$

This allows us to define the **RMSE** (standard deviation of the error) to be

$$\mathbb{E}_{U(0,1)}\left[(\varepsilon_N(f))^2\right]$$

As with any numerical scheme, the size of the associated errors is a chief concern. The Central Limit Theorem (CLT) describes the statistical properties of the errors involved in Monte Carlo integration. Assuming we are sampling from a distribution with a finite second moment, the CLT asserts that

$$\lim_{N \longrightarrow \infty} \varepsilon_N(f) \sim \frac{1}{\sqrt{N}} \phi \sigma_f$$

where $\phi \sim N(0,1)$ is a standard normal and σ_f the standard deviation of f, given by

$$\sigma_f^2 = \int_{\mathbf{I}^3} \left(f(x) - \underbrace{\mathbb{E}_{U(0,1)}[f]}_{\overline{f}} \right)^2 dx.$$

Transformation Methods

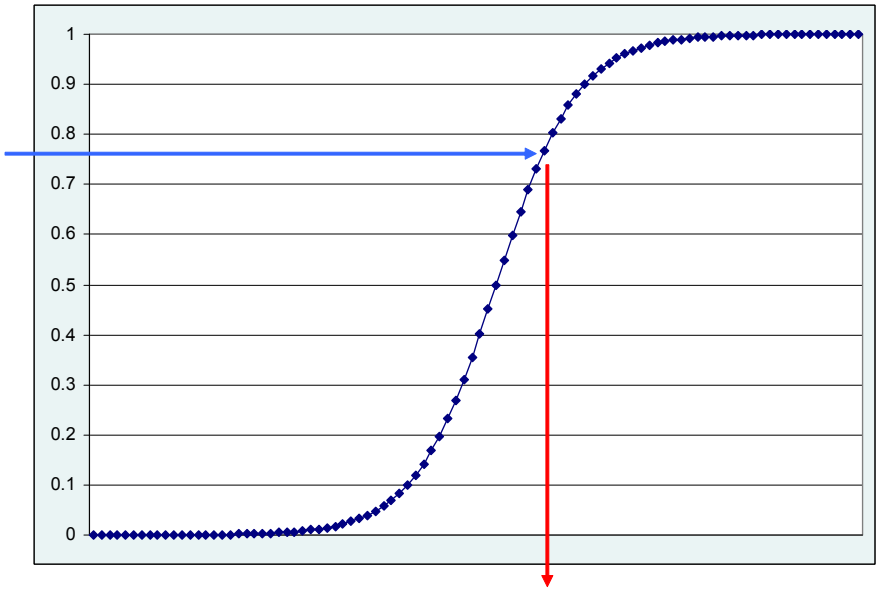
Most programming languages have random number generators that produce uniformly distributed random numbers. Then applying various transformations, these can be conveniently converted to non-uniformly distributed random numbers. We know that given a non-uniform random variable with density, $p(x)$ its expectation and error in turn are

$$\begin{aligned} \mathbb{E}[f(x)] &= \int_{\mathbb{R}} f(x)\, p(x)\, dx \\ \varepsilon_N(f) &= \int_{\mathbb{R}} f(x)\, p(x)\, dx - \frac{1}{N} \sum_{n=1}^{N} f(x_n) \end{aligned}$$

The CLT gives

$$\begin{aligned}\varepsilon_N(f) &\sim \frac{1}{\sqrt{N}}\phi\sigma_f \\ \sigma_f^2 &= \int_{\mathbb{R}} \left(f(x)-\overline{f}\right)^2 dx\end{aligned}$$

Earlier we discussed how to create random variables following $N(0,1)$ in excel using **NORMSINV(RAND())**. This represents the inverse cumulative density function with a uniformly distributed RV $U(0,1)$ as the function parameter.



Suppose y is a RV which is $U(0,1)$ and is to be converted into another variable x with pdf $p(x)$. If its CDF $F(x)$ is defined as

$$F(x) = \int_{-\infty}^{x} p(s)\, ds$$

then we wish to invert this to obtain

$$x = F^{-1}(y) : y \sim U(0,1)$$

in a computationally efficient manner, assuming of course that F^{-1} exists. Recall that the **NORMSINV()** function is an accurate but slow mode of numerically inverting the integral.

Justification: To turn a $y \sim U(0,1)$ RV into a random variable with density $p(x)$, we know the CDF $F(x)$ is

$$F(x) = \int_{-\infty}^{x} p(s)\, ds.$$

Introduce $x = X(y)$ where X is to be determined. We know

$$\mathbb{E}[f(x)] = \mathbb{E}_{U(0,1)}[f(X(y))]$$

which in integral form

$$\int f(x)\, p(x)\, dx = \int f(X(y))\, dy$$

The second integral becomes (using $x = X(y)$)

$$\int f(x)\, p(x)\, dx = \int f(x)\, \frac{dy}{dx} dx$$

where $\frac{dx}{dy} = X'(y) = X_y$ so $\frac{dy}{dx} = X_y^{-1}$.
Thus

$$p(x) = \frac{dy}{dx} = X_y^{-1}$$

hence $p(x)\, dx = dy$.
This gives

$$F(X(y)) = y$$

i.e. $X(y) = F^{-1}(y)$.
In summary, to transform a $U(0,1)$ RV to x from a density $p(x)$ by the inverse function of the CDF F^{-1}

$$x = F^{-1}(y) : Y \sim U(0,1)$$

we are assumming that F^{-1}. We will only use this method if computing F^{-1} is indeed practical.
Consider the following simple case of generating x from a unit Cauchy distribution $p(x) = \frac{\pi^{-1}}{1+x^2}$.
Firstly write

$$\begin{aligned} F(x) &= \tfrac{1}{\pi}\int_{-\infty}^{x} \frac{1}{1+s^2} ds \\ &= \tfrac{1}{\pi}\left(\arctan x + \tfrac{\pi}{2}\right) \end{aligned}$$

upon rearranging we have

$$\begin{aligned} F(x) - \tfrac{1}{2} &= \tfrac{1}{\pi} \arctan x \\ x &= \tan\left(\pi\left(F(x) - \tfrac{1}{2}\right)\right) \end{aligned}$$

i.e.

$$\begin{aligned} x &= F^{-1}(y) \\ &= \tan\left(\pi\left(y - \tfrac{1}{2}\right)\right). \end{aligned}$$

As a second example we wish to generate from a Normal distribution.
If $p(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$, then the CDF $N(x)$ is given by

$$N(x) = \tfrac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-t^2/2} dt$$

and related to the error function

$$\operatorname{erf}(x) = \frac{2}{\sqrt{\pi}} \int_{0}^{x} e^{-t^2/2} dt$$

through

$$N(x) = \frac{1}{2} + \frac{1}{2} \operatorname{erf}\left(x/\sqrt{2}\right).$$

So to sample a standard normal x, using a uniform variable $y \sim U(0,1)$, write

$$y = \frac{1}{2} + \frac{1}{2} \operatorname{erf}\left(x/\sqrt{2}\right)$$

and rearrange to get

$$x = \sqrt{2}\,\mathrm{erf}^{-1}\left(2y-1\right).$$

Numerical Algorithms for evaluating $N^{-1}\left(x\right)$ are available in some languages. In excel this is NORMSINV(RAND (which we have already seen.

We need a numerical technique for the conversion. Most programming languages generate uniformly distributed random variables over 0 and 1. How can these be transformed to standard normals $N\left(0,1\right)$?

The Box Müller Method

Recall the CDF for the standardized normal distribution is defined as

$$N\left(x\right) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} e^{-s^2/2}ds.$$

In this technique the inefficient inversion process of $N\left(x\right)$ is not required as the random number $x \sim N\left(0,1\right)$ can be directly captured using the following algorithm:

1. generate two independent uniformly distributed random variables $y_1, y_2 \overset{\text{i.i.d}}{\sim} U\left(0,1\right)$
2. compute $x_1, x_2 \sim N\left(0,1\right)$ by

$$\begin{aligned} x_1 &= \sqrt{-2\log y_1}\cos\left(2\pi y_2\right) \\ x_2 &= \sqrt{-2\log y_1}\sin\left(2\pi y_2\right). \end{aligned}$$

   which are independent standard varaibles.

The procedure underlying this technique is as follows.

Consider the joint distribution of two independent normal variables $\left(X,Y\right) \overset{\text{i.i.d}}{\sim} N\left(0,1\right)$ i.i.d given by

$$\begin{aligned} F\left(x,y\right) &= \int_{-\infty}^{x}\tfrac{1}{\sqrt{2\pi}}e^{-x'^2/2}dx'\int_{-\infty}^{y}\tfrac{1}{\sqrt{2\pi}}e^{-y'^2/2}dy' \\ &= \tfrac{1}{2\pi}\int_{-\infty}^{y}\int_{-\infty}^{x} e^{-\frac{1}{2}\left(x'^2+y'^2\right)}dx'dy' \end{aligned}$$

$$\Phi\left(x_1,x_2\right) = \frac{1}{2\pi}\exp\left(-\tfrac{1}{2}\left(x_1^2+x_2^2\right)\right).$$

Now, define random variables R, θ as

$$R = \sqrt{X^2+Y^2};\ \theta = \arctan\frac{Y}{X} \text{ with } \theta \in [0,2\pi]$$

$$X = R\cos\theta;\ Y = R\sin\theta$$

Then express $F\left(x,y\right)$ using r,θ and $dxdy = rdrd\theta$

$$F\left(x,y\right) = F\left(r,\theta\right) = \tfrac{1}{2\pi}\int_0^{\theta}\int_0^{r} e^{-\frac{1}{2}r'^2}r'dr'd\theta'$$

where we know $\int_0^r e^{-\frac{1}{2}r'^2} r'dr' = 1 - e^{-\frac{1}{2}r^2}$. Hence the double integral simplifies to

$$\begin{aligned} \tfrac{1}{2\pi}\int_0^\theta e^{-\frac{1}{2}r'^2} d\theta' &= \tfrac{1}{2\pi}\left(1 - e^{-\frac{1}{2}r^2}\right)\theta'\Big|_0^\theta \\ &= \left(\tfrac{\theta}{2\pi}\right)\left(1 - e^{-\frac{1}{2}r^2}\right) \\ &= F(r)\,F(\theta). \end{aligned}$$

Now draw $U_1, U_2 \overset{\text{i.i.d}}{\sim} U(0,1)$. Then we say as follows

$$F(\theta) \leq u_1,\ F(r) \leq u_2.$$

Since

$$U_2 \in [0,1] \sim U(0,1)$$

we can say

$$1 - U_2 \in [0,1] \sim U(0,1).$$

Then

$$F(\theta) \leq u_1;\ F(r) \leq 1 - u_2$$

Therefore we can get $u_1,\ u_2$ as follows. Firstly

$$u_1 = \tfrac{\theta}{2\pi} \Rightarrow \theta = 2\pi u_1.$$

Secondly

$$\begin{aligned} 1 - u_2 &= 1 - e^{-\frac{1}{2}r^2}, \\ u_2 &= e^{-\frac{1}{2}r^2} \Rightarrow r = \sqrt{-2\log u_2} \end{aligned}$$

Now we can generate X, Y using $U_1,\ U_2$

$$\begin{aligned} X &= R\cos\theta = \sqrt{-2\log U_2}\cos 2\pi U_1 \\ Y &= R\sin\theta = \sqrt{-2\log U_2}\sin 2\pi U_1 \end{aligned}$$

The disadvantage with this method lies in the computation of the trigonometric and transcendental functions sin, cos and log. This leads on to a more efficient scheme which employs an *acceptance-rejection* method.

Polar Marsaglia Method

1. Generate $U_1, U_2 \overset{\text{i.i.d}}{\sim} U(0,1)$
2. Set
$$\left.\begin{array}{l} V_1 = 2U_1 - 1 \\ V_2 = 2U_2 - 1 \end{array}\right\} \quad R = V_1^2 + V_2^2$$
where $(V_1, V_2) \sim U(-1,1)$.
3. While $R \leq 1$, draw $U \sim U(0,1)$ and returnset
$$X = \sqrt{\frac{-2\log U}{R}}V_1\ ;\ Y = \sqrt{\frac{-2\log U}{R}}V_2$$
Else go back to step 1. Can also use U in place of R.

Here

$$\cos(2\pi U_1) \text{ is replaced with } \frac{V_1}{\sqrt{R}}$$
$$\sin(2\pi U_1) \text{ is replaced with } \frac{V_2}{\sqrt{R}}$$

Then $X, Y \sim N(0,1)$. The probability of of S being accepted (i.e. area of unit circle to area of square) is

$$\mathbb{P}(S \leq 1) = \frac{\text{Area of circle}}{\text{Area of square}} = \frac{\pi \times 1^2}{2 \times 2} = \frac{\pi}{4} \approx 0.785$$

which means less than 21.5% of uniform deviates $V_1,\ V_2$ are rejected for which $S > 1$. This is far more efficient than the BM Method.

Generating Correlated Normal Random Variables

Earlier we looked at how to obtain correlated random variables given a pair of uncorrelated ones that follow $N(0,1)$. Recall, if X, Y are random variables then the covariance written σ_{XY} is

$$\begin{aligned}\mathbf{Cov}(X,Y) &= \sigma_{XY} \\ &= \mathbb{E}[X,Y] - \mathbb{E}[X]\,\mathbb{E}[Y]\end{aligned}$$

and the correlation of X and Y written ρ_{XY}

$$\rho_{XY} = \frac{\sigma_{XY}}{\sigma_X \sigma_Y},$$

where $-1 \leq \rho_{XY} \leq 1$ (due to Cauchy-Schwartz for random variables).

If X, Y are independent then $\rho_{XY} = 0$, the converse is not generally true. This is a theorem which most students encounter for the first time in linear algebra. Let's start off with the version for random variables (RVs) X and Y, then the Cauchy-Schwartz inequality is

$$[\mathbb{E}[XY]]^2 \leq \mathbb{E}\left[X^2\right]\mathbb{E}\left[Y^2\right].$$

We know that the covariance of X, Y is

$$\sigma_{XY} = \mathbb{E}[(X-\mu_X)(Y-\mu_Y)]$$

If we put

$$\begin{aligned}\mathbb{V}[X] &= \sigma_X^2 = \mathbb{E}\left[(X-\mu_X)^2\right] \\ \mathbb{V}[Y] &= \sigma_Y^2 = \mathbb{E}\left[(Y-\mu_Y)^2\right].\end{aligned}$$

From Cauchy-Schwartz we have

$$\left(\mathbb{E}[(X-\mu_X)(Y-\mu_Y)]\right)^2 \leq \mathbb{E}\left[(X-\mu_X)^2\right]\mathbb{E}\left[(Y-\mu_Y)^2\right]$$

or we can write

$$\sigma_{XY}^2 \leq \sigma_X^2 \sigma_Y^2$$

Divide through by $\sigma_X^2 \sigma_Y^2$

$$\frac{\sigma_{XY}^2}{\sigma_X^2 \sigma_Y^2} \leq 1$$

and we know that the left hand side above is ρ_{XY}^2, hence

$$\rho_{XY}^2 = \frac{\sigma_{XY}^2}{\sigma_X^2 \sigma_Y^2} \leq 1$$

and since ρ_{XY} is a real number, this implies $|\rho_{XY}| \leq 1$ which is the same as

$$-1 \leq \rho_{XY} \leq +1.$$

Suppose now that $\mathbf{Z} = (z_1, z_2, ..., z_n)^{\mathrm{T}}$ is a random vector where each $z_i \sim N(0,1)$ and are i.i.d.

Then define the *covariance matrix* Σ of $\mathbf{Z}$ as the $n \times n$ matrix that has element at (i,j) given by

$$\begin{aligned}\Sigma_{ij} &= \mathbf{Cov}(Z_i, Z_j) \\ &= \sigma_{Z_i Z_j}\end{aligned}$$

So for example, in the case of a 3×3, we have

$$\begin{aligned} \Sigma &= \begin{pmatrix} \sigma_1^2 & \rho_{12}\sigma_1\sigma_2 & \rho_{13}\sigma_1\sigma_3 \\ \rho_{21}\sigma_2\sigma_1 & \sigma_2^2 & \rho_{23}\sigma_2\sigma_3 \\ \rho_{31}\sigma_3\sigma_1 & \rho_{32}\sigma_3\sigma_2 & \sigma_3^2 \end{pmatrix} \\ \Sigma_C &= \begin{pmatrix} 1 & \rho_{12} & \rho_{13} \\ \rho_{21} & 1 & \rho_{23} \\ \rho_{31} & \rho_{32} & 1 \end{pmatrix} \end{aligned}$$

The properties of Σ are

- It is a symmetric matrix $\Sigma^{\mathrm{T}} = \Sigma$
- The leading diagonal elements are non-negative $\Sigma_{ii} \geq 0$
- It is *positive semi-definite,* (PSD) i.e. $\forall \mathbf{y} \in \mathbb{R}^n$, $\mathbf{y}^{\mathrm{T}}\Sigma\mathbf{y} \geq 0$.

The new vector $\mathbf{X} = (x_1, x_2, ..., x_n)^{\mathrm{T}}$ where $\mathbf{X} \sim \mathbf{MN}(0, \Sigma)$, i.e. a multivariate normal distribution. We write

$$\mathbf{X} = L\mathbf{Z}$$

such that

$$\boxed{\Sigma = LL^{\mathrm{T}}}$$

Cholesky Decomposition

Here we perform Cholesky factorisation of a symmetric positive-definite matrix, M. Such a matrix can be written as

$$M = lDl^{\mathrm{T}},$$

where

l is a lower triangular matrix

D is a diagonal matrix with positive elements

So we can write

$$\begin{aligned} \Sigma &= lDl^{\mathrm{T}} \\ &= \left(l\sqrt{D}\right)\left(\sqrt{D}l^{\mathrm{T}}\right) \\ &= \underbrace{\left(l\sqrt{D}\right)}_{L}\underbrace{\left(l\sqrt{D}\right)^{\mathrm{T}}}_{L^{\mathrm{T}}} \end{aligned}$$

Therefore $L = l\sqrt{D}$ satisfies $\Sigma = LL^{\mathrm{T}}$. It is called the Cholesky Decomposition of Σ.

Consider a 2×2 case, and write

$$\begin{aligned} \Sigma &= \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_2\sigma_1 & \sigma_2^2 \end{pmatrix} = LL^{\mathrm{T}} \\ &= \begin{pmatrix} a & 0 \\ b & c \end{pmatrix}\begin{pmatrix} a & b \\ 0 & c \end{pmatrix} \end{aligned}$$

Equating elements gives

$$L = \begin{pmatrix} \sigma_1 & 0 \\ \rho\sigma_2 & \sigma_2\sqrt{1-\rho^2} \end{pmatrix}.$$

Now put

$$\begin{aligned} \mathbf{X} &= \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = L\mathbf{Z} \\ &= \begin{pmatrix} \sigma_1 & 0 \\ \rho\sigma_2 & \sigma_2\sqrt{1-\rho^2} \end{pmatrix} \begin{pmatrix} z_1 \\ z_2 \end{pmatrix} \\ &= \begin{pmatrix} \sigma_1 z_1 \\ \rho\sigma_2 z_1 + \sigma_2\sqrt{1-\rho^2} z_2 \end{pmatrix}. \end{aligned}$$

In the case of a standard normal distribution we have $\begin{pmatrix} z_1 \\ \rho z_1 + \sqrt{1-\rho^2} z_2 \end{pmatrix}$.

Relationship between derivative values and simulations

The fair value of an option is the present value of the expected payoff at expiry under the risk-neutral random walk for the underlying. (Phelim Boyle 1977)

Recall an amount of cash $M = M(t)$ in the bank grows according to

$$\frac{dM}{dt} = r(t) M$$

where $r(t)$ is the variable (risk-free) interest rate. This differential equation has solution

$$M(t) = M(T) \exp\left(-\int_t^T r(\tau)\, d\tau\right)$$

i.e. the present value (time t) of a future cash flow (time T). The exponential term is the discount factor. In the simple case of a fixed rate of interest it becomes $e^{(-r(T-t))}$.

This gives the fair price of an option V to be

$$V = \mathbb{E}^{\mathbb{Q}}\left[e^{-\int_t^T r(\tau)d\tau} \text{Payoff}(S)\right]$$

where

S =asset price, r = stochastic domestic interest rate, T = expiry, t = current time, $\mathbb{Q}$= risk neutral density.

Scheme: The Monte Carlo method consists of the following steps

1. Simulate sample paths/realizations for the underlying asset price (e.g. equities or interest rates) over the relevant time horizon, according to the risk-neutral measure. Here we use the discretized SDE

$$S_{i+1} = S_i\left(1 + r\delta t + \sigma\phi\sqrt{\delta t}\right).$$

2. Evaluate the discounted cashflows (using domestic rate of interest) of a derivative on each sample path, as determined by the structure of the security being priced.

3. Average the discounted cashflows over sample paths.

So the option price becomes

$$e^{(-r(T-t))}.\frac{1}{N}\sum_{n=1}^{N}\text{Payoff}(S)$$

The payoff for a European Call Option with strike E is $C = \max(S(T) - E,\ 0)$, where $S(T)$ is obtained from

$$S_T = S_0 e^{\left(r-0.5\sigma^2\right)T+\sigma W_T}$$

in discrete form, for each value $1 \leq n \leq N$.

Although we are not concerned with the path followed by the process $S(t)$ in getting to $S(T)$ we will nevertheless simulate this as we can price other options which are *path dependent*.

Based upon the N realizations an estimate for the price of an option becomes $\overline{C}(S,\ t)$

$$\overline{C}(S,\ t) = \frac{1}{N}\sum_{n=1}^{N} C^{(n)}(S,\ T)$$

which is equivalent to

$$e^{(-r(T-t))}.\frac{1}{N}\sum_{n=1}^{N} \max\left(S^{(n)}(T) - E,\ 0\right).$$

If we put $S(T) = S_T$, where

$$S_T = S_0 e^{\left(r-0.5\sigma^2\right)T+\sigma W_T}$$

then an algorithm for the estimator $\widehat{C}$ of the BS option price can be written as

set $sum = 0$
for $i = 1$ **to** n
simulate S_T
set $sum = sum + \max(S_T - E, 0)$
End for
set $\widehat{C} = e^{-r(T-t)}\frac{sum}{n}$

Now consider the example of an Asian call option with arithmetic averaging (fixed strike). Recall at expiry T, the payoff is

$$h(Y_1, ..., Y_n) = \max\left(\tfrac{1}{m}\sum_{i=1}^{m} S_i - E, 0\right).$$

As earlier we can write the option price as $C = \mathbb{E}^{\mathbb{Q}}\left[e^{-r(T-t)}h(\mathbf{Y})\right]$ with the following algorithm

set $sum = 0$
for $i = 1$ **to** n
simulate S_i
set $sum = sum + \max\left(\frac{1}{m}\sum_{i=1}^{m} S_i - E, 0\right)$
End for
set $\widehat{C} = e^{-r(T-t)}\frac{sum}{n}$

Variance Reduction

As before the idea is to estimate $\mathbb{E}[Y]$ where Y is an output random variable obtained from a simulation. We are motivated by the standard error $\varepsilon = \frac{\sigma}{\sqrt{N}}$, where N is the number of sample paths. The main idea is to reduce σ^2. Earlier the method was to simulate Y such that $\mathbb{E}[X] = \mathbb{E}[Y]$ and $\mathbb{V}[X] \leq \mathbb{V}[Y]$. The idea now is to work with the original output variable Y.

Instead of generating a sequence of Y_i random variables in an i.i.d fashion; correlation will be induced to reduce the variance.

For the simulation of an asset price, samples are drawn from a probability (normal) distribution. If these samples are generated in a fashion, which is not entirely random, but in a manner that reduces the fluctuations (i.e. volatility) of the resulting samples, computational time can be reduced considerably to obtain the desired degree of accuracy. A similar effect can be obtained by performing suitable transformations on the function, which forms the basis of the simulation, so that dependency upon the fluctuations arising in the samples is reduced. The disadvantage is that correlations are introduced. It then becomes a choice, whether to compromise computational time over the risk of correlations being introduced. Recall that the standard error ε associated with the Monte Carlo method is

$$\varepsilon = \frac{\sigma}{\sqrt{N}}.$$

Increasing the number of sample paths generated,by increasing N, leads to a reduction in ε. In addition we are able to manipulate the variance, i.e. reduce the value of σ. For this reason a whole area of Monte Carlo, namely *variance reduction techniques* has been developed.

Antithetic Variable Technique

This method attempts to reduce the variance by introducing negative correlation between pairs of observations. The estimation of $\mathbb{E}[X]$ is the main problem, for the ouptut variable (from a simulation) X. So suppose

$$X = h(\mathbf{Z})$$

where the components of the m dimensional vector $\mathbf{Z}$ are independent. The function $h(\cdot)$ is

$$h : \mathbb{R}^m \to \mathbb{R}.$$

As a motivating example suppose X represents the payoff of a path dependent option with sampling dates $0 < t_1 < \cdots < t_{m-1} < t_m = T$. Then note that $X = h(Z_1, Z_2, \cdots, Z_m)$, for some function $h(Z_1, Z_2, \cdots, Z_m)$, where the Z_i terms are i.i.d and $N(0,1)$.

When a sequence of $U_i \sim U(0,1)$ are used to generate $Z \sim N(0,1)$, many simulations can be represented as

A very simple technique is by use of *antithetic variates*, which can reduce computational time and be implemented at no additional effort, was introduced to option pricing by Boyle (1977).

The method, which was initially used in the pricing of a European call option on a dividend paying stock, is outlined below. It is based upon the observation that if $\phi^{(n)} \sim N(0,1)$, then $-\phi^{(n)}$ also has a standard Normal distribution.

In this technique, by using the one set of random numbers generated, two estimates for an option are calculated. If ϕ_i is used to obtain $\overline{C}$, then $-\phi_i$ gives $\widehat{S}(t)$ and hence a second approximation for the option price $\widehat{C}$ where

$$\widehat{C} = e^{(-r(T-t))}.\frac{1}{N}\sum_{n=1}^{N} \max\left(\widehat{S}^{(n)}(T) - E,\ 0\right)$$

The estimate for the option C_μ is now the average of the two values, $\overline{C}$ & $\widehat{C}$, so

$$C_\mu = \frac{\overline{C} + \widehat{C}}{2}$$

The technique converges because of the symmetry of the Normal Distribution. Justification for obtaining C_μ is based upon the distribution of the antithetic variates.

The pairs $\left\{\left(\phi^{(n)}, -\phi^{(n)}\right)\right\}$are distributed more regularly than a collection of $2n$ independent samples with the sample mean over the antithetic pairs always equal to the population mean of 0. The data set has a lower variance.