Chapter 5

Simulating the value of Options

Asian Options

An Asian option, at expiration T, has value determined not by the closing price of the underlying asset as for a European option, but on an average price of the asset over an interval. For example a *discretely sampled Asian call option* on an asset with price process $S(t)$ pays an amount on maturity equal to $\max(0, \bar{S}_k - K)$ where $\bar{S}_k = \frac{1}{k}\sum_{i=1}^{k} S(iT/k)$ is the average asset price at k equally spaced time points in the time interval $(0, T)$. Here, k depends on the frequency of sampling (e.g. if $T = .25$ (years) and sampling is weekly, then $k = 13$. If $S(t)$ follows a geometric Brownian motion, then $\bar{S}_k$ is the sum of lognormally distributed random variables and the distribution of the sum or average of lognormal random variables is very difficult to express analytically. For this reason we will resort to pricing the Asian option using simulation. Notice, however that in contrast to the arithmetic average, the distribution of the *geometric average* has a distribution which can easily be obtained. The geometric mean of n values $X_1, ..., X_n$ is $(X_1 X_2 ... X_n)^{1/n} = \exp\{\frac{1}{n}\sum_{i=1}^n \ln(X_i)\}$ and if the random variables X_n were each lognormally distributed then this results adding the normally distributed random variables $\ln(X_i)$ in the exponent, a much more familiar operation. In fact the sum in the exponent $\frac{1}{n}\sum_{i=1}^n \ln(X_i)$ is normally distributed so the geometric average will have a lognormal distribution.

Our objective is to determine the value of the Asian option $E(V_1)$ with

$$V_1 = e^{-rT} max(0, \bar{S}_k - K)$$

Since we expect geometric means to be close to arithmetic means, a reasonable control variate is the random variable $V_2 = e^{-rT} max(0, \tilde{S}_k - K)$ where $\tilde{S}_k = \{\prod_{i=1}^k S(iT/k)\}^{1/k}$ is the geometric mean. Assume that V_1 and V_2 obtain from the same simulation and are therefore possibly correlated. Of course V_2 is only useful as a control variate if its expected value can be determined analytically or numerically more easily than V_1 but in view of the fact that V_2 has a known lognormal distribution, the prospects of this are excellent. Since $S(t) = S_0 e^{Y(t)}$ where $Y(t)$ is a Brownian motion with $Y(0) = 0$, drift $r - \sigma^2/2$ and diffusion σ, it follows that $\tilde{S}_k$ has the same distribution as does

$$S_0 \exp\{\frac{1}{k}\sum_{i=1}^k Y(iT/k)\}. \tag{5.1}$$

This is a weighted average of the independent normal increments of the process and therefore normally distributed. In particular if we set

$$\begin{aligned}\bar{Y} &= \frac{1}{k}\sum_{i=1}^k Y(iT/k) \\ &= \frac{1}{k}[k(Y(T/k)) + (k-1)\{Y(2T/k) - Y(T/k)\} + (k-2)\{Y(3T/k) - Y(2T/k)\} \\ &+ ... + \{Y(T) - Y((k-1)T/k)\}],\end{aligned}$$

then

$$\begin{aligned} E(\bar{Y}) &= \frac{r-\sigma^2/2}{k}\sum_{i=1}^{k} iT/k \\ &= (r-\frac{\sigma^2}{2})\frac{k+1}{2k}T \\ &= \widetilde{\mu}T, \text{ say,} \end{aligned}$$

and

$$\begin{aligned} var(\bar{Y}) &= \frac{1}{k^2}\{k^2 var(Y(T/k)) + (k-1)^2 var\{Y(2T/k) - Y(T/k)\} + ...\} \\ &= \frac{T\sigma^2}{k^3}\sum_{i=1}^{k} i^2 = \frac{T\sigma^2(k+1)(2k+1)}{6k^2} \\ &= \widetilde{\sigma}^2 T, \quad \text{say.} \end{aligned}$$

The closed form solution for the price $E(V_2)$ in this case is therefore easily obtained because it reduces to the same integral over the lognormal density that leads to the Black-Scholes formula. In fact

$$\begin{aligned} E(V_2) &= E\{e^{-rT}(S_0 e^{\overline{Y}} - K)^+\}, \text{ where } \overline{Y} \sim N(\widetilde{\mu}, \widetilde{\sigma}^2 T) \text{ so} \\ &= E[e^{-rT+\widetilde{\mu}T} S_0 e^{\overline{Y}-\widetilde{\mu}T} - e^{-rT}K]^+ \\ &= E[S_0 e^{(-r+\widetilde{\mu}+\frac{1}{2}\widetilde{\sigma}^2)T}\exp\{\overline{Y} - \widetilde{\mu}T - \frac{1}{2}\widetilde{\sigma}^2 T\} - e^{-rT}K]^+ \\ &= E[S_0 e^{(-r+\widetilde{\mu}+\frac{1}{2}\widetilde{\sigma}^2)T}\exp\{N(-\frac{\widetilde{\sigma}^2 T}{2}, \widetilde{\sigma}^2 T)\} - Ke^{-rT}]^+. \end{aligned}$$

where we temporarily denote a random variable with the Normal(μ, σ^2) distribution by $N(\mu, \sigma^2)$. Recall that the Black-Scholes formula gives the price at time $t = 0$ of a European option with exercise price K, initial stock price S_0,

$$BS(S_0, K, r, T, \sigma) = E(e^{-rT}(S_0 \exp\{N((r - \frac{\sigma^2}{2})T, \sigma^2 T)\} - K)^+ \tag{5.2}$$

$$= E(S_0 \exp\{N(-\frac{\sigma^2 T}{2}, \sigma^2 T)\} - Ke^{-rT})^+ \tag{5.3}$$

$$= S_0\Phi(d_1) - Ee^{-rT}\Phi(d_2) \tag{5.4}$$

where

$$d_1 = \frac{\log(S_0/K) + (r + \sigma^2/2)T}{\sigma\sqrt{T}}, d_2 = d_1 - \sigma\sqrt{T}.$$

Thus $E(V_2)$ is given by the Black-Scholes formula with S_0 replaced by

$$\widetilde{S_0} = S_0 \exp\{T(\frac{\widetilde{\sigma}^2}{2} + \widetilde{\mu} - r)\} = S_0 \exp\{-rT(1 - \frac{1}{k}) - \frac{\sigma^2 T}{12}(1 - \frac{1}{k^2})\}$$

and σ^2 by $\widetilde{\sigma}^2$. Of course when $k = 1$, this gives exactly the same result as the basic Black-Scholes because in this case, the Asian option corresponds to the average of a single observation. For $k > 1$ the effective initial stock price is reduced $\widetilde{S_0} < S_0$ and the volatility parameter is also smaller $\widetilde{\sigma}^2 < \sigma^2$. With lower initial stock price and smaller volatility the price of a European call will decrease, indicating that if an asian option priced using a geometric mean has price lower than a similar European option on the same stock.

Recall from our discussion of a control variate estimators that we can estimate $E(V_1)$ unbiasedly using

$$V_1 - \beta(V_2 - E(V_2)) \tag{5.5}$$

where

$$\beta = \frac{cov(V_1, V_2)}{var(V_2)}. \tag{5.6}$$

In practice, of course, we simulate many values of the random variables V_1, V_2 and replace V_1, V_2 by their averages $\overline{V_1}, \overline{V}_2$ so the resulting estimator is

$$\overline{V_1} - \beta(\overline{V}_2 - E(V_2)). \tag{5.7}$$

Table 4.1 is similar to that in Boyle, Broadie and Glasserman(1997) and compares the variance of the crude Monte Carlo estimator with that of an estimator using a simple control variate,

$$E(V_2) + \overline{V_1} - \overline{V}_2,$$

a special case of (5.7) with $\beta = 1$.We chose $K = 100, k = 50, r = 0.10, T = 0.2$, a variety of initial asset prices S_0 and two values for the volatility parameter

$\sigma = 0.2$ and $\sigma = 0.4$. The efficiency depends only on S_0 and K through the ratio K/S_0 or alternatively the *moneyness* of the option, the ratio $e^{rT}S_0/K$ of the value on maturity of the current stock price to the strike price. Standard errors are estimated from $N = 10,000$ simulations. Since the efficiency is the ratio of the number of simulations required for a given degree of accuracy, or alternatively the ratio of the variances, this table indicates efficiency gains due to the use of a control variate of several hundred. Of course using the control variate estimator (5.7) described above could only improve the efficiency further.

Table 4.1. Standard Errors for Arithmetic Average Asian Options.

σ	Moneyness=$e^{rT}S_0/K$	STANDARD ERROR USING CRUDE MC	STANDARD ERROR USING CONTROL VARIATE
0.2	1.13	0.0558	0.0007
	1.02	0.0334	0.00064
	0.93	0.00636	0.00046
0.4	1.13	0.105	0.00281
	1.02	0.0659	0.00258
	0.93	0.0323	0.00227

The following function implements the control variate for an Asian option and was used to produce the above table.

```
function [v1,v2,sc]=asian(r,S0,sig,T,K,k,n)
% computes the value of an asian option V1 and control variate V2
% S0=initial price, K=strike price
% sig = sigma, k=number of time increments in interval [0.T]
% sc is value of the score function for the normal inputs with respect to
% r the interest rate parameter. Repeats for a total of n simulations.
v1=[]; v2=[]; sc=[]; mn=(r-sig^2/2)*T/k;
sd=sig*sqrt(T/k); Y=normrnd(mn,sd,k,n);
```

```
sc= (T/k)*sum(Y-mn)/(sd^2);       Y=cumsum([zeros(1,n); Y]);
S = S0*exp(Y);                v1= exp(-r*T)*max(mean(S)-K,0);
v2=exp(-r*T)*max(S0*exp(mean(Y))-K,0);
disp(['standard errors ' num2str(sqrt(var(v1)/n)) ' num2str(sqrt(var(v1-v2)/n))])
```

For example if we use $K = 100$, we might confirm the last row of the above table using the command

asian(.1,100/1.1,.4,.2,100,50,10000);

Asian Options and Stratified Sampling

For many options, the terminal value of the stock has a great deal of influence on the option price. Although it is difficult in general to stratify samples of stock prices, it fairly easy to stratify along a single dimension, for example the dimension defined by the stock price at time T. In particular we may stratify the generation of

$$S_t = S_0 \exp(Z_t)$$

where Z_t can be written in terms of a standard Brownian motion

$$Z_t \ = \mu t + \sigma W_t, \quad \text{with } \mu = r - \sigma^2/2.$$

To stratify into K strata of equal probability for S_T we may generate Z_T using

$$Z_T = rT + \sqrt{rT - \sigma^2 T/2}\ \Phi^{-1}(i - 1 + \frac{U_i}{K}), i = 1, 2, ...K$$

for Uniform[0,1] random variables U_i and then randomly generate the rest of the path interpolating the value of S_0 and S_T using Brownian Bridge interpolation. To do this we use the fact that for a standard Brownian motion and $s < t < T$ we have that the conditional distribution of W_t given W_s, W_T is normal with mean a weighted average of the value of the process at the two endpoints

$$\frac{T-t}{T-s} W_s + \frac{t-s}{T-s} W_T$$

and variance

$$\frac{(t-s)(T-t)}{T-s}.$$

Thus given the value of S_T (or equivalently the value of $W(T)$) the increments of the process at times $\varepsilon, 2\varepsilon, ...N\epsilon = T$ say can be generated sequentially so that the j'th increment $W(j\varepsilon) - W((j-1)\varepsilon)$ conditionally on the value of $W((j-1)\varepsilon)$ and of $W(T)$ has a Normal distribution with mean

$$(\frac{N-j}{N}W((j-1)\varepsilon) + \frac{j}{N}W(T)$$

and with variance

$$\frac{N-j}{N-j+1}.$$

Use of Girsanov's Lemma.

There are many other variance reduction schemes that one can apply to valuing an Asian Option. However prior to attacking this problem by other methods, let us consider a simpler example.

Importance Sampling and Pricing a European Call Option

Suppose we wish to estimate the value of a call option using Monte Carlo methods which is well "out of the money", one with a strike price K far above the current price of the stock S_0 so that if we were to attempt to evaluate this option using crude Monte Carlo, the majority of the values randomly generated for S_T would fall below the strike and contribute zero to the option price. One possible remedy is to generate values of S_T under a distribution that is more likely to exceed the strike, but of course this would increase the simulated value of the option. We can compensate for changing the underlying distibution by multiplying by a factor adjusting the mean as one does in importance sampling.

More specifically, we wish to estimate

$$E_Q[e^{-rT}(S_0e^{Z_T} - K)^+], \text{ where } Z_T \sim N(rT - \sigma^2T/2, \sigma^2T)$$

where E_Q indicates that the expectation is taken under a risk neutral distribution or probability measure Q for and K is large. Suppose that we modify the underlying probability measure of Z_T to Q_0, say a normal distribution with mean value $\ln(K/S_0) - \sigma^2 T/2$ but the same variance $\sigma^2 T$, then the expected stock price under this new distribution

$$E_{Q_0} S_0 e^{Z_T} = S_0 \exp(E_{Q_0} Z_T + \sigma^2 T/2) = K$$

so there is a much greater probability that the strike price is attained. The importance sampling adjustment that insures that the estimator continues to be an unbised estimator of the option price is the ratio of two probability densities. Denote the normal probability density function by

$$\varphi(x, \mu, \sigma^2) = \frac{1}{\sqrt{2\pi}\sigma} \exp\{-\frac{(x-\mu)^2}{2\sigma^2}\}.$$

Then the Radon-Nikodym derivative

$$\frac{dQ}{dQ_0}(z_T) = \frac{\varphi(z_t; rT - \frac{\sigma^2 T}{2}, \sigma^2 T)}{\varphi(z_t; \ln(K/S_0) - \frac{\sigma^2 T}{2}, \sigma^2 T)}$$

is simply the ratio of the two normal density functions with the two different means, and

$$\begin{aligned} E_Q(e^{-rT}(S_T - K)^+) &= E_{Q_0}(e^{-rT}(S_T - K)^+ \frac{dQ}{dQ_0}(z_T)) \\ &= E_{Q_0}(e^{-rT}(S_0 e^{Z_T} - K)^+ \frac{\varphi(Z_T; rT - \frac{\sigma^2 T}{2}, \sigma^2 T)}{\varphi(Z_T; \ln(K/S_0) - \frac{\sigma^2 T}{2}, \sigma^2 T)}) \end{aligned}$$

so the importance sample estimator is the average of terms of the form

$$e^{-rT}(S_0 e^{Z_T} - K)^+ \frac{\varphi(Z_T; rT - \frac{\sigma^2 T}{2}, \sigma^2 T)}{\varphi(Z_T; \ln(K/S_0) - \frac{\sigma^2 T}{2}, \sigma^2 T)}, \text{ where } Z_T \sim N(\ln(K/S_0) - \frac{\sigma^2 T}{2}, \sigma^2 T).$$

The new simulation generates paths that are less likely to produce options expiring with zero value, and in a sense has thus eliminated some computational waste. What gains in efficiency result from this use of importance sampling? Let us consider a three month ($T = 0.25$) call option with $S_0 = 10$, $K = 15$,

$\sigma = 0.2$, $r = .05$. We determined the efficiency of the importance sampling estimator relative to using crude Monte Carlo in this situation using the function below. Running this using the command [eff,m,v]=importance2(10,.05,15,.2,.25) results in an efficiency gain of around 73, in part because very few of the crude estimates of S_T exceed the exercise price.

```
function [eff,m,v]=importance2(S0,r,K,sigma,T,N)
% simple importance sampling estimator of call option price
% outputs efficiency relative to crude. Run using
% [eff,m,v]=importance2(10,.05,15,.2,.25)
Z=randn(1,N);
%first do crude
ZT=(r-.5*sigma^2)*T+sigma*sqrt(T).*Z;
est1=exp(-r*T)*max(0,S0*exp(ZT)-K);
% now do importance
ZT=(log(K/S0)-.5*sigma^2)*T+sigma*sqrt(T).*Z;
ST=S0*exp(ZT);
est2=exp(-r*T)*max(0,ST-K).*normpdf(ZT,(r-.5*sigma^2)*T,sigma*sqrt(T))./normpdf(ZT,(log(K/S0)-
.5*sigma^2)*T,sigma*sqrt(T));
v=[var(est1) var(est2)];
m=[mean(est1) mean(est2)];
eff=v(1)/v(2);
```

Importance Sampling and Pricing an Asian Call Option

Let us now return to the price of an Asian option. We wish to use a variety of variance reduction techniques including importance sampling as in the above example, but in this case the relevant observation is not a simple stock price at one instant but the whole stock price history from time 0 to T. An Asian option to have a payoff related to the closing value of the stock $S(T)$. It might be reasonable to stratify the sample; i.e. sample more often when $S(T)$ is large or to use importance sampling and generate $S(T)$ from a geometric Brownian motion with drift larger than rS_t so that it is more likely that $S(T) > K$. As before if we do this we need to then multiply by the ratio of the two probability density functions, or the Radon Nikodym derivative of one process with respect to the other. This density is given by a result called Girsanov's Theorem (see Appendix B). The idea is as follows: Suppose P is the probability measure induced on the paths on $[0, T]$ by an Ito process

$$dS_t = \mu(S_t)dt + \sigma(S_t)dW_t. \tag{5.8}$$

Similarly suppose P_0 is the probability measure on path generated by a similar process with the same diffusion term but different drift term

$$dS_t = \mu_0(S_t)dt + \sigma(S_t)dW_t. \tag{5.9}$$

Assume that in both cases, the process starts at the same initial value S_0.Then the "likelihood ratio" or the Radon-Nikodym $\frac{dP}{dP_0}$ of P with respect to P_0 is

$$\frac{dP}{dP_0} = \exp\{\int_0^T \frac{\mu(S_t) - \mu_0(S_t)}{\sigma^2(S_t)} dS_t - \int_0^T \frac{\mu^2(S_t) - \mu_0^2(S_t)}{2\sigma^2(S_t)} dt\} \tag{5.10}$$

We do not attempt to give a technical proof of this result, either here or in the appendix, since real "proofs" can be found in a variety of texts, including Steele (2004) and Karatzas and Shreve, (xxx). Instead we provide heuristic justification of (5.10). Let us consider the conditional distribution of a small increment in the process S_t under the model (5.8). Since this distribution is conditionally normal distributed it has conditional probability density function given the past

$$\frac{1}{\sqrt{2\pi dt}} \exp\{-(dS_t - \mu(S_t)dt)^2/(2\sigma^2(S_t)dt) \tag{5.11}$$

and under the model (5.9), it has the conditional probability density

$$\frac{1}{\sqrt{2\pi dt}} \exp\{-(dS_t - \mu_0(S_t)dt)^2/(2\sigma^2(S_t)dt) \tag{5.12}$$

The ratio of these two probability density functions is

$$\exp\{\frac{\mu(S_t)-\mu_0(S_t)}{\sigma^2(S_t)}dS_t-\frac{\mu^2(S_t)-\mu_0^2(S_t)}{2\sigma^2(S_t)}dt\}$$

But the joint probability density function over a number of disjoint intervals is obtained by multiplying these conditional densities together and this results in

$$\Pi_t\exp\{\frac{\mu(S_t)-\mu_0(S_t)}{\sigma^2(S_t)}dS_t-\frac{\mu^2(S_t)-\mu_0^2(S_t)}{2\sigma^2(S_t)}dt\}$$
$$=\exp\{\int_0^T\frac{\mu(S_t)-\mu_0(S_t)}{\sigma^2(S_t)}dS_t-\int_0^T\frac{\mu^2(S_t)-\mu_0^2(S_t)}{2\sigma^2(S_t)}dt\}$$

where the product of exponentials results in the sum of the exponents, or, in the limit as the increment dt approaches 0, the corresponding integrals.

Girsanov's result is very useful in conducting simulations because it permits us to change the distribution under which the simulation is conducted. In general, if we wish to determine an expected value under the measure P, we may conduct a simulation under P_0 and then multiply by $\frac{dP}{dP_0}$ or if we use a subscript on E to denote the measure under which the expectation is taken,

$$E_PV(S_T)=E_{P_0}[V(S_T)\frac{dP}{dP_0}].$$

Suppose, for example, we wish to determine by simulation the expected value of $V(r_T)$ for an interest rate model

$$dr_t=\mu(r_t)dt+\sigma dW_t \tag{5.13}$$

for some choice of function $\mu(r_t)$. Then according to Girsanov's theorem, we may simulate r_t under the Brownian motion model $dr_t=\mu_0dt+\sigma dW_t$ (having the same initial value r_0 as in our original simulation) and then average the values of

$$V(r_T)\frac{dP}{dP_0}=V(r_T)\exp\{\int_0^T\frac{\mu(r_t)-\mu_0}{\sigma^2}dr_t-\int_0^T\frac{\mu^2(r_t)-\mu_0^2}{2\sigma^2}dt\} \tag{5.14}$$

So far, the constant μ_0 has been arbitrary and we are free to choose it in order to achieve as much variance reduction as possible. Ideally we do not want to get too far from the original process so μ_0 should not be too far from the values of $\mu(r_t)$. In this case we hope that the term $\frac{dP}{dP_0}$ is not too variable (note that $c\frac{dP}{dP_0}$ would be the estimator if $V(S_T) = c$ were constant). On the other hand, the term $V(r_T)$ cannot generally be ignored, and there is no formula or simple rule for choosing parameters which minimize the variance of $V(r_T)\frac{dP}{dP_0}$. Essentially we need to resort to choosing μ_0 to minimize the variance of $V(r_T)\frac{dP}{dP_0}$ by experimentation, usually using with preliminary simulations.

Pricing a Call option under stochastic interest rates.

(REVISE MODEL)Again we consider pricing a call option, but this time under a more realistic model which permits stochastic interest rates. We will use the method of conditioning, although there are many other potential variance reduction tools here. Suppose the asset price, (under the risk-neutral probability measure, say) follows a geometric Brownian motion model of the form

$$dS_t = r_t S_t dt + \sigma S_t dW_t^{(1)} \tag{5.15}$$

where r_t is the spot interest rate. We assume r_t is stochastic and follows the Brennan-Schwartz model,

$$dr_t = a(b - r_t)dt + \sigma_0 r_t dW_t^{(2)} \tag{5.16}$$

where $W_t^{(1)}, W_t^{(2)}$are both Brownian motion processes and usually assumed to be correlated with correlation coefficient ρ. The parameter b in (5.16) can be understood to be the long run average interest rate (the value that it would converge to in the absence of shocks or resetting mechanisms) and the parameter $a > 0$ governs how quickly reversion to b occurs.

It would be quite remarkable if a stock price is completely independent of interest rates, since both will depend on an overlaping set of factors. However we begin by assuming something a little less demanding, that the random noise processes driving the asset price and stock price are independent or that $\rho = 0$.

Control Variates.

The first method might be to use crude Monte Carlo; i.e. to simulate both the process S_t and the process r_t, evaluate the option at expiry, say $V(S_T, T)$ and then discount to its present value by multiplying by $exp\{-\int_0^T r_t dt\}$. However, in this case we can exploit the assumption that $\rho = 0$ so that interest rates are independent of the Brownian motion process $W_t^{(1)}$ which drives the asset price process. For example, suppose that the interest rate function r_t *were known* (equivalently: condition on the value of the interest rate process so that in the conditional model it is known). While it may be difficult to obtain the value of an option under the model (5.15), (5.16) it is usually much easier under a model which assumes constant interest rate c. Let us call this constant interest rate model for asset prices with the same initial price S_0 and driven by the equation

$$dZ_t = cZ_t dt + \sigma Z_t dW_t^{(1)}, Z_0 = S_0 \tag{5.17}$$

model "0" and denote the probability measure and expectations under this distribution by P_0 and E_0 respectively. The value of the constant c will be determined later. Assume that we simulated the asset prices under this model and then valued a call option, say. Then since

$$\ln(Z_T/Z_0) \text{ has a } N((c - \frac{\sigma^2}{2})T, \sigma^2 T) \quad \text{distribution}$$

we could use the Black-Scholes formula to determine the conditional expected value

$$E_0[exp\{-\int_0^T r_t dt\}(Z_T-K)^+|r_s, 0<s<T] \tag{5.18}$$

$$= EE_0[(S_0e^{(c-\overline{r})T}e^W - e^{-\overline{r}T}K)^+|r_s, 0<s<T],$$

where W has a $N(-\sigma^2T/2, \sigma^2T)$

$$= E[BS(S_0e^{(c-\overline{r})T}, K, \overline{r}, T, \sigma)], \text{ with } \overline{r} = \frac{1}{T}\int_0^T r_t dt.$$

Here, $\overline{r}$ is the average interest rate over the period and the function BS is the Black-Scholes formula (5.2). In other words by replacing the interest rate by its average over the period and the initial value of the stock by $S_0e^{(c-\overline{r})T}$, the Black-Scholes formula provides the value for an option on an asset driven by (5.17) conditional on the value of $\overline{r}$. The constant interest rate model is a useful control variate for the more general model (5.16). The expected value $E[BS(S_0e^{(c-\overline{r})T}, K, \overline{r}, T, \sigma)]$ can be determined by generating the interest rate processes and averaging values of $BS(S_0e^{(c-\overline{r})T}, K, \overline{r}, T, \sigma)$. Finally we may estimate the required option price under (5.15),(5.16) using an average of values of

$$exp\{-\int_0^T r_t dt\}[(S_T-K)^+ - (Z_T-K)^+]\} + E\{BS(S_0e^{(c-\overline{r})T}, K, \overline{r}, T, \sigma)\}$$

for S_T and Z_T generated using common random numbers.

We are still able to make a choice of the constant c. One simple choice is $c \approx E(\overline{r})$ since this means that the second term is approximately $E\{BS(S_0, K, \overline{r}, T, \sigma)\}$. Alternatively we can again experiment with small numbers of test simulations and various values of c in an effort to roughly minimize the variance

$$var(exp\{-\int_0^T r_t dt\}[(S_T-K)^+ - (Z_T-K)^+]\}).$$

Evidently it is fairly easy to arrive at a solution in the case $\rho = 0$ since we really only need to average values of the Black Scholes price under various randomly generated interest rates. This does not work in the case $\rho \neq 0$ because the conditioning involved in (5.18) does not result in the Black Scholes formula. Nevertheless we could still use common random numbers to generate two interest rate paths, one corresponding to $\rho = 0$ and the other to $\rho \neq 0$ and use the former as a control variate in the estimation of an option price in the general case.

Importance Sampling

The expectation under the correct model could also be determined by multiplying this random variable by the ratio of the two likelihood functions and then taking the expectation under E_0. By Girsanov's Theorem, $E\{V(S_T,T)\} = E_0\{V(S_T,T)\frac{dP}{dP_0}\}$ where P is the measure on the set of stock price paths corresponding to (5.15),(5.16) and P_0 that measure corresponding to (5.17). The required Radon-Nykodym derivative is

$$\frac{dP}{dP_0} = exp\{\int_0^T \frac{(r_t-c)S_t}{S_t^2\sigma^2}dS_t - \int_0^T \frac{(r_t^2-c^2)S_t^2}{2\sigma^2 S_t^2}dt\} \tag{5.19}$$

$$= exp\{\int_0^T \frac{r_t-c}{S_t\sigma^2}dS_t - \int_0^T \frac{r_t^2-c^2}{2\sigma^2}dt\} \tag{5.20}$$

The resulting estimator of the value of the option is therefore an average over all simulations of the value of

$$V(S_T,T)exp\{-\int_0^T r_t dt + \int_0^T \frac{r_t-c}{\sigma^2 r_t}dS_t - \int_0^T \frac{r_t^2-c^2}{2\sigma^2}dt\} \tag{5.21}$$

where the trajectories $r_t$are simulated under interest rate model (5.16).

As discussed before, we can attempt to choose the drift parameter c to approximately minimize the variance of the estimator (5.21).

Simulating Barrier and lookback options

For a financial times series X_t observed over the interval $0 \leq t \leq T$, what is recorded in newspapers is often just the initial value or *open* of the time series $O = X_0$, the terminal value or *close* $C = X_T$, the maximum over the period or the *high,* $H = max\{X_t; 0 \leq t \leq T\}$ and the minimum or the *low* $L = min\{X_t; 0 \leq t \leq T\}$. Very few uses of the highly informative variables H and L are made, partly becuase their distribution is a bit more complicated than that of the normal distribution of returns. Intuitively, however, the difference between H and L should carry a great deal of information about one of the most important parameters of the series, its volatility. Estimators of volatility obtained from the range of prices $H - L$ or H/L will be discussed in Chapter 6. In this section we look at how simple distributional properties of H and L can be used to simulate the values of certain exotic path-dependent options.

Here we consider options such as barrier options, lookback options and hindsight options whose value function depends only on the four variables (O, H, L, C) for a given process. Barrier options include knock-in and knock-out call options and put options. Barrier options are simple call or put options with a feature that should the underlying cross a prescribed barrier, the option is either knocked out (expires without value) or knocked in (becomes a simple call or put option). Hindsight options, also called fixed strike lookback options are like European call options in which we may use any price over the interval $[0, T]$ rather than the closing price in the value function for the option. Of course for a call option, this would imply using the high H and for a put the low L. A few of these path-dependent options are listed below.

Option	Payoff
Knock-out Call	$(C - K)^+ I(H \leq m)$
Knock-in Call	$(C - K)^+ I(H \geq m)$
Look-back Put	$H - C$
Look-Back Call	$C - L$
Hindsight (fixed strike lookback) Call	$(H - K)^+$
Hindsight (fixed strike lookback) put	$(K - L)^+$

Table XX: Value Function for some exotic options

For further details, see Kou et. al. (1999) and the references therein.

Simulating the High and the Close

All of the options mentioned above are functions of two or three variables O, C, and H or O, C, and L and so our first challenge is to obtain in a form suitable to calculation or simulation the joint distribution of these three variables. Our argument will be based on one of the simplest results in combinatorial probability, the reflection principle. We would like to be able to handle more than just a Black-Scholes model, both discrete and continuous distributions, and we begin with the simple discrete case.

In the real world, the market does not rigorously observe our notions of the passage of time. Volatility and volume traded vary over the day and by day of the week. A successful model will permit some variation in clock speed and volatility, and so we make an attempt to permit both in our discrete model.

In the discrete case, we will assume that the stock price S_t forms a trinomial tree, taking values on a set set $D = \{\cdots d_{-1} < d_0 < d_1 \cdots\}$... At each time point t, the stock may either increase, decrease, or stay in the same place and the probability of these movements may depend on the time. Specifically we assume that if $S_t = d_i$, then for some parameters $\theta, p_t, t = 1, 2, ...,$

$$P(S_{t+1} = d_j | S_t = d_i) = \frac{1}{k_t(\theta)} \times \begin{cases} p_t e^{\theta} & \text{if} \quad j = i+1 \\ 1 - 2p_t & \text{if} \quad j = i \\ p_t e^{-\theta} & \text{if} \quad j = i-1 \\ 0 & \text{otherwise} \end{cases} \tag{5.22}$$

where $k_t(\theta) = 1 + p_t(e^{\theta} + e^{-\theta} - 2)$ and $p_t \le \frac{1}{2}$ for all t. If we choose all $p_t = \frac{1}{2}$, then this is a model of a simple binomial tree which either steps up or down at each time point. The increment in this process $X_{t+1} = S_{t+1} - S_t$ has mean which depends on the time t except in the case $\theta = 0$

$$E(X_{t+1} | X_t = d_i) = \frac{p_t}{k_t(\theta)} \{(d_{i+1} - d_i)e^{\theta} - (d_i - d_{i-1})e^{-\theta}),$$

and variance, also time-dependent except in the case $\theta = 0$. The parameter θ describes one feature of this process which is not dependent on the time or the

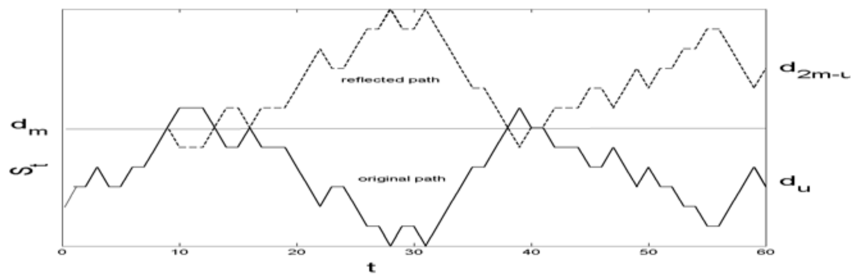

Figure 5.1: Illustration of the Reflection Principle

location of the process, since the log odds of a move up versus a move down is

$$2\theta = \log[\frac{P[\text{UP}]}{P[\text{DOWN}]}].$$

Now suppose we label the states of the process so that $S_0 = d_0$ and there is a barrier at the point d_m where $m > 0$. We wish to count the number of paths over an interval of time $[0, T]$ which touch or cross this barrier and end at a particular point $d_u, u < m$. Such a path is shown as a solid line in Figure 5.1 in the case that the points d_i are all equally spaced. Such a path has a natural "reflection" about the horizontal line at d_m. The reflected path is identical up to the first time τ that the original path touches the point d_m, and after this time, say at time $t > \tau$, the relected path takes the value d_{2m-i} where $S_t = d_i$. This path is the dotted line in Figure 5.1. Notice that if the original path ends at $d_u < d_m$, below the barrier, the reflected path ends at $d_{2m-u} > d_m$ or above the barrier. Each path touching the barrier at least once and ending below it at d_u has a reflected path ending above it at d_{2m-u}, and of course each path that ends above the barrier must touch the barrier for a first time at some point and has a reflection that ends below the barrier. This establishes a one-one correspondence useful for counting these paths. Let us denote by the symbol

"#" the "number of paths such that". Then:

$$\#\{H \geq d_m \text{ and } C = d_u < d_m\} = \#\{C = d_{2m-u}\}.$$

Now consider the probability of any path ending at a particular point d_u,

$$(S_0 = d_0, S_1, ..., S_T = d_u).$$

To establish this probability, each time the process jumps up in this interval we must multiply by the factor $\frac{p_t e^{\theta}}{k_t(\theta)}$ and each time there is a jump down we multiply by $\frac{p_t e^{-\theta}}{k_t(\theta)}$. If the process stays in the same place we multiply by $\frac{1-2p_t}{k_t(\theta)}$. The reflected path has exactly the same factors except that after the time τ at which the barrier is touched, the "up" jumps are replaced by "down" jumps and vice versa. For an up jump in the original path multiply by $e^{-2\theta}$. For a down jump in the original path, multiply by $e^{2\theta}$. Of course this allows us to compare path probabilities for an arbitrary value of the parameter θ, say with P_0, the probability under $\theta = 0$ since, if the path ends at $C = d_u$,

$$\begin{aligned} P_\theta(\text{path}) &= \frac{e^{N_U\theta}e^{-N_D\theta}}{\prod_t k_t(\theta)} P_0[\text{path}] \\ &= \frac{e^{u\theta}}{\prod_t k_t(\theta)} P_0[\text{path}] \end{aligned} \tag{5.23}$$

where N_U and N_D are the number of up jumps and down jumps in the path. Note that we have subscripted the probability measure by the assumed value of the parameter θ.

This makes it easy to compare the probabilities of the original and the reflected path, since

$$\frac{P_\theta[\text{original path}]}{P_\theta[\text{reflected path}]} = e^{-2\theta N_U} e^{2\theta N_D}$$

where now the number of up and down jumps N_U and N_D are counted following time τ. However, since $S_T = d_u$ and $S_\tau = d_m$, it follows that $N_D - N_U = m - u$ and that

$$\frac{P[\text{original path}]}{P[\text{reflected path}]} = e^{2\theta(u-m)}$$

which is completely independent of how that path arrived at the closing value d_u, depending only on the close. This makes it easy to establish the probability of paths having the property that $H \geq d_m$ and $C = d_u < d_m$ since there are exactly the same number of paths such that $C = d_{2m-u}$ and the probabilities of these paths differ by a constant factor $e^{2\theta(u-m)}$. Finally this provides the useful result for $u < m$.

$$P_\theta[H \geq d_m \text{ and } C = d_u] = e^{2\theta(u-m)} P_\theta[C = d_{2m-u}],$$

or, on division by $P[C = d_u]$,

$$\begin{aligned} P_\theta[H \geq d_m | C = d_u] &= \frac{e^{2\theta(u-m)} P_\theta[C = d_{2m-u}]}{P_\theta[C = d_u]} \\ &= \frac{e^{2\theta(u-m)} e^{\theta(2m-u)} P_0[C = d_{2m-u}]}{e^{\theta u} P_0[C = d_u]} \\ &= \frac{P_0[C = d_{2m-u}]}{P_0[C = d_u]} \end{aligned}$$

where we have used (5.23). This rather simple formula completely descibes the conditional distribution of the high under an arbitrary value of the parameter θ in terms of the value of the close under parameter value $\theta = 0$. In fact we have proved the following proposition.

Proposition 43 *Suppose a stock price S_t has dynamics determined by (5.22), and $S_0 = d_0$. Define*

$$H = \max_{0 \le t \le T} S_t \text{ and } C = S_T$$

Then for $u < m$,

$$P_\theta(H \geq d_m | C{=}d_u) {=} \frac{P_0[C = d_{2m-u}]}{P_0[C = d_u]}, \text{ for } min(u, 0) < m, \tag{5.24}$$

$$= 1, \text{ for } min(u,0) \geq m$$

Thus, the conditional distribution of the high of a process given the open and close can be determined easily without knowledge of the underlying parameter and is related to the distribution of the close when the "drift" $\theta = 0$. This result also gives the expected value of the high in fairly simple form if the points d_j are equally spaced. Suppose $d_j = j\Delta$ for $j = 0, \pm 1, \pm 2, ...$ Then for $u = j\Delta$,with $j \geq 0$ and $k \geq 1$, (see Problem 1)

$$P_\theta[H - C \geq k\Delta | C = j\Delta] =$$

$$E[H|C = u] = u + \Delta \frac{P[C > u \text{ and } \frac{C-u}{\Delta} \text{ is even}]}{P[C = u]}.$$

Roughly, (5.24) indicates that if you can simulate the close under θ, then you use the properties of the close under $\theta = 0$ to simulate the high of the process. Consider the problem of simulating the high for a given value of the close $C = S_T = d_u$ and again assuming that $S_0 = d_0$. Suppose we use inverse transform from a uniform random variable U to solve the inequalities

$$P_\theta(\max_{0 \le t \le T} S_t \geq d_{m+1} | S_T = d_u) < U \leq P_\theta(\max_{0 \le t \le T} S_t \geq d_m | S_T = d_u)$$

for the value of d_m. In this case the value of

$$d_m = \sup\{d_j; U P_0[S_T = d_u] \leq P_0[S_T = d_{2j-u}]\}$$

is the generated value of the high. This inequality is equivalent to

$$P_0[S_T = d_{2m+2-u}] < U P_0[S_T = d_u] \leq P_0[S_T = d_{2m-u}].$$

Graphically this inequality is demonstrated in Figure 5.2 which shows the probability histogram of the distribution S_T under $\theta = 0$. The value $U P_0[S_T = d_u]$ is the y-coordinate of a point P randomly chosen from the bar corresponding to the point d_u. The high d_m is generated by moving horizontally to the right an even number of steps until just before exiting the histogram. This is above the value d_{2m-u} and d_m is between d_u and d_{2m-u}.

A similar result is available for Brownian motion and Geometric Brownian motion. A justification of these results can be made by taking a limit in the

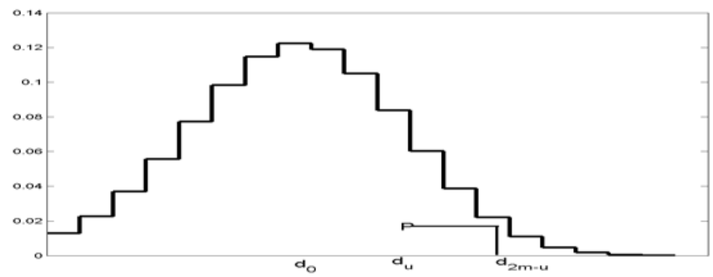

Figure 5.2: Generating a High, discrete distributions

discrete case as the time steps and the distances $d_j - d_{j-1}$ all approach zero. If we do this, the parameter θ is analogous to the drift of the Brownian motion. The result for Brownian motion is as follows:

Theorem 44 *Suppose S_t is a Brownian motion process*

$$dS_t = \mu dt + \sigma dW_t,$$

$$S_0 = 0,\ S_T = C,$$

$$H = \max\{S_t; 0 \le t \le T\} \text{ and}$$

$$L = \min\{S_t; 0 \le t \le T\}.$$

If f_0 denotes the Normal(0,$\sigma^2 T$) probability density function, the distribution of C under drift $\mu = 0$, then

$$U_H = \frac{f_0(2H-C)}{f_0(C)} \text{ is distributed as } U[0,1] \text{ independently of } C,$$

$$U_L = \frac{f_0(2L-C)}{f_0(C)} \text{ is distributed as } U[0,1] \text{ independently of } C.$$

$$Z_H = H(H-C) \text{ is distributed as Exponential } (\frac{1}{2}\sigma^2 T) \text{ independently of } C,$$

$$Z_L = L(L-C) \text{ is distributed as Exponential } (\frac{1}{2}\sigma^2 T) \text{independently of } C.$$

We will not prove this result since it is a special case of Theorem 46 below. However it is a natural extension of Proposition 43 in the special case that

$d_j = j\Delta$ for some Δ and so we will provide a simple sketch of a proof using this proposition. Consider the ratio

$$\frac{P_0[C = d_{2m-u}]}{P_0[C = d_u]}$$

on the right side of (5.24). Suppose we take the limit of this as $\Delta \to 0$ and as $m\Delta \to h$ and $u\Delta \to c$. Then this ratio approaches

$$\frac{f_0(2h-c)}{f_0(c)}$$

where f_0 is the probability density function of C under $\mu = 0$. This implies for a Brownian motion process,

$$P[H \geq h|C = c] = \frac{f_0(2h-c)}{f_0(c)} \text{ for } h \geq c. \tag{5.25}$$

If we temporarily denote the cumulative distribution function of H given $C = c$ by $G_c(h)$ then (5.25) gives an expression for $1 - G_c(h)$ and recall that since the sumulative distribution function is continuous, when we evaluate it at the observed value of a random variable we obtain a $U[0,1]$ random variable e.g. $G_c(H) \sim U[0,1]$. In other words conditional on $C = c$ we have

$$\frac{f_0(2H-c)}{f_0(c)} \sim U[0,1].$$

This result verifies a simple geometric procedure, directly analogous to that in Figure 5.2, for generating H for a given value of $C = c$. Suppose we generate a point $P_H = (c, y)$ under the graph of $f_0(x)$ and uniformly distributed on $\{(c, y); 0 \leq y \leq f_0(c)\}$. This point is shown in Figure **??**. We regard the $y-$coordinate of this point as the generated value of $f_0(2H - c)$. Then H can be found by moving from P_H horizontally to the right until we strike the graph of f_0 and then moving vertically down to the axis (this is now the point $2H - c$) and finally taking the midpoint between this coordinate $2H - c$ and the close c to obtain the generated value of the high H. The low of the process can be generated in the same way but with a different point P_L uniform on the set

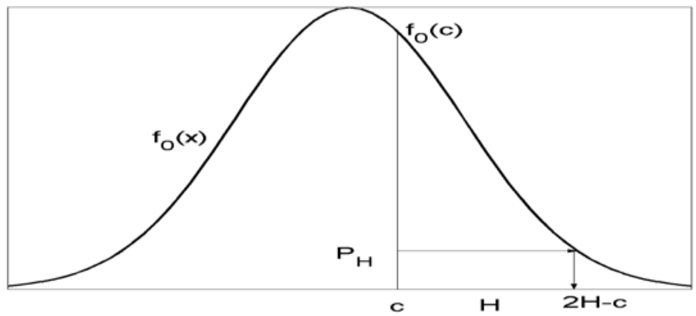

Figure 5.3: Generating H for a fixed value of C for a Brownian motion.

$\{(c,y); 0 \le y \le f_0(c)\}$. The algorithm is the same in this case except that we move horizontally to the left.

There is a similar argument for generating the high under a geometric Brownian motion as well, since the logarithm of a geometric Brownian motion is a Brownian motion process.

Corollary 45 *For a Geometric Brownian motion process*

$$dS_t = \mu S_t dt + \sigma S_t dW_t,$$
$$S_0 = O \text{ and } S_T = C$$

with f_0 the normal$(0, \sigma^2 T)$ probability density function, we have

$$\ln(H/O)\ln(H/C) \sim \exp(\frac{1}{2}\sigma^2 T) \text{ independently of } O, C \text{ and}$$
$$\ln(L/O)\ln(L/C) \sim \exp(\frac{1}{2}\sigma^2 T) \text{ independently of } O, C.$$
$$U_H = \frac{f_0(\ln(H^2/OC))}{f_0(\ln(C/O))} \sim U[0,1] \text{ independently of } O, C \text{ and}$$
$$U_L = \frac{f_0(\ln(L^2/OC))}{f_0(\ln(C/O))} \sim U[0,1] \text{ independently of } O, C.$$

Both of these results are special cases of the following more general Theorem. We refer to McLeish(2002) for the proof. As usual, we consider a price process S_t and define the high $H = max\{S_t; 0 \le t \le T\}$, and the open and close $O = S_0$, $C = S_T$.

Theorem 46 *Suppose the process S_t satisfies the stochastic differential equation:*

$$dS_t = \{\nu + \frac{1}{2}\sigma'(S_t)\}\sigma(S_t)\lambda^2(t)dt + \sigma(S_t)\lambda(t)dW_t \tag{5.26}$$

where $\sigma(x) > 0$ and $\lambda(t)$ are positive real-valued functions such that $g(x) = \int^x \frac{1}{\sigma(y)}dy$ and $\theta = \int_0^T \lambda^2(s)ds < \infty$ are well defined on $\Re^+$.
(a) Then with f_0 the $N(0,\theta)$ probability density function we have

$$U_H = \frac{f_0\{2g(H) - g(O) - g(C)\}}{f_0\{g(C) - g(O)\}} \sim U[0,1]$$

and U_H is independent of C.
(b) For each value of T, $Z_H = (g(H) - g(O))(g(H) - g(C))$ is independent of O, C, and has an exponential distribution with mean $\frac{1}{2}\theta$.

A similar result holds for the low of the process over the interval, namely that

$$U_L = \frac{f_0\{2g(L) - g(O) - g(C)\}}{f_0\{g(C) - g(O)\}} \sim U[0,1]$$

and $Z_H = \{g(L) - g(O)\}\{g(L) - g(C)\}$ is independent of O, C, and has an exponential distribution with mean $\frac{1}{2}\theta$.

Before we discuss the valuation of various options, we examing the significance of the ratio appearing in on the right hand side of (5.25) a little more closely. Recall that f_0 is the $N(0, \sigma^2 T)$ probability density function and so we can replace it by

$$\frac{f_0(2h - c)}{f_0(c)} = \frac{\exp\{-\frac{(2h-c)^2}{2\sigma^2 T}\}}{\exp\{-\frac{c^2}{2\sigma^2 T}\}} = \exp\{-2\frac{z_h}{\sigma^2 T}\} \tag{5.27}$$

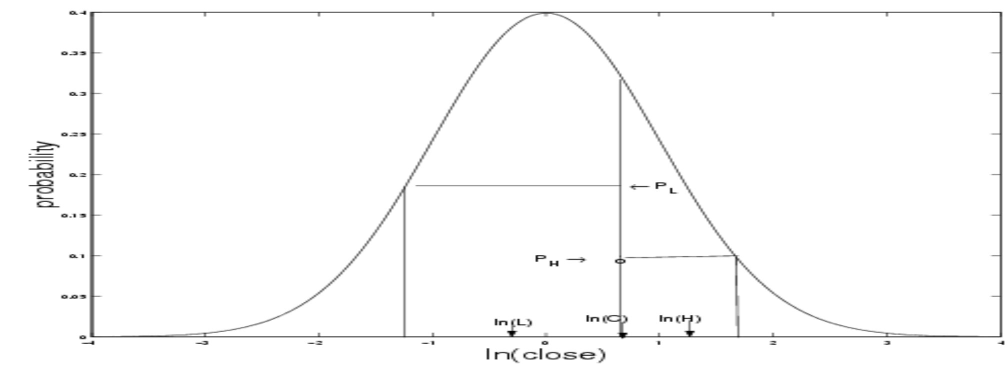

Figure 5.4: Simulating from the joint distribution of (H, C) or from (L, C)

where $z_h = h(h - c)$ or in the more general case where $S(0) = O$ may not be equal to zero,

$$z_h = (h - O)(h - c). \tag{5.28}$$

This ratio $\frac{f_0(2h-c)}{f_0(c)}$ represents the probability that a particular process with close c breaches a barrier at h and so the exponent

$$2\frac{z_h}{\sigma^2 T}$$

in the right hand side of (5.27) controls the probability of this event.

Of course we can use the above geometric algorithm for Brownian motion to generated highs and closing prices for a geometric Brownian motion, for example, S_t satisfying $d\ln(S_t) = \sigma dW_t$ (minor adjustments required to accommodate nonzero drift). The graph of the normal probability density function $f_0(x)$ of $\ln(C)$ is shown in Figure **??**.

If a point P_H is selected at random uniformly distributed in the region below the graph of this density, then, by the usual arguments supporting the acceptance rejection method of simulation, the x-coordinate of this point is a variate generated from the probability density function $f_0(x)$, that is, a simulated value from the distribution of $\ln(C)$. The y-coordinate of such a randomly selected point also generates the value of the high as before.If we extend a line horizontally to the right from P_H until it strikes the graph of the probability density and then consider the abscissa, of this point, this is the simulated value of $\ln(H^2/C)$, and $\ln(H)$ the average the simulated values of $\ln(H^2/C)$ and $\ln(C)$.

A similar point P_L uniform under the probability density function f_0 can be used to generate the low of the process if we extend the line from P_L to the left until it strikes the density. Again the abscissa of this point is $\ln(L^2/C)$ and the average with $\ln(C)$ gives a simulated value of $\ln(L)$. Although the $y-$coordinate of both P_H and P_L are uniformly distributed on $[0, f_0(C)]$ conditional on the value of C they are not independent.

Suppose now we wish to price a barrier option whose payoff on maturity depends on the value of the close C but provided that the high H did not exceed a certain value, the barrier. This is an example of an knock-out barrier but other types are similarly handled. Once again we assume the simplest form of the geometric Brownian motion $d\ln(S_t) = \sigma dW_t$ and assume that the upper barrier is at the point Oe^b so that the payoff from the option on maturity T is

$$\psi(C)I(H < Oe^b)$$

for some function ψ. It is clear that the corresponding value of H does not exceed a boundary at Oe^b if and only if the point P_H is below the graph of the probability density function but **not** in the shaded region obtained by reflecting the right hand tail of the density about the vertical line $x = b - \ln(O)$ in Figure 5.5. To simulate the value of the option, choose points uniformly under the graph of the probability density $f_0(x)$. For those points in the non-shaded region under f_0 (the x-coordinate of these points are simulated values $\psi(C)$of $\ln(C)$ under the condition that the barrier is not breached) we average the values of $\psi(C)$ and for those in the shaded region we average 0.

Equivalently,

$$E\psi(C)I(H < Oe^b) = E\psi^*(C)$$

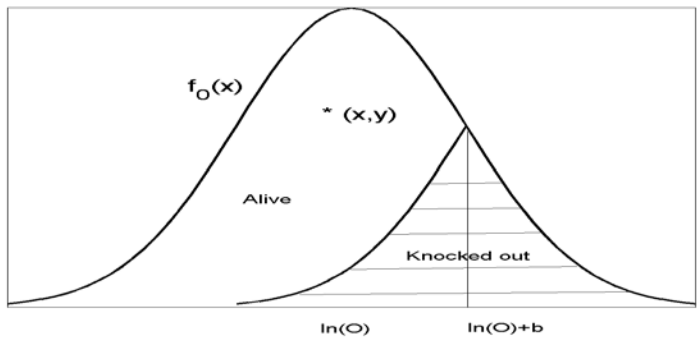

Figure 5.5: Simulating a knock-out barrier option with barrier at Oe^b

where

$$\psi^*(C) = \begin{cases} \psi(C) & \text{for} \quad C \leq Oe^b \\ -\psi(2b + \ln(O^2/C)) & \text{for} \quad C > Oe^b \end{cases}.$$

and so the barrier option can be priced as if it were a vanilla European option with payoff function $\psi^*(C)$.

Any option whose value depends on the high and the close of the process (or (L, C)) can be similarly valued as a European option. If an option becomes worthless whenever an upper boundary at Oe^b is breached, we need only multiply the payoff from the option ignoring the boundary by the factor

$$1 - \exp\{-2\frac{z_h}{\sigma^2 T}\}$$

with

$$z_h = b(b + \ln(O/C))$$

to accommodate the filtering effect of the barrier and then value the option as if it were a European option.

There is a a variety of distributional results related to H, some used by Redekop (1995) to test the local Brownian nature of various financial time series. These are easily seen in Figure 5.7. For example, for a Brownian motion process

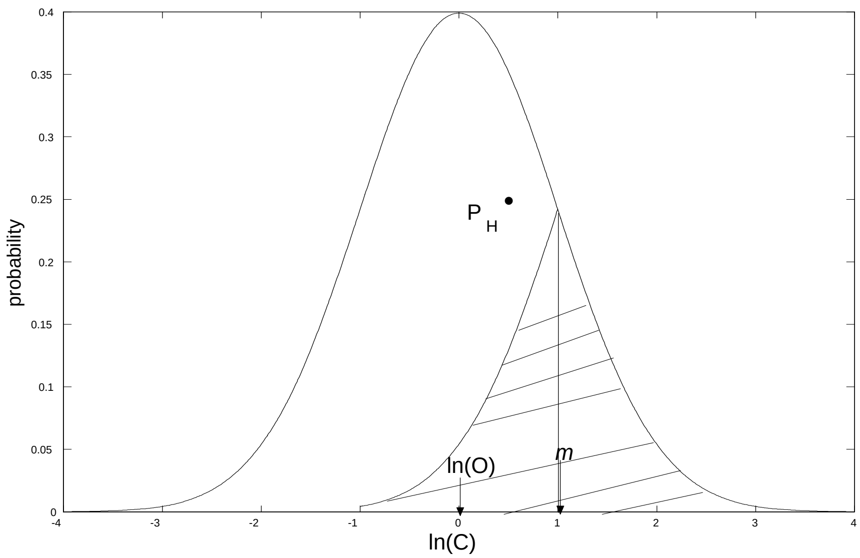

Figure 5.6: Simulating a Barrier Option with barrier at e^m

with sero drift, suppose we condition on the value of $2H-O-C$. Then the point P_H must lie (uniformly distributed) on the line $\mathcal{L}_1$ and therefore the point H lies uniformly on this same line but to the right of the point O. This shows that conditional on $2H-O-C$ the random variable $H-O$ is uniform or,

$$\frac{H-O}{2H-O-C} \sim U[0,1].$$

Similarly, conditional on the value of H, the point P_H must fall somewhere on the curve labelled $\mathcal{C}_2$ whose y-coordinate is uniformly distributed showing that

$$\frac{C-O}{2H-O-C} \sim U[0,1].$$

Redekop shows that for a Brownian motion process, the statistic

$$\frac{H-O}{2H-O-C} \tag{5.29}$$

is supposed to be uniformly $[0,1]$ distributed but when evaluated using real financial data, is far too often close to or equal the extreme values 0 or 1.

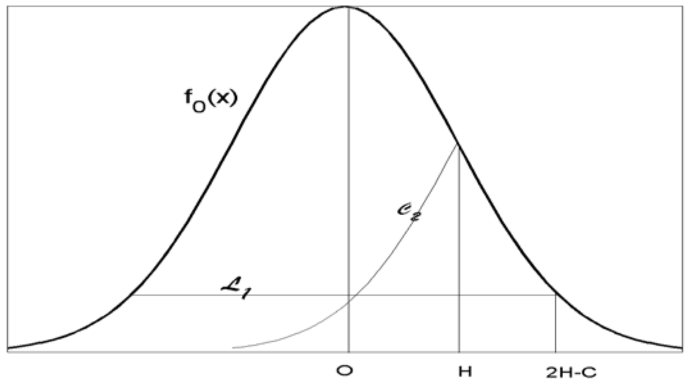

Figure 5.7: Some uniformly distributed statistics for Brownian Motion

The joint distribution of (C, H) can also be seen from Figure 5.8. Note that the rectangle around the point (x, y) of area $\Delta x \Delta y$ under the graph of the density, when mapped into values of the high results in an interval of values for $(2H - C)$ of width $-\Delta y/\phi'(2y - x)$ where ϕ' is the derivative of the standard normal probability density function (the minus sign is to adjust for the negative slope of the density here). This interval is labelled $\Delta(2H - C)$. This, in turn generates the interval ΔH of possible values of H, of width exactly half this, or

$$\frac{-\Delta y}{2\phi'(2y - x)}.$$

Inverting this relationship between (x, y) and (H, C),

$$P[H \in \Delta H, C \in \Delta C] = -2\phi'(2y - x)\Delta x \Delta y$$

confirming that the joint density of (H, C) is given by $-2\phi'(2y - x)$ for $x < y$.

In order to get the joint density of the High and the Close when the drift is non-zero, we need only multiply by the ratio of the two normal density functions of the close

$$\frac{f_\mu(x)}{f_0(x)}$$

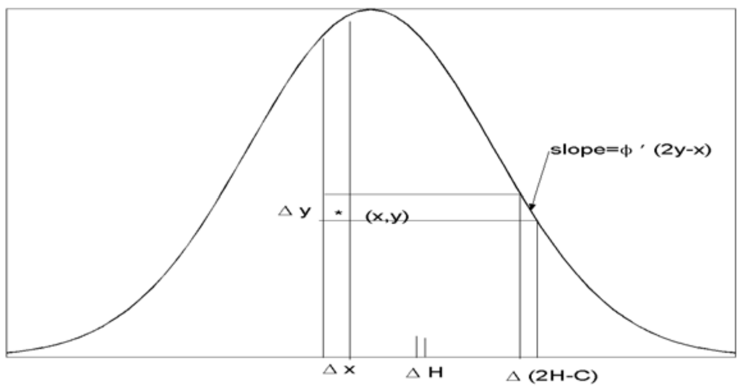

Figure 5.8: Confirmation of the joint density of (H, C)

and this gives the more general result in the table below.

The table below summarizes many of our distributional results for a Brownian motion process with drift on the interval $[0, 1]$,

$$dS_t = \mu dt + \sigma dW_t, \text{ with } S_0 = O.$$

Statistic	Density	Conditions
$X = C - O$, $Y = H - O$	$f(y, x) = -2\phi'(2y - x)\exp(\mu x - \mu^2/2)$	$-\infty < x < y$, and $y > 0, \sigma = 1$ given O
$Y\|X$	$f_{Y\|X}(y\|x) = 2(2y - x)e^{-2y(y-x)}$	$y > x, \sigma = 1$
$Z = Y(Y - X)$	$\exp((\sigma^2/2)$	given O, X
$(L - O)(L - C)$	$\exp(\sigma^2/2)$	given (O, C)
$(H - O)(H - C)$	$\exp(\sigma^2/2)$	given (O, C)
$\frac{H-O}{2H-O-C}$	$U[0, 1]$	drift $\nu = 0$, given $O, 2H - O - C$
$\frac{L-O}{2L-O-C}$	$U[0, 1]$	drift $\nu = 0$, given $O, 2L - O - C$
$\frac{C-O}{2H-O-C}$	$U[-1, 1]$	drift $\nu = 0$, given H, O

TABLE 5.1: Some distributional results for High, Close and Low.

We now consider briefly the case of non-zero drift for a geometric Brownian motion. Fortunately, all that needs to be changed in the results above is the marginal distribution of $\ln(C)$ since all conditional distributions given the value of C are the same as in the zero-drift case. Suppose an option has payoff on maturity $\psi(C)$ if an upper barrier at level $Oe^b, b > 0$ is not breached. We have already seen that to accommodate the filetering effect of this knock-out barrier we should determine, numerically or by simulation, the expected value

$$E[\psi(C)(1 - \exp\{-2\frac{b(b + \ln(O/C))}{\sigma^2 T}\})]$$

the expectation conditional (as always) on the value of the open O. The effect of a knock-out lower barrier at Oe^{-a} is essentially the same but with b replaced by a, namely

$$E[\psi(C)(1 - \exp\{-2\frac{a(a + \ln(C/O))}{\sigma^2 T}\})].$$

In the next section we consider the effect of two barriers, both an upper and a lower barrier.

One Process, Two barriers.

We have discussed a simple device above for generating jointly the high and the close or the low and the close of a process given the value of the open. The joint distribution of H, L, C given the value of O or the distribution of C in the case of upper and lower barriers is more problematic. Consider a single factor model and two barriers- an upper and a lower barrier. Note that the high and the low in any given interval is dependent, but if we simulate a path in relatively short segments, by first generating n increments and then generating the highs and lows within each increment, then there is an extremely low probability that the high and low of the process will both lie in the same short increment. For example for a Brownian motion with the time interval partitioned into 5 equal subintervals, the probability that the high and low both occur in the same increment is less than around 0.011 whatever the drift. If we increase the number of subintervals to 10, this is around 0.0008. This indicates that provided we are willing to simulate highs, lows and close in ten subintervals, pretending that within subintervals the highs and lows are conditionally independent, the error in our approximation is very small.

An alternative, more computationally intensive, is to differentiate the infinite series expression for the probability $P(H \leq b, L \geq a, C = u|O = 0]$ A first step in this direction is the the following result, obtained from the reflection principle with two barriers.

Theorem 47 *For a Brownian motion process*

$$dS_t = \mu dt + dW_t, S_0 = 0$$

defined on $[0, 1]$ *and for* $-a < u < b$,

$$\begin{aligned} &P(L < -a \text{ or } H > b|C = u) \\ &\quad = \frac{1}{\phi(u)} \sum_{n=1}^{\infty} [\phi\{2n(a+b) + u\} + \phi\{2n(a+b) - 2a - u\} \\ &\qquad\qquad + \phi\{-2n(b+a) + u\} + \phi\{2n(b+a) + 2a + u\}] \end{aligned}$$

where ϕ *is the* $N(0, 1)$ *probability density function.*

Proof. The proof is a well-known application of the reflection principal. It is sufficient to prove the result in the case $\mu = 0$ since the conditional distribution of L, H given C does not depend on μ (A statistician would say that C is a sufficient statistic for the drift parameter). Denote the following paths determined by their behaviour on $0 < t < 1$. All paths are assumed to end at $C = u$.

$$\begin{aligned}
A_{+1} &= H > b \text{ (path goes above } b) \\
A_{+2} &= \text{path goes above } b \text{ and then falls below } -a \\
A_{+3} &= \text{goes above } b \text{ then falls below } -a \text{ then rises above } b \\
&\text{etc.} \\
A_{-1} &= L < -a \\
A_{-2} &= \text{path falls below } -a \text{ then rises above } b \\
A_{-3} &= \text{falls below } -a \text{ then rises above } b \text{ then falls below } -a \\
&\text{etc.}
\end{aligned}$$

For an arbitrary event A, denote by $P(A|u)$ probability of the event conditional on $C = u$. Then according to the reflection principal the probability that the Brownian motion leaves the interval $[-a, b]$ is given from an inclusion-exclusion argument by

$$\begin{aligned}
&P(A_{+1}|u) - P(A_{+2}|u) + P(A_{+3}|u) - \cdots \\
&+P(A_{-1}|u) - P(A_{-2}|u) + P(A_{-3}|u) \cdots
\end{aligned} \tag{5.30}$$

This can be verified by considering the paths in Figure 5.9. (It should be noted here that, as in our application of the reflection principle in the one-barrier case, the reflection principle allows us to show that the number of paths in two sets is the same, and this really only translates to probability in the case of a discrete sample space, for example a simple random walk that jumps up or down by a fixed amount in discrete time steps. This result for Brownian motion obtains if we take a limit over a sequence of simple random walks approaching a Brownian motion process.)

Note that

$$\begin{aligned}
P(A_{+1}|u) &= \frac{\phi(2b-u)}{\phi(u)} \\
P(A_{+2n}|u) &= \frac{\phi\{2n(a+b)+u\}}{\phi(u)} \\
P(A_{+(2n-1)}|u) &= \frac{\phi\{2n(a+b)-2a-u\}}{\phi(u)}
\end{aligned}$$

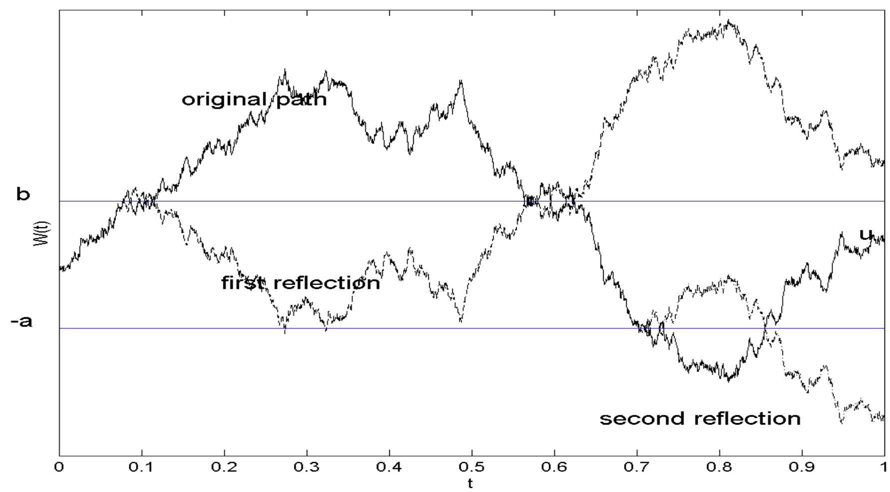

Figure 5.9: The Reflection principle with Two Barriers

and

$$P(A_{-1}|u) = \frac{\phi(-2a-u)}{\phi(u)}$$
$$P(A_{-2n}|u) = \frac{\phi\{-2n(b+a)+u\}}{\phi(u)}$$
$$P(A_{-(2n+1)}|u) = \frac{\phi\{2n(b+a)+2a+u\}}{\phi(u)}.$$

The result then obtains from substitution in (5.30). ■

As a consequence of this result we can obtain an expression for $P(a < L \leq H < b, u < C < v)$ (see also Billingsley, (1968), p. 79) for a Brownian motion on $[0,1]$ with zero drift:

$$\begin{aligned} P(a,b,u,v) &= P(a < L \leq H < b, u < C < v) \\ &= \sum_{k=-\infty}^{\infty} \Phi[v+2k(b-a)] - \Phi[u+2k(b-a)] \\ &- \sum_{k=-\infty}^{\infty} \Phi[2b-u+2k(b-a)] - \Phi[2b-v+2k(b-a)]. \end{aligned} \quad (5.31)$$

where Φ is the standard normal cumulative distribution function. From (5.31) we derive the joint density of (L, H, C) by taking the limit $P(a, b, u, u+\delta)/\delta$ as $\delta \to 0$, and taking partial derivatives with respect to a and b:

$$\begin{aligned} f(a,b,u) &= 4\sum_{k=-\infty}^{\infty} k^2\phi''[u+2k(b-a)] - k(1+k)\phi''[2b-u+2k(b-a)] \\ &= 4\sum_{k=1}^{\infty} k^2\phi''[u+2k(b-a)] - k(1+k)\phi''[2b-u+2k(b-a)] \\ &\quad + k^2\phi''[u-2k(b-a)] + k(1-k)\phi''[2b-u-2k(b-a)] \end{aligned} \tag{5.32}$$

for $a < u < b$.

From this it is easy to see that the conditional cumulative distribution function of L given $C = u, H = b$ is given by on $a \le u \le b$ (where $-2\phi'(2b-u)$ is the joint p.d.f. of H, C) by

$$\begin{aligned} F(a|b,u) &= 1 + \frac{\frac{\partial^2}{\partial b \partial v} P(a,b,u,v)|_{v=u}}{2\phi'(2b-u)} \qquad (5.33) \\ &= \frac{-1}{\phi'(2b-u)} \sum_{k=1}^{\infty} \{-k\phi'[u+2k(b-a)] + (1+k)\phi'[2b-u+2k(b-a)] \\ &\quad + k\phi'[u-2k(b-a)] + (1-k)\phi'[2b-u-2k(b-a)]\} \end{aligned}$$

This allows us to simulate both the high and the low, given the open and the close by first simulating the high and the close using $-2\phi'(2b-u)$ as the joint p.d.f. of (H, C) and then simulating the low by inverse transform from the cumulative distribution function of the form (5.33).

Survivorship Bias

It is quite common for retrospective studies in finance, medicine and to be subject to what is often called "survivorship bias". This is a bias due to the fact that only those members of a population that remained in a given class (for example the survivors) remain in the sampling frame for the duration of the study. In general, if we ignore the "drop-outs" from the study, we do so at risk of introducing substantial bias in our conclusions, and this bias is the survivorship bias.

Suppose for example we have hired a stable of portfolio managers for a large pension plan. These managers have a responsibility for a given portfolio over a period of time during which their performance is essentially under continuous review and they are subject to one of several possible decisions. If returns below a given threshhold, they are deemed unsatisfactory and fired or converted to another line of work. Those with exemplary performance are promoted, usually to an administrative position with little direct financial management. And those between these two "absorbing" barriers are retained. After a period of time, T, an amibitious graduate of an unnamed Ivey league school working out of head office wishes to compare performance of those still employed managing portfolios. How are should the performance measures reflect the filtering of those with unusually good or unusually bad performance? This is an example of a process with upper and lower absorbing barriers, and it is quite likely that the actual value of these barriers differs from one employee to another, for example the son-in-law of the CEO has a substantially different barriers than the math graduate fresh out of UW. However, let us ignore this difference, at least for the present, and concentrate on a difference that is much harder to ignore in the real world, the difference between the volatility parameters of portfolios, possibly in different sectors of the market, controlled by different managers. For example suppose two managers were responsible for funds that began and ended the year at the same level and had approximately the same value for the lower barrier as in Table 5.2. For each the value of the volatility parameter σ was estimated using individual historical volatilities and correlations of the component investments.

Portfolio	Open price	Close Price	Lower Barrier	Volatility
1	40	$56\frac{5}{8}$	30	.5
2	40	$56\frac{1}{4}$	30	.2

Table 5.2

Suppose these portfolios (or their managers) have been selected retrospectively from a list of "survivors" which is such that the low of the portfolio value never crossed a barrier at $l = Oe^{-a}$ (bankruptcy of fund or termination or demotion of manager, for example) and the high never crossed an upper barrier at $h = Oe^{b}$. However, for the moment let us assume that the upper barrier is so high that its influence can be neglected, so that the only absorbtion with any substantial probability is at the lower barrier. We interested in the estimate of return from the two portfolios, and a preliminary estimate indicates a continuously compounded rate of return from portfolio 1 of $R_1 = \ln(56.625/40) = 35\%$ and from portfolio two of $R_2 = \ln(56.25/40) = 34\%$. Is this difference significant and are these returns reasonably accurate in view of the survivorship bias?

We assume a geometric Brownian motion for both portfolios,

$$dS_t = \mu S_t dt + \sigma S_t dW_t, \tag{5.34}$$

$$\text{and define } O = S(0), \qquad C = S(T),$$

$$H = \max_{0 \le t \le T} S(t), \quad L = \min_{0 \le t \le T} S(t)$$

with parameters μ, σ possibly different.

In this case it is quite easy to determine the expected return or the value of any performance measure dependent on C conditional on survival, since this is essentially the same as a problem already discussed, the valuation of a barrier option. According to (5.27), the probability that a given Brownian motion process having open 0 and close c strikes a barrier placed at $l < \min(0, c)$ is

$$\exp\{-2\frac{z_l}{\sigma^2 T}\}$$

with

$$z_l = l(l - c).$$

Converting this statement to the Geometric Brownian motion (5.34), the probability that a geometric Brownian motion process with open O and close c breaches a lower barrier at l is

$$P[L \leq l|O, C] = \exp\{-2\frac{z_l}{\sigma^2 T}\}$$

with

$$z_l = \ln(O/l)\ln(C/l) = a(a + \ln(C/O)).$$

Of course the probability that a particular path with this pair of values (O, C) is a "survivor" is 1 minus this or

$$1 - \exp\{-2\frac{z_l}{\sigma^2 T}\}. \tag{5.35}$$

When we observe the returns or the closing prices C of survivors only, the results have been filtered with probability (5.35). In other words if the probability density function of C without any barriers at all is $f(c)$ (in our case this is a lognormal density with parameters that depend on μ and σ) then the density function of C of the survivors in the presence of a lower barrier is proportional to

$$\begin{aligned} & f(c)[1 - \exp\{-2\frac{\ln(O/l)\ln(c/l)}{\sigma^2 T}\}] \\ & = f(c)(1 - (\frac{l}{c})^\lambda), \text{ with } \lambda = \frac{2\ln(O/l)}{\sigma^2 T} = \frac{2a}{\sigma^2 T} > 0. \end{aligned}$$

It is interesting to note the effect of this adjustment on the moments of C for various values of the parameters. For example consider the expected value of C conditional on survival

$$\begin{aligned} E(C|L \geq l] &= \frac{\int_l^\infty cf(c)(1 - (\frac{l}{c})^\lambda)dc}{\int_l^\infty f(c)(1 - (\frac{l}{c})^\lambda)dc} \\ &= \frac{E[CI(C \geq l)] - l^\lambda E[C^{1-\lambda}I(C \geq l)]}{P[C \geq l] - l^\lambda E[C^{-\lambda}I(C \geq l)]} \end{aligned} \tag{5.36}$$

and this is easy to evaluate in the case of interest in which C has a lognormal distribution. In fact the same kind of calculation is used in the development of the Black-Scholes formula. In our case $C = \exp(Z)$ where Z is $N(\mu T, \sigma^2 T)$ and so for any p and $l > 0$, we have from (3.11), using the fact that $E(C|O) = O\exp\{\mu T + \sigma^2 T/2\}$, (and assuming O is fixed),

$$E[C^p I(C > l)] = O^p \exp\{p\mu T + p^2\sigma^2 T/2\}\Phi(\frac{1}{\sigma\sqrt{T}}(a + \mu T) + \sigma\sqrt{T}p)$$

To keep things slightly less combersome, let us assume that we observe the geometric Brownian motion for a period of $T = 1$. Then (5.36) results in

$$\frac{Oe^{\mu+\sigma^2/2}\Phi(\frac{1}{\sigma}(a+\mu)+\sigma) - Oe^{-a\lambda+(1-\lambda)\mu+(1-\lambda)^2\sigma^2/2}\Phi(\frac{1}{\sigma}(a+\mu)+\sigma(1-\lambda))}{\Phi(\frac{1}{\sigma}(a+\mu)) - e^{-\lambda a-\lambda\mu+\lambda^2\sigma^2/2}\Phi(\frac{1}{\sigma}(a+\mu)-\sigma\lambda)}$$

Let there be no bones about it. At first blush this is still a truly ugly and opaque formula. We can attempt to beautify it by re-expressing it in terms more like those in the Black-Scholes formula, putting

$$d_2(\lambda) = \frac{1}{\sigma}(\mu - a), \text{and } d_2(0) = \frac{1}{\sigma}(a + \mu),$$
$$d_1(\lambda) = d_2(\lambda) + \sigma, \quad d_1(0) = d_2(0) + \sigma.$$

These are analogous to the values of $d_{1,}d_2$ in the Black-Scholes formula in the case $\lambda = 0$. Then

$$E[C|L \geq l] = O\frac{e^{\mu+\sigma^2/2}\Phi(d_1(0)) - e^{-\lambda a+(1-\lambda)\mu+(1-\lambda)^2\sigma^2/2}\Phi(d_1(\lambda))}{\Phi(d_2(0)) - e^{-\lambda a-\lambda\mu+\lambda^2\sigma^2/2}\Phi(d_2(\lambda))}. \quad (5.37)$$

What is interesting is how this conditional expectation, the expected close for the survivors, behaves as a function of the volatility parameter σ. Although this is a rather complicated looking formula, we can get a simpler picture (Figure 5.10) using a graph with the drift parameter μ chosen so that $E(C) = 56.25$ is held fixed. We assume $a = -\ln(30/40)$ (consistent with Table 5.2)and vary the value of σ over a reasonable range from $\sigma = 0.1$ (a very stable investment) through $\sigma = .8$ (a highly volatile investment). In Figure 5.10 notice that for small volatility, e.g. for $\sigma \leq 0.2$, the conditional expectation $E[C|L \geq 30]$

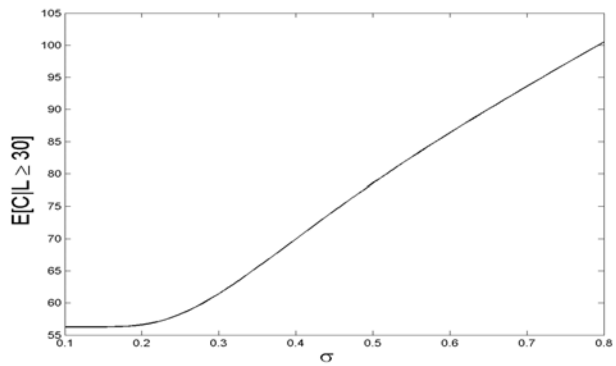

Figure 5.10: $E[C|L \geq 30]$ for various values of (μ, σ) chosen such that $E(C) = 56.25$.

remains close to its unconditional value $E(C)$ but for $\sigma \geq 0.3$ it increases almost linearly in σ to around 100 for $\sigma = 0.8$. The intuitive reason for this dramatic increase is quite simple. For large values of σ the process fluctuates more, and only those paths with very large values of C have abeen able to avoid the absorbing barrier at $l = 30$. Two comparable portfolios with unconditional return about 40% will show radically different apparent returns in the presence of an absorbing barrier. If $\sigma = 20\%$ then the survivor's return will still average around 40%, but if $\sigma = 0.8$, the survivor's returns average close to 150%. The practical implications are compelling. *If there is any form of survivorship bias (as there usually is), no measure of performance should be applied to the returns from different investments, managers, or portfolios without an adjustment for the risk or volatility.*

In the light of this discussion we can return to the comparison of the two portfolios in Table 5.2. Evidently there is little bias in the estimate of returns for portfolio 2, since in this case the volatility is small $\sigma = 0.2$. However there is very substantial bias associated with the estimate for portfolio 1, $\sigma = 0.5$. In fact if we repeat the graph of Figure 5.10 assuming that the unconditional

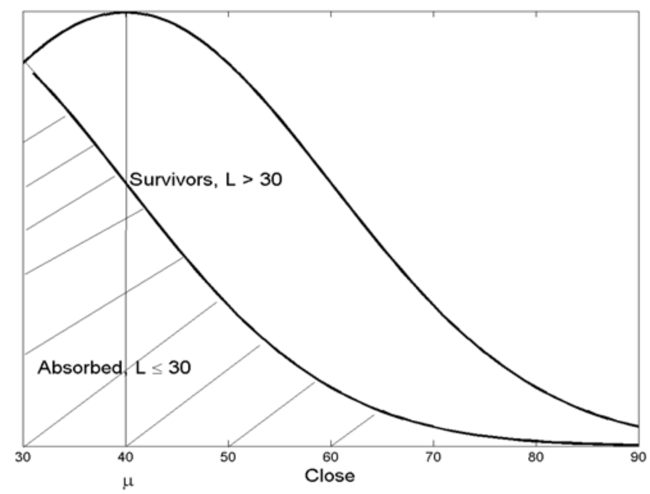

Figure 5.11: The Effect of Surivorship bias for a Brownian Motion

return is around 8% we discover that $E[C|L \geq 30]$ is very close to $56\frac{5}{8}$ when $\sigma = 0.5$ indicating that this is a more reasonable estimator of the performance of portfolio 1.

For a Brownian motion process it is easy to demonstrate graphically the nature of the surivorship bias. In Figure 5.11, the points under the graph of the probability density which are shaded correspond to those which whose low fell below the absorbing barrier at $l = 30$. The points in the unshaded region correspond to the survivors. The expected value of the return conditional on survival is the mean return (x-cooredinate of the center of mass) of those points chosen uniformly under the density but above the lower curve, in the region labelled "survivors". Note that if the mean μ of the unconditional density approaches the barrier (here at 30) , this region approaches a narrow band along the top of the curve and to the right of 30. Similarly if the unconditional standard deviation or volatility increases, the unshaded region stretches out to the right in a narrow band and the conditional mean increases.

We arrive at the following seemingly paradoxical conclusions which make it imperative to adjust for survivorship bias *The conditional mean, conditional on survivorship, may increase as the volatility increases even if the unconditional mean decreases.*

Let us return to the problem with both an upper and lower barrier and consider the distribution of returns conditional on the low never passing a barrier Oe^{-a} and the high never crossing a barrier at Oe^{b} (representing a fund buyout, recruitment of manager by competitor or promotion of fund manager to Vice President).It is common in process control to have an upper and lower barrier and to intervene if either is crossed, so we might wish to study those processes for which no intervention was required. Similarly, in a retrospective study we may only be able to determine the trajectory of a particle which has not left a given region and been lost to us. Again as an example, we use the following data on two portfolio managers, both observed conditional on survival, for a period of one year.

Portfolio	Open price	Close Price	Lower Barrier	Upper Barrier	Volatility
1	40	$56\frac{5}{8}$	30	100	.5
2	40	$56\frac{1}{4}$	30	100	.2

If ϕ denotes the standard normal p.d.f., then the conditional probability density function of $\ln(C/O)$ given that $Oe^{-a} < L < H < Oe^{b}$ is proportional to $\frac{1}{\sigma}\phi(\frac{u-\mu}{\sigma})w(u)$ where, as before

$$w(u) = 1 - e^{-2b(b-u)/\sigma^2} + e^{-2(a+b)(a+b-u)/\sigma^2} - e^{-2a(a+u)/\sigma^2} + e^{-2(a+b)(a+b+u)/\sigma^2} - E(W),$$

$$W = I[frac1(\frac{\ln(H)}{a+b}) > \frac{b}{a+b}] + I[frac1(\frac{-\ln(L)}{a+b}) > \frac{a}{a+b}], \quad \text{and}$$

$$b = \ln(100/40), \quad a = -ln(30/40).$$

The expected return conditional on survival when the drift is μ is given by

$$E(\ln(C/O)|30 < L < H < 100) = \frac{1}{\sigma}\int_{-a}^{b} uw(u)\phi(\frac{u-\mu}{\sigma})du.$$

where $w(u)$ is the weight function above. Therefore a moment estimator of the drift for the two portfolios is determined by setting this expected return equal to the observed return, and solving for μ_i the equation

$$\frac{1}{\sigma_i}\int_{-a}^{b} uw(u)\phi(\frac{u-\mu_i}{\sigma_i})du = R_i, \qquad i = 1, 2.$$

The solution is, for portfolio 1, $\mu_1 = 0$ and for portfolio 2, $\mu_2 = 0.3$. Thus the observed values of C are completely consistent with a drift of 30% per annum for portfolio 2 and a zero drift for portfolio 1. The bias again very strongly effects the portfolio with the greater volatility and estimators of drift should account for this substantial bias. Ignoring the survivorship bias has led in the past to some highly misleading conclusions about persistence of skill among mutual funds.

Problems

1. If the values of d_j are equally spaced, i.e. if $d_j = j\Delta, j = ..., -2, -1, 0, 1, ...$and with $S_0 = 0, S_T = C$ and $M = \max(S_0, S_T)$, show that

$$E[H|C = u] = M + \Delta\frac{P[C > u \text{ and } \frac{C-M}{\Delta} \text{ is even}]}{P[C = u]}.$$

2. Let $W(t)$ be a standard Brownian motion on $[0, 1]$ with $W_0 = 0$. Define $C = W(1)$ and $H = \max\{W(t); 0 \le t \le 1\}$. Show that the joint probability density function of (C, H) is given by

$$f(c, h) = 2\phi(c)(2h - c)e^{-2h(h-c)}, \text{ for } h > \max(0, c)$$

where $\phi(c)$ is the standard normal probability density function.

3. Use the results of Problem 2 to show that the joint probability density function of the random variables

$$Y = \exp\{-(2H - C)^2/2\}$$

and C is a uniform density on the region $\{(x, y); y < \exp(x^2/2)\}$.

4. Let $X(t)$ be a Brownian motion on $[0, 1]$, i.e. X_t satisfies

$$dX_t = \mu dt + \sigma dW_t, \text{ and } X_0 = 0.$$

Define $C = X(1)$ and $H = \max\{X(t); 0 \leq t \leq 1\}$. Find the joint probability density function of (C, H).