MONTE CARLO METHODS IN FINANCE

Don L. McLeish

August, 2000

Contents

Chapter 0

Introduction

This book concerns the analysis of models for financial markets, particularly the assets traded there. We pay particular attention to financial derivatives such as options and futures. These are financial instruments which derive their value from some associated asset. For example a call option is written on a particular stock, and its value at expiry depends on the price of the stock at expiry. But there are many other types of financial derivatives, traded on assets such as bonds, currency markets or foreign exchange markets, and commodities. Indeed there is a growing interest in so-called "real options", those written on some real-world physical process such as the temperature or the amount of rainfall.

In general, an option gives the holder a right, not an obligation, to sell or buy a prescribed asset (the underlying asset) at a price determined by the contract (the exercise or strike price). For example if you own a call option on shares of IBM with expiry date Oct. 20, 2000 and exercise price $120, then on October 20, 2000 you have the right to purchase a fixed number , say 100 shares of IBM at the price $120. If IBM is selling for $130 on that date, then your option is worth $10 per share on expiry. If IBM is selling for $120 or less, then your option is worthless. We need to know what a fair value would be for this option when it is sold, say on February 1, 2000. Determining this fair value relies on sophisticated models both for the movements in the underlying asset and the relationship of this asset with the derivative, and is the subject of a large part of this book. You may have bought an IBM option for two possible reasons, either because you are speculating on an increase in the stock price, or to hedge a promise that you have made to deliver IBM stocks to someone in the future against possible increases in the stock price. The second use of derivatives is similar to the use of an insurance policy against movements in an asset price that could damage or bankrupt the holder of a portfolio. It is this second use of derivatives that has fueled most of the phenomenal growth in their trading. With the globalization of economies, industries are subject to more and more economic forces that they are unable to control but nevertheless wish some form of insurance against. This requires a hedges against a whole

litany of disadvantageous moves of the market such as increases in the cost of borrowing, decreases in the value of assets held, changes in a foreign currency exchange rates, etc.

The advanced theory of finance, like many areas where advanced mathematics plays an important part, is undergoing a revolution aided and abetted by the computer and the proliferation of powerful simulation and symbolic mathematical tools. This is the mathematical equivalent of the invention of the printing press. The numerical and computational power once reserved for the most highly trained mathematicians, engineers is now available to all.

One of the first hurdles faced before adopting stochastic or random models in finance is the recognition that for all practical purposes, the prices of equities in an efficient market are *random variables*, that is while they may show some dependence on fiscal and economic processes and policies, they have a component of randomness that makes them unpredictable. This appears on the surface to be contrary to the training we all receive that every effect has a cause, and every change in the price of a stock must be driven by some factor in the company or the economy. But we should remember that random models are often applied to systems that are essentially causal when measuring and analyzing the various factors influencing the process and their effects is too monumental a task. Even in the simple toss of a fair coin, the result is predetermined by the forces applied to the coin during and after it is tossed. In spite of this, we model it as a random variable because we have insufficient information on these forces to make a more accurate prediction of the outcome. Most financial processes in an advanced economy are of a similar nature. Exchange rates, interest rates and equity prices are subject to the pressures of a large number of traders, government agencies, speculators, as well as the forces applied by international trade and the flow of information. In the aggregate there is an extraordinary number of forces and information that influence the process. While we might hope to predict some features of the process such as the average change in price or the volatility, a precise estimate of the price of an asset one year from today is clearly impossible. This is the basic argument necessitating stochastic models in finance. A stochastic model does not militate against some ability to forecast. It is adopted whenever we acknowledge that a process is not perfectly predictable and the non-predictable component of the process is of sufficient importance to attempt to model.

Now if we accept that the price of a stock is a random variable, what are the constants in our model? Is a dollar of constant value, and if so, the dollar of which nation? Or should we accept one unit of a index what in some sense represents a share of the global economy as the constant? This question concerns our choice of what is called the "numeraire" in deference to the French influence on the theory of probability, or the process against which the value of our assets will be measured. We will see that there is not a unique answer to this question, nor does that matter for most purposes. We can use a bond denominated in Canadian dollars as the numeraire or one in US dollars. Provided we account for the variability in the exchange rate, the price of an asset will be the same. So to some extent our choice of numeraire is arbitrary- we may pick whatever

is most convenient for the problem at hand.

One of the most important modern tools for analyzing a stochastic system is simulation. Simulation is the imitation of a real-world process or system. It is essentially a model, often a mathematical model of a process. In finance, a basic model for the evolution of stock prices, interest rates, exchange rates etc. would be necessary to determine a fair price of a derivative security. Simulations, like purely mathematical models, usually make assumptions about the behaviour of the system being modelled. This model requires inputs, often called the parameters of the model and outputs a result which might measure the performance of a system, the price of a given financial instrument, or the weights on a portfolio chosen to have some desirable property. We usually construct the model in such a way that inputs are easily changed over a given set of values, as this allows for a more complete picture of the possible outcomes.

Why use simulation? The simple answer is that is that it transfers work to the computer. When compared with a purely mathematical analysis, models with more complexity and fewer assumptions, models that are closer to the real-world, are possible. By changing parameters we can examine interactions, and sensitivities of the system to various factors. Experimenters may either use a simulation to provide a numerical answer to a question, assign a price to a given asset, identify optimal settings for controllable parameters, examine the effect of exogenous variables or identify which of several schemes is more efficient or more profitable. The variables that have the greatest effect on a system can be isolated. We can also use simulation to verify the results obtained from an analytic solution. For example many of the tractable models used in finance to select portfolios and price derivatives are wrong. They put too little weight on the extreme observations, the large positive and negative movements (crashes), which have the most dramatic effect on the results. Is this lack of fit of major concern when we use a standard model such as the Black-Scholes model to price a derivative? Questions such as this one can be answered in part by examining simulations which accord more closely with the real world, but which are intractable to mathematical analysis.

Simulation is also used to answer questions starting with "what if". For example, What would be the result if interest rates rose 3 percentage points over the next 12 months? In engineering, determining what would happen under more extreme circumstances is often referred to as stress testing and simulation is a particularly valuable tool here since the scenarios we are concerned about are those that we observe too rarely to have a substantial experience of. Simulations are used, for example, to determine the effect of an aircraft of flying under extreme conditions and is used to analyse the flight data information in the event of an accident. Simulation often provides experience at a lower cost than the alternatives.

But these advantages are not without some sacrifice. Two individuals may choose to model the same phenomenon in different ways, and as a result, may have quite different simulation results. Because the output from a simulation is random, it is sometimes harder to analysis- some statistical experience and tools are a valuable asset. Building models and writing simulation code is not

always easy- time is required both to construct the simulation, validate it, and to analysis the results. And simulation does not render all mathematical analysis unnecessary- if a reasonably simple analytic expression for a solution exists it is always preferable to a simulation. While a simulation may provide an approximate numerical answer at one or more possible parameter values, only an expression for the solution provides insight to the way in which it responds to the individual parameters, the sensitivities of the solution.

In constructing a simulation, you should be conscious of a number of distinct steps;

1. Formulate the problem at hand. Why do we need to use simulation?
2. Set the objectives as specifically as possible. This should include what measures on the process are of most interest.
3. Suggest candidate models. Which of these are closest to the real-world? Which are fairly easy to write computer code for? What parameter values are of interest?
4. If possible, collect real data and identify which of the above models is most appropriate. Which does the best job of generating the general characteristics of the real data?
5. Implement the model. Write computer code to run simulations.
6. Verify (debug) the model. Using simple special cases, insure that the code is doing what you think it is doing.
7. Validate the model. Ensure that it generates data with the characteristics of the real data.
8. Determine simulation design parameters. How many simulations are to be run and what alternatives are to be simulated?
9. Run the simulation. Collect and analyse the output.
10. Are there surprises? Do we need to change the model or the parameters? Do we need more runs?
11. Finally we document the results and conclusions in the light of the simulation results. Tables of numbers are to be avoided. Well-chosen graphs are often better ways of gleaning qualitative information from a simulation.

In this book, we will not always follow our own advice, leaving some of the above steps for the reader to fill in. Nevertheless, the importance of model validation, for example, cannot be overstated. Particularly in finance where data is often plentiful, highly complex mathematical models are too often applied without any evidence that they fit the observed data adequately. The reader is advised to consult and address the points in each of the steps above with each new simulation (and many of the examples in this text).

Example

Let us consider the following example illustrating a simple use for a simulation model. We are considering a buy-out bid for the shares of a company. Although the company's stock is presently valued at around $11.50 per share, a careful analysis has determined that it fits sufficiently well with our current assets that if the buy-out were successful, it would be worth approximately $14.00 per share in our hands. We are considering only three alternatives. An immediate cash offer of $12.00, $13.00 or $14.00 per share for outstanding shares of the company. Naturally we would like to bid as little as possible. Unfortunately we expect a competitor to also make a bid for the company, and the competitor values the shares differently. Moreover there are costs associated with losing a given bid to the competitor. Suppose the payoff to our firm depends on the amount bid by the competitor and the possible scenarios are as given in the following table.

			Competitor		
	Bid	A	B	C	
Your	12	3	2	-2	
Co.	13	1	-4	4	
	14	0	-5	5	

The payoffs to the competitor are somewhat different and given below

		Competitor	Bid	
		A	B	C
Your	12	-1	-2	3
Bid	13	0	4	-6
	14	0	5	-5

Define the 3×3 matrix of payoffs listed in Table 1 by A and that in Table 1 by B. Provided that you play strategy $i = 1, 2, 3$ (i.e. bid $12,$13,$14 with probabilities p_1, p_2, p_3 respectively and the probabilities of the competitor's strategies are $q_j, j = 1, 2, 3$, then we can write the expected payoff to you in the form $\sum_{i=1}^{3} \sum_{j=1}^{3} p_i A_{ij} q_j.$, which, when written as a vector-matrix product, takes the form $p^T Aq$. This can be thought of as the average return to your firm in the long run if this game were repeated many times. In this case you would clearly choose $p_i = 1$ for the row i corresponding to the maximum component of Aq, if the vector q were known to you. Similarly your competitor would choose $q_j = 1$ for the column j corresponding to the maximum component of $p^T B$. Over the long haul, if this game were indeed repeated may times, you would likely keep track of your opponent's frequencies and replace the unknown probabilities by the frequencies. This would seem to be a reasonable approach to building a simulation model for this game. Play the game repeatedly with each of the two players updating the estimated probabilities with which their opponent uses their available strategies and record the number of times each strategy is used. In this case we run the function

function [p,q]=nonzerosum(A,B,nsim)

% A and B are payoff matrices to the two participants in a game. Outputs

%mixed strategies p and q determined by simulation conducted nsim times

```
n=size(A);                          % A and B have the same size
p=ones(1,n(1)); q=ones(n(2),1);  % initialize with positive weights on all strategies
for i=1:nsim
[m,s]=max(A*q); p(s)=p(s)+1;
[m,s]=max(p*B); q(s)=q(s)+1;
end
p=p-ones(1,n(1)); p=p/sum(p);
q=q-ones(n(2),1); q=q/sum(q);
```

The following output results from running this function for 50,000 simulations.

```
[p,q]=nonzerosum(A,B,50000)
p =  0.6667    0.2222     0.1111
q = 0  0.5000   0.5000
```

and this seems to indicate that the strategies will be "mixed" or random. You should choose a bid of \$12.00 with probability around 2/3, \$13.00 with probability about 2/9 and \$14.00 with probability 1/9. It appears that the competitor need only toss a fair coin and select between B and C based on its outcome. Why randomize your choice? If you were to choose a single strategy as your "best" then your competitor could presumably determine what your "best" strategy is and act to reduce your return while increasing theirs. Only randomization provides the necessary insurance that neither player can guess the strategy to be employed by the other. This is a rather simple example of a two-person game with non-constant sum (in the sense that A+B is not a constant matrix). Mathematical analysis of such games can be quite complex. Participants may compete or cooperate for a greater total return.

While there is no assurance that the solution is optimal, it is in this case easily seen to be sensible solution, achieved with little effort. Indeed in a game such as this, there is no clear definition of what an optimal strategy would be. Do you plan your play based on the worst case or the best case scenario or something in between such as some form of average? Do you attempt to collaborate with your competitor for greater total return and then subsequently divide this in some fashion? Here the simulation has emulated a simple form of competitor behaviour and arrived at a reasonable solution.

There remains the question of how we actually select a bid with probabilities 2/3, 2/9 and 1/9 respectively. First let us assume that we are able to choose a "random number" U in the interval [0,1] so that the probability that it falls in any given subinterval is proportional to the length of that subinterval. This means that the random number has a uniform distribution on the interval [0,1]. Then we could determine our bid based on the value of this random number from the following table;

If	$U < 2/3$	$2/3 \leq U < 8/9$	$8/9 \leq U < 1$
Bid	12	13	14

The way in which U is generated on a computer will be discussed in more detail in chapter 2, but for the present note that each of the three alternative bids have the correct probability.

Chapter 1

Some Basic Theory of Finance.

1.1 Introduction to Pricing: Single Period Models.

Let us begin with a very simple example designed to illustrate the no-arbitrage approach to pricing derivatives. Consider a stock whose price at present is $\$s$. Over a given period, the stock may move either up or down, up to a value su where $u > 1$ with probability p or down to the value sd where $d < 1$ with probability $1-p$. In this model, these are the only moves possible for the stock in a single period. Over a longer period, of course, many other values are possible. In this market, we also assume that there is a so-called risk-free bond available returning a guaranteed rate of $r\%$ per period. Such a bond cannot default; there is no random mechanism governing its return which is known upon purchase. An investment of \$1 at the beginning of the period returns a guaranteed $\$(1+r)$ at the end. Then a portfolio purchased at the beginning of a period consisting of y stocks and x bonds will return at the end of the period an amount $\$x(1+r) + ysZ$ where Z is a random variable taking values u or d with probabilities p and $1-p$ respectively. We permit owning a negative amount of a stock or bond, corresponding to shorting or borrowing the correspond asset for immediate sale.

An ambitious investor might seek a portfolio whose initial cost is zero (i.e. $x + ys = 0$) such that the return is greater than or equal to zero with positive probability. Such a strategy is called an *arbitrage* since with a net investment of \$0, we are able to achieve non-negative return (i.e. the possibility of future profits with no down-side risk). In mathematical terms, the investor seeks a point (x, y) such that $x + ys = 0$ (net cost of the portfolio is zero) and

$$\begin{aligned} x(1+r) + ysu &\geq 0, \\ x(1+r) + ysd &\geq 0 \end{aligned}$$

with at least one of the two inequalities strict (so there is never a loss and a non-zero chance of a positive return). Alternatively, is there a point on the line $y = -\frac{1}{s}x$ which lies *above both* of the two lines

$$\begin{aligned} y &= -\frac{1+r}{su}x \\ y &= -\frac{1+r}{sd}x \end{aligned}$$

and strictly above one of them? Since all three lines pass through the origin, we need only compare the slopes; an arbitrage will NOT be possible if

$$-\frac{1+r}{sd} \leq -\frac{1}{s} \leq -\frac{1+r}{su} \tag{1.1}$$

and otherwise there is a point (x, y) permitting an arbitrage. The condition for no arbitrage **??** reduces to

$$\frac{d}{1+r} < 1 < \frac{u}{1+r} \tag{1.2}$$

So the condition for no arbitrage demands that $(1 + r - u)$ and $(1 + r - d)$ have opposite sign or $d \leq (1 + r) \leq u$. Unless this occurs, the stock *always* has either better or worse returns than the bond, which makes no sense in a free market where both are traded without compulsion. Under a no arbitrage assumption since $d \leq (1 + r) \leq u$, the bond payoff is a *convex combination* or a weighted average of the two possible stock payoffs; i.e. there are probabilities $0 \leq q \leq 1$ and $(1 - q)$ such that $(1 + r) = qu + (1 - q)d$. In fact it is easy to solve this equation to determine the values of q and $1 - q$.

$$q = \frac{(1+r)-d}{u-d}, \qquad \text{and} \quad 1-q = \frac{u-(1+r)}{u-d}.$$

Denote the probability measure which puts probabilities q and $1 - q$ on the same points su, sd by Q. Then note that if S_1 is the value of the stock at the end of the period,

$$\frac{1}{1+r}E^Q(S_1) = \frac{1}{1+r}(qsu + (1-q)sd) = \frac{1}{1+r}s(1+r) = s$$

where E^Q denotes the expectation assuming that Q describes the probabilities of the two outcomes.

In other words, *if there is to be no arbitrage, there exists a probability measure Q such that the expected price of future value of the stock S_1 discounted to the present using the return from a risk-free bond is exactly the present value of the stock.* The measure Q is called the *risk-neutral* measure and the probabilities that it assigns to the possible outcomes of S are not necessarily those that determine the future behaviour of the stock. The risk neutral measure embodies both the current consensus beliefs in the future value of the stock and the

consensus investors' attitude to risk avoidance. It is not necessarily true that $\frac{1}{1+r}E^P(S_1) = s$ with P denoting the actual probability distribution describing the future probabilities of the stock. Indeed it is highly unlikely that an investor would wish to purchase a risky stock if he or she could achieve exactly the same expected return with no risk at all using a bond. We generally expect that to make a risky investment attractive, its expected return should be greater than that of a risk-free investment. Notice in this example that the risk-neutral measure Q did not use the probabilities p, and $1-p$ that the stock would go up or down and this seems contrary to intuition. Surely if a stock is more likely to go up, then a call option on the stock should be valued higher! Let us suppose for example that we have a friend willing to value a stock using the actual distribution P different from Q. Then discounted to the present, the friend believes that the stock is worth

$$\frac{1}{1+r}E^P S_1 = \frac{psu + (1-p)sd}{1+r} \neq s \quad \text{since } p \neq q.$$

Such a friend offers their assets as a sacrifice to any investor. If the friend's assessed price is greater than the current market price, buy on the market and sell to the friend. Otherwise, do the reverse. In any case, you are richer (except of course by the loss of one friend)!

So why should we use the Q measure to determine the price of a given asset in a market (assuming, of course, there is a risk-neutral Q measure and we are able to determine it)? Not because it precisely describes the future behaviour of the stock, *but because if we use any other distribution, we offer an intelligent investor (there are many!) an arbitrage opportunity, or an opportunity to make money at no risk and at our expense.*

Derivatives are investments which derive their value from that of a corresponding asset, such as a stock. A *European call option* is an option which permits you (but does not compel you) to purchase the stock at a future time for a given predetermined price, the exercise price of the option). For example a call option with exercise price \$10 on a stock whose future value is denoted S_1, is worth on expiry $S_1 - 10$ if $S_1 > 10$. This is the difference between the value of the stock on expiry and the exercise price of the option or your profit on purchasing the stock for \$10 and selling it on the open market at $\$S_1$. However, if $S_1 < 10$, there is no point whatever in exercising your option as you are not compelled to do so and your return is \$0. In general, your payoff from purchasing the option is a simple function of the future price of the stock, such as $V(S_1) = \max(S_1 - 10,\ 0)$. The future value of the option is itself a random variable but it derives its value from that of the stock, hence it is called a *derivative.*

Now consider an arbitrary function of the stock, $X = V(S_1)$ representing the payoff to an investor from a certain financial instrument (derivative) when the stock price at the end of the period is S_1. Any function of the stock price is called a *contingent claim.* In our example above, the random variable takes only two possible values $V(su)$ and $V(sd)$. We will show that there is a portfolio, called a *replicating* portfolio, consisting of an investment solely in the above

stock and bond which reproduces these values $V(su)$ and $V(sd)$ exactly. We can determine the corresponding weights on the bond and stocks (x, y) simply by solving the two equations in two unknowns

$$\begin{aligned} x(1+r) + ysu &= V(su) \\ x(1+r) + ysd &= V(sd) \end{aligned}$$

Solving: $y^* = \frac{V(su)-V(sd)}{su-sd}$ and $x^* = \frac{V(su)-y^*su}{1+r}$. Upon solving these two equations we are able to replicate the contingent claim $V(S_1)$ exactly- i.e. buy x^* units of bond and y^* of stock, we produce a portfolio of stocks and bonds with exactly the same return as the contingent claim. So in this case at least, there can be only one possible present value for the contingent claim and that is the present value of the replicating portfolio $x^* + y^*s$. If the market placed any other value on the contingent claim, then a trader could guarantee a positive return by a simple trade, shorting the contingent claim and buying the equivalent portfolio or buying the contingent claim and shorting the replicating portfolio. Thus this is the only price that precludes an arbitrage opportunity. There is a simpler expression for the current price of the contingent claim in this case: Note that

$$\begin{aligned} \frac{1}{1+r}E^Q V(S_1) &= \frac{1}{1+r}(qV(su) + (1-q)V(sd)) \\ &= \frac{1}{1+r}(\frac{1+r-d}{u-d}V(su) + \frac{u-(1+r)}{u-d}V(sd)) \\ &= x^* + y^*s. \end{aligned}$$

In words, *the discounted expected value of the contingent claim is equal to the no-arbitrage price of the derivative where the expectation is taken using the* $Q-measure$. Indeed any contingent claim that is attainable must have its price determined in this way. While we have developed this only in an extremely simple case, it extends more generally to complete markets, or markets in which any contingent claim is attainable by an investment in other marketable instruments. The following theorem provides a more general proof of this result, the proof due to Chris Rogers.

Suppose we hae a total of N risky assets whose prices at times $t = 0, 1$, are given by $(S_0^j(\omega), S_1^j(\omega)), j = 1, 2, ..., N$ for possible states ω. For simplicity assume that these states have positive probability $P(\omega) > 0$. Assume also a riskless asset (a bond) paying interest rate r over one unit of time. Suppose we borrow (i.e. short bonds) at the risk-free rate to buy w_j units of stock j at time 0 for a total cost of $\sum w_j S_0^j(\omega)$. The value of this portfolio at time $t = 1$ is $\sum w_j(S_1^j(\omega) - (1+r)S_0^j(\omega))$. Then we say there are *no arbitrage opportunities if* for all weights w_j this is either identically 0 for all ω or it takes both positive and negative values.

Theorem 1 *A necessary and sufficient condition that there be no arbitrage opportunities is that there exists a measure Q equivalent to P such that $E_Q(S_1^j) = \frac{1}{1+r}S_0^j$ for all $j = 1, ..., N$.*

Problem 2 Proof. *Define $M(w) = E[\exp(\sum w_j(S_1^j - (1+r)S_0^j))]$ and consider the problem*

$$\min_w \ln(M(w)).$$

If there is no arbitrage opportunity then there is a unique minimum satisfying $\frac{\partial M}{\partial w_j} = 0$ or

$$E[\exp(\sum w_j(S_1^j - (1+r)S_0^j))S_1^j] = (1+r)S_0^j E[\exp(\sum w_j(S_1^j - (1+r)S_0^j))].$$

or

$$S_o^j = \frac{E[\exp(\sum w_j S_1^j)S_1^j]}{(1+r)E[\exp(\sum w_j S_1^j)]}.$$

Define a measure Q equivalent to the original probability measure such that

$$Q(\omega) = P(\omega)\frac{\exp(\sum w_j S_1^j(\omega))]}{E[\exp(\sum w_j S_1^j)]}$$

Then note that for each j,

$$E_Q(S_1^j) = \frac{1}{1+r}S_0^j$$

so the current price of each stock is the discounted expected value of the future price under the risk-neutral measure Q. Conversely if

$$E_Q(S_1^j) = \frac{1}{1+r}S_0^j$$

holds for some measure Q then $E_Q[\sum w_j(S_1^j - (1+r)S_0^j)] = 0$ for all w_j and this implies that the random variable $\sum w_j(S_1^j - (1+r)S_0^j)$ is either identically 0 or admits both posiive and negative values. Therefore the existence of the measure Q implies that there are not arbitrage opportunities. ■

So the theory of pricing derivatives in a complete market is based on a rather trivial observation. If we can reproduce exactly the same (random) returns as the derivative provides using a linear combination of other marketable securities (which have prices signed by the market) then the derivative must have the same price as the linear combination of other securities. Any other price would provide arbitrage opportunities.

Of course in the real world, there are costs associated with trading, these costs usually represented by a bid-ask spread. In other words there is a different

price for buying a security and for selling it. The argument above assumes a frictionless market with no trading costs, with borrowing any amount at the risk-free bond rate possible, and a completely liquid market- any amount of any security can be bought or sold. Moreover it assumes a complete market and it is questionable whether such a market can exist. For example if a derivative security can be perfectly replicated using other marketable instruments, then what is the purpose of the derivative security in the market? So like all models, this one has its deficiencies and its critics. Like all reasonable models, its merit is that it provides an approximation to a real-world phenomenon, one that permits further study and improvement.

1.2 Multiperiod Models.

When an asset price evolves over time, we normally allow the investor to make decisions about an investment at various periods during the life of that investment. Such decisions are made with the benefit of information, and this information, whether used or not, includes the price of the asset and any related assets at all previous time periods, beginning at some time $t = 0$ when we began observation of the process. We denote this information available for use at time t as H_t. Formally, H_t is what is called a *sigma-field* generated by the past, and there are two fundamental properties of this sigma-field that will use. The first is that the sigma-fields increase over time. In other words, our information about this and related processes increases over time because we have observed more of the relevant history. In the mathematical model, we do not "forget" relevant information, though regrettably this is an increasingly important factor in real-life. The second property of H_t that we use is that it includes the value of every asset price at time t, or, in measure-theoretic language, S_t *is adapted to or measurable with respect to* H_t. Now the analysis above shows that when our investment life began at time $t = 0$, and we were planning for the next period of time, there was defined a risk-neutral measure Q such that $E^Q(\frac{1}{1+r}S_1) = S_0$. Imagine now that we are in a similar position at time t, planning our investment for the next unit time. All expected values should be taken in the light of our current knowledge, i.e. given the information H_t. An identical analysis to that above shows that under the risk neutral measure Q, if S_t represents the price of the stock after t periods, and r_t the risk-free one-period interest rate offered that time, then

$$E^Q(\frac{1}{1+r_t}S_{t+1}|H_t) = S_t. \tag{1.3}$$

Suppose we let B_t be the value of \$1 invested at time $t = 0$ after a total of t periods. Then $B_1 = (1 + r_0)$, $B_2 = (1 + r_0)(1 + r_1)$, and in general $B_t = (1 + r_0)(1 + r_1)...(1 + r_{t-1})$. If I were to promise you exactly \$1.00 payable at time t (and if you believed me), then to cover this promise I would require an investment at time $t = 0$ of $\$1/B_t$, which we might call (at least in

the case when the interest rates r_t are known) the *present value* of the promise. In general, at time t, the present value of a certain amount $\$V$ promised at time T (i.e. the present value or the value discounted to the present of this payment) is $V\frac{B_t}{B_T}$. Now suppose we divide (**??**) above by B_t. We obtain

$$E^Q(\frac{S_{t+1}}{B_{t+1}}|H_t) = E^Q(\frac{1}{B_t(1+r_t)}S_{t+1}|H_t) = \frac{1}{B_t}E^Q(\frac{1}{1+r_t}S_{t+1}|H_t) = \frac{S_t}{B_t}. \tag{1.4}$$

Notice that we are able to take the divisor B_t outside the expectation since B_t is known at time t and therefore a constant with respect to the history H_t. This equation (**??**) describes an elegant mathematical property shared by all marketable securities in a complete market. Under the risk-neutral measure, the discounted price $Y_t = S_t/B_t$ forms a *martingale*. A *martingale* is a process Y_t for which the expectation of a future value given the present is equal to the present i.e.

$$E(Y_T|H_t) = Y_t \text{ for all } T > t. \tag{1.5}$$

A martingale is a fair game in a world with no inflation, no need to consume and no mortality. Your future fortune if you play the game is a random variable whose expectation, given everything you know at present, is your present fortune. Now the condition (**??** implies of the process $Y_t = S_t/B_t$ that $E(Y_{t+1}|H_t) = Y_t$ for all $T > t$ and this implies the martingale condition (**??**) since $E(Y_T|H_t) = E[...E[E(Y_T|H_{T-1})|H_{T-2}]...|H_t] = Y_t$.

Thus, under a risk-neutral measure Q in a complete market, all marketable securities discounted to the present form martingales. For this reason, we often refer to the risk-neutral measure as a martingale measure. And the fact that prices of marketable commodities must be martingales under the risk neutral measure has many consequences to the canny investor. Suppose, for example, you believe that you are able to model the history of the price process nearly perfectly, and it tells you that the price of a share of XXX computer systems increases on average 2% per year. Should you use this $P-$measure in valuing a derivative, even if you are confident it is absolutely correct, in pricing a call option on XXX computer systems with maturity one year from now? If you do so, you are offering some arbitrager another free lunch at your expense. The measure Q, not the measure P, determines derivative prices in a no-arbitrage market. This also means that there is no advantage when pricing derivatives in using some elaborate statistical method to estimate the expected rate of return.

What have we discovered? In general, prices in a market are determined as expected values, but expected values with respect to the measure Q. This is true in any complete market, regardless of the number of assets traded in the market; for any future time $T > t$, and for any derivative defined on the traded assets in a market, $E^Q(\frac{B_t}{B_T}V(S_T)|H_t] = V_t =$ the market price of the asset at time t. So in theory, determining a reasonable price of a derivative should be a simple task, one that could be easily handled by simulation. In order to

determine a suitable price for a derivative simply generate many simulations of the future value $V(S_T)$ of the derivative given the current store of information H_t, simulations conducted under the measure Q, and then average the values, discounted to the present, over all simulations. The catch is that the Q measure is neither obvious from the present market prices nor statistically estimable from its past. It is given implicitly by the fact that the expected value of the discounted future value of traded assets must produce the present market price. In other words, a first step in valuing any asset is to determine a measure Q which has this property. Now in a simple model involving a single stock, this is a fairly simple job, and there is a unique such measure Q. This is the case, for example, for the stock model above in which the stock moves in simple steps, either increasing or decreasing at each step. But as the number of traded assets increases, and as the number of possible jumps per period changes, a measure Q which completely describes the stock dynamics and which has the necessary properties for a risk neutral measure becomes potentially much more complicated.

Solving for the Q Measure.

Let us consider the following simple example. Over each period, a stock price provides greater or less or the same return as a risk free investment. Assume for simplicity that increases are by the factor $u(1+r)$ and decreases by factor $(1+r)/u$ where $u > 1$ and otherwise the stock price increases by the risk free rate factor $(1+r)$. The Q probability of increases and decreases is unknown, and may vary from one period to the next. Over two periods, the possible paths executed by this stock price process are displayed below assuming that the stock begins at time $t = 0$ with price s_0.

In general in such a tree there are three branches from each of the nodes at times $t = 0, 1$ and there are a total of $1 + 3 = 4$ such nodes. Thus, even if we assume that probabilities of up and down movements do not depend on

how the process arrived at a given node, there is a total of $3 \times 4 = 12$ unknown parameters. Of course there are constraints; for example the sum of the three probabilities on branches exiting a given node must add to zero and the price process must form a martingale. For each of the four nodes, this provides two constraints for a total of 8 constraints, leaving 4 parameters to be estimated. We would need the market price of 4 different derivatives or other contingent claims to be able to generate 4 equations in these 4 unknowns and solve for them. Provided we are able to obtain prices of four such derivatives, then we can solve these equations. Consider the following special case, with the risk-free interest rate per period=$r = 1\%$, $u = 1.089$, $s_0 = \$1.00$. We also assume that we are given the price of four call options expiring at time $T = 2$. The possible values of the price at time $T = 2$ corresponding to two steps up, one step up and one constant, one up one down, are the values of $S(T)$ in the set

$\{\ 1.1859, 1.0890, 1.0000, 0.9183, 0.8432\}$. Recall that a call option expiring at time $T = 2$ has price the value $E^Q(S_2 - K)^+$ discounted to the present, where K is the exercise price of the option and S_2 is the price of the stock at time 2. We have market prices $E^Q(S_2 - K)^+/(1 + r)^2$ of four call options with the same expiry and different exercise prices in the following table;

The price of the call options are in the following table

Exercise Price	Maturity	Call Option Price
0.867	2	0.154
0.969	2	.0675
1.071	2	.0155
1.173	2	.0016

Since in general the price of a call option with exercise price K and maturity date $T = 2$ is given by $E^Q(S_2-K)^+/(1+r)^2$, The equations to be solved equate the observed price of the options to their theoretical price $E(S_2 - K)^+/(1 + r)^2$ and are as follows;

$$
\begin{aligned}
0.0016 &= \frac{1}{(1.01)^2}(1.186 - 1.173)p_1p_2 \\
0.0155 &= \frac{1}{(1.01)^2}[(1.186 - 1.071)p_1p_2 + (1.089 - 1.071)\{p_1(1 - 2p_2) + (1 - 2p_1)p_2\}] \\
0.0675 &= \frac{1}{(1.01)^2}[0.217p_1p_2 + 0.12\{p_1(1 - 2p_2) + (1 - 2p_1)p_2\} \\
&\quad +0.031\{(1 - 2p_1)(1 - 2p_2) + p_1p_2 + p_1p_4)\} \\
0.154 &= \frac{1}{(1.01)^2}[0.319p_1p_2 + 0.222\{p_1(1 - 2p_2) + (1 - 2p_1)p_2\} \\
&\quad +0.133\{(1 - 2p_1)(1 - 2p_2) + p_1p_2 + p_1p_4)\} \\
&\quad +0.051\{\{p_1(1 - 2p_4) + (1 - 2p_1)p_3\}]
\end{aligned}
$$

While it is not too difficult to solve this system in this particular case (in this case the solution is given by $p_1 = 0.2, p_2 = 0.22, p_3 = 0.2, p_4 = 0.3$) one can see that

with more branches and more derivatives, this non-linear system of equations becomes difficult very quickly. What do we do if we only have market prices for two derivatives defined on this stock, and therefore only two parameters which can be obtained from the market information? This is an example of what is called an incomplete market, a market in which the risk neutral distribution is not uniquely specified by market information. In general when we have fewer equations than parameters in a model, there are really only two choices

(a) simplify the model so that the number of parameters and the number of equations match.

(b) Determine additional natural criteria or constraints that the parameters must satisfy.

In this case, for example, one might prefer a model in which the probability of a step up or down depends on the time, but not on the current price of the stock. This assumption would force equal all of $p_2 = p_3 = p_4$ and simplify the system of equations above. For example using only the prices of the first two derivatives, we obtain equations, which, when solved, determine the probabilities on the other branches as well.

$$\begin{aligned} 0.0016 &= \frac{1}{(1.01)^2}(1.186 - 1.173)p_1p_2 \\ 0.0155 &= \frac{1}{(1.01)^2}[(1.186 - 1.071)p_1p_2 + (1.089 - 1.071)\{p_1(1 - 2p_2) + (1 - 2p_1)p_2\}] \end{aligned}$$

This example reflects a basic problem which occurs often when we build a reasonable and flexible model in finance. Frequently there are more parameters than there are marketable securities. It is quite common to react by simplifying the model. For example, it is for this reason that binomial trees (with only two branches emanating from each node) are often preferred to the trinomial tree example we use above, even though they provide a substantially worse approximation to the actual distribution of stock returns.

In general if there are n different securities (excluding derivatives whose value is a function of one or more of these) and if each security can take any one of m different values, then there are a total of m^n possible states of nature at time $t = 1$. The Q measure must assign a probability to each of them. This results in a total of m^n unknown probability values, which, of course must add to one, and result in the right expectation for each of n marketable securities. To uniquely determine Q we would require a total of $m^n - n - 1$ equations or $m^n - n - 1$ different derivatives. For example for $m = 10$, $n = 100$, approximately one with a hundred zeros, a prohibitive number, are required to uniquely determine Q. But in a complete market, Q is uniquely determined by marketable securities. No real market can be complete. And in real markets, one asset is not perfectly replicated by a combination of other assets. This is true whether one asset is a derivative defined as a function of another marketed security (and interest rates and volatilities). The most we can probably hope for in practice is to find a model or measure Q in a subclass of measures with

desirable features under which

$$E^Q(\frac{B_t}{B_T}V(S_T)|H_t] \approx V_t \text{ for all marketable } V$$

Now if these were equalities, this would represent a number of equations in the unknown Q probabilities, typically fewer equations than unknowns so some simplification of the model is required before settling on a measure Q. One could, at one's peril, ignore the fact that certain factors in the market depend on others. Similar stocks behave similarly, few are really independent. Can we, with any reasonable level of confidence, accurately predict the effect that a lowering of interest rates will have on a given bank stock? Perhaps the best model for the future behaviour of most processes is the past, except that as we have seen the historical distribution of stocks do not generally produce a risk-neutral measure. Even if historical information provided a flawless guide to the future, there is too little of it to accurately estimate the large number of parameters required for a simulation of a market of reasonable size. Some simplification of the model is clearly necessary. Are some baskets of stocks independent of other combinations? What independence can we reasonably assume over time?

As a first step in simplifying a model, consider some of the common measures of behaviour. Stocks can go up, or down. The drift of a stock is a tendency in one or other of these two directions. But it can also go *up and down*- by a lot or a little. The measure of this, the variance or variability in the stock returns is called the *volatility* of the stock. Our model should have as ingredients these two quantities. It should also have as much dependence over time and among different asset prices as we have evidence to support.

1.3 Determining the Process B_t.

We have seen in the last section that given the Q or risk-neutral measure, we can (at least in theory) determine the price of a derivative if we are given the "numeraire" or the price B_t of a risk-free investment at time t. Unfortunately no such investment is traded on the open market. There are government treasury bills which, depending on the government, one might wish to assume are almost risk-free, and there are government bonds, usually with longer terms, which complicate matters by paying dividends periodically. The question dealt with in this section is whether we can estimate the process B_t given information on the prices of these bonds.

We begin with what we know. We assume we know the current prices, components of the vector, S_t of marketable securities. We also know the price of certain risk-free bonds with face value F , the value of the bond on maturity at time T. These pricesP_t provide some information on the bank account process B_t. In particular since a dividend-paying bond is a linear combination of payments at certain times $t < T$ plus a final payment of F, each current bond

price provides a value of the form $P_t = \sum_{T>s>t} d_s B_t/B_s + FB_t/B_T$. This can be written as a system of linear equations

$$P_t/B_t = \sum_{T>s>t} d_s/B_s + F/B_T$$

and provided that we have a sufficient number of bond prices P_t, possibly with different maturities, this system permits solving for certain values of $1/B_s$, $s > t$. Now the catch here is that there are typically too few risk-free bond maturities to get a detailed picture of the process $1/B_s, s > 0$. We could use government bonds for this purpose. But are these genuinely risk-free? Might not the additional use of bonds in large highly rated companies provide a more detailed picture of the bank account process B_s.

Can incorporate information on bond prices from lower grade debt? To do so, we need a simple model linking the debt rating of a given bond and the probability of default and payoff to the bond-holders in the event of default. To begin with, let us assume that a given basket of companies, say those with a common debt rating from one of the major bond rating organisations, have a common distribution of default time. We will also assume in this preliminary model that once default occurs, provided it occurs before the maturity date T of the bond, the payoff is a constant proportion p of the principal amount F owing. Then if τ denotes the time of default, a bond with face value $\$F$ which promises dividend payments d_s at time $s < T$ has price at time t given by

$$\begin{aligned} P_t &= \sum_{t<s<T} \frac{B_t}{B_s} d_s P(\tau > s|\tau > t) + \frac{pFB_t}{B_T} P(\tau \le T|\tau > t) + \frac{FB_t}{B_T} P(\tau > T|\tau > t) \\ &= \sum_{t<s<T} \frac{B_t}{B_s} d_s P(\tau > s|\tau > t) + \frac{pFB_t}{B_T} + \frac{(1-p)FB_t}{B_T} P(\tau > T|\tau > t). \end{aligned}$$

All probabilities are conditional on the event $[\tau > t]$ because unless this is true the debt has already defaulted and therefore its value is known. Unknowns in this equation are $P(\tau > s|\tau > t)/B_s$, $t < s < T$, $\frac{1}{B_T}$ and $pP(\tau > T|\tau > t)$. Now if we died and went to investors' heaven, a bond of every maturity T would be sold and we could solve this system of equations simply using the given bond prices. We might also hope that the probabilities of default are very small and follow a simple pattern. If the pattern is not perfect, then little harm results provided that indeed the default probabilities are small. Suppose for example that the time of default follows a geometric or exponential distribution so that the probability of a default occurring in any period of fixed length is constant. Then $P(\tau > s|\tau > t) = \exp\{-k(s-t)\}$ for some $k > 0$. Suppose we define a new bank account process

$$\widetilde{B_s} = \frac{B_s}{P[\tau > s]} = B_s \exp\{ks\} \text{ for } s > t.$$

Clearly this bank account grows faster than the original, and it grows faster as the probability of default increases. The effective interest rate on this account is k units per period higher. Then rewriting the above equation for P_t,

$$P_t - \frac{pFB_t}{B_T} = \sum_{t<s<T} \frac{\widetilde{B_t}}{\widetilde{B_s}} d_s + (1-p)F\frac{\widetilde{B_t}}{\widetilde{B_T}}$$

This equation has a simple interpretation. The left side is the price of the bond reduced by the present value of the guaranteed payment on maturity Fp. The right hand side is the current value of a risk-free bond paying the same dividends, with interest rates augmented by k and with face value $F(1-p)$. *So to value a defaultable bond, augment the interest rate, change the face value to the potential loss of face value on default and then add the present value of the guaranteed payment on maturity.* Given only three bond prices with the same default characteristics, for example, and assuming constant interest rates so that $B_s = \exp(rs\}$, we may solve for the values of the three unknown parameters (r, k, p).

1.4 Minimum Variance Portfolios and the Capital Asset Pricing Model.

Let us begin by building a model for portfolios of securities that capture more or less of the major market movements. We have solved above for the values of $1/B_s$ only for certain values of s, but let us assume for the present that by interpolation, B_s is known for all $s > t$.

To begin with, define a common measure on investments that is equivalent to price, but from many perspectives, more convenient and stable statistically. For a security that has price $S(t)$ and $S(t+1)$ at times t and $t+1$, we define the return $R_i(t+1)$ on the security over this increment by

$$R_i(t+1) = \frac{S_i(t+1) - S_i(t)}{S_i(t)}.$$

Returns can be measured in units that are easily understood (for example 5% or 10% per unit time) and independent of the amount invested. It is also easy to obtain the price at time t from the initial price at time 0 and the sequence of returns.

$$S_i(t) = S_i(0)(1 + R_i(1))(1 + R_i(2))...(1 + R_i(t)).$$

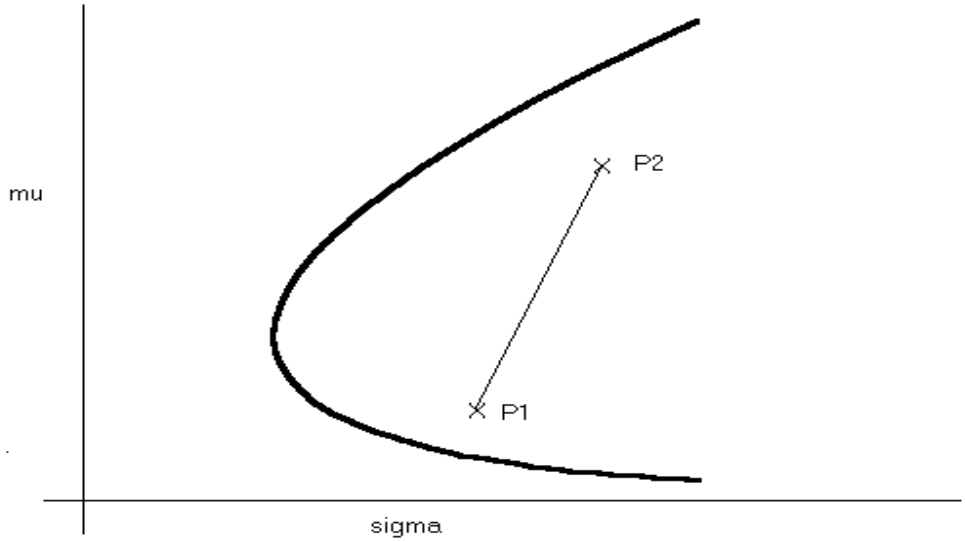
When we buy a portfolio the return on the portfolio is simply a weighted average of the individual stock returns and the weights are the relative amounts invested in each stock. For example if our portfolio is such that $w_i(t)$ is the proportion of our total investment of $\$I$ invested in stock i at time t then the number of shares of stock i purchased is $\frac{w_i(t)}{S_i(t)}I$ and the portfolio return over the next period is

$$\frac{\sum \frac{w_i(t)}{S_i(t)} I[S_i(t+1) - S_i(t)]}{\sum w_i(t) I} = \sum w_i(t) R_i(t+1).$$

When time is measured continuously, the instantaneous return process might be defined by a limit of the above form $R_i(t) = \lim_{h\to 0} \frac{S_i(t+h)-S_i(t)}{hS_i(t)}$ or formally $dS_i/S_i = d(\ln(S_i(t))$ provided these are well-defined. More generally, the returns process is a process whose product integral results in the original stock price. process.

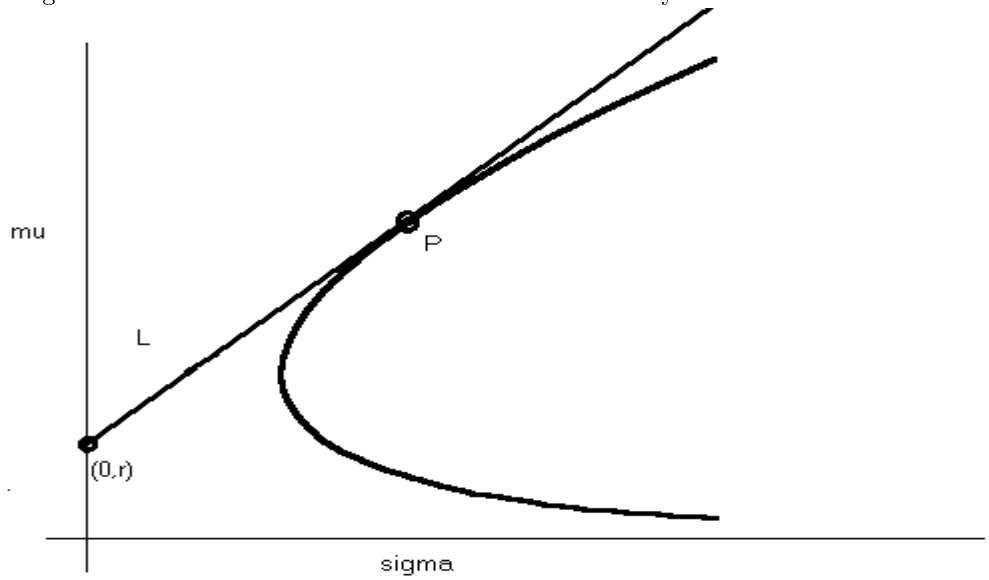
1.4.1 The Capital Asset Pricing Model (CAPM)

In the Capital Asset Pricing model it is assumed at the outset that investors concentrate on two measures of return, its expected value μ and its standard deviation σ. We consider the expected values and variances under the real-world probability measure P not under the risk-neutral Q measure. Suppose, for example we were to plot the set of all possible pairs (σ, μ) for portfolios of risky stocks. Let us assume for the present that the vector of all stock returns has mean return column vector given by η and covariance matrix of returns given by Σ. If a portfolio is such that we invest proportion w_i of our wealth in stock i, then defining $w = (w_1, ..., w_n)^T$, the total return on the portfolio has mean $\eta^T w$ and covariance matrix $w^T \Sigma w$. The set of all possible pairs of standard deviation and mean return $(\sqrt{w^T \Sigma w}, \eta^T w)$ has a semi-elliptical boundary as in the following figure.



Now assuming that investors prefer higher expected return for the same standard deviation, only the upper envelope or roof of this region is efficient in the sense that no other portfolio has higher expected return for the same standard deviation. This is called the efficient frontier. Now the picture changes substantially if there is also a risk-free investment that all investors are able to include, because in this case there is also a point on the $\mu-$axis corresponding to

$\sigma = 0$. If this point is added, then the efficient frontier is now the region below the line L in the following figure. The point $(0, r)$ corresponds to the risk free investment whose return is r, and the point P is the point at which this line is tangent to the efficient frontier determined from the risky investments.



If all investors have access to the same risk free rate, then the line L is the unique efficient frontier for all investors, and P is the only efficient point in the risky portfolio. It must therefore represent the total market in the sense that the proportions βw_i represents the proportion of the total market invested in stock i. Suppose the market portfolio P has standard deviation σ_P and mean μ_P. Then this line is described by the relation

$$\mu = r + \frac{\mu_P - r}{\sigma_P}\sigma.$$

For any stock i with mean and standard deviation of return (μ_i, σ_i) to be competitive, it must lie on this efficient frontier, i.e. it must satisfy the relation

$$\mu_i - r = \beta_i(\mu_P - r), \quad \text{where} \quad \beta_i = \frac{\sigma_i}{\sigma_P}.$$

This is the most important result in the capital asset pricing model. The constant β_i called the *beta* of a stock is both the change in the expected stock return for each unit change in the market expected return and also the ratio of the standard deviations of return of the stock and the market. Such a model is a reasonable basis for simplifying the covariance structure of stock returns to some manageable number of parameters. Analogous to the relation above describing the expected returns is a regression model relating the returns from the stock R_i and from the market portfolio R_P.

$$R_i - r = \beta_i(R_P - r) + \epsilon_i$$

where ϵ_i is a zero-mean random error uncorrelated with the market return. Taking variance on both sides, we obtain

$$var(R_i) = \beta_i^2 var(R_P) + var(\epsilon_i) = \sigma_i^2 + var(\epsilon) > \sigma_i^2$$

which appears to indicate that the variance of the return from stock i is greater than the value σ_i^2 assumed earlier. What is the cause of this contradiction? We assumed earlier that the stock i lay on the efficient frontier, but this is not a necessary condition for investors to choose it. All that is required is that it form a part of a portfolio which lies on the efficient frontier. We cannot expect a higher rate of return to compensate for additional risk that can be diversified away. In an efficient market, not all risk need be rewarded with additional return. Suppose, for example, we had many stocks with similar β, then we could presumably invest equally in all of them and end up with the average of many returns $\bar{R}$. Notationally,

$$\bar{R} - r = \beta(R_P - r) + \bar{\epsilon}$$

where, provided that we have sufficiently many such investments to average over, $\bar{\epsilon}$ has mean zero and variance close to 0 and is itself almost 0. By averaging or diversifying, we are able to provide an investment with the same average return characteristics but smaller variance than the original stock, implying that the original stocks could not have been on the efficient frontier. We say that the specific risk (i.e. $var(\epsilon_i)$) associated with stock i can be diversified away, and should not therefore be rewarded with increased return. Only the systematic risk σ_i is rewarded with increased expected return.

The capital asset pricing model provides a simplified form of the covariance matrix Σ of the vector of stock returns. Notice that under the model

$$R_i - r = \beta_i(R_P - r) + \epsilon_i, \qquad var(\epsilon_i) = \delta_i$$

we have

$$cov(R_i, R_j) = \beta_i \beta_j \sigma_P^2, i \neq j, \qquad var(R_i) = \beta_i^2 \sigma_P^2 + \delta_i.$$

Whereas N stocks would otherwise require a total of $N(N+1)/2$ parameters in the covariance matrix Σ of returns, the Capital Asset Pricing Model allows us to reduce this to the $N+1$ parameters σ_P^2, and δ_i, $i = 1, ..., N$. There is the disadvantage in this formula however that every pair of stocks in the same market must be *positively correlated,* a feature that seems to contradict some observations at least over substantial periods of time.

Minimum Variance under Q.

Suppose we wish to find a portfolios of securities which has the smallest possible variance under Q. . For example for a given set of weights, define the portfolio

$\Pi(t) = \sum w_i(t)S_i(t)$. For any weights $w_i(t)$ this produces a portfolio with exactly the same conditional rate of return under the Q measure as any one of the constituent stocks, for

$$E_Q[\Pi(t+1)|H_t] = \sum_i w_i(t)E_Q[S_i(t+1)|H_t] = \sum_i w_i(t)\frac{B(t+1)}{B(t)}S_i(t) = \frac{B(t+1)}{B(t)}\Pi(t).$$

In terms of returns, this equation says that all securities have the same expected return under Q as does the portfolio Π. If we are interested in hedging our investment, we wish to minimize the instantaneous variance of the return of the portfolio since the conditional mean is unaffected by the choice of weights. The natural constraint is that the cost of the portfolio is determined by the amount $c(t)$, say, that we presently have to invest. Several choices of the function $c(t)$ present themselves as possible. For example you might wish to compare, at time t, the benefits of liquidating a risk free investment of $B(t)$ for investment in this portfolio, in which case we could take $c(t) = B(t)$. Since $c(t)$ is purely a matter of rescaling the results, one might also scale so that $c(t) = 1$. Alternatively, we might wish to study a self-financing portfolio $\Pi(t)$, one for which past gains (or perish the thought, past losses) only are available to pay for the current portfolio. In this case $c(t) = \Pi(t)$. We wish to minimise

$$var_Q[\Pi(t+1)|H_t] \text{ subject to the contraint } \sum_i w_i(t)S_i(t) = c(t).$$

The solution is quite easy to obtain, and in fact the weights are given by the vector

$$W_1(t) = \begin{pmatrix} w_1(t) \\ w_2(t) \\ . \\ . \\ . \\ w_n(t) \end{pmatrix} = \frac{c(t)}{S'(t)\Sigma_t^{-1}S(t)}\Sigma_t^{-1}S(t).$$

where $\Sigma_t = var_Q(S(t+1)|H_t)$ is the instantaneous conditional covariance of $S(t)$ under the measure Q. If my objective were to minimize risk under the Q measure, then this portfolio is optimal for fixed cost. The conditional variance of this portfolio is given by

$$var_Q(\Pi(t+1)|H_t) = W_1'(t)\Sigma_t W_1(t) = \frac{c^2(t)}{S'(t)\Sigma_t^{-1}S(t)}.$$

Once again we express this relation in terms of the return $R_\Pi(t+1) = \frac{\Pi(t+1)-\Pi(t)}{\Pi(t)}$ from the portfolio Π. Assume for the present that the portfolio is self-financing so that $c(t) = \Pi(t)$. Then the above relation states that the conditional variance of the return $R_\Pi(t+1)$ given the past is simply

$$var_Q(R_\Pi(t+1)|H_t) = \frac{1}{S'(t)\Sigma_t^{-1}S(t)}.$$

The covariances between returns for individual stocks and the portfolio Π as it turns out are given by exactly the same quantity, namely

$$cov(R_i(t+1), R_\Pi(t+1)|H_t) = \frac{1}{S'(t)\Sigma_t^{-1}S(t)}.$$

By this formula, note that all such covariances are positive and therefore in the definition of $\Pi(t)$, all weights are positive and so no stocks are shorted. Now suppose that we wish to model the individual stock returns using a model for the minimum variance portfolio Π. Under the risk neutral measure Q, recall that every stock (and portfolio) has the risk-free rate of return $r(t+1) = \frac{B(t+1)-B(t)}{B(t)}$. Then we have by standard regression formulae,

$$R_i(t+1) - r(t+1) = \beta(R_\Pi(t+1) - r(t+1)) + \varepsilon_i(t+1)$$

where $\varepsilon_i(t+1)$ is a (random) error whose expected values under Q are all zero. The regression coefficient $\beta = \beta(t) = \frac{1}{S'(t)\Sigma_t^{-1}S(t)}$ is the same for all stocks. Of course the usual "beta" in the CAPM model in finance is the regression coefficient with a market rate of return under the actual measure P and these do vary from stock to stock, as do the expected returns.

Let us summarize our findings so far. We assume that the conditional covariance matrix Σ_t of the vector of stock prices is non-singular. *Under the risk neutral measure, all stocks have exactly the same expected returns equal to the risk-free rate. There is a unique self-financing minimum-variance portfolio* $\Pi(t)$ *and all stocks have exactly the same conditional covariance* β *with* Π. *All stocks have exactly the same regression coefficient* β *when we regress on the minimum variance portfolio.*

The question arises whether there are other minimum variance portfolios uncorrelated with this one. Suppose we define $\Pi_2(t)$ similarly to minimize the variance subject to the condition that the weights $W_2(t)$ satisfy $W_2'(t)\Sigma_t W_1(t) = 0$. This implies that the corresponding portfolio $\Pi_2(t)$ satisfies $Cov_Q(\Pi_2(t+1), \Pi(t+1)|H_t) = 0$. In view of the definition of $W_1(t)$, this implies that $W_2'(t)S(t) = 0$ or that the cost of such a portfolio at the beginning of periods is 0. This means that the portfolio is such that there is a perfect balance between long and short stocks, or that the value of the long and short stocks are equal. Subject to this restriction it is possible to make the conditional variance as large or as small as we wish, simply by scaling up or down by a common factor the amount of our investments.

The above analysis assumes that our objective is minimizing the variance of the portfolio. However, under the risk neutral measure Q, every stock has the same risk-free rate of return and so it is equally logical to minimize the variance of the *portfolio return.* By the same analysis as above, this is achieved when the proportion of our total investment at each time period in stock i is chosen as components of the vector $\frac{\Sigma_t^{-1}1}{1'\Sigma_t^{-1}1}$ where now Σ_t is the conditional covariance matrix of the stock *returns*. This may appear to be a different criterion and hence a different solution, but since at each step the stock price is a linear

function of the return $S_i(t+1) = S_i(t)(1+R_i(t))$ the variance minimizing portfolios are essentially the same.

Before we continue, let us examine how practical the above decomposition is for a large market. Note that it requires knowledge of Σ_t and worse still, it requires inverting this matrix. We will try to avoid estimating all $\frac{n(n+1)}{2}$ parameters in Σ_t by using the form implied by the Capital Asset Pricing Model $\Sigma_t = \beta\beta'\sigma_P^2(t) + \Delta(t)$, where $\sigma_P^2(t)$ is the market volatility at time t, $\Delta(t)$ is the diagonal matrix with the $\delta_i(t)$ along the diagonal and β is the vector of individual stock betas. In this case $\Sigma_t^{-1} = \Delta^{-1} + c\Delta^{-1}\beta\beta'\Delta^{-1}$ where

$$c = \frac{-1}{\sigma_p^{-2} + \sum_i \beta_i^2/\delta_i} = -\sigma_p^2 \frac{1}{1 + \sum_i \beta_i^2\sigma_p^2/\delta_i}$$

and consequently the optimal investment in each stock i is proportional to

$$\frac{1}{\delta_i} + c\beta_i(\sum_j \beta_j/\delta_j)$$

$$\text{or } \beta_i + \frac{1}{c\delta_i(\sum_j \beta_j/\delta_j)}$$

The conditional variance of $R_i(t+1)$ given the market at time $t+1$ is δ_i. Let us call this the excess volatility for stock i. *Then the weights for the optimal portfolio are linear in the beta for the stock and the reciprocal of the excess volatility.*

1.5 Entropy: choosing a Q measure

Typically market information does not completely determine the risk-neutral measure Q . We will argue that while the historical data should not strictly determine the Q measure, it should be used to fill in the information that is not dictated by no-arbitrage considerations. In order to relate the real world to the risk-free world, we need either sufficient market data to completely describe a risk-neutral measure Q (such a model is called a *complete market)* or we need to limit our candidate class of Q measures somewhat. We may either define the joint distributions of the stock prices or their returns, since from one we can pass to the other. For convenience, suppose we describe the joint distribution of the returns process. The conditions we impose on the martingale measure are the following;

1. Under Q, each normalized stock price $S_j(t)/B_t$ and derivative price V_t/B_t forms a martingale. Equivalently, $E^Q[S_i(t+1)|H_t] = S_i(t)r(t+1)$ where $r(t+1)$ is the risk free interest rate over the interval $(t, t+1)$. (Recall that this risk-free interest rate $r(t+1)$ is defined by the equation $B(t+1) = (1+r(t+1))B(t)$.)

2. Q is a probability measure.

A slight revision of notation is necessary here. We will build our joint distributions conditionally on the past and if P denotes the joint distribution stock prices $S(1), S(2), ...S(T)$ over the whole period of observation $0 < t < T$ then P_{t+1} denotes the conditional distribution of $S(t+1)$ given H_t. Let us denote the conditional moment generating function of the vector $S(t+1)$ under the measure P_{t+1} by

$$m_t(u) = E_P[\exp(u'S(t+1)|H_t] = E_P[\exp(\sum_i u_i S_i(t+1))|H_t]$$

We implicitly assume, of course, that this moment generating function exists. Suppose, for some vector of parameters η we choose Q_{t+1} to be the exponential tilt of P_{t+1}, i.e.

$$dQ_{t+1}(s) = \frac{\exp(\eta' s)}{m_t(\eta)} dP_{t+1}(s)$$

The division by $m_t(\eta)$ is necessary to ensure that Q_{t+1} is a probability measure.

Why transform a density by multiplying by an exponential in this way? There are many reasons for such a transformation. Exponential families of distributions are built in exactly this fashion and enjoy properties of sufficiency, completeness and ease of estimation. But we also argue that the measure Q is the probability measure which is closest to P in a certain sense while still satisfying the required moment constraint. We must begin with the notion of entropy which underlies considerable theory in Statistics and elsewhere in Science.

1.5.1 Cross Entropy

Consider two probability measures P and Q. Then the cross entropy or Kullbach-Leibler distance between the two measures is given by

$$H(Q|P) = \sup_{\{E_i\}} \sum Q(E_i) \log \frac{Q(E_i)}{P(E_i)}$$

where the supremum is over all finite partitions $\{E_i\}$ of the probability space. It is not hard to show that this measure is always non-negative and if Q is absolutely continuous with respect to P this can be rewritten in the form

$$H(Q|P) = E^Q \log(\frac{dQ}{dP}).$$

If, however, Q is not absolutely continuous with respect to P then the cross entropy is infinite. We should also remark that the cross entropy is not really a metric in the usual sense (although we unashamedly use the term distance in reference to it) since in general $H(Q|P) \neq H(P|Q)$. Now the following result asserts that the probability measure Q which is closest to P but satisfies a constraint on its mean is generated by an exponential tilt of the distribution of P.

Theorem 3 *: Minimizing cross-entropy.*

Consider the problem

$$\min_Q H(Q|P)$$

subject to the constraint $E^Q(f(X)) = \mu$. *Then the solution, if it exists, is given by*

$$dQ = \frac{\exp(\eta' f(X))}{m(\eta)} dP$$

where $m(\eta) = E^P[\exp(\eta' f(X))]$ *and* η *is chosen so that* $\frac{m'(\eta)}{m(\eta)} = \mu$.

The proof of this result, in the case of a discrete distribution P is a straightforward use of Lagrange multipliers. We leave it as a problem at the end of the chapter.

Now let us return to the constraints on the vector of stock prices. In order that $E^Q[S(t+1)|H_t] = (1 + r(t+1))S(t)$ we require that $\eta = \eta_t$ be chosen so that

$$\int s \frac{\exp(\eta' s)}{m_t(\eta)} dP_{t+1}(s) = \frac{m_t'(\eta)}{m_t(\eta)} = \frac{d}{d\eta} \log(m_t(\eta)) = (1 + r(t+1))S(t).$$

Of course, the parameter $\eta = \eta_t$ is dependent on time since it depends on the conditional distribution given H_t. Now consider a measure Q given by

$$dQ = \frac{\exp(\sum_t \eta_t' dx_t)}{\Pi_t m_t(\eta_t)} dP$$

Theorem 1.4.5 shows that this exponentially tilted distribution has the property of being the closest to the original measure P while satisfying the condition that the normalized sequence of stock prices forms a martingale. There is a continuous time analogue of this result, which, for completeness, we simply state below and reserve the proof for the problem set. The statement requires understanding of the continuous time models of the next section, and so can be skipped on first reading.

Theorem 4 .

Suppose under a probability measure P, the stock price process S_t satisfies an Ito equation of the form

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

Then the measure Q which satisfies

$$\min_Q H(Q|P)$$

subject to the constraint that the drift of the process is given by the risk free interest rate, (formally written $E^Q[dS_t|H_t] = r_t S_t dt$) is such that

$$dQ = \exp(\int \eta_t dS_t - \int (\eta_t \mu_t + \eta_t^2 \sigma_t^2/2)dt)dP$$

where

$$\eta_t = \frac{S_t r_t - \mu_t}{\sigma_t^2}$$

The effect of this change of measure is such that under Q, S_t satisfies an Ito equation with the same diffusion term and drift determined by the risk-free interest rate

$$dS_t = r_t S_t dt + \sigma_t dW_t$$

1.5.2 Maximum Entropy

In 1948 in a fundamental paper on the transmission of information, C. E. Shannon proposed the following idea of *entropy*. The entropy of a distribution attempts to measure the expected number of steps required to determine a given outcome of a random variable with a given distribution when using a simple binary poll. For example suppose that a random variable X has distribution given by

x	0	1	2
$P[X = x]$	.25	.25	.5

In this case, if we ask first whether the random variable is ≥ 2 and then, provided the answer is no, if it is ≥ 1, the expected number of queries to ascertain the value of the random variable is $1 + 1(1/2) = 1.5$. There is no more efficient scheme for designing this binary poll in this case so we will take 1.5 to be a measure of entropy of the distribution of X. In general for a discrete distribution, such that $P[X = x] = p(x)$, the entropy may be defined to be

$$H(p) = E\{-\ln(p(X)) = -\sum_x p(x)\ln(p(x)).$$

In the case of the above distribution, if we were to replace the natural logarithm by the log base 2, (ln and $\log_2$ differ only by a scale factor and are therefore the corresponding measures of entropy are equivalent up a scale multiple) notice that $-\sum_x p(x)\log_2(p(x)) = .5(1) + .5(2) = 1.5$, so this formula correctly measures the difficulty in ascertaining a random variable from a sequence of questions with yes-no or binary answers. This is true in general in fact. The complexity of a distribution as measured by the expected number of questions in a binary poll to determine the value of a random variable with that distribution can be measured by the

Many statistical distributions have an interpretation in terms of maximizing entropy. For example, what discrete distribution p has values on a certain set and yet maximizes the entropy $H(p)$? First notice that if p is uniform on n points, $p(x) = 1/n$ for all x and so the entropy is $-\sum_x \frac{1}{n}\ln(\frac{1}{n}) = \ln(n)$. Now consider the problem of maximizing the entropy $H(p)$ for any distribution on n points (subject to the constraint, of course, that the probabilities add to one). The Lagrangian for this problem is $-\sum_x p(x)\ln(p(x)) - \lambda\{\sum_x p(x) - 1\}$ where λ is a Lagrange multiplier. Upon differentiating with respect to $p(x)$ we obtain $-\ln(p(x)) - 1 - \lambda = 0$ or $p(x) = e^{-(1+\lambda)}$. Applying the constraint that the sum of the probabilities is one results in $p(x) = 1/n$ for all x. This shows that the discrete distribution with maximum entropy is the uniform distribution. What if we repeat this analysis using additional constraints, for example on the moments of the distribution? Suppose for example that we require that the mean of the distribution is some fixed constant μ and the variance fixed at σ^2. The problem is similar to that treated above but with two more terms in the Lagrangian for each of the additional constraints. The Lagrangian becomes

$$-\sum_x p(x)\ln(p(x)) - \lambda_1\{\sum_x p(x) - 1\} - \lambda_2\{\sum_x xp(x) - \mu\} - \lambda_3\{\sum x^2 p(x) - \mu^2 - \sigma^2\}$$

whereupon setting the derivative with respect to $p(x)$ equal to zero and applying the constraints we obtain

$$p(x) = \exp\{-\lambda_1 - \lambda_2 x - \lambda_3 x^2\},$$

with constants $\lambda_1, \lambda_2, \lambda_3$ chosen to satisfy the three constraints. Since the exponent is a quadratic function of x, this is analogous to the normal distribution except that we have required that it be supported on a discrete set of points x. Let us call such a distribution the discrete normal distribution. In fact if we drop the requirement that the distribution is discrete, the same kind of argument shows that the maximum entropy distribution is the normal distribution.

So here, at least, are two simple distributions arising out of maximum entropy considerations. *The maximum entropy distribution on a discrete set of points is the uniform distribution. The maximum entropy subject to a constraint on the mean and the variance is a (discrete) normal distribution.*

1.6 Models in Continuous Time

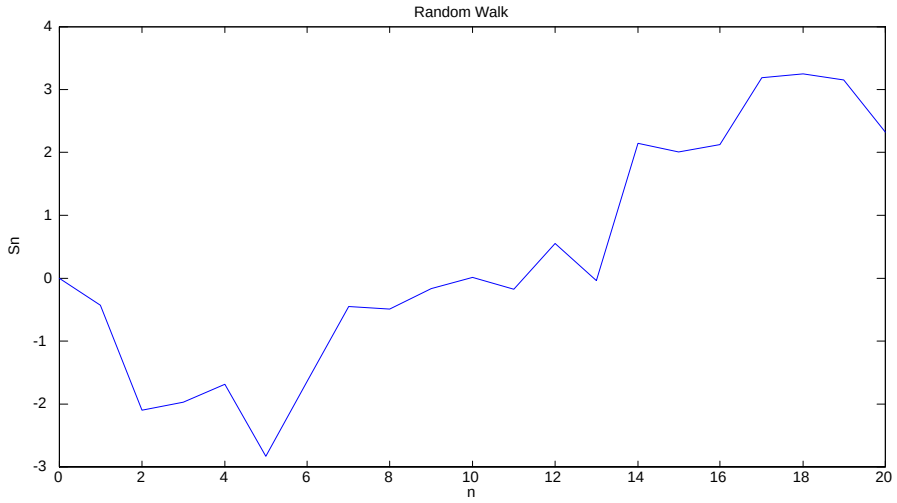
We begin with some oversimplified rules of stochastic calculus which can be omitted by those with a background in Brownian motion and diffusion. First, we define a stochastic process W_t called the *standard Brownian motion* or *Wiener process* having the following properties;

1. *For each $h > 0$, the increment $W(t+h) - W(t)$ has a $N(0,h)$ distribution and is independent of all preceding increments $W(u) - W(v), t > u > v > 0$.*

2. $W(0) = 0$.

The Standard Brownian Motion Process

The fact that such a process exists is by no means easy to see. It has been an important part of the literature in Physics, Probability and Finance at least since the papers of Bachelier and Einsten, about 100 years ago. A Brownian motion process also has some interesting and remarkable theoretical properties; it is continuous with probability one but the probability that the process has finite variation in any interval is 0. With probability one it is *nowhere differentiable*. Of course one might ask how a process with such apparently bizarre properties can be used to approximate real-world phenomena, where we expect functions to be built either from continuous and differentiable segments or jumps in the process. The answer is that a very wide class of functions constructed from those that are quite well-behaved (e.g. step functions) and that have independent increments converge as the scale on which they move is refined either to a Brownian motion process or to a process defined as an integral with respect to a Brownian motion process and so this is a useful approximation to a broad range of continuous time processes. For example, consider a random walk process $S_n = \sum_{i=1}^{n} X_i$ where the random variables X_i are independent identically distributed with expected value $E(X_i) = 0$ and $var(X_i) = 1$. Suppose we plot the graph of this random walk (n, S_n) as below. Notice that we have linearly interpolated the graph so that the function is defined for all n, whether integer or not.



Now if we increase the sample size and decrease the scale appropriately on both axes, the result is, in the limit, a Brownian motion process. The vertical scale is to be decreased by a factor $1/\sqrt{n}$ and the horizontal scale by a factor n^{-1} . The theorem concludes that the sequence of processes

$$Y_n(t) = \frac{1}{\sqrt{n}} S_{nt}$$

converges weakly to a standard Brownian motion process as $n \to \infty$. In practice this means that a process with independent stationary increments tends to look like a Brownian motion process. As we shall see, there is also a wide variety of non-stationary processes that can be constructed from the Brownian motion process by integration. Let us use the above limiting result to render some of the properties of the Brownian motion more plausible, since a serious proof is beyond our scope. Consider the question of continuity, for example. Since $|Y_n(t+h) - Y_n(t)| \approx |\frac{1}{\sqrt{n}} \sum_{i=nt}^{n(t+h)} X_i \ |$ and this is the absolute value of an asymptotically normally$(0, h)$ random variable by the central limit theorem, it is plausible that the limit as $h \to 0$ is zero so the function is continuous at t. On the other hand note that

$$\frac{Y_n(t+h) - Y_n(t)}{h} \approx \frac{1}{h} \frac{1}{\sqrt{n}} \sum_{i=nt}^{n(t+h)} X_i$$

should by analogy behave like h^{-1} times a $N(0, h)$ random variable which blows up as $h \to 0$ so it would appear that the derivative at t does not exist. To obtain the total variation of the process in the interval $[t, t+h]$, consider the lengths of the segments in this interval, i.e.

$$\frac{1}{\sqrt{n}} \sum_{i=nt}^{n(t+h)} |X_i|$$

and notice that since the law of large numbers implies that $\frac{1}{nh}\sum_{i=nt}^{n(t+h)}|X_i|$ converges to a positive constant, namely $E|X_i|$, if we multiply by $\sqrt{n}h$ the limit must be infinite, so the total variation of the Brownian motion process is infinite.

Continuous time process are usually built one small increment at a time and defined to be the limit as the size of the time increment is reduced to zero. Let us consider for example how we might define a stochastic (*Ito*) integral of the form $\int_0^T h(t)dW_t$. An approximating sum takes the form

$$\int_0^T h(t)dW_t \approx \sum_{i=0}^{n-1} h(t_i)(W(t_{i+1}) - W(t_i)), 0 = t_0 < t_1 < ... < t_n = T.$$

Note that the function $h(t)$ is evaluated at the left hand end-point of the intervals $[t_i, t_{i+1}]$, and this is characteristic of the Ito calculus, and an important feature distinguishing it from the usual Riemann calculus studied in undergraduate mathematics courses. There are some simple reasons why evaluating the function at the left hand end-point is necessary for stochastic models in finance. For example let us suppose that the function $h(t)$ measures how many shares of a stock we possess and $W(t)$ is the price of one share of stock at time t. It is clear that we cannot predict precisely future stock prices and our decision about investment over a possibly short time interval $[t_i, t_{i+1}]$ must be made at the *beginning* of this interval, not at the end or in the middle. Second, in the case of a Brownian motion process $W(t)$, it *makes a difference* where in the interval $[t_i, t_{i+1}]$ we evaluate the function h to approximate the integral, whereas it makes no difference for Riemann integrals. As we refine the partition of the interval, the approximating sums $\sum_{i=0}^{n-1} h(t_{i+1})(W(t_{i+1}) - W(t_i))$, for example, approach a completely different limit. This difference is essentially due to the fact that $W(t)$, unlike those functions studied before in calculus, is of infinite variation. As a consequence, there are other important differences in the Ito calculus. Let us suppose that the increment dW is used to denote small increments $W(t_{i+1}) - W(t_i)$ involved in the construction of the integral. If we denote the interval of time $t_{i+1} - t_i$ by dt, we can loosely assert that dW has the normal distribution with mean 0 and variance dt. If we add up a large number of independent such increments, since the variances add, the sum has variance the sum of the values dt and standard deviation the square root. Very roughly, we can assess the size of dW since its standard deviation is $(dt)^{1/2}$. Now consider defining a process as a function both of the Brownian motion and of time, say $V_t = g(W_t, t)$. If W_t represented the price of a stock or a bond, V_t might be the price of a derivative on this stock or bond. Expanding the increment dV using a Taylor series expansion gives

$$\begin{aligned} dV_t &= \frac{\partial}{\partial W}g(W_t,t)dW + \frac{\partial^2}{\partial W^2}g(W_t,t)\frac{dW^2}{2} + \frac{\partial}{\partial t}g(W_t,t)dt \qquad (1.6)\\ &\quad +(\text{stuff}) \times (dW)^3 + (\text{more stuff}) \times (dt)(dW)^2 + ... \end{aligned}$$

Loosely, dW is normal with mean 0 and standard deviation $(dt)^{1/2}$ and so dW is non-negligible compared with dt as $dt \to 0$. We can define each of the

differentials dW and dt essentially by reference to the result when we integrate both sides of the equation. If I were to write an equation in differential form

$$dX_t = h(t)dW_t$$

then this only has real meaning through its integrated version

$$X_t = X_0 + \int_0^t h(t)dW_t.$$

What about the terms involving $(dW)^2$? What meaning should we assign to a term like $\int h(t)(dW)^2$? Consider the approximating function $\sum h(t_i)(W(t_{i+1})-W(t_i))^2$. Notice that, at least in the case that the function h is non-random we are adding up independent random variables $h(t_i)(W(t_{i+1})-W(t_i))^2$ each with expected value $h(t_i)(t_{i+1}-t_i)$ and when we add up these quantities the limit is $\int h(t)dt$ by the law of large numbers. Roughly speaking, as differentials, we should interpret $(dW)^2$ as dt because that is the way it acts in an integral. Subsequent terms such as $(dW)^3$ or $(dt)(dW)^2$ are all $o(dt)$, i.e. they all approach 0 faster than does dt as $dt \to 0$. So finally substituting for $(dW)^2$ in **??** and ignoring all terms that are $o(dt)$, we obtain a simple version of *Ito's lemma*

$$dg(W_t,t) = \frac{\partial}{\partial W}g(W_t,t)dW + \{\frac{1}{2}\frac{\partial^2}{\partial W^2}g(W_t,t) + \frac{\partial}{\partial t}g(W_t,t)\}dt.$$

This rule results, for example, when we put $g(W_t,t) = W_t^2$ in

$$d(W_t^2) = 2W_t dW_t + dt$$

or on integrating both sides and rearranging,

$$\int_a^b W_t dW_t = \frac{1}{2}(W_b^2 - W_a^2) - \frac{1}{2}\int_a^b dt.$$

The term $\int_a^b dt$ above is what distinguishes the Ito calculus from the Riemann calculus, and is a consequence of the nature of the Brownian motion process, a continuous function of infinite variation.

There is one more property of the stochastic integral that makes it a valuable tool in the construction of models in finance, and that is that a stochastic integral with respect to a Brownian motion process is *always a martingale*. To see this, note that in an approximating sum

$$\int_0^T h(t)dW_t \approx \sum_{i=0}^{n-1} h(t_i)(W(t_{i+1}) - W(t_i))$$

each of the summands has conditional expectation 0 given the past, i.e.

$$E[h(t_i)(W(t_{i+1}) - W(t_i))|H_{t_i}] = h(t_i)E[(W(t_{i+1}) - W(t_i))|H_{t_i}] = 0$$

since the Brownian increments have mean 0 given the past and since $h(t)$ is measurable with respect to H_t.

We begin with an attempt to construct the model for an Ito process or diffusion process in continuous time. We construct the price process one increment at a time and it seems reasonable to expect that both the mean and the variance of the increment in price may depend on the current price but does not depend on the process before it arrived at that price. This is a loose description of a Markov property. The conditional distribution of the future of the process depends only on the current time t and the current price of the process. Let us suppose in addition that the increments in the process are, conditional on the past, normally distributed. Thus we assume that for small values of h, conditional on the current time t and the current value of the process X_t, the increment $X_{t+h} - X_t$ can be generated from a normal distribution with mean $a(X_t, t)h$ and with variance $\sigma^2(X_t, t)h$ for some functions a and σ^2 called the drift and diffusion coefficients respectively. Such a normal random variable can be formally written as $a(X_t, t\ )dt + \sigma^2(X_t, t)dW_t$. Since we could express X_T as an initial price X_0 plus the sum of such increments, $X_T = X_0 + \sum_i (X_{t_{i+1}} - X_{t_i})$.

The single most important model of this type is called the *Geometric Brownian motion or Black-Scholes model.* Since the actual value of stock, like the value of a currency or virtually any other asset is largely artificial, depending on such things as the number of shares issued, it is reasonable to suppose that the *changes in a stock price* should be modeled relative to the current price. For example rather than model the increments, it is perhaps more reasonable to model the relative change in the process. The simplest such model of this type is one in which both the mean and the standard deviation of the increment in the price are linear multiples of price itself; viz. dX_t is approximately normally distributed with mean $aX_t dt$ and variance $\sigma^2 X_t^2 dt$. In terms of stochastic differentials, we assume that

$$dX_t = aX_t dt + \sigma X_t dW_t. \tag{1.7}$$

Now consider the relative return from such a process over the increment $dY_t = dX_t/X_t$. Putting $Y_t = g(X_t) = \ln(X_t)$ note that analogous to our derivation of Ito's lemma

$$\begin{aligned} dg(X_t) &= g'(X_t)dX_t + \frac{1}{2}g''(X_t)(dX)^2 + ... \\ &= \frac{1}{X_t}\{aX_t dt + \sigma X_t dW_t.\} - \frac{1}{2X_t^2}\sigma^2 X_t^2 dt \\ &= (a - \frac{\sigma^2}{2})dt + \sigma dW_t. \end{aligned}$$

which is a description of a general Brownian motion process, a process with increments dY_t that are normally distributed with mean $(a - \frac{\sigma^2}{2})dt$ and with variance $\sigma^2 dt$. This process satisfying $dX_t = aX_t dt + \sigma X_t dW_t$ is called the *Geometric Brownian motion* process (because it can be written in the form $X_t = e^{Y_t}$ for a Brownian motion process Y_t) or a Black-Scholes model.

Many of the continuous time models used in finance are described as Markov diffusions or Ito processes which permits the mean and the variance of the increments to depend more generally on the present value of the process and the time. The integral version of this relation is of the form

$$X_T = X_0 + \int_0^T a(X_t, t)dt + \int_0^T \sigma(X_t, t)dW_t.$$

We often write such an equation with differential notation,

$$dX_t = a(X_t, t)dt + \sigma(X_t, t)dW_t. \tag{1.8}$$

but its meaning should always be sought in the above integral form. The coefficients $a(X_t, t)$ and $\sigma(X_t, t)$ vary with the choice of model. As usual, we interpret **??** as meaning that a small increment in the process, say $dX_t = X_{t+h} - X_t$ (h very small) is approximately distributed according to a normal distribution with conditional mean $a(X_t, t)dt = a(X_t, t)h$ and conditional variance given by $\sigma^2(X_t, t)var(dW_t) = \sigma^2(X_t, t)var(W_{t+h} - W_t) = \sigma^2(X_t, t)h$. Here the mean and variance are conditional on H_t, the history of the process X_t up to time t.

Various choices for the functions $a(X_t, t), \sigma(X_t, t)$ are possible. For the Black-Scholes model or geometric Brownian motion, $a(X_t, t) = aX_t$ and $\sigma(X_t, t) = \sigma X_t$ for constant drift and volatility parameters a, σ. The *Cox-Ingersoll-Ross model,* usesd to model spot interest rates, corresponds to $a(X_t, t) = A(b - X_t)$ and $\sigma(X_t, t) = c\sqrt{X_t}$ for constants A, b, c. The Vasicek model, also a model for interest rates, has $a(X_t, t) = A(b - X_t)$ and $\sigma(X_t, t) = c$. There is a large number of models for most continuous time processes observed in finance which can be written in the form **??**. So called multi-factor models are of similar form where X_t is a vector of financial time series and the coefficient functions $a(X_t, t)$ is vector valued, $\sigma(X_t, t)$ is replaced by a matrix-valued function and dW_t is interpreted as a vector of independent Brownian motion processes. For technical conditions on the coefficients under which a solution to **??** is guaranteed to exist and be unique, see Karatzas and Shreve, sections 5.2, 5.3.

As with any differential equation there may be initial or boundary conditions applied to **??** that restrict the choice of possible solutions. Solutions to the above equation are difficult to arrive at, and it is often even more difficult to obtain distributional properties of them. Among the key tools are the *Kolmogorov differential equations* (see Cox and Miller, p. 215). Consider the transition probability kernel

$$p(s, z, t, x) = P[X_t = x | X_s = z]$$

in the case of a *discrete Markov Chain.* If the Markov chain is continuous (as it is in the case of diffusions), that is if the conditional distribution of X_t given X_s is absolutely continuous with respect to Lebesgue measure, then we can define $p(s, z, t, x)$ to be the *conditional probability density function* of X_t given $X_s = z$. The two equations, for a diffusion of the above form, are:

Kolmogorov's backward equation

$$\frac{\partial}{\partial s}p = -a(z,s)\frac{\partial}{\partial z}p - \frac{1}{2}\sigma^2(z,s)\frac{\partial^2}{\partial z^2}p \tag{1.9}$$

and the *forward equation*

$$\frac{\partial}{\partial t}p = -\frac{\partial}{\partial x}(a(x,t)p) + \frac{1}{2}\frac{\partial^2}{\partial x^2}(\sigma^2(x,t)p) \tag{1.10}$$

Note that if we were able to solve these equations, this would provide the transition density function p, giving the conditional distribution of the process. It does not immediately provide other characteristics of the diffusion, such as the distribution of the maximum or the minimum, important for valuing various exotic options such as look-back and barrier options. However for a European option defined on this process, knowledge of the transition density would suffice at least theoretically for valuing the option. Unfortunately these equations are often very difficult to solve explicitly.

Besides the Kolmogorov equations, we can use simple ordinary differential equations to arrive at some of the basic properties of a diffusion. To illustrate, consider one of the simplest possible forms of a diffusion, where $a(X_t,t) = \alpha(t) + \beta(t)X_t$ where the coefficients $\alpha(t)$, $\beta(t)$ are deterministic (i.e. non-random) functions of time. Note that the integral analogue of **??** is

$$X_t = X_0 + \int_0^t a(X_s,s)ds + \int_0^t \sigma(X_s,s)dW_s \tag{1.11}$$

and by construction that last term $\int_0^t \sigma(X_s,s)dW_s$ is a zero-mean martingale. For example its small increments $\sigma(X_t,t)dW_s$ are approximately $N(0,\sigma(X_t,t)dt)$. Therefore, taking expectations on both sides conditional on the value of X_0, and letting $m(t) = E(X_t)$, we obtain:

$$m(t) = X_0 + \int_0^t [\alpha(s) + \beta(s)m(s)]ds \tag{1.12}$$

and therefore $m(t)$ solves the ordinary differential equation

$$m'(t) = \alpha(t) + \beta(t)m(t). \tag{1.13}$$

$$m(0) = X_0 \tag{1.14}$$

Thus, in the case that the *drift term* a is a linear function of X_t, the mean or expected value of a diffusion process can be found by solving a similar ordinary differential equation, similar except that the diffusion term has been dropped.

These are only two of many reasons to wish to solve both ordinary and partial differential equations in finance. The solution to the Kolmogorov partial differential equations provides the conditional distribution of the increments of a process. And when the drift term $a(X_t,t\ )$ is linear in X_t, the solution of an ordinary differential equation will allow the calculation of the expected value of

the process and this is the first and most basic description of its behaviour. The appendix provides an elementary review of techniques for solving partial and ordinary differential equations.

However, that the information about a stochastic process obtained from a deterministic object such as a ordinary or partial differential equation is necessarily limited. For example, while we can sometimes obtain the marginal distribution of the process at time t it is more difficult to obtain quantities such as the joint distribution of variables which depending on the path of the process, and these are important in valuing certain types of exotic options such as lookback and barrier options. For such problems, we often use Monte Carlo methods.

The Black-Scholes Formula

Before discussing methods of solution in general, we develop the Black-Scholes equation in a general context. Suppose that a security price is an Ito process satisfying the equation

$$dS_t \ = a(S_t,\ t\ )\ dt \ + \ \sigma(S_t,\ t)\ dW_t \tag{1.15}$$

Assumed the market allows investment in the stock as well as a risk-free bond whose price at time t is B_t. It is necessary to make various other assumptions as well and strictly speaking all fail in the real world, but they are a resonable approximation to a real, highly liquid and nearly frictionless market:

1. partial shares may be purchased
2. there are no dividends paid on the stock
3. There are no commissions paid on purchase or sale of the stock or bond
4. There is no possibility of default for the bond
5. Investors can borrow at the risk free rate governing the bond.
6. All investments are liquid- they can be bought or sold instantaneously.

Since bonds are assumed risk-free, they satisfy an equation

$$dB_t = r_t B_t dt$$

where r_t is the risk-free (spot) interest rate at time t.

We wish to determine $V(S_t, t)$, the value of an option on this security when the security price is S_t, at time t. Suppose the option has expiry date T and a general payoff function which depends only on S_T, the process at time T.

Ito's lemma provides the ability to translate an a relation governing the differential dS_t into a relation governing the differential of the process $dV(S_t, t)$. In this sense it is the stochastic calculus analogue of the chain rule in ordinary calculus. It is one of the most important single results of the twentieth century

in finance and in science. The stochastic calculus and this mathematical result concerning it underlies the research leading to 1997 Nobel Prize to Merton and Black for their work on hedging in financial models. We saw one version of it at the beginning of this section and here we provide a more general version.

Ito's lemma.

Suppose S_t *is a diffusion process satisfying*

$$dS_t = a(S_t, t)dt + \sigma(S_t, t)dW_t$$

and suppose $V(S_t, t)$ *is a smooth function of both arguments. Then* $V(S_t, t)$ *also satisfies a diffusion equation of the form*

$$dV = [a(S_t, t)\frac{\partial V}{\partial S} + \frac{\sigma^2(S_t, t)}{2}\frac{\partial^2 V}{\partial S^2} + \frac{\partial V}{\partial t}]dt + \sigma(S_t, t)\frac{\partial V}{\partial S}dW_t. \qquad (1.16)$$

Proof. The proof of this result is technical but the ideas behind it are

simple. Suppose we expand an increment of the process $V(S_t, t)$ (we write V in place of $V(S_t, t)$ omitting the arguments of the function and its derivatives. We will sometimes do the same with the coefficients a and σ.)

$$V(S_{t+h}, t+h) \approx V + \frac{\partial V}{\partial S}(S_{t+h} - S_t) + \frac{1}{2}\frac{\partial^2 V}{\partial S^2}(S_{t+h} - S_t)^2 + \frac{\partial V}{\partial t}h \qquad (1.17)$$

where we have ignored remainder terms that are $o(h)$. Note that substituting from **??** into **??**, the increment $(S_{t+h} - S_t)$ is approximately normal with mean $a(S_t,\ t\ )\ h$ and variance $\sigma^2(S_t,\ t\ )\ h$. Consider the term $(S_{t+h} - S_t)^2$. Note that it is the square of the above normal random variable and has expected value $\sigma^2(S_t, t)h + a^2(S_t, t)h^2$. The variance of this random variable is $O(h^2)$ so if we ignore all terms of order $o(h)$ the increment $V(S_{t+h}, t+h) - V(S_t, t)$ is approximately normally distributed with mean

$$[a(S_t,\ t\ )\frac{\partial V}{\partial S} + \frac{\sigma^2(S_t,\ t)}{2}\frac{\partial^2 V}{\partial S^2} + \frac{\partial V}{\partial t}]h$$

and standard deviation $\sigma(S_t,\ t)\frac{\partial V}{\partial S}\sqrt{h}$ justifying (but not proving!) the relation **??**. ■

By Ito's lemma, provided V is smooth, it also satisfies a diffusion equation of the form **??**. We should note that when V represents the price of an option, some lack of smoothness in the function V is inevitable. For example for a European call option with exercise price K, $V(S_T, T) = \max(S_T - K, 0)$ does not have a derivative with respect to S_T at $S_T = K$, the exercise price. Fortunately, such

exceptional points can be worked around in the argument, since the derivative does exist at values of $t < T$.

The basic question in building a replicating portfolio is: for hedging purposes, is it possible to find a *self-financing* portfolio consisting only of the security and the bond which exactly replicates the option price process $V(S_t, t)$? The self-financing requirement is the analogue of the requirement that the net cost of a portfolio is zero that we employed when we introduced the notion of arbitrage. The portfolio is such that no funds are needed to be added to (or removed from) the portfolio during its life, so for example any additional amounts required to purchase equity is obtained by borrowing at the risk free rate. Suppose the self-financing portfolio has value at time t equal to $V_t = u_t S_t + w_t B_t$ where the (predictable) functions u_t, w_t represent the number of shares of stock and bonds respectively owned at time t. Since the portfolio is assumed to be self-financing, all returns obtain from the changes in the value of the securities and bonds held, i.e. it is assumed that $dV_t = u_t dS_t + w_t dB_t$. Substituting from **??**,

$$dV_t = u_t dS_t + w_t dB_t = [u_t a(S_t, t) + w_t r_t B_t]dt + u_t \sigma(S_t, t) dW_t \tag{1.18}$$

If V_t is to be exactly equal to the price $V(S_t, t\ )$ of an option, it follows on comparing the coefficients of dt and dW_t in **??** and **??**, that $u_t = \frac{\partial V}{\partial S}$, called the *delta* corresponding to *delta hedging*. Consequently,

$$V_t = \frac{\partial V}{\partial S} S_t + w_t B_t$$

and solving for w_t we obtain:

$$w_t = \frac{1}{B_t}[V - \frac{\partial V}{\partial S} S_t].$$

The conclusion is that it is possible to dynamically choose a trading strategy, i.e. the weights w_t, u_t so that our portfolio of stocks and bonds perfectly replicates the value of the option. If we own the option, then by shorting (selling) Delta= $\frac{\partial V}{\partial S}$ units of stock, we are **perfectly** hedged in the sense that our portfolio replicates a risk-free bond. Surprisingly, in this ideal word of continuous processes and continuous time trading commission-free trading, the perfect hedge is possible. In the real world, it is said to exist only in a Japanese garden. The equation we obtained by equating both coefficients in **??** and **??** is;

$$-r_t V + r_t S_t \frac{\partial V}{\partial S} + \frac{\partial V}{\partial t} + \frac{\sigma^2(S_t, t)}{2} \frac{\partial^2 V}{\partial S^2} = 0. \tag{1.19}$$

Rewriting this allows an interpretation in terms of our hedged portfolio. If we own an option and are short delta units of stock our net investment at time t is given by $(V - S_t \frac{\partial V}{\partial S})$ where $V = V_t = V(S_t, t)$. Our return over the next time increment dt if the portfolio were liquidated and the identical amount invested

in a risk-free bond would be $r_t(V_t - S_t\frac{\partial V}{\partial S})dt$. On the other hand if we keep this hedged portfolio, the return over an increment of time dt is

$$\begin{aligned} d(V - S_t\frac{\partial V}{\partial S}) &= dV - (\frac{\partial V}{\partial S})dS \\ &= (\frac{\partial V}{\partial t} + \frac{\sigma^2}{2}\frac{\partial^2 V}{\partial S^2} + a\frac{\partial V}{\partial S})dt + \sigma\frac{\partial V}{\partial S}dW_t \\ &\quad -\frac{\partial V}{\partial S}[adt + \sigma dW_t] \\ &= (\frac{\partial V}{\partial t} + \frac{\sigma^2}{2}\frac{\partial^2 V}{\partial S^2})dt \end{aligned}$$

Therefore

$$r_t(V - S_t\frac{\partial V}{\partial S}) = \frac{\partial V}{\partial t} + \frac{\sigma^2(S_t,t)}{2}\frac{\partial^2 V}{\partial S^2}.$$

The left side $r_t(V - S_t\frac{\partial V}{\partial S})$ represents the amount made by the portion of our portfolio devoted to risk-free bonds. The right hand side represents the return on a hedged portfolio long one option and short delta stocks. Since these investments are at least in theory identical, so is their return. This fundamental equation is evidently satisfied by any option price process where the underlying security satisfies a diffusion equation and the option value at expiry depends only on the value of the security at that time. The type of option determines the terminal conditions and usually uniquely determines the solution.

It is extraordinary that this equation in no way depends on the drift coefficient $a(S_t, t)$. This is a remarkable feature of the arbitrage pricing theory. *Essentially, no matter what the drift term for the particular security is, in order to avoid arbitrage, all securities and their derivatives are priced as if they had as drift the spot interest rate.This is the effect of calculating the expected values under the martingale measure Q.*

This PDE governs most derivative products, European call options, puts, futures or forwards. However, the boundary conditions and hence the solution depends on the particular derivative. The solution to such an equation is possible analytically in a few cases, while in many others, numerical techniques are necessary. One special case of this equation deserves particular attention. In the case of geometric Brownian motion, $a(S_t, t) = \mu S_t$ and $\sigma(S_t, t) = \sigma S_t$ for constants μ, σ. Assume that the spot interest rate is a constant r and that a constant rate of dividends D_0 is paid on the stock. In this case, the equation specializes to

$$-rV + \frac{\partial V}{\partial t} + (r - D_0)S\frac{\partial V}{\partial S} + \frac{\sigma^2 S^2}{2}\frac{\partial^2 V}{\partial S^2} = 0. \tag{1.20}$$

Note that we have not used *any* of the properties of the particular derivative product yet, nor does this differential equation involve the drift coefficient μ. The assumption that there are no transaction costs is essential to this analysis, as we have assumed that the portfolio is continually rebalanced.

We have now seen two derivations of parabolic partial differential equations, so-called because like the equation of a parabola, they are first order (derivatives) in one variable (t) and second order in the other (x). Usually the solution of such an equation requires reducing it to one of the most common partial differential equations, the heat or diffusion equation, which models the diffusion of heat along a rod. This equation takes the form

$$\frac{\partial}{\partial t}u = k\frac{\partial^2}{\partial x^2}u \tag{1.21}$$

A solution of **??** with appropriate boundary conditions can sometime be found by the separation of variables. We will later discuss in more detail the solution of parabolic equations, both by analytic and numerical means. First, however, when can we hope to find a solution of **??** of the form $u(x,t) = g(x/\sqrt{t})$. By differentiating and substituting above, we obtain an ordinary differential equation of the form

$$g''(\omega) + \frac{1}{2k}\omega g'(\omega) = 0, \omega = x/\sqrt{t} \tag{1.22}$$

Let us solve this using MAPLE.

```
eqn := diff(g(w),w,w)+(w/(2*k))*diff(g(w),w)=0;
dsolve(eqn,g(w));
```

and because the derivative of the solution is slightly easier (for a statistician) to identify than the solution itself,

```
> diff(%,w);
```

giving

$$\frac{\partial}{\partial w}g(\omega) = C_2 \exp\{-w^2/4k\} = C_2 \exp\{-x^2/4kt\} \tag{1.23}$$

showing that a constant plus a constant multiple of the Normal $(0, 2kt)$ cumulative distribution function or

$$u(x,t) = C_1 + C_2\frac{1}{2\sqrt{\pi kt}}\int_{-\infty}^{x} \exp\{-z^2/4kt\}dz \tag{1.24}$$

is a solution of this, the heat equation for $t > 0$. The role of the two constants is simple. Clearly if a solution to **??** is found, then we may add a constant and/or multiply by a constant to obtain another solution. The constant in general is determined by initial and boundary conditions. Similarly the integral can be removed with a change in the initial condition for if u solves **??** then so does $\frac{\partial u}{\partial x}$. For example if we wish a solution for the half real $x > 0$ with initial condition $u(x,0) = 0, u(0,t) = 1$ all $t > 1$, we may use

$$u(x,t) = 2P(N(0,2kt) > x) = \frac{1}{\sqrt{\pi kt}}\int_{x}^{\infty} \exp\{-z^2/4kt\}dz, t > 0, x \geq 0.$$

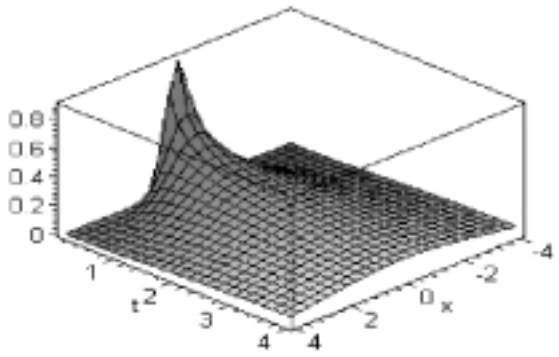


Figure 1.1: The function $u(x,t)$

Let us consider a basic solution to **??**:

$$u(x,t) = \frac{1}{2\sqrt{\pi kt}} \exp\{-x^2/4kt\} \tag{1.25}$$

This connection between the heat equation and the normal distributions is fundamental and the wealth of solutions depending on the initial and boundary conditions is considerable. We plot a fundamental solution of the equation as follows:

```
> u(x,t) := (.5/sqrt(Pi*t))*exp(-x^2/(4*t));
> plot3d(u(x,t),x=-4..4,t=.02..4,axes=boxed);
```

As $t \to 0$, the function approaches a spike at $x = 0$, usually referred to as the "Dirac delta function" (although it is no function at all) and symbolically representing the derivative of the "Heaviside function". The Heaviside function is defined as $H(x) = 1, x \geq 0$ and is otherwise 0 and is the cumulative distribution function of a point mass at 0. Suppose we are given an initial condition of the form $u(x,0) = u_0(x)$. To this end, it is helpful to look at the solution $u(x,t)$ and the initial condition $u_0(x)$ as a distribution or measure (in this case described by a density) over the space variable x. For example the density $u(x,t)$ corresponds to a measure for fixed t of the form $\nu_t(A) = \int_A u(x,t)dx$. Note that the initial condition compatible with the above solution **??** can be described somewhat clumsily as "$u(x,0)$ corresponds to a measure placing all mass at $x = x_0 = 0$ ".In fact as $t \to 0$, we have in some sense the following convergence $u(x,t) \to \delta(x) = dH(x)$, the Dirac delta function. We could just as easily construct solve the heat equation with a more general initial condition of

the form $u(x,0) = dH(x - x_0)$ for arbitrary x_0 and the solution takes the form

$$u(x,t) = \frac{1}{2\sqrt{\pi kt}} \exp\{-(x-x_0)^2/4kt\}. \tag{1.22}$$

Indeed sums of such solutions over different values of x_0, or weighted sums, or their limits, integrals will continue to be solutions to **??**. In order to achieve the initial condition $u_0(x)$ we need only pick a suitable weight function. Note that

$$u_0(x) = \int u_0(z) dH(z - x)$$

Note that the function

$$u(x,t) = \frac{1}{2\sqrt{\pi kt}} \int_{-\infty}^{\infty} \exp\{-(z-x)^2/4kt\} u_0(z) dz \tag{1.22}$$

solves **??** subject to the required boundary condition.

Solution of the Diffusion Equation.

We now consider the general solution to the diffusion equation of the form **??**, rewritten as

$$\frac{\partial V}{\partial t} = r_t V - r_t S_t \frac{\partial V}{\partial S} - \frac{\sigma^2(S_t,t)}{2} \frac{\partial^2 V}{\partial S^2} \tag{1.26}$$

where S_t is an asset price driven by a diffusion equation

$$dS_t = a(S_t,t)dt + \sigma(S_t,t)dW_t, \tag{1.27}$$

$V(S_t,t)$ is the price of an option on that asset at time t, and $r_t = r(t)$ is the spot interest rate at time t. We assume that the price of the option at expiry T is a known function of the asset price

$$V(S_T,T) = V_0(S_T). \tag{1.28}$$

Somewhat strangely, the option is priced using a related but not identical process (or, equivalently, the same process under a different measure). Recall from the backwards Kolmogorov equation **??** that if a related process X_t satisfies the stochastic differential equation

$$dX_t = r(X_t,t)X_t dt + \sigma(X_t,t)dW_t \tag{1.29}$$

then its transition kernel $p(t,s,T,z) = \frac{\partial}{\partial z} P[X_T \le z | X_t = s]$ satisfies a partial differential equation similar to **??**;

$$\frac{\partial p}{\partial t} = -r(s,t)s \frac{\partial p}{\partial s} - \frac{\sigma^2(s,t)}{2} \frac{\partial^2 p}{\partial s^2} \tag{1.30}$$

For a given process X_t this determines one solution. For simplicity, consider the case (natural in finance applications) when the spot interest rate is a function of time, not of the asset price; $r(s,t) = r(t)$. To obtain the solution so that terminal conditions is satisfied, consider a product

$$f(t,s,T,z) = p(t,s,T,z)q(t,T) \tag{1.31}$$

where

$$q(t,T) = exp\{-\int_t^T r(v)dv\}$$

is the discount function or the price of a zero-coupon bond at time t which pays 1$ at maturity.

Let us try an application of one of the most common methods in solving PDE's, the "lucky guess" method. Consider a linear combination of terms of the form **??** with weight function $w(z)$. i.e. try a solution of the form

$$V(s,t) = \int p(t,s,T,z)q(t,T)w(z)dz \tag{1.32}$$

for suitable weight function $w(z)$. In view of the definition of p as a transition probability density, this integral can be rewritten as a conditional expectation:

$$V(t,s) = E[w(X_T)q(t,T)|X_t = s] \tag{1.33}$$

the discounted conditional expectation of the random variable $w(X_T)$ given the current state of the process, where the process is assumed to follow (2.18). Note that in order to satisfy the terminal condition **??**, we choose $w(x) = V_0(x)$. Now

$$\begin{aligned}
\frac{\partial V}{\partial t} &= \frac{\partial}{\partial t}\int p(t,s,T,z)q(t,T)w(z)dz \\
&= \int [-r(S_t,t)S_t\frac{\partial p}{\partial s} - \frac{\sigma^2(S_t,t)}{2}\frac{\partial^2 p}{\partial s^2}]q(t,T)w(z)dz \\
&\quad +r(S_t,t)\int p(t,S_t,T,z)q(t,T)w(z)dz \text{ by } \textbf{??} \\
&= -r(S_t,t)S_t\frac{\partial V}{\partial S} - \frac{\sigma^2(S_t,t)}{2}\frac{\partial^2 V}{\partial S^2} + r(S_t,t)V(S_t,t)
\end{aligned}$$

where we have assumed that we can pass the derivatives under the integral sign. Thus the process

$$V(t,s) = E[V_0(X_T)q(t,T)|X_t = s] \tag{1.34}$$

satisfies both the partial differential equation **??** and the terminal conditions **??** and is hence the solution. Indeed it is the unique solution satisfying certain

regularity conditions. The result asserts that the value of any European option is simply the conditional expected value of the *discounted payoff* (discounted to the present) assuming that the distribution is that of the process **??**. This result is a special case when the spot interest rates are functions only of time of the following more general theorem.

Theorem 5 *(Feynman-Kac)*

Suppose the conditions for a unique solution to (**??**,**??**) (see for example Duffie, appendix E) are satisfied. Then the general solution to (2.15) under the terminal condition **??** is given by

$$V(S,t) = E[V_0(X_T)exp\{-\int_t^T r(X_v, v)dv\} \mid X_t = S] \tag{1.35}$$

This represents the discounted return from the option under the distribution of the process X_t . The distribution induced by the process X_t is referred to as the *equivalent martingale measure* or *risk neutral measure*. Notice that when the original process is a diffusion, the equivalent martingale measure shares the same diffusion coefficient but has the drift replaced by $r(X_t, t)X_t$. The option is priced as if the drift were the same as that of a risk-free bond i.e. as if the instantaneous rate of return from the security if identical to that of bond. Of course, in practice, it is not. A risk premium must be paid to the stock-holder to compensate for the greater risk associated with the stock.

There are some cases in which the conditional expectation **??** can be determined explicitly. In general, these require that the process or a simple function of the process is Gaussian.

For example, suppose that both $r(t)$ and $\sigma(t)$ are deterministic functions of time only. Then we can solve the stochastic differential equation (2.22) to obtain

$$X_T = \frac{X_t}{q(t,T)} + \int_t^T \frac{\sigma(u)}{q(u,T)} dW_u \tag{1.36}$$

The first term above is the conditional expected value of X_T given X_t. The second is the random component, and since it is a weighted sum of the normally distributed increments of a Brownian motion with weights that are non-random, it is also a normal random variable. The mean is 0 and the (conditional) variance is $\int_t^T \frac{\sigma^2(u)}{q^2(u,T)} du$. Thus the conditional distribution of X_T given X_t is normal with conditional expectation $\frac{X_t}{q(t,T)}$ and conditional variance $\int_t^T \frac{\sigma^2(u)}{q^2(u,T)} du$.

The special case of **??** of most common usage is the Black-Scholes model: suppose that $\sigma(S,t) = S\sigma(t)$ for $\sigma(t)$ some deterministic function of t. Then the distribution of X_t is not Gaussian, but fortunately, its logarithm is. In this case we say that the distribution of X_t is lognormal.

Lognormal Distribution

Suppose Z is a normal random variable with mean μ and variance σ^2. Then we say that the distribution of $X = e^Z$ is lognormal with mean $\eta = \exp\{\mu+\sigma^2/2\}$ and volatility parameter σ. The lognormal probability density function with mean $\eta > 0$ and volatility parameter $\sigma > 0$ is given by the probability density function

$$g(x|\eta,\sigma) = \frac{1}{x\sigma\sqrt{2\pi}} exp\{-(log\ x - log\ \eta\ - \sigma^2/2)^2/2\sigma^2\}. \tag{1.37}$$

The solution to (2.18) with non-random functions $\sigma(t), r(t)$ is now

$$X_T = X_t exp\{\int_t^T (r(u) - \sigma^2(u)/2)du + \int_t^T \sigma(u)dW_u\}. \tag{1.38}$$

Since the exponent is normal, the distribution of X_T is lognormal with mean $log(X_t) + \int_t^T (r(u) - \sigma^2(u)/2)du$ and variance $\int_t^T \sigma^2(u)du$. It follows that the conditional distribution is lognormal with mean $\eta = X_t q(t,T)$ and volatility parameter $\sqrt{\int_t^T \sigma^2(u)du}$.

We now derive the well-known Black-Scholes formula as a special case of **??**. For a call option with exercise price E, the payoff function is $V_0(S_T) = max(S_T - E, 0)$. Now it is helpful to use the fact that for a standard normal random variable Z and arbitrary $\sigma > 0, -\infty < \mu < \infty$ we have the expected value of $max(e^{\sigma Z+\mu}, 0)$ is

$$e^{\mu+\sigma^2/2}\Phi(\frac{\mu}{\sigma} + \sigma) - \Phi(\frac{\mu}{\sigma}) \tag{1.39}$$

where $\Phi(.)$ denotes the standard normal cumulative distribution function. As a result, in the special case that r and σ are constants, (**??**) results in the famous Black-Scholes formula which can be written in the form

$$V(S,t) = S\Phi(d_1) - Ee^{-r(T-t)}\Phi(d_2) \tag{1.40}$$

where $d_1 < d_2$ are the values $\pm\sigma^2/2$ standardized by adding $log(S/E)+r(T-t)$ and dividing by $\sigma\sqrt{T-t}$. This may be derived by the following device; Assume (i.e. pretend) that, given current information, the distribution of $S(T)$ at expiry is lognormally distributed with the mean $\eta = S(t)e^{r(T-t)}$.

The mean of the log-normal in the risk neutral world $S(t)e^{r(T-t)}$ is exactly the future value of our current stocks $S(t)$ if we were to sell the stock and invest the cash in a bank deposit. Then the future value of an option with payoff function given by $V_0(S_T)$ is the expected value of this function against this lognormal probability density function, then discounted to present value

$$e^{-r(T-t)} \int_0^\infty V_0(x) g(x|S(t)e^{r(T-t)},\ \sigma\sqrt{T-t})dx. \tag{1.41}$$

Notice that the Black-Scholes derivation covers any diffusion process governing the underlying asset which is driven by a stochastic differential equation of the form

$$dS = a(S)dt + \sigma S dW_t \tag{1.42}$$

regardless of the nature of the drift term $a(S)$. For example a non-linear function $a(S)$ can lead to distributions that are not lognormal and yet the option price is determined as if it were.

Example: Pricing Call and Put options.

Consider pricing an index option on the S&P 500 index an January 11, 2000 (the index SPX closed at 1432.25 on this day). The option SXZ AE-A is a January call option with strike price 1425. The option matures (as do equity options in general) on the third Friday of the month or January 21, a total of 7 trading days later. Suppose we wish to price such an option using the Black-Scholes model. In this case, $T-t$ measured in years is $7/252 = 0.027778$. The annual volatility of the Standard and Poor 500 index is around 19.5 percent or 0.195 and assume the very short term interest rates approximately 3%. In *Matlab* we can value this option using

[CALL,PUT] = BLSPRICE(1432.25,1425,0.03,7/252,0.195,0)

CALL = 23.0381

PUT = 14.6011

Arguments of the function BLSPRICE are, in order, the current equity price, the strike price, the annual interest rate r, the time to maturity $T-t$ in years, the annual volatility σ and the last argument is the dividend yield in percent which we assumed 0. Thus the Black-Scholes price for a call option on SPX is around 23.03. Indeed this call option did sell on Jan 11 for \$23.00. and the put option for \$14 5/8. From the put call parity relation (see for example Wilmott, Howison, Dewynne, page 41) $S + P - C = Ee^{-r(T-t)}$ or in this case $1432.25 + 14.625 - 23 = 1425e^{-r(7/252)}$. We might solve this relation to obtain the spot interest rate r. In order to confirm that a different interest rate might apply over a longer term, we consider the September call and put options (SXZ) on the same day with exercise price 1400 which sold for \$152 and 71\$ respectively. In this case there are171 trading days to expiry and so we need to solve $1432.25 + 71 - 152 = 1400e^{-r(171/252)}$, whose solution is $r = 0.0522$. This is close to the six month interest rates at the time, but 3% is low for the very short term rates. The discrepancy with the actual interest rates is one of several modest failures of the Black-Scholes model to be discussed further later. The low implied interest rate is influenced by the cost of handling and executing an option, which are non-negligible fractions of the option prices, particularly with short term options such as this one.

1.7 Review Problems

1. It is common for a stock whose price has reached a high level to *split* or issue shares on a two-for-one or three-for-one basis. What is the effect of a stock split on the price of an option?

2. If a stock issues a dividend of exactly D (known in advance) on a certain date, provide a no-arbitrage argument for the change in price of the stock at this date.

3. Suppose Σ is a positive definite covariance matrix and η a column vector. Show that the set of all possible pairs of standard deviation and mean return $(\sqrt{w^T \Sigma w}, \eta^T w)$ for weight vector w such that $\sum_i w_i = 1$ is a convex region with an elliptical boundary.

4. The current rate of interest is 5% per annum and you are offered a random bond which pays either \$210 or \$0 in one year. You believe that the probability of the bond paying \$210 is one half. How much would you pay now for such a bond? Suppose this bond is publicly traded and a large fraction of the population is risk averse so that it is selling now for \$80. Does your price offer an arbitrage to another trader? What is the risk-neutral measure for this bond?

5. Which would you prefer, a gift of \$100 or a 50-50 chance of making \$200? A fine of \$100 or a 50-50 chance of losing \$200? Are your preferences self-consistent and consistent with the principle that individuals are risk-averse?

6. Compute the stochastic differential dX_t (assuming W_t is a Wiener process) when

   (a) $X_t = \exp(rt)$

   (b) $X_t = \int_0^t h(t) dW_t$

   (c) $X_t = X_0 \exp\{at + bW_t\}$

   (d) $X_t = \exp(Y_t)$ where $dY_t = \mu dt + \sigma dW_t$.

7. Show that if X_t is a geometric Brownian motion, so is X_t^β for any real number β.

8. Suppose a stock price follows a geometric Brownian motion process

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

   Find the diffusion equation satisfied by the processes (a) $f(S_t) = S_t^n$,(b) $\log(S_t)$, (c) $1/S_t$. Find a combination of the processes S_t and $1/S_t$ that does not depend on the drift parameter μ. How does this allow constructing estimators of σ that do not require knowledge of the value of μ?

9. Consider an Ito process of the form

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

Is it possible to find a function $f(S_t)$ which is also an Ito process but with zero drift?

10. Consider an Ito process of the form

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

Is it possible to find a function $f(S_t)$ which has constant diffusion term?

11. Consider approximating an integral of the form $\int_0^T g(t)dW_t \approx \sum g(t)\{W(t+h) - W(t)\}$ where $g(t)$ is a non-random function and the sum is over values of $t = nh, n = 0, 1, 2, ...T/h - 1$. Show by considering the distribution of the sum and taking limits that the random variable $\int_0^T g(t)dW_t$ has a normal distribution and find its mean and variance.

12. Give an example of a function $g(t, W_t)$ such that the random variable $\int_0^1 g(t, W_t)dW_t$ does not have a normal distribution but has larger tails than does the normal distribution.

13. Consider two geometric Brownian motion processes X_t and Y_t both driven by the same Wiener process

$$\begin{aligned} dX_t &= aX_t dt + bX_t dW_t \\ dY_t &= \mu Y_t dt + \sigma Y_t dW_t. \end{aligned}$$

Derive a stochastic differential equation for the ratio $Z_t = X_t/Y_t$. Suppose for example that X_t models the price of Telecom stock in \$NZ and Y_t is the exchange rate ($\$NZ/\$US$) at time t. Then what is the process Z_t?

14. Verify that for any pair of constants $a \neq 0$ and $b > 0$

$$dX_t = (X_t^{-1} + ab)X_t dt + bX_t dW_t$$

does not have a solution in the form $X_t = f(t, Y_t)$, where $f(t, y)$ is, say, a real function and Y_t is a Gaussian process.

15. Consider solving the problem

$$\min_q \sum q_i \log(\frac{q_i}{p_i})$$

subject to the constraints $\sum_i q_i = 1$ and $\sum q_i f(i) = \mu$. Show that the solution, if it exists, is given by

$$q_i = \frac{\exp(\eta f(i))}{m(\eta)} p_i$$

where $m(\eta) = \sum_i p_i \exp(\eta f(i))]$ and η is chosen so that $\frac{m'(\eta)}{m(\eta)} = \mu$.

16. Consider a defaultable bond which pays a fraction of its face value Fp on maturity in the event of default. Suppose the risk free interest rate continuously compounded is r so that $B_s = \exp(sr)$. Suppose also that a constant coupon $\$d$ is paid at the end of every period $s = t+1, ..., T-1$. Then show that the value of this bond at time t is

$$\begin{aligned} P_t &= d\frac{\exp\{-(r+k)\} - \exp\{-(r+k)\{T-t)\}}{1-\exp\{-(r+k)\}} \\ &\quad +pF\exp\{-r(T-t)\} + (1-p)F\exp\{-(r+k)(T-t)\} \end{aligned}$$

17. (a) Show that entropy is always positive and if $Y = g(X)$ is a function of X then Y has smaller entropy than X, i.e. $H(p_Y) \leq H(p_X)$.

    (b) Show that if X has any discrete distribution over n values, then its entropy is $\leq log(n)$.

Chapter 2

Basic Monte Carlo Methods

2.1 Simulation and Monte Carlo Methods

Consider as an example the following very simple problem. We wish to price a European call option with exercise price \$22 and payoff function $V(S_T) = (S_T - 22)^+$. Assume for the present that the interest rate is 0% and S_T can take only the following five values with corresponding (Q) probabilities

s	20	21	22	23	24
$Q[S_T = s]$	1/16	4/16	6/16	4/16	1/16

In this case, since the distribution is very simple, we can price the call option explicitly;

$$E^Q V(S_T) = E^Q (S_T - 22)^+ = (23 - 22)\frac{4}{16} + (24 - 22)\frac{1}{16} = \frac{3}{8}.$$

However, the ability to value an option explicitly is a rare luxury. An alternative would be to generate a large number (say $n = 1000$) independent simulations of the stock price S_T under the measure Q and average the returns from the option. Say the simulations yielded values for S_T of $22, 20, 23, 21, 22, 23, 20, 24, ...$ then the estimated value of the option is

$$\begin{aligned} \overline{V(S_T)} &= \frac{1}{1000}[(22 - 22)^+ + (20 - 22)^+ + (23 - 22)^+ + ...]. \\ &= \frac{1}{1000}[0 + 0 + 1 + ...] \end{aligned}$$

The law of large numbers assures us for a large number of simulations n, the estimator $\overline{V(S_T)}$ will approximate the true expectation $E^Q V(S_T)$. Now while it would be foolish to use simulation a simple problem like this, there are many models in which it is much easier to randomly generate values of the process S_T than it is to establish its exact distribution. In such a case, simulation is the method of choice.

Randomly generating a value of S_T in the above discrete distribution is easy, provided that we can produce independent random uniform random numbers

on a computer. For example, if we were able to generate a random number Y_i which has a uniform distribution on the integers $\{0, 1, 2, ..., 15\}$ then we could define S_T for the i'th simulation as follows:

If Y_i is in set	$\{0\}$	$\{1,2,3,4\}$	$\{5,6,7,8,9,10\}$	$\{11,12,13,14\}$	$\{15\}$
define $S_T =$	20	21	22	23	24

Of course, to get a reasonably accurate estimate of the price of a complex derivative may well require a large number of simulations, but this is decreasingly a problem with increasingly fast computer processors. The first ingredient in a simulation is a stream of uniform random numbers Y_i used above. In practice all other distributions are generated by processing discrete uniform random numbers. Their generation is discussed in the next section.

2.2 Uniform Random Number Generation

The first requirement of a stochastic model is the ability to generate "random" variables or something resembling them. Early such generators attached to computers exploited physical phenomena such as the least significant digits in an accurate measure of time, or in the amount of background cosmic radiation as the basis for such a generator, but these suffer from a number of disadvantages. While they may well be "random" in some more general sense than are the pseudo-random number generators that we use now, their properties are difficult to establish, and the sequences are impossible to reproduce, the latter being important for debugging a simulation program and for reducing the variance therein. Quite remarkably, it was discovered that very simple recursion formulae defined sequences that for practical purposes looked like sequences of independent random numbers and seemed (although the theorems rarely allow a proof of this fact) to more or less obey the major laws of probability such as the law of large numbers, the central limit theorem, the Glivenko-Cantelli theorem, etc. This would seem to indicate that the conclusions of probability hold under much more general circumstances than the relatively restrictive conditions on these theorems indicate. Indeed, one would intuitively expect an enormous difference between the behaviour of independent random variables X_n and a sequence satisfying a recursion of the form $x_n = g(x_{n-1})$ for a simple function g and so it is surprising that for a large class of such functions g it is quite difficult to determine the difference between such a sequence and an independent sequence. Of course, any sequence of numbers generated from a simple recursion such as this is neither random, nor are x_{n-1} and x_n independent. Often we will emphasize this failure by referring to the sequence of *pseudo-random numbers.* While they are in no case independent, we will nevertheless attempt to find simple functions g which provide behaviour *similar* to that of independent uniform random numbers.

Definition: reduction modulo m. For positive integers x and m, the value $a \bmod(m)$ is the remainder (between 0 and $m-1$) obtained when a is divided by m. So for example $7 mod(3) = 1$ since $7 = 2 \times 3 + 1$.

The single most common class of random number generators are of the form

$$x_n := (ax_{n-1} + c) \ mod(m)$$

for given integers $a, c,$ and m which we select in advance. This generator is initiated with a "seed" x_0 and then run to produce a whole sequence of values. When $c = 0$, these generators are referred to as *multiplicative congruential generators* and in general as *mixed congruential generators* or *linear congruential generators*. The "seed", x_0, is usually updated by the generator with each call to it. There are two common choices of m, either m prime or $m = 2^k$ for some k (usually 31 for 32 bit machines).

Example: Mixed Congruential generator

Define $x_n = (5x_{n-1} + 3) \ mod \ 8$, and the seed $x_0 = 3$. Note that by this recursion

$$\begin{aligned} x_1 &= (5 \times 3 + 3) \bmod 8 = 18 \bmod 8 = 2 \\ x_2 &= 13 \bmod 8 = 5 \\ x_3 &= 28 \bmod 8 = 4 \\ \text{and } x_4, x_5, x_6, x_7, x_8 &= 7, 6, 1, 0, 3 \text{ respectively} \end{aligned}$$

and after this point (for $n > 8$) the recursion will simply repeat again the pattern already established, $3, 2, 5, 4, 7, 6, 1, 0, 3, 2, 5, 4, ...$

The above repetition is inevitable for a linear congruential generator. There are at most m possible numbers after reduction mod m and once we arrive back at the seed the sequence is destined to repeat itself. In the example above, the sequence cycles after 8 numbers. The length of one cycle, before the sequence begins to repeat itself again, is called the *period* of the generator. For a mixed generator, the period must be less than or equal to m. For multiplicative generators, the period is shorter, and often considerably shorter.

Multiplicative Generators.

For multiplicative generators, $c = 0$. Consider for example the generator $x_n = 5x_{n-1} \ mod \ 8$ and $x_0 = 3$. This produces the sequence $3, 7, 3, 7, ...$ In this case, the period is only 2, but in general it is clear that the maximal possible period is $m-1$ because it generates values in the set $\{1, ..., m-1\}$. The generator cannot generate the value 0 because if it did, all subsequent values generated are identically 0. Therefore the maximum possible period corresponds to a cycle through non-zero integers exactly once. But in the example here, the period is far from attaining its theoretical maximum, $m - 1$. When is this maximal period achieved? The following Theorem shows that the period of a multiplicative generator is maximal when m is a prime number and a satisfies some conditions.

Theorem 6 *: period of multiplicative generator.*

If m is prime, the multiplicative congruential generator $x_n = ax_{n-1} \, mod \, m$, $a \neq 0$, has maximal period $m-1$ if and only if $a^{m-1} = 1(mod\ m)$ and $a^i \neq 1(mod\ m)$ for all $i < m-1$.

Consider the multiplicative congruential generator $x_n = 2x_{n-1} mod 11$. It is easy to check that $2^i \bmod 11 = 2, 4, 8, 5, 10, 9, 7, 3, 6, 1$ as $i = 1, 2, ...10$. Since the value $i = m-1$ is the first for which $2^i \bmod 11 = 1$, this is a maximal period generator having period 10. When $m = 11$, only the values $a = 2, 6, 7, 8$ produce full period (10) generators.

One of the more common moduli on 32 bit machines is the prime $m = 2^{31}-1$. In this case, the following values of a (among many others) all produce full period generators:

$$a = 7,\ 16807,\ 48271,\ 69621,\ 630360016,\ 742938285,\ 950706376,$$

$$1226874159,\ 62089911,\ 1343714438,\ 39373$$

Let us suppose now that m is prime and a_2 is the multiplicative inverse $(\mathrm{mod}(m))$ of a_1 by which we mean $a_1 a_2 \bmod(m) = 1$. When m is prime, the set of integers $\{0, 1, 2, ..., m-1\}$ together with the operations of addition and multiplication $\mathrm{mod}(m)$ forms a finite field and this allows essentially the same operations as we enjoy in the real number system. Suppose for two integers $x_1, x_2 \in \{0, 1, 2, ..., m-1\}$, $x_2 = a_1 x_1 \bmod(m)$. Then multiplying by a_2 we have

$$a_2 x_2 \bmod(m) = a_2 a_1 x_1 \bmod(m) = a_2 a_1 \bmod(m) x_1 \bmod(m) = x_1$$

and this shows that $\mathrm{x}_1 = a_2 x_2 \bmod(m)$. In other words, if a_2 is the multiplicative inverse of $a_1(\mathrm{mod}\ m)$, then the multiplicative generator with multiplier a_2 generates exactly the same sequence as that with multiplier a_1 except in reverse order.

Theorem 7 *Period of Multiplicative Generators with $m = 2^k$*

If $m = 2^k \geq 8$, the multiplicative congruential generator has maximal period $m/4$ if $a \bmod 8 = 3$ or 5 and if x_0 is odd. T

For the proof of these results, see Ripley(1987), chapter 2. The following simple Matlab code allows us to compare linear congruential generators with small values of m. It generates a total of n such values for user defined $a, c, m, x0 = seed$. The efficient implementation of a generator for large values of m of approximately the same size as the machine precision depends very much on the architecture of the computer.

```
function x=lcg(x0,a,c,m,n)
y=x0;
x=x0;
for i=1:n
 y=rem(a*y+c,m);
 x=[x y];
end
```

Theorem 8 *Period of Mixed Congruential Generators.*

The Mixed Congruential Generator,

$$x_n = (ax_{n-1} + c) mod(m) \tag{2.1}$$

has full period m *if and only if*
(i) c *and* m *are relatively prime.*
(ii) Each prime factor of m *is also a factor of* $a-1$.
(iii) If 4 divides m *it also divides* $a-1$.

Consider $m = 2^{31} - 1$ which is a prime. When m is prime, by (i) together with the assumption that $a < m$, m must divide $a-1$ which implies $a = 1$. So for prime m the only full-period generators correspond to $a = 1$.

Also by Theorem 3.2.5, if $m = 2^k, k \geq 2$, the conditions become that c is odd, and 4 divides $a-1$. Then the generator $x_n = (ax_{n-1} + c) mod(m)$ has full period 2^k.

Among the common linear or multiplicative generators are the following:

m	a	c	
$2^{31}-1$	$7^5 = 16807$	0	Lewis,Goodman, Miller (1969)IBM,
$2^{31}-1$	630360016	0	Fishman (Simscript II)
2^{31}	65539	0	RANDU
2^{32}	69069	1	Super-Duper (Marsaglia)
2^{35}	$5^{13} = 1220703125$	0	APPLE
2^{32}	134775813	1	Turbo-Pascal,Version 7.(period=2^{32})
2^{59}	13^{13}	0	NAG
$2^{31}-1$	630360016	0	Fishman (Simscript II)
$2^{31}-1$	742938285	0	Fishman and Moore
2^{32}	3934873077		Fishman and Moore
$10^{12}-11$	427419669081	0	MAPLE
2^{32}	3141592653	1	DERIVE
2^{32}	663608941	0	Ahrens (C-RAND)

Matrix Congruential Generators.

We consider a generator of $k-$dimensional vectors X. Suppose the components of X are to be integers between 0 and $m-1$ where m is a power of a prime number. If A is an arbitrary $k \times k$ matrix with integral elements also in the range $\{0, 1, ..., m-1\}$ then one simple generator is to begin with a seed X_0 a vector, a constant vector C and define recursively

$$X_n := (AX_{n-1} + C) \ mod(m)$$

Such generators are most common when C = the zero vector and called *matrix multiplicative congruential generators* .

In many cases, the uniform random number generator in packages such as *Splus* and *Matlab* are not completely described in the package documentation For example, in *Splus*, the multiplicative congruential generator is used, and then the sequence is " shuffled" using a Shift-register generator (a special case of the matrix congruential generator described above). This secondary processing of the sequence can increase the period. In general, shuffling is conducted according to the following steps

1. Generate a sequence X_i using $X_{i+1} = a_1 X_i (mod\ m_1)$.
2. For fixed k put $(T_1, \ldots, T_k) = (X_1, \ldots, X_k)$.
3. Generate, using a different generator, a sequence $Y_{i+1} = a_2 Y_i (mod m_2)$.
4. Output the random number T_I where $I = ceiling(Y_i k/m_2)$.
5. Increment i, replace T_I by the next value of X, and return to 3.

One generator is used to produce the sequence as numbers are needed to fill k holes. The other generator is then used select which hole to draw the next number from.

Example: A shuffled generator

Consider a generator described by the above steps with $k = 4, m_1 = 19, m_2 = 29$

```
X(i) = 2   14   7   13   11   18   6   3   9

Y(i) = 11  25  17    4   16   28  14   0   3
```

We start by filling four pigeon-holes with the numbers produced by the first generator so that $(T_1, \ldots, T_4) = (2, 14, 7, 13)$. Then use the second generator to select a random index I telling us which pigeon-hole to draw the next number from. Since these holes are numbered from 1 through 4, we use $I = \lceil 4 \times 11/29 \rceil = 2$. Then the first number in our random sequence is drawn from box 2, i.e. $z_1 = T_2 = 14$, so $z_1 = 14$. This element of the vector is now replaced by 11, the next number in the X sequence. Proceeding in this way, the next index is $I = \lceil 4 \times 25/29 \rceil = 4$ and so the next number drawn is $z_2 = T_4 = 13$. Of course, when we have finished generating the values $z_1, z_2, \ldots$ all of which lie between 0 and $m_1 = 19$, we will usually transform them in the usual way (e.g. z_i/m_1) to produce something approximating continuous uniform random numbers on [0,1]. Because of the small value we chose for m_1, this approximation will not be very good in this case. But the advantage of shuffling is that the period of the generator is greatly extended.

There is another approach, summing pseudo-random numbers, which is also used to extend the period of a generator. This is based on the following theorem (see L'Ecuyer (1988)). For further discussion of the effect of taking linear combinations of the output from two or more random number generators, see Fishman (1995, Section 7.13).

Theorem 9 *summing mod m.*

If X is random variable uniform on the integers $\{0, \ldots, m-1\}$ and if Y is any integer-valued random variable independent of X, then the random variable $W = (X+Y)(mod\ m)$ is uniform on the integers $\{0, \ldots, m-1\}$.

Theorem 10 *(period of generator summed mod m_1)*

If $X_{i+1} = a_1 X_i (mod\ m_1)$ has period $m_1 - 1$ and $Y_{i+1} = a_2 Y_i (mod\ m_2)$ has period $m_2 - 1$, then $(X_i + Y_i)(mod\ m_1)$ has period the least common multiple of $(m_1 - 1, m_2 - 1)$.

Example: a shuffled generator

If $X_{i+1} = 16807 X_i mod(2^{31} - 1)$ and $Y_{i+1} = 40692 Y_i mod(2^{31} - 249)$, then the period of $(X_i - Y_i) mod(2^{31} - 1)$ is

$$\frac{(2^{31} - 2)(2^{31} - 250)}{2 \times 31} \approx 7.4 \times 10^{16}$$

This is much greater than the period of either of the two constituent generators.

Other generators.

There is, in addition to those mentioned above, a wide variety of generators in the literature that have been proposed. Some, like the *Tausworthe* generators, generate pseudo-random bits $\{0, 1\}$ according based on a primitive polynomial over a Galois Field and then map these bits into uniform (0,1) numbers. Others use a non-linear map to replace a linear one. For example we might define $x_{n+1} = x_n^2 \bmod(m)$ (called a *quadratic residue generator*) or $x_{n+1} = h(x_n) \bmod(m)$ for any function h designed to result in large values and more or less random low order bits.

Uniform $(0, 1)$ generators:

In general, random integers should be mapped into the unit interval in such a way that the values 0 and 1, each of which have probability 0 for a continuous distribution are avoided. For a multiplicative generator, since values lie between 1 and $m-1$, we may divide the random number by m. For a linear congruential generator taking possible values $\{0, 1, ..., m-1\}$, it is suggested that we use $(x + 0.5)/m$.

2.3 Apparent Randomness of Pseudo-Random Number Generators

In order that one of the above generators be reasonable approximations to independent uniform variates it should satisfy a number of statistical tests. Suppose we reduce the uniform numbers on $\{0, 1, \ldots m-1\}$ to values approximately uniformly distributed on the unit interval $[0, 1]$ by dividing through by m (it

may sometimes be better to eliminate 0 by first adding 1/2 and then dividing through by m but for a multiplicative generator, since 0 does not occur, this is unnecessary). There is a large number of tests that can be applied to determine whether the hypothesis of independent uniform variates is credible (not, of course, whether the hypothesis is true since we know in advance it is not!).

Runs Test

Consider the hypothesis $H_0 : \{U_i, \quad i = 1, 2, ...\}$ are independent identically distributed random variables. The *runs test* seeks runs, either in the original sequence or in its differences. For example, suppose we denote a positive difference between consecutive elements of the sequence by + and a negative difference by −. Then we regard may regard a sequence of the form .21 .24 .34 .37 .41 .49 .56 .51 .21 .25 .28 .56 .92 .96 as unlikely because the corresponding differences $+ + + + + + - - + + + + +$ has too few "runs" (here $R = 3$). Under the assumption of independence, $E(R) = \frac{2n-1}{3}$ and $var(R) = \frac{16n-29}{90}$ and we may approximate the distribution of R with the normal distribution for $n \geq 25$. We can look for runs of length n in batches in a longer sequence of variates.

Another alternative is the *serial correlation test*. The above test checks for the uniformity of the marginal distribution of $(x_n \ , \ x_{n+1})$ and this could obviously be generalized to variables separated by any number; say $(x_i, \ x_{i+j})$. One could also use the sample correlation or covariance as the basis for such a test. For example, for $j \geq 0$,

$$C_j = \frac{1}{n}(x_1 x_{1+j} + x_2 x_{2+j} + \ . \ . \ x_{n-j} x_n + x_{n+1-j} x_1 + \ . \ . \ . \ x_n x_j) \qquad (2.2)$$

The test may be based on the normal approximation to the distribution of C_j with mean $E(C_j) = 1/3, \ j = 0, \ 1/4$ for $j \geq 1$. Also $var(C_j) = \frac{13}{144n}, \ j \geq 1, \quad var(C_0) = 4/45n$.

Chi-squared test.

The chi-squared test can be applied to the sequence in any dimension, for example $k = 2$. We use the generator to produce a sequence of uniform$(0, 1)$ variables, $U_j, j = 1, 2, ...2n$, and then for a partition $\{A_i; i = 1, ..., K\}$ of the unit square, we count N_i, the number of pairs of the form $(U_{2j-1}, U_{2j}) \in A_i$. Clearly this should be related to the area or probability $P(A_i)$ of the set A_i. Pearson's chi-squared statistic is

$$\chi^2 = \sum_{i=1}^{K} \frac{[N_i - nP(A_i)]^2}{nP(A_i)} \qquad (2.3)$$

which should be compared with a chi-squared distribution with degrees of freedom $K - 1$ or one less than the number of sets in the partition. Observed values

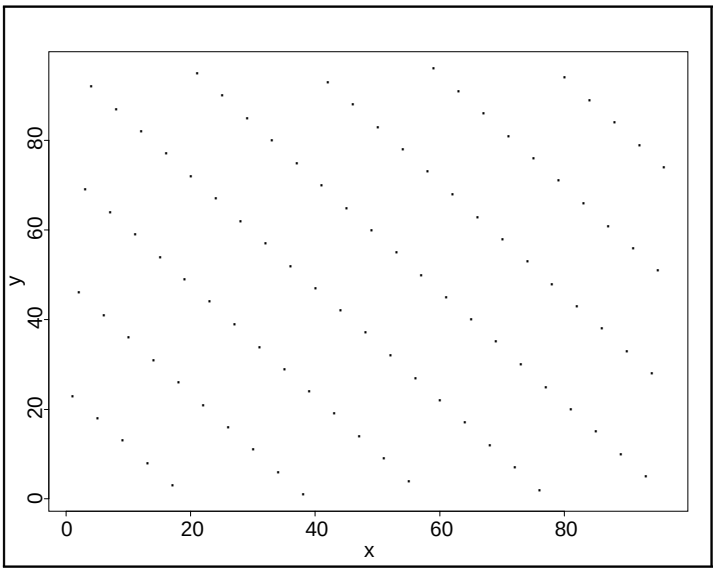


Figure 2.1:

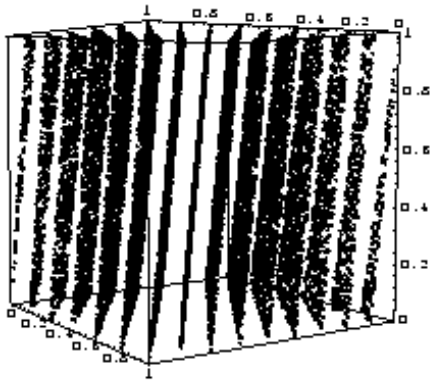
of the statistic that are unusually large for this distribution should lead to rejection of the uniformity hypothesis. The partition usually consists of squares of identical area but could, in general, be of arbitrary shape.

Spectral Test

Consecutive values plotted as pairs (x_n, x_{n+1}), when generated from a multiplicative congruential generator $x_{n+1} = ax_n \; mod \; m$ fall on a *lattice*. A lattice is a set of points of the form $t_1e_1 + t_2e_2$ where t_1, t_2 range over all integers and e_1, e_2 are vectors, (here two dimensional vectors since we are viewing these points in pairs of consecutive values (x_n, x_{n+1})) called the "basis" for the lattice. A given lattice, however, has many possible different bases, and in order to analyze the lattice structure, we need to isolate the most "natural" basis, e.g. the one that we tend to see in viewing a lattice in two dimensions. Consider, for example, the lattice formed by the generator $x_n = 23x_{n-1} mod \; 97$ when we plot adjacent pairs in 2-dimensional space. This plot is given below in Figure ??.

We could use $e_1 = (1, 23) \; and \; e_2 = (4, -6)$, or we could replace e_1 by $(5, 18)$ or $(9, 13)$ etc. Beginning at an arbitrary point on the lattice as origin (in this case, since the original point (0,0) is on the lattice, we will leave it unchanged), we choose an unambiguous definition of e_1 to be the *shortest* vector in the lattice, and then define e_2 as the shortest vector in the lattice which is not of the form te_1 for integer t. Such a basis will be called a *natural basis*. The extension to a lattice in$k$-dimensions is done similarly. All linear

congruential random number generators result in points which when plotted as consecutive k-tuples lie on a lattice. The best generators are those for which the cells in the lattice are as close as possible to squares so that e_1 and e_2 are approximately the same length. Note that the area of the parallelogram with sides e_1 and e_2 is approximately a constant $(1/m)$ whatever the multiplier a so that a longer vector e_1 is associated with a shorter vector e_2 and therefore the two vectors of reasonably similar length. The *spectral test statistic* ν is the renormalized length of the first basis vector $||e_1||$. In general, for k consecutive points, it is equal to $min\ (b_1^2 + b_2^2 + . + . + . + b_k^2)^{1/2}$ under the constraint $b_1 + b_2 a + \ldots b_k a^{k-1} = mq, \quad q \neq 0$. Large values of the statistic indicate that the generator is adequate and Knuth suggests as a minimum threshold the value $\pi^{-1/2}[(k/2)!m/10]^{1/k}$. One of the generators that fails this test most spectacularly with $k = 3$ is the generator RANDU, used commonly in simulations until the 1980's and notorious for the fact that very few hyperplanes fit through all of the points (see Marsaglia, 1968). For RANDU, successive triplets tend to remain on the plane $x_n = 6x_{n-1} - 9x_{n-2}$. This may be seen by rotating a 3-dimensional graph of the sequence of triplets of the form $\{(x_{n-2}, x_{n-1}, x_n); n = 2, 3, 4, ...N\}$ as in figure **??**



Lattice Structure of RANDU

For example consider the following plot of 5000 consecutive triplets from a linear congruential random number generator with $a = 383, c = 263, m = 10,000$.

Linear planes are evident from some angles in this view, but not from others. In many problems, particularly ones in which random numbers are processed in groups of three or more, this phenomenon can lead to highly misleading results. The spectral test is the most widely used test which attempts to insure against lattice structure. The table below gives some values of the spectral test statistic for some linear congruential random number generators in dimension $k \leq 7$.

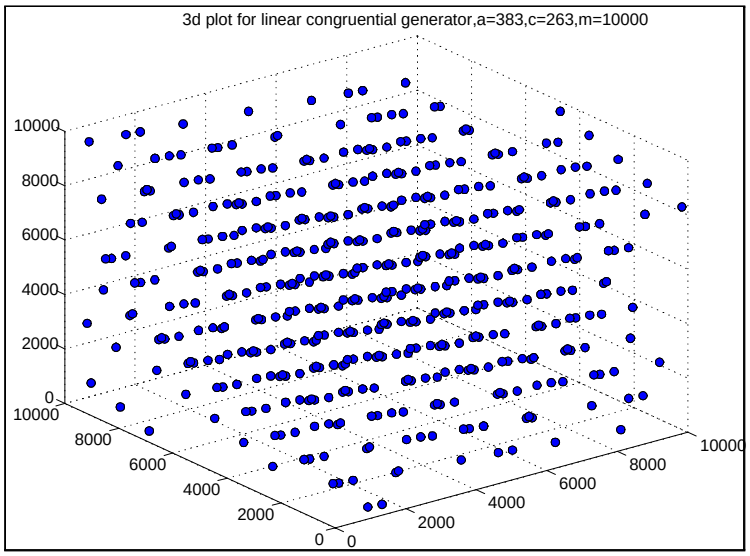


Figure 2.2:

m	a	c	$k=2$	$k=3$	$k=4$	$k=5$	$k=6$	$k=7$
$2^{31}-1$	7^5	0	0.34	0.44	0.58	0.74	0.65	0.57
$2^{31}-1$	630360016	0	0.82	0.43	0.78	0.80	0.57	0.68
$2^{31}-1$	742938285	0	0.87	0.86	0.86	0.83	0.83	0.62
2^{31}	65539	0	0.93	0.01	0.06	0.16	0.29	0.45
2^{32}	69069	0	0.46	0.31	0.46	0.55	0.38	0.50
2^{32}	3934873077	0	0.87	0.83	0.83	0.84	0.82	0.72
2^{32}	663608941	0	0.88	0.60	0.80	0.64	0.68	0.61
2^{35}	5^{13}	0	0.47	0.37	0.64	0.61	0.74	0.68
2^{59}	13^{13}	0	0.84	0.73	0.74	0.58	0.64	0.52

2.4 Non-Uniform Random Number Generation

By far the simplest and most common method for generating non-uniform variates is based on the inverse cumulative distribution function. For arbitrary c.d.f. $F(x)$, define $F^{-1}(y) = min\ \{x; F(x) \geq y\}$. This defines a pseudo-inverse function which is a real inverse (i.e. $F(F^{-1}(y)) = F^{-1}(F(y)) = y$) only in the case that the c.d.f. is continuous and strictly increasing. However, in the general case of a possibly discontinuous non-decreasing c.d.f. the function continues to enjoy some of the properties of an inverse. In particular, in the general case,

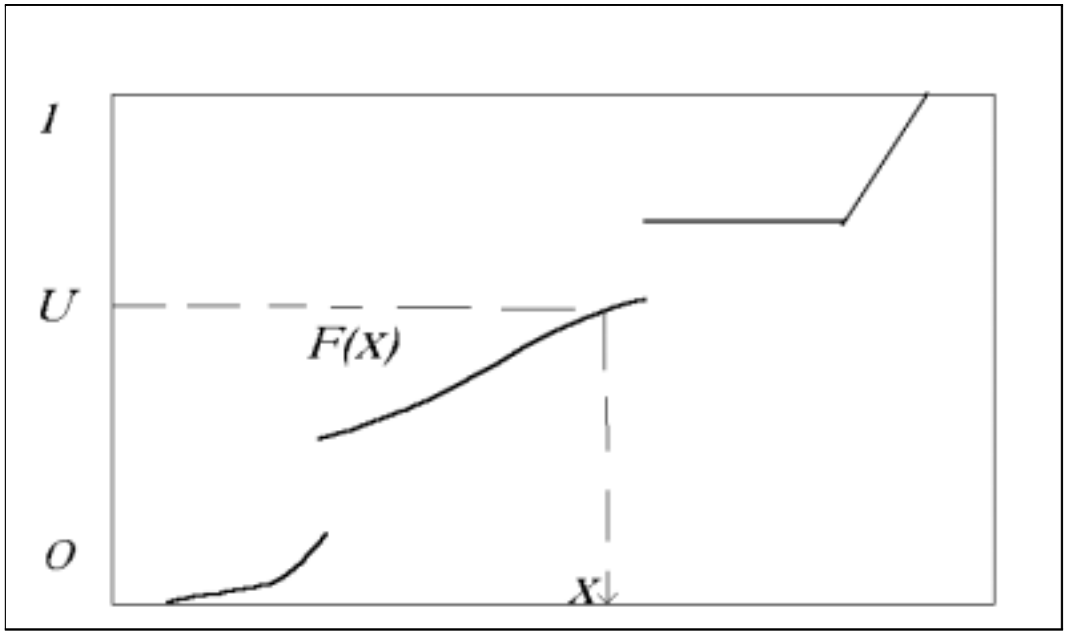


Figure 2.3:

Theorem 11 *(: inverse transform)*

If F is an arbitrary c.d.f. and U is uniform$[0,1]$ then $X = F^{-1}(U)$ has c.d.f. $F(x)$.

Proof:

The proof is a simple consequence of the fact that

$$[U < F(x)] \subset [X \leq x] \subset [U \leq F(x)] \text{ for all x,} \tag{2.4}$$

evident from Figure **??**. Taking probabilities throughout **??**, and using the continuity of the distribution of U so that $P[U = F(x)] = 0$, we obtain

$$F(x) \leq P[X \leq x] \leq F(x).$$

Examples of Inverse Transform

Exponential (θ)

This distribution, a special case of the gamma distributions, is common in most applications of probability. For example in risk management, it is common to model the time between defaults on a contract as exponential (so the default times follow a Poisson process). In this case the probability density function is

$f(x) = \frac{1}{\theta}e^{-x/\theta}, x \geq 0$ and $f(x) = 0$ for $x < 0$. The cumulative distribution function is $F(x) = 1 - e^{-x/\theta}$, $x \geq 0$. Then taking its inverse,

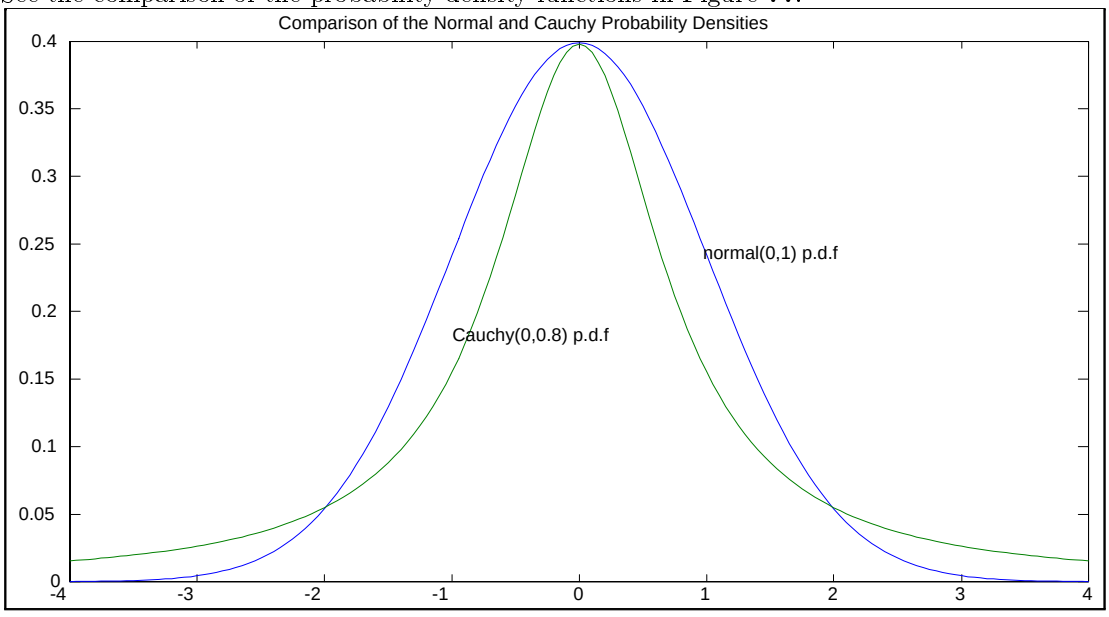
$$\begin{aligned} X &= -\theta ln(1-U) \quad \text{or} \\ X &= -\theta lnU \text{ since } U \text{ and } 1-U \text{ have the same distribution.} \end{aligned}$$

Cauchy (a, b)

This distribution is a member of the "stable family " of distributions. It is similar to the normal only substantially more peaked in the center and with more area in the extreme tails of the distribution. The probability density function is

$$f(x) = \frac{b}{\pi(b^2 + (x-a)^2)}, -\infty < x < \infty.$$

See the comparison of the probability density functions in Figure **??**.



The cumulative distribution function is $F(x) = \frac{1}{2} + \frac{1}{\pi}\, arctan(\frac{x-a}{b})$. Then the inverse transform generator is

$$X = a + b\tan\{\pi(U - \frac{1}{2})\} \quad \text{or} \quad X = a + b/tan(\pi U)$$

where the second expression follows from the fact that $\tan(\pi(x - \frac{1}{2})) = (\tan \pi x)^{-1}$.

Geometric (p)

The probability function is $f(x) = p(1-p)^x, x = 1, 2, 3, ...$ and the cumulative distribution function is $F(x) = 1 - (1-p)^{[x]}$, $x \geq 0$ where $[x]$ denotes the

integer part of x. Then

$$X \;=\; 1 + [\frac{log(1-U)}{log(1-p)}] \;\text{ or }\; 1 + [\frac{E}{log(1-p)}]$$

where we write $\log(1-U) = E$, an exponential(1) random variable.

Pareto (a, b)

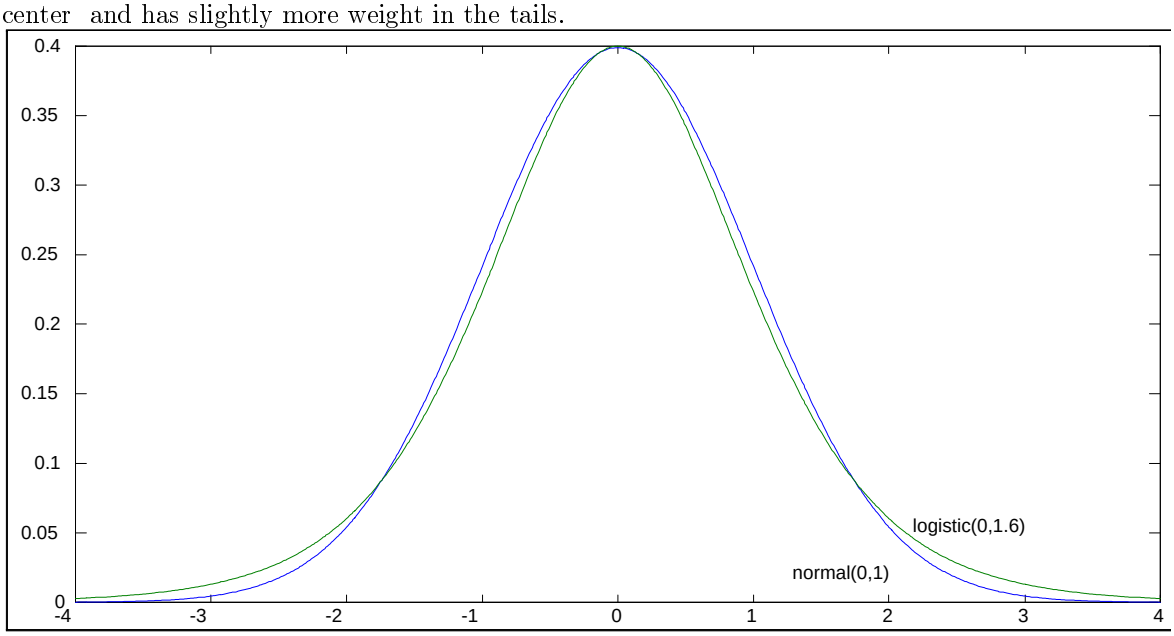
This is one of the simpler families of distributions used in econometrics for modeling quantities with lower bound b (often equal to 0).

$F(x) \;=\; 1 - \left[\frac{b}{x}\right]^a, \quad x \geq b > 0$. Then

$$X \;=\; \frac{b}{(1-U)^{1/a}} \quad \text{or} \quad \frac{b}{U^{1/a}}$$

Logistic

This is again a distribution with shape similar to the normal but closer than is the Cauchy. Indeed as can be seen in Figure **??**, the two densities are almost indistinguishable, except that the logistic is very slightly more peaked i the center and has slightly more weight in the tails.



The logistic cumulative distribution function is

$$F(x) = \frac{1}{1 + exp\{-(x-a)/b\}}.$$

and on taking its inverse, the logistic generator is

$$X = a + b\; ln(U/(1-U)).$$

Extreme Value

$F(x) = 1 - exp\{-exp[(x-a)/b]\}$. Then

$$X = a + b\ log\ log(U)$$

In *Matlab*, the exponential and geometric random number generators are called *exprnd, geornd* respectively and the Cauchy distribution can be generated using the central student's t generator *trnd*. For example, *trnd(V,m,n)* generates an $m \times n$ matrix of student's t random variables having V degrees of freedom.

The generators of certain distributions are as described below. In each case we produce a vector of length n with the associated parameter values.

DISTRIBUTION	SPLUS	MATLAB
normal	rnorm(n, μ, σ)	normrnd$(\mu, \sigma, 1, n)$ or randn$(1, n)$ if $\mu = 1, \sigma = 1$
Student's t	$rt(n, \nu)$	trnd$(\nu, 1, n)$
exponential	rexp(n, λ)	exprnd$(\lambda, 1, n)$
uniform	runif(n, a, b)	unifrnd$(a, b, 1, n)$ or rand$(1, n)$ if $a = 0, b = 1$
Weibull	rweibull(n, a, b)	weibrnd$(a, b, 1, n)$
gamma	rgamma(n, a, b)	gamrnd$(a, b, 1, n)$
Cauchy	rcauchy(n, a, b)	
binomial	rbinom(n, m, p)	binornd$(m, p, 1, n)$
Poisson	rpois(n, λ)	poissrnd$(\lambda, 1, n)$

Inversion performs reasonably well for any distribution for which the cumulative distribution function and its inverse can be found in closed form and computed reasonably efficiently. This includes, as well as the distributions above, the Weibull, the logistic distribution and most discrete distributions which are reasonably well concentrated about the mode. However, for other distributions such as the Poisson with large mean, or the normal, chi-squared, beta etc. other methods need to be used.

In some circumstances, when both the c.d.f. and the probability density are known, we might attempt to invert the c.d.f. by numerical methods. For example, if we use the Newton-Ralphson method, we would iterate until convergence the equation

$$X = X - \frac{F(X) - U}{f(X)} \tag{2.5}$$

beginning with a good approximation to X . For example we might choose the initial value of $X = X(U)$ by using an easily inverted c.d.f. to approximate the true c.d.f.

Suppose $F(x)$ is a cumulative distribution function and $f(x)$ is the corresponding probability density function. Consider the transformation

$$x(u,v) = F^{-1}(u), \quad y(u,v) = vf(F^{-1}(u)), 0 < u < 1, 0 < v < 1$$

This maps a pair of random variables (U, V) which is uniform on the unit square into points uniformly distributed under the graph of the probability density f. This is most easily seen by examining the inverse transformation: $U = F(X), V = Y/f(X)$. The variate V is not needed here if our only objective is producing X with given c.d.f. This is standard inversion. Nevertheless, the fact that the point (X, Y) is uniform under the density underlies one of the simplest yet most useful methods of generating non-uniform variates, the *rejection* or acceptance-rejection method. It is based on the following simple result.

Theorem 12 . *(Acceptance-Rejection)*

(X, Y) is uniformly distributed in the region between the probability density function $y = f(x)$ and the axis $y = 0$ if and only if the marginal distribution of X has density $f(x)$ and the conditional distribution of Y given X is uniform on $[0, f(X)]$.

Proof

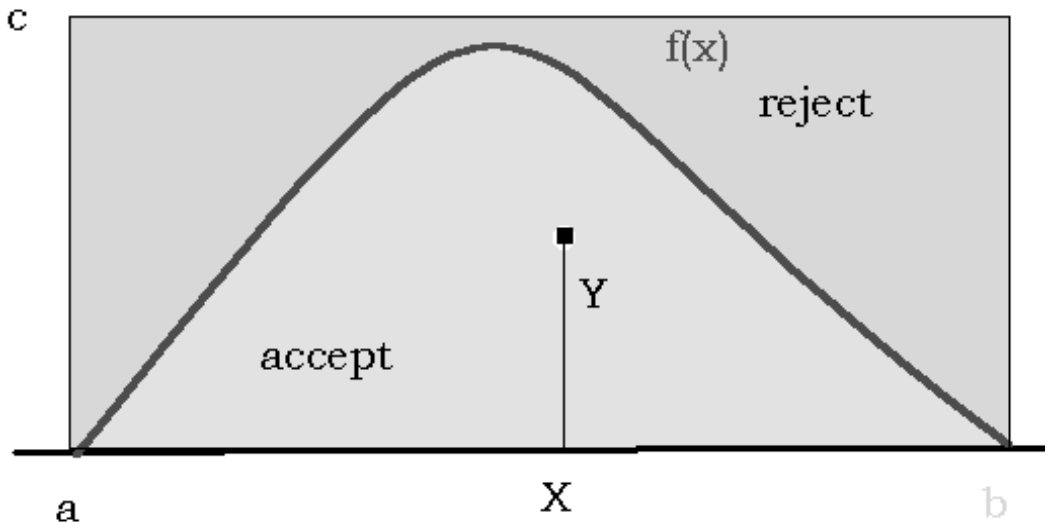
If a point (X, Y) is uniformly distributed under the graph of $f(x)$ notice that the probability $P[a < X < b]$ is proportional to the area under the graph betwen the lines at $x = a$ and $x = b$. In other words $P[a < X < b]$ is proportional to $\int_a^b f(x)dx$. This implies that $f(x)$ is proportional to the probability density function of X and provided that $\int_{-\infty}^{\infty} f(x)dx = 1$, $f(x)$ is the probability density function of X. The converse and the rest of the proof is similar.

The *acceptance-rejection method* works as follows. Suppose $g(x)$ is some easy density for which we can generate variates say by inversion. Suppose we wish to generate a variate X from the harder density $f(x)$ where $f(x) \leq cg(x)$ for some $c > 1$ and for all x. We generate a point uniformly under the graph of $cg(x)$ and then accept that point (in particular X , the x-coordinate of the point) if it turns out to be also below the graph of $f(x)$. Otherwise generate a new point (X, Y), repeating until the condition is satisfied. See Figure **??** where it is assumed that $g(x)$ is the uniform probability density function on the interval $[a, b]$. In algorithmic form, the acceptance-rejection method is;

REPEAT: Generate independent random variables X, U where X has density g and U is uniform on $[0, 1]$.

UNTIL: $U \leq \frac{f(X)}{cg(X)}$

THEN RETURN X.



The Acceptance-Rejection Algorithm

The rejection method is useful if the density g is considerably simpler than f both to evaluate and to generate distributions from and if the constant c is close to 1. The number of iterations through the above loop until a point satisfies the condition has a geometric distribution with parameter $p = 1/c$ and mean c so when c is large, the rejection method is not very effective.

Most schemes for generating non-uniform variates are based on a transformation of uniform with or without some rejection step. We have seen that the rejection algorithm is a special case. Suppose, for example, that $T = (u(x,y),\ v(x,y))$ is a one-one area-preserving transformation of the region $-\infty < x < \infty, 0 < y < f(x)$ into a subset A of the unit square $[0,1]^2$. Notice that any such transformation defines a random number generator for the density $f(x)$ since we can generate a uniform variable in the set A by rejection and when the point falls inside A apply the inverse transformation T^{-1} to this point. The first coordinate X will then have density f. We can think of inversion as a mapping on $[0,1]$ and acceptance-rejection algorithms as an area preserving mapping on $[0,1]^2$.

The most common distribution required for simulations in finance and elsewhere is the *normal distribution*. Recall that if (X,Y) are independent standard normal variates, then expressed in polar coordinates,

$$(R,\Theta) = (\sqrt{X^2+Y^2},\ arctan(Y/X)) \tag{2.6}$$

are distributed as independent variates. $\sqrt{X^2+Y^2}$ has the distribution of the square root of a chi-squared(2) or exponential(2) variable. The distribution of $arctan(Y/X))$ is uniform on $[0,2\pi]$.

It is easy to show that if (X, Y) *are independent standard normal variates, then* $\sqrt{X^2 + Y^2}$ *has the distribution of the square root of a chi-squared(2) (i.e. exponential(2)) variable and* $arctan(Y/X))$ *is uniform on* $[0, 2\pi]$. This result is left as a problem.

This observation is the basis of two of the most common normal generators. The *Box-Muller* algorithm uses two uniform$[0, 1]$ variates U, V to generate R and Θ with the above distributions as

$$R = \{-2ln(U)\}^{1/2}, \quad \Theta = 2\pi V \tag{2.7}$$

and then defines two independent normal(0,1) variates as

$$(X, Y) = R(cos\Theta, sin\Theta) \tag{2.8}$$

Note that normal variates must be generated in pairs, which makes simulations involving an even number of normal variates convenient. If an odd number are required, we will generate one more than required and discard one.

Theorem 13 *(Normal Random Number generator)*

Suppose (R, Θ) *are independent random variables such that* R^2 *has an exponential distribution with mean* 2 *and* Θ *has a* $U[0, 2\pi]$ *distribution. Then* $(X, Y) = R(cos\ \Theta, sin\ \Theta)$ *is distributed as a pair of independent normal variates.*

Proof. Since R^2 has an exponential distribution, R has probability density function

$$f_R(r) = \frac{d}{dr}(1 - e^{-r^2/2}) = re^{-r^2/2}, \quad \text{for } r > 0.$$

and Θ has probability density function $f_\Theta(\theta) = \frac{1}{2\pi}$ for $0 < \theta < 2\pi$. The Jacobian of the transformation is

$$|\frac{\partial(r, \theta)}{\partial(x, y)}| = \begin{vmatrix} \frac{\partial r}{\partial x} & \frac{\partial r}{\partial y} \\ \frac{\partial \theta}{\partial x} & \frac{\partial \theta}{\partial y} \end{vmatrix} = (x^2 + y^2)^{-1/2}.$$

Consequently the joint probability density function of (X, Y) is given by

$$f_\Theta(\arctan(y/x)) f_R((x^2 + y^2)^{-1/2}) |\frac{\partial(r, \theta)}{\partial(x, y)}| = \frac{1}{2\pi} e^{-(x^2+y^2)/2}$$

and this is joint probability density function of two independent standard normal random variables. ■

An alternative algorithm for generating standard normal variates is the *Marsaglia polar* method. This is a modification of the Box-Muller generator and avoids the calculation of *sin* or *cos*. Here we generate a point (Z_1, Z_2) from the uniform distribution on the unit circle. This is done by rejection, generating the point initially from the square $-1 \le z_1 \le 1, \quad -1 \le z_2 \le 1$ and accepting it when it falls in the unit circle or if $z_1^2 + z_2^2 \le 1$. Note that we

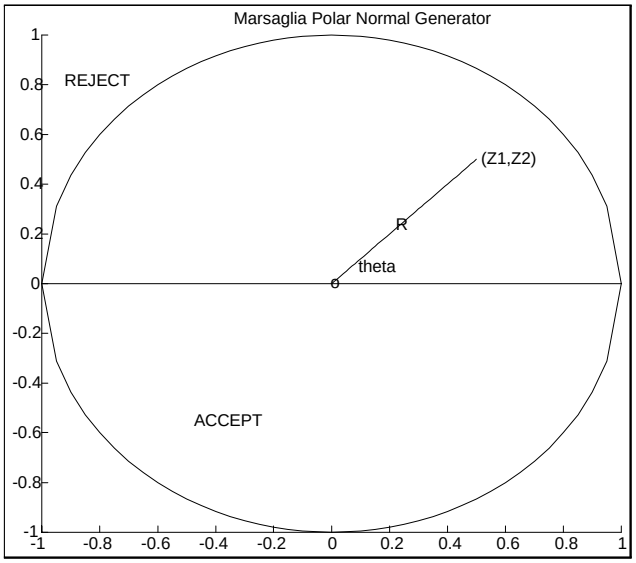


Figure 2.4:

can replace R^2 by $-2log(Z_1^2+Z_2^2)$ and $cos(\Theta), sin(\Theta)$ by $\frac{Z_1}{\sqrt{Z_1^2+Z_2^2}}$ and $\frac{Z_2}{\sqrt{Z_1^2+Z_2^2}}$ respectively. Thus the pair of independent standard normal variables is given by

$$(X,Y) = \sqrt{-2log(Z_1^2+Z_2^2)}(\frac{Z_1}{\sqrt{Z_1^2+Z_2^2}}, \frac{Z_2}{\sqrt{Z_1^2+Z_2^2}}) \text{ when } Z_1^2+Z_2^2 < 1$$

The probability that a point generated inside the square falls inside the unit circle is $\pi/4,$ so that on average around $4/\pi \approx 1.27$ pairs of uniforms are needed to generate a pair of normal variates.

The speed of the Marsaglia polar algorithm compared to that of the Box-Muller algorithm depends on the relative speeds of generating uniform variates versus the sine and cosine transformations. The Box-Muller and Marsaglia polar method are illustrated in Figure **??**:

The normal random number generator in *Matlab* is called *normrnd* or for standard normal *randn*. For example $normrnd(\mu, \sigma, m, n)$ generates a matrix of $m \times n$ pseudo-independent normal variates with mean μ and standard deviation σ and *rand(m,n)* generates an $m \times n$ matrix of standard normal random numbers.

The Lognormal Distribution

If Z is a normal random variable with mean μ and variance σ^2, then we say that the distribution of $X = e^Z$ is lognormal with mean $\eta = \exp\{\mu + \sigma^2/2\}$ and volatility parameter σ. Note that a random variable with a lognormal distribution is strictly positive, making it a good candidate for modelling stock prices. The lognormal probability density function with mean $\eta > 0$ and volatility parameter $\sigma > 0$ is given by the probability density function

$$g(x|\eta,\sigma) = \frac{1}{x\sigma\sqrt{2\pi}} exp\{-(log\ x - log\ \eta\ - \sigma^2/2)^2/2\sigma^2\}. \tag{2.9}$$

A random variable with a lognormal distribution is easily generated by generating an appropriate normal random variable Z and then exponentiating.

A Discrete Time Black-Scholes Model

Suppose that a stock price $S_t, t = 1, 2, 3, ...$ has a lognormal distribution such that the returns over non-overlapping independent periods are independent. Let us assume that there is a total of N such periods in a year. In other words, we require that $S_t = S_0 \exp\{\sum_{i=1}^t Z_i\}$ for independent normal random variables Z_i which have an expected value which may depend on i. We assume that $var(Z_i) = \sigma^2/N$ so that the parameter σ^2 represents the volatility parameter of the stock price after one year. This is a fairly natural model since $S_t = S_0 \Pi_{i=1}^t e^{Z_i}$ is the product of independent returns over the t periods. Assume that the annual interest rate on a risk-free bond is r (so that the interest rate per period is r/N). Recall that the risk-neutral measure Q is such that the stock price discounted to the present forms a martingale, and let us assume for the present that under the risk neutral measure, the stock price process has a similar lognormal representation $S_t = S_0 \exp\{\sum_{i=1}^t Z_i\}$ for independent normal random variables Z_i where the Z_i may have a different mean. Full justification of this process really relies on the continuous time version of the Black Scholes described in Section 1.6. Note that if the process

$$e^{-rt/N} S_t = S_0 \exp\{\sum_{i=1}^t (Z_i - \frac{r}{N})\}$$

is to form a martingale, it is necessary that

$$\begin{aligned} E^Q[S_0 \exp\{\sum_{i=1}^{t+1}(Z_i - \frac{r}{N})\}|H_t] &= S_0 \exp\{\sum_{i=1}^t (Z_i - \frac{r}{N})\} E^Q[\exp\{Z_{t+1} - \frac{r}{N}\}] \\ &= S_0 \exp\{\sum_{i=1}^t (Z_i - \frac{r}{N})\} \end{aligned}$$

and so $\exp\{Z_{t+1} - \frac{r}{N}\}$ must have a lognormal distribution with expected value 1. In other words, for each i the expected value of $Z_i - \frac{r}{N}$ is, under Q, equal to

$-\sigma^2/2N$ and variance σ^2/N. Notice then that under Q, S_T has a lognormal distribution with mean

$$S_0 e^{rT/N}$$

and volatility parameter $\sigma\sqrt{T/N}$. We can price a call option with maturity T periods from now by generating the random path $S_t, t = 1, 2, ...T$ using the lognormal distribution for S_T and then discounting the payoffs to the present and then averaging the results; i.e. the value of a call with exercise price K is an average of simulated values of

$$e^{-rT/N}(S_0 \exp\{\sum_{t=1}^{T} Z_i\} - K)^+, \text{ where } Z_i \text{ are independent } N(\frac{r}{N} - \sigma^2/(2N), \sigma^2/N).$$

The following function simulates the stock price over the whole period until maturity and then values a European call option on the stock by averaging.

Example 14 *(simulating a call option)*

Consider simulating a call option on a stock whose current value is \$1.00. The option expires in T days and the strike price is K. We assume constant spot interest rate r and the stock price follows a lognormal distribution with annual volatility σ. The following Matlab function provides a simple simulation and graph of the path of the stock over the life of the option and then outputs the discounted payoff from the option.

```
function z=plotlogn(r,sigma,T, K)
% outputs the discounted simulated return on expiry of a call option (per dollar pv of stock).
% Expiry =T days from now,
% current stock price=$1.
% r = annual spot interest rate
% sigma=annual vol. K= strike price.
N=250 ; % N is the assumed number of periods in a year.
s = sigma/sqrt(N); %s is volatility per period
mn = r/N - s^2/2; % mean of the normal increments per period
y=exp(cumsum(normrnd(mn,s,T,1)));
y=[1 y'];
x = (0:T)/N;
plot(x,y,'-',x,K*ones(1,T+1),'y')
xlabel('time (in years)')
ylabel('value of stock')
title('SIMULATED RETURN FROM CALL OPTION')
z = exp(-r*T/N)*max(y(T+1)-K, 0);        % payoff from option discounted to present
```

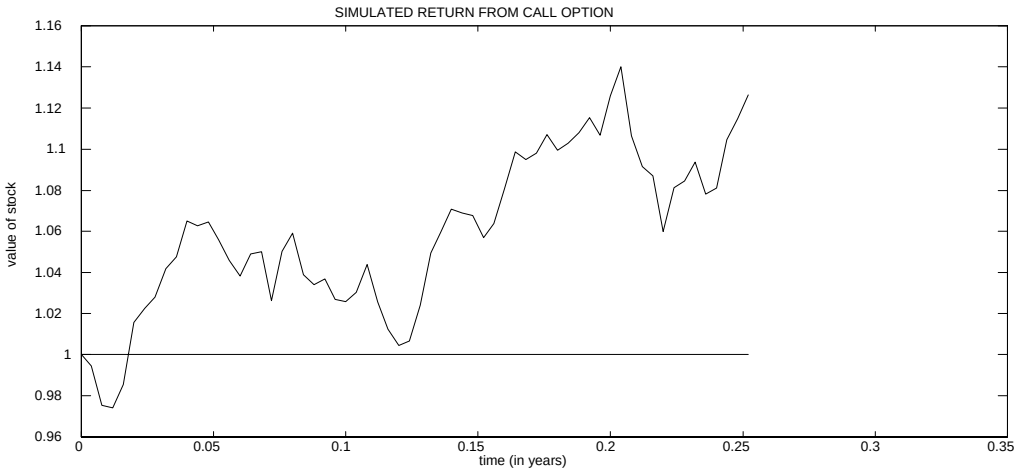


Figure 2.5:

Figure **??** resulted from one simulation run with $r = .05, T = 63$ (about 3 months), $\sigma = .20, K = 1$.

The return on this run was the discounted difference between the terminal value of the stock and the strike price or 0.113. We may repeat this many times, averaging the result and then discounting these returns to the present as in the following function.

```
function z = simcall(r, sigma, T, K, m)
% repeats plotlogn a total of m times and averages discounted return to the present.
z=[];
hold on
for i=1:m
z = [z plotlogn( r, sigma, T, K )];
end
optionval=mean(z);
disp([' option val= 'num2str(optionval)])
hold off
```

For example to value an at the money call option with exercise price=the initial price of the stock=\$1, we type

simcall(.05,.2,63,1,100); and obtain the output *option val= 0.044978.* If we repeat the identical statement, the output is different, i.e. *option val= 0.049117* because each is an average obtained from only 100 simulations. Averaging over more simulations would result in greater precision, but this function is not written with computational efficiency in mind. We will provide more efficient simulations for this problem later. For the moment we can compare the price of this option as determined by simulation with the exact price according to the

Black-Scholes formula. This formula was developed in Section 1.6. The price of a call option at time $t = 0$ given by

$$V(S_T, T) = S_T\Phi(d_1) - Ke^{-rT/N}\Phi(d_2)$$

where

$$d_1 = \frac{\log(S_T/K) + (r + \frac{\sigma^2}{2})T/N}{\sigma\sqrt{T/N}} \quad \text{and} \quad d_2 = \frac{\log(S_T/K) + (r - \frac{\sigma^2}{2})T/N}{\sigma\sqrt{T/N}}$$

and the Matlab function which evaluates this is the function *blsprice* which gives, in this example, and exact price on entering

[CALL,PUT] =BLSPRICE(1,1,.05,63/250,.2,0)

of CALL=0.0464. With these parameters, 4.6 cents on the dollar allows us to lock in the present price of a stock (or commodity if the lognormal model fits) for a period of about three months. The fact that this can be done cheaply and with ease is part of the explanation for the popularity of derivatives as tools for hedging.

We turn now to algorithms for generating the *Gamma distribution* with density

$$f(x|a, b) = \frac{x^{a-1}e^{-x/b}}{\Gamma(a)b^a}, \quad x > 0 \tag{2.10}$$

Recall that the exponential ($a = 1$) and the chi-squared ($a = \nu/2$, $b = 2$, ν *integer*) are special cases of the Gamma distribution. The following result lists some of the properties of the Gamma distributions.

Theorem 15 *(Gamma distribution)*

If X_1, X_2 are independent Gamma (a_1, b) and Gamma (a_2, b) variates, then $Z = \frac{X_1}{X_1+X_2}$ and $Y = X_1 + X_2$ are independent variates with the beta (a_1, a_2) and the Gamma $(a_1 + a_2,\ b)$ distributions respectively. Conversely, if $(Z,\ Y)$ are independent variates with the latter pair of distributions, then $X_1 = YZ$, $X_2 = Y(1 - Z)$ have the indicated Gamma distributions.

Proof. *Assume that X_1, X_2 are independent Gamma (a_1, b) and Gamma (a_2, b) variates. Then their joint probability density function is*

$$f_{X_1X_2}(x_1, x_2) = kx_1^{a_1-1}x_2^{a_2-1}e^{-(x_1+x_2)/b}, x_1 > 0, x_2 > 0$$

where k is the constant $[\Gamma(a_1)\Gamma(a_2)]^{-1}$. Consider the change of variables $x_1 = zy, x_2 = (1 - z)y$. Then the Jacobian $\begin{vmatrix} \frac{\partial x_1}{\partial z} & \frac{\partial x_1}{\partial y} \\ \frac{\partial x_2}{\partial z} & \frac{\partial x_2}{\partial y} \end{vmatrix} = y$. The joint probability density function of (z, y) is given by

$$\begin{aligned} f_{z,y}(z, y) &= f_{X_1X_2}(zy, (1 - z)y) \begin{vmatrix} \frac{\partial x_1}{\partial z} & \frac{\partial x_1}{\partial y} \\ \frac{\partial x_2}{\partial z} & \frac{\partial x_2}{\partial y} \end{vmatrix} \\ &= kz^{a_1-1}(1 - z)^{a_2-1}y^{a_1+a_2-1}e^{-y/b}, 0 < z < 1, y > 0. \end{aligned}$$

and this is easily seen to be the product of two probability density functions, the $Beta(a_1, a_2)$ *density for* Z *and the* $Gamma(a_1 + a_2, b)$ *probability density function for* Y. ■

This result is a basis for generating gamma variates with integral shape parameter a since this can be done by adding independent exponential variates. Thus $-log(\prod_{i=1}^{n} U_i)$ generates a gamma $(n, 1)$ variate for independent uniform U_i . The computation required for this algorithm, however, increases linearly in the parameter $a = n$, and therefore alternatives are required, especially for large a .

For large a one successful algorithm is due to Cheng (1977) and involves rejection from the *Burr XII* density of the form

$$g(x) = \lambda\mu \frac{x^{\lambda-1}}{(\mu + x^{\lambda})^2} \tag{2.11}$$

generated by inverse transform as $\{\frac{\mu U}{1-U}\}^{1/\lambda}$. Assume that the scale parameter of the Gamma $b = 1$. Matching the modes of these two distributions for large a results in choosing $\mu = a^{\lambda}$ and choosing λ to minimize $max\{f(x)/g(x)| - \infty < x < \infty\}$ results in $\lambda = \sqrt{2a-1}$. In this case, $c = \frac{4a^a e^{-a}}{\lambda\Gamma(a)}$ and this approaches $\sqrt{4/\pi}$ as $a \to \infty$.

A much simpler function for dominating the gamma densities is a minor extension of that proposed by Ahrens and Dieter (1974). It corresponds to using

$$cg(x) = \frac{x^{\alpha-1}}{\Gamma(\alpha)} , \quad x \le b \tag{2.12}$$

and

$$cg(x) = \frac{b^{\alpha-1}e^{-x}}{\Gamma(\alpha)} , \quad x > b \tag{2.13}$$

where the efficiency is determined by $c = \int cg(x) = \frac{b^{\alpha}}{\Gamma(\alpha+1)} + f(b)$ and we would clearly try to choose b corresponding as closely as possible to a minimum of this quantity. Ahrens and Dieter use $b = 1$. Other distributions that have been used as dominating functions for the Gamma are the Cauchy (Ahrens and Dieter), the Laplace (Tadakamalla), the exponential (Fishman), the Weibull, the relocated and scaled t distribution with 2 degrees of freedom (Best), a combination of normal density (left part) and exponential density (right part) (Ahrens and Dieter), and a mixture of two Erlang distributions (Gamma with integral shape parameter α).

Best's algorithm generates a Student's t variate as

$$Y = \frac{\sqrt{2}(U - 1/2)}{\sqrt{U(1-U)}} \tag{2.14}$$

where $U \sim U[0,1]$. Then Y has the Student(2) density

$$g(y) = \frac{1}{(2+y^2)^{3/2}}. \tag{2.15}$$

We then generate X as $(\alpha-1)+Y\sqrt{3\alpha/2 - 3/8}$ and apply a rejection step to X . See Devroye (p. 408) for details.

Most of the above algorithms are reasonably efficient only for $\alpha > 1$ with the one main exception being the combination of power of x and exponential density suggested by Ahrens and Dieter above. Cheng and Feast (1979) also suggest a ratio of uniforms algorithm for the gamma distribution, $\alpha > 1$.

A final alternative for the case $\alpha < 1$ is the use of Stuart's theorem which states that $XU^{1/\alpha}$ *has a gamma* $(\alpha,1)$ *distribution when* U *is uniform* $[0,1]$ *and* X *is Gamma* $(\alpha+1,1)$. The Matlab function *gamrnd* uses Best's algorithm and acceptance rejection for $\alpha > 1$ and for $\alpha < 1$, it uses Johnk's generator. This consists of generating U and V both independent U[0,1] and setting

$$X = \frac{U^{1/\alpha}}{U^{1/\alpha}+V^{1/(1-\alpha)}}$$

conditional on the denominator $U^{1/\alpha}+V^{1/(1-\alpha)} < 1$. Multiplying by an independent exponential (1) results in a Gamma$(\alpha,1)$ random variable.

We now turn to generating the *beta distribution* which has density given by

$$f(x) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}x^{a-1}(1-x)^{b-1}, \quad 0 \le x \le 1 \tag{2.16}$$

The beta density obtains as a transformation of an F-distribution (the basis of the transformation is the theorem on page ..), or as the distribution of order statistics in a sample from independent uniform $[0,1]$ variates. The variable $Z = \frac{X_1}{X_1+X_2}$ indicates one method of using a gamma generator to produce beta variates, and this is highly competitive as long as the gamma generator is reasonably fast. The MATLAB generator is *betarnd(a,b,1,n)* Alternatives are, as with the gamma density, rejection from a Burr XII density (Cheng, 1978) and use of the following theorem as a generator (due to Johnk).

Theorem 16 *(Beta distribution)*

Suppose U, V *are independent uniform* $[0,1]$ *variates. Then the conditional distribution of*

$$X = \frac{U^{1/a}}{U^{1/a}+V^{1/b}} \tag{2.17}$$

given that $U^{1/a}+V^{1/b} \le 1$ *is Beta* (a,b). *Similarly the conditional distribution of* $U^{1/a}$ *given that* $U^{1/a}+V^{1/b} \le 1$ *is Beta* $(a+1,b)$.

Proof. *Define a change of variables*

$$\begin{aligned} X &= \frac{U^{1/a}}{U^{1/a}+V^{1/b}}, Y = U^{1/a}+V^{1/b} \\ \text{or } U &= (YX)^a \text{ and } V=[(1-X)Y]^b \end{aligned}$$

so that the joint probability density function of (X,Y) *is given by*

$$\begin{aligned} f_{X,Y}(x,y) &= f_{U,V}((yx)^a,[(1-x)y]^b)\left|\begin{array}{cc}\frac{\partial u}{\partial x} & \frac{\partial u}{\partial y}\\ \frac{\partial v}{\partial x} & \frac{\partial v}{\partial y}\end{array}\right| \\ &= aby^{a+b-1}x^{a-1}(1-x)^{b-1} \quad \text{provided } 0<x<1 \text{ when } y<1 \\ \text{or provided } 1-\frac{1}{y} &< x < \frac{1}{y} \text{ when } 1<y<2. \end{aligned}$$

Notice that in the case $y<1$, *the range of values of* x *is the unit interval and does not depend on* y *and so the conditional probability density function of* X *given* $Y=y$ *is a constant times* $x^{a-1}(1-x)^{b-1}$, *i.e. is the Beta*(a,b) *probability density function. The rest of the proof is similar.* ■

A generator exploiting this theorem produces pairs (U,V) until the condition is satisfied and then transforms to the variable X. However, the probability that the condition is satisfied is $\frac{\Gamma(a+1)\Gamma(b+1)}{\Gamma(a+b+1)}$ which is close to 0 unless a,b are small, so this procedure should be used only for small values of both parameters. Theorems 3.3.12 and 3.3.15 together provide an algorithm for generating Gamma variates with non-integral α from variates with integral ones. For example if X is Gamma $(4,1)$ and Z is independent Beta $(3.4,.6)$ then XZ is Gamma $(3.4,1)$.

There are various other continuous distributions commonly associated with statistical problems. For example the *Student's t-distribution* with ν degrees of freedom is defined as a ratio $\sqrt{\frac{2\nu}{X}}Z$ where Z is standard normal and X is gamma $(\frac{\nu}{2},\ 2)$. Alternatively, we may use $\sqrt{\nu}\frac{X-1/2}{\sqrt{X(1-X)}}$ where X is generated as a symmetric beta $(\nu/2,\ \nu/2)$ variate.

The Symmetric Stable Laws.

A final family of distributions of increasing importance in modelling is the *symmetric stable family*. These are unimodal densities, symmetric about their mode, and roughly similar in shape to the normal or Cauchy distribution (both special cases). They are of considerable importance in finance as an alternative to the normal distribution, because they tend to fit observations better in the tail of the distribution than does the normal. However, this is a more complicated family of densities to work with; neither the density function nor the cumulative distribution function can be expressed in a simple closed form. Both require a series expansion. They are most easily described by their *characteristic function*, which, upon setting location equal to 0 and scale equal to 1 is $Ee^{iXt} = e^{-|t|^\alpha}$ where here i is the complex number $i^2=-1$. The parameter

$0 < \alpha \leq 2$ indicates what moments exist, for except in the special case $\alpha = 2$ (the normal distribution), moments of order less than α exists while moments of order α or more do not. Of course, for the normal distribution, moments of all orders exist. The stable laws are useful for modelling in situations in which variates are thought to be approximately normalized sums of independent identically distributed random variables. To determine robustness against heavy-tailed departures from the normal distribution, tests and estimators can be computed with data simulated from a symmetric stable law with α near 2. The probability density function does not have a simple closed form except in the case $\alpha = 1$ (Cauchy) and $\alpha = 2$ but can be determined from the series expansion of the probability density

$$f_c(x) = \sum_{k=0}^{\infty} (-1)^k \frac{\Gamma((k+1)/2)}{\pi c \alpha k!} \cos(\frac{k\pi}{c})(\frac{x}{c})^k$$

where c is a scale parameter. Especially for large values of x, this probability density function converges extremely slowly. According to Chambers, Mallows and Stuck, (1976), such a variate can be generated in the case $\alpha \neq 1$,

$$X = sin(\alpha U) \left[\frac{cos(U(1-\alpha))}{E} \right]^{\frac{1}{\alpha}-1} (cos\ U\ )^{-1/\alpha} \tag{2.18}$$

where U is uniform $[-\pi/2,\ \pi/2]$ and E is standard exponential and independent. The case $\alpha = 1$ is the Cauchy $X = tan(U)$. Since the tan is a relatively slow operation, this is sometimes replaced by the ratio of Normal variates produced by Marsaglia's polar algorithm. In other words we use $X = V_1/V_2$ where $V_i \sim U[-1, 1]$ conditional on $V_1^2 + V_2^2 \leq 1$ as a standard Cauchy variate.

Example: Stable random walk.

Generate a random walk with 10,000 time steps where each increment is distributed as independent stable random variables having parameter 1.7.

The following *Matlab* function was used

```
function s=stabrnd(a,n)
u=(unifrnd(0,1,n,1)*pi)-.5*pi;
e = exprnd(1,n,1);
s=sin(a*u).*(cos((1-a)*u)./e).^(1/a-1).*(cos(u)).^(-1/a)
```

Then the command

plot(1:10000, cumsum(stabrnd(1.7,10000)));

resulted in the Figure **??**. Note the occasional very large jump which dominates the history of the process up to that point.

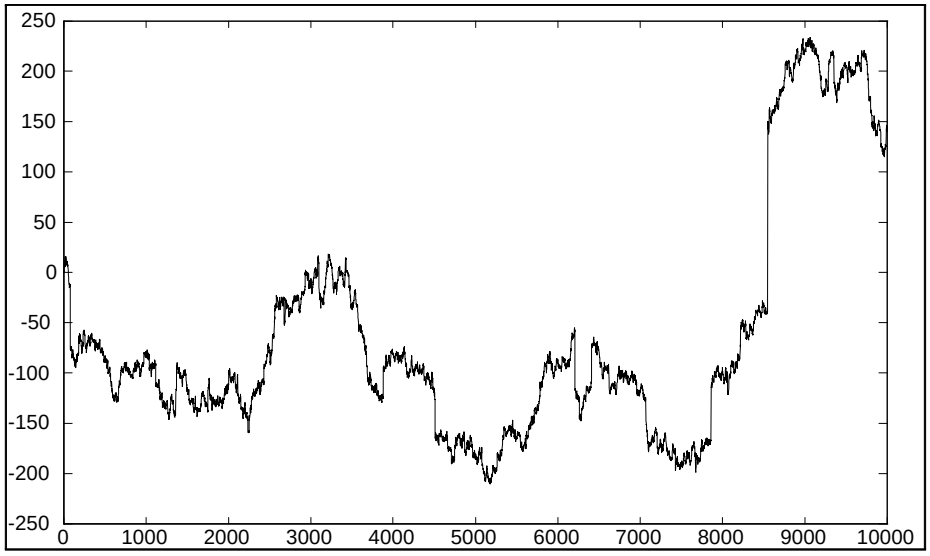


Figure 2.6:

2.5 Generating Discrete Distributions

Many of the methods described above such as inversion and rejection for generating continuous distributions work well for discrete random variables. For example, if X is a discrete distribution taking values on the integers with probability function $f(x)$, $x = 0, 1, \ldots$ we may use rejection to generate a continuous variate Y which has the same cumulative distribution function at the integers $F_Y(j) = F_X(j)$ and then set $X = \lfloor Y \rfloor$ the integer part of Y.

Inversion for discrete variates often requires, for reasonable efficiency, some setup costs. For example if X has c.d.f. $F(x)$, $x = 0, 1, \ldots$ we wish to output an integer X satisfying $F(X-1) < U \leq F(X)$ and the most obvious technique for finding such a value of X is to search sequentially through the potential values $0, 1, 2, \ldots$. Figure **??** is the search tree for inversion for the distribution on the integers $1, \ldots 5$ given by $(p_1, p_2, p_3, p_4, p_5) = (0.11, 0.30, 0.25, 0.21, 0.13)$ We generate an integer by repeatedly comparing a uniform [0,1] variate U with the value at each node, taking the right branch if it is greater than this threshold value, the left if it is smaller. The number of values searched will average to $E(X)$ which for many discrete distributions can be unacceptably large.

An easy alternative is to begin the search at the median m (or mode or mean) of the distribution, searching to the left or right depending on the value of U as in Figure **??**.

This results in searching an average of $E[|X + 1 - m|]$ before obtaining the generated variable often substantially smaller than $E(X)$ when the mean is

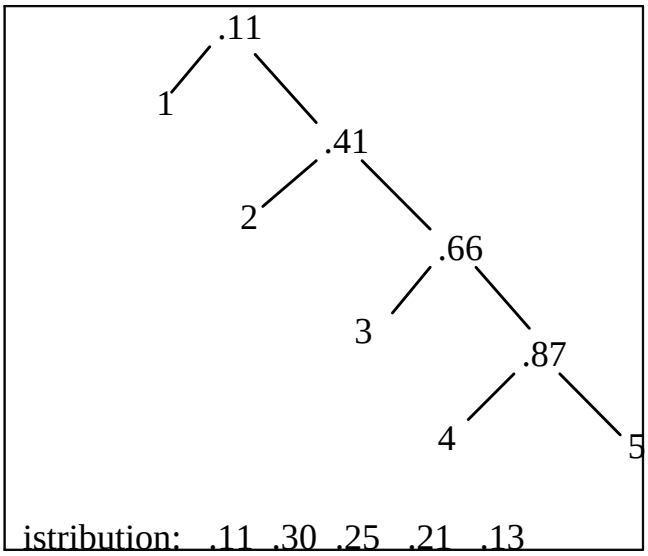


Figure 2.7:

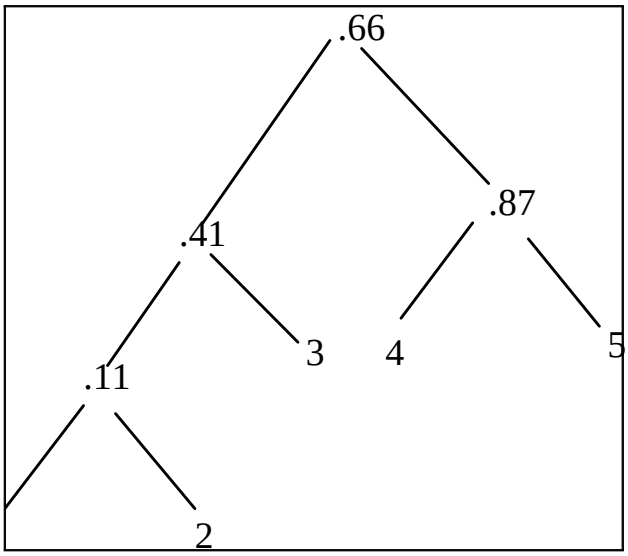


Figure 2.8:

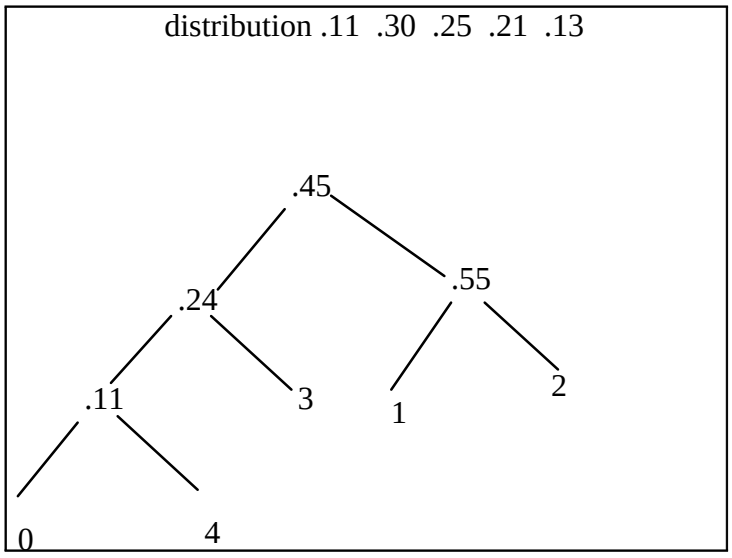


Figure 2.9:

large but still unacceptably large when the distribution has large variance. An optimal binary search tree for this distribution is graphed in Figure **??**. This tree has been constructed by joining the two smallest probabilities to form a new node with weight a combination of the two, and hence working from the leaves to the root of the tree. Equivalently, we use inversion after a re-ordering of the values from those with smallest to those with largest probability.

The leaves of the tree are the individual p_i and the internal nodes are sums of the weights of the children. If D_i represents the depth of the $i'th$ leaf, then the average number of comparisons to generate a single X is $\sum_i p_i D_i$ and the procedure for constructing this tree provides an optimal algorithm in the sense that this quantity is minimized. It is interesting to note that an optimal binary search tree will reduce the average number of comparisons from $E(X)$ for ordinary inversion to less than $1 + 4\,[log_2(1 + E(X))]$.

Another general method for producing variates from a discrete distribution was suggested by Walker (1974, 1977) and is called the *alias method*. Apart from the time required to set up an initial table of aliases and aliasing probabilities, the time required to generate values from a discrete distribution with K supporting points is bounded in K, unlike inversion or binary search which increase as $E(X)$ increases. The idea is to reduce any discrete distribution to a uniform mixture of two-point distributions. For a discrete distribution of the form $p_1, p_2, \ldots p_K$ on the integers $1,\ 2,\ \ldots K$, we seek a table of values of $A(i)$ and associated alias probabilities $q(i)$ so that the following algorithm generates the desired discrete distribution.

GENERATE I UNIFORM ON $\{1, \ldots K\}$.

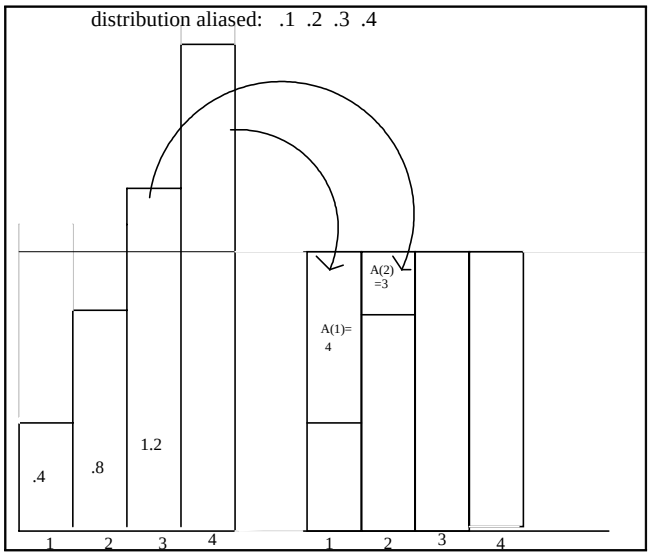


Figure 2.10:

WITH PROBABILITY $q(I)$, OUTPUT $X = I$, OTHERWISE, $X = A(I)$.

An algorithm for producing these values of $A(i)$, $q(i)$ is suggested by Walker(1977) and proceeds by reducing the number of non-zero probabilities one at a time.

1. Put $q_i = Kp_i$ for all i.

2. LET m be the index of $min\{q_i; q_i > 0\}$ and let M be the index of the maximum.

3. SET $A(m) = M$ and fix $q(m)$.

4. Replace $(q_1, \ldots q_K)$ by $(q_1, ..., q_{m-1}, q_{m+1}, ...q_M - (1 - q_m)\ldots$ (so the component with index m is removed).

5. Return to 2 unless all remaining $q_i = 1$ or the vector of q_i's is empty.

Figure **??** shows the way in which aliasing iteratively adjusts a probability histogram to form a rectangle with base K. We construct the vector of aliasing probabilities and aliases for the distribution vector

$$p_i = (.1,\ .2,\ .3,\ .4), \quad i = 1, \ldots 4.$$

This results in $A(i) = (4, 3, x, x)$ and $q_i = (.4, .8, 1, 1)$ respectively.

The Poisson Distribution.

Consider the *Poisson distribution*

$$f(x) = \frac{\lambda^x e^{-\lambda}}{x!}, \; x = 0, 1, \ldots \tag{2.19}$$

The simplest generator is to use the Poisson process. Recall that if points are distributed on the line in such a way that the spacings between consecutive points are independent exponential(1), then the resulting process is a Poisson process with rate 1. Thus, the number of points in an interval of length λ has the desired Poisson (λ) distribution. So for small λ, we could use $inf\{X; \sum_{i=1}^{X+1}(-ln\ U_i) > \lambda\}$ or equivalently

$$inf\{X; \prod_{i=1}^{X+1} U_i < e^{-\lambda}\} \tag{2.20}$$

Once again this generator requires time which grows linearly with λ and so an alternative for large λ is to use rejection. Various possibilities have been suggested for dominating function, from the logistic distribution(Atkinson (1979)) to a normal distribution with exponential right tail (cf. Devroye, lemma 3.8, page 509). A simple all-purpose alternative seems to be a table-mountain function (cf. Stadlober (1989)), essentially a function with a flat top and tails that decrease as $1/x^2$.

A simple alternative for generating Poisson variates that is less efficient but simpler to implement is to use the *Lorentzian*, or truncated Cauchy distribution with probability density function

$$g(x|a, b) = \frac{c_0}{b^2 + (x - a)^2}, \quad x > 0 \tag{2.21}$$

where c_0 is the normalizing constant. A random variable is generated from this distribution using the inverse transform method; $X = a + b\ tan(\pi U)$, , where $U \sim U[0, 1]$. Provided that we match the modes of the distribution $a = \lambda$ and put $b = \sqrt{2\lambda}$, this function may be used to dominate the Poisson distribution and provide a simple rejection generator. The *Matlab* Poisson random number generator is $poissrnd(\lambda, m, n)$ which generates an $m \times n$ matrix of Poisson(λ) variables. This uses the simple generator **??** and is not computationally efficient for large values of λ.

The Binomial Distribution

For the *Binomial* distribution, we may use any one of the following alternatives:

(1) $\sum_{i=1}^{n} I(U_i < p), \quad U_i \sim uniform[0, 1]$

(2) $inf\{X; \sum_{i=1}^{X+1} G_i > n\}, \quad G_i \sim Geometric(p)$

(3) $inf\{X; \sum_{i=1}^{X+1} \frac{E_i}{n-i+1} > -log(1-p)\}, \quad E_i \sim Exponential(1).$

Methods (1) and (2) are more efficient when n is large but np fairly small. Of course for large n and np sufficiently small (e.g. <1), we can replace the Binomial distribution by its Poisson ($\lambda = np$) approximation. For large mean, a rejection method is required. Again we may use rejection beginning with a Lorentzian distribution, choosing $a = np$, $b = \sqrt{2np(1-p)}$ in the case $p < 1/2$. When $p > 1/2$, we simply reverse the roles of "failures" and "successes". Alternatively, a dominating table-mountain function may be used (Stadlober (1989)). The binomial generator in *Matlab* is the function *binornd(n,p,j,k)* which generates an $n \times k$ matrix of binomial(n, p) random variables. This uses the simplest form (1) of the binomial generator and is not computationally efficient for large n.

2.6 Simulating Stochastic Partial Differential Equations.

Consider a derivative product whose underlying asset has price X_t which satisfies a diffusion equation. Then if the derivative payoff function depends only on the current time and the current value of the asset, the Feynman- Kac theorem indicates that its value is an expectation of the form

$$V(S,t) = E[V_0(X_T)exp\{-\int_t^T r(X_v, v)dv\} \mid X_t = S] \tag{2.22}$$

where $r(X_t, t)$ is the current spot interest rate at time t. In most cases, this expectation is impossible to evaluate analytically and so we need to resort to numerical methods. If the spot interest rate is function of *both arguments* (X_v, v) and not just a function of time, then this integral is over the *whole joint distribution of the process* X_v, $0 < v < T$ and simple one-dimensional methods of numerical integration do not suffice. In such cases, we will usually resort to a careful simulation. The simplest version requires simulating a number of sample paths for the process X_v , evaluating **??** and averaging the results over all simulations. We begin by discussing the simulation of the process X_v.

Many of the stochastic models in finance reduce to simple diffusion equation (which may have more than one *factor* or dimension). Since most of the models in finance are Markovian, we restrict to the Markov diffusion model of the form

$$dX_t = a(X_t, t)dt + \sigma(X_t, t)dW_t \tag{2.23}$$

with initial value for X_0 where W_t is a driving standard Brownian motion process. Solving deterministic differential equations can sometimes provide a solution to a specific problem such as finding the arbitrage-free price of a derivative. In general, for more complex features of the derivative such as the distribution of return, important for considerations such as the *Value at Risk*, we need to obtain a solution $\{X_t,\ 0 < t < T\}$ to an equation of the above form which is a stochastic process. Typically this can only be done by simulation. One of the

simplest methods of simulating such a process is to adopt the crudest interpretation of the above equation, that is that a small increment $X_{t+h} - X_t$ in the process is approximately normally distributed with mean given by $a(X_t, t)h$ and variance given by $\sigma^2(X_t, t)h$. We generate these increments sequentially, beginning with an assumed value for X_0, and then adding to obtain an approximation to the value of the process at discrete times $t = 0, h, 2h, 3h, \ldots$. Between these discrete points, we can linearly interpolate the values. Approximating the process by assuming that the conditional distribution of $X_{t+h} - X_t$ is $N(a(X_t, t)h, \sigma^2(X_t, t)h)$ is called *Euler's method* by analogy to a simple method by the same name for solving ordinary differential equations. Given simulations of the process satisfying (3.2) together with some initial conditions, we might average the returns on a given derivative for many such simulations, (provided the process is expressed with respect to the risk-neutral distribution), to arrive at an arbitrage-free return for the derivative.

In this section we will discuss the numerical solution, or simulation of the solution to stochastic differential equations.

Letting $t_i = i\Delta x$, the equation **??** in integral form implies

$$X_{t_{i+1}} = X_{t_i} + \int_{t_i}^{t_{i+1}} a(X_s, s)ds + \int_{t_i}^{t_{i+1}} \sigma(X_s, s)dW_s \tag{2.24}$$

Ito's lemma can be written in terms of two operators on twice (with respect to x) differentiable functions f: In particular,

$$df(X_t, t) = L^o f dt + L^1 f dW_t \quad \text{where}$$

$$L^o = a\frac{\partial}{\partial x} + \frac{1}{2}\sigma^2\frac{\partial^2}{\partial x^2} + \frac{\partial}{\partial t},$$

and

$$L^1 = \sigma\frac{\partial}{\partial x}.$$

Then for any twice differentiable function f,

$$f(X_{t_{i+1}}, t_{i+1}) = f(X_{t_i}, t_i) + \int_{t_i}^{t_{i+1}} L^0 f(X_s, s)ds + \int_{t_i}^{t_{i+1}} L^1 f(X_s, s)dW_s. \tag{2.25}$$

By substituting in each of the integrands in **??** using the above identity and iterating this process we arrive at the Ito-Taylor expansions (e.g. Kloeden and Platen, 1992). For example,

$$\int_{t_i}^{t_{i+1}} a(X_s, s)ds = \int_{t_i}^{t_{i+1}} \{a(X_{t_i}, t_i) + \int_{t_i}^{s} L^0 a(X_u, u)du + \int_{t_i}^{s} L^1 a(X_u, u)dW_u\}ds$$

$$\approx a(X_{t_i},t_i)\Delta t + L^0 a(X_{t_i},t_i)\int_{t_i}^{t_{i+1}}\int_{t_i}^{s} duds + L^1 a(X_{t_i},t_i)\int_{t_i}^{t_{i+1}}\int_{t_i}^{s} dW_u ds \tag{2.26}$$

The first term in **??**, $a(X_{t_i})\Delta t$, is an initial approximation to the desired integral. The rest, we may regard as an error term for the moment, and it has smaller order. For example it is easy to see that the second term is $O(\Delta t)^2$ because the integral $\int_{t_i}^{t_{i+1}}\int_{t_i}^{s} duds = (\Delta t)^2/2$. The third term in **??** is $O(\Delta t)^{3/2}$ since $\int_{t_i}^{t_{i+1}}\int_{t_i}^{s} dW_u ds = \int_{t_i}^{t_{i+1}}(t_{i+1}-u)dW_u$ and this is a normal random variable with mean 0 and variance $\int_{t_i}^{t_{i+1}}(t_{i+1}-u)^2 du = (\Delta t)^3/3$. The simplest *Euler approximation* to the distribution of the increment assumes that ΔX has conditional mean $a(X_{t_i},t_i)\Delta t$. Similarly

$$\int_{t_i}^{t_{i+1}} \sigma(X_s,s)dW_s = \int_{t_i}^{t_{i+1}} \{\sigma(X_{t_i},t_i) + \int_{t_i}^{s} L^0\sigma(X_u,u)du + \int_{t_i}^{s} L^1\sigma(X_u,u)dW_u\}dW_s$$

$$\approx \sigma(X_{t_i},t_i)\Delta W_t + L^0\sigma(X_{t_i},t_i)\int_{t_i}^{t_{i+1}}\int_{t_i}^{s} dudW_s + L^1\sigma(X_{t_i},t_i)\int_{t_i}^{t_{i+1}}\int_{t_i}^{s} dW_u dW_s$$

$$= \sigma(X_{t_i},t_i)\Delta W_t + \frac{\sigma(X_{t_i},t_i)\frac{\partial}{\partial x}\sigma(X_{t_i},t_i)}{2}[(\Delta W_t)^2 - \Delta t] + O(\Delta t)^{3/2}$$

Since the integral $\int_{t_i}^{t_{i+1}}\int_{t_i}^{s} dW_u dW_s = \frac{1}{2}[(\Delta W_t)^2 - \Delta t]$. and $\int_{t_i}^{t_{i+1}}\int_{t_i}^{s} dudW_s = O(\Delta t)^{3/2}$. Putting these terms together, we arrive at an approximation to the increment of the form

$$\Delta X_t = a(X_{t_i},t_i)\Delta t + \sigma(X_{t_i},t_i)\Delta W_t + \frac{\sigma(X_{t_i},t_i)\frac{\partial}{\partial x}\sigma(X_{t_i},t_i)}{2}[(\Delta W_t)^2 - \Delta t] + O(\Delta t)^{3/2} \tag{2.27}$$

which allow an explicit representation of the increment in the process X in terms of the increment of a Brownian motion process $\Delta W_t \sim N(0,\Delta t)$. This is called the *Milstein approximation*. After the Euler scheme, it is the second Ito-Taylor approximation to a diffusion process. Obviously, the increments of the process are quadratic functions of a normal random variable and are no longer normal. The error approaches 0 at the rate $O(\Delta t)^{3/2}$ *in probability* only. This does not mean that the trajectory is approximated to this order but that the difference between the Milstein approximation to a diffusion and the diffusion itself is bounded in probability when divided by $\Delta t^{3/2}$ and as we let $\Delta t \to 0$. Higher order Taylor approximations are also possible, although they grow excessively complicated very quickly. See the book by Kloeden and Platten for details.

Example: Down-and-out-Call.

Consider an asset whose price follows a geometric Brownian motion process

$$dS_t = \mu S_t dt + \sigma S_t dW_t \tag{2.28}$$

for a standard Brownian motion process W_t . A down-and-out call option with exercise price E provides the usual payment of a call option if the asset never falls below a given *out barrier* b . Use simulation to price such an option with exercise price E , current asset price S , time to maturity $T - t$, and out barrier $b < S$. Assume constant interest rate r, and evaluate the expected return from the *risk neutral process*, the one with the drift term replaced by $rS_t dt$. Then discount this return to the present by multiplying by $e^{-r(T-t)}$. Compare with the Black-Scholes formula as $b \to 0$.

A geometric Brownian motion is most easily simulated by taking logarithms. For example if S_t satisfies the risk-neutral specification

$$dS_t = rS_t dt + \sigma S_t dW_t \tag{2.29}$$

then $Y_t = log(S_t)$ satisfies

$$dY_t = (r - \sigma^2/2)dt + \sigma dW_t. \tag{2.30}$$

This is a Brownian motion and is simulated with a normal random walk. Independent normal increments are generated $\Delta Y_t \sim N((r - \sigma^2/2)\Delta t,\ \sigma^2 \Delta t)$ and their partial sums used to simulate the process Y_t . The return for those options that are *in the money* is the average of the values of $(e^{Y_T} - E)^+$ over those paths for which $min\{Y_s; t < s < T\} \geq log(b)$. The following *Matlab* functions were used.

```
function z=barrier(r,sigma,dt,T, e,b)
n=T/dt;
sigma = sigma*sqrt(dt);
mn = r*dt - sigma^2/2;
y=exp(cumsum(normrnd(mn,sigma,n,1)));
x = (1:n).*dt;
plot(x,y,'-',x,e*ones(1,n),'y',x,b*ones(1,n),'b')
z = max(y(n)-e, 0);
if min(y) < b
  z=0;
end;

function z = simbarr(r, sigma, dt,T, e,b, m)
hold on
for i=1:m
   z(i) = barrier( r, sigma,dt,T,e,b);
end
disp([‘mean value=’  num2str(exp(-r*T)*sum(z)/m)])
hold off
```

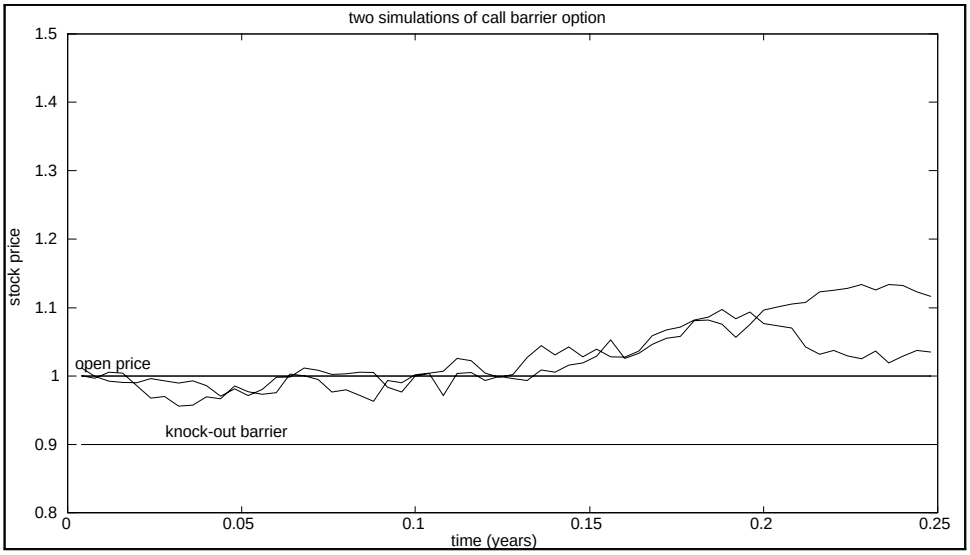


Figure 2.11:

For example when $r = .05, \sigma = .2, \Delta t = .004, T - t = .25, E = 1, b = .9$, this is run with the command

```
simbarr(.05, .2, .004, .25, 1, .9, 2)
```

for a total of 2 simulations, both, in this case, (see Figure **??**) ending in the money.

2.7 Problems

1. Consider the mixed generator $x_n = (ax_{n-1} + 1) mod(m)$ with $m = 64$. What values of a result in the maximum possible period and which generators appears more random?

2. Consider the multiplicative generator with $m = 64$. What values of a result in the maximum possible period and which generators appears more random?

3. Consider a shuffled described in section XXX with $k = 4, m_1 = 19, m_2 = 29$ X(i) = 2 14 7 13 11 18 6 3 9, Y(i) = 11 25 17 4 16 28 14 0 3 .

   We start by filling four pigeon-holes with the numbers produced by the first generator so that $(T_1, \ldots, T_4) = (2, 14, 7, 13)$. Then use the second generator to select a random index I telling us which pigeon-hole to draw the next number from. Since these holes are numberd from 1 through 4, we use $I = \lceil 4 \times 11/29 \rceil = 2$. Then the first number in our random

sequence is drawn from box 2, i.e. $z_1 = T_2 = 14$, so $z_1 = 14$. This element of the vector is now replaced by 11, the next number in the X sequence. Proceeding in this way, the next index is $I = \lceil 4 \times 25/29 \rceil = 4$ and so the next number drawn is $z_2 = T_4 = 13$. Of course, when we have finished generating the values $z_1, z_2, ...$ all of which lie between 0 and $m_1 = 19$, we will usually transform them in the usual way (e.g. z_i/m_1) to produce something approximating continuous uniform random numbers on [0,1]. Because of the small value we chose for m_1, this approximation will not be very good in this case. But the advantage of shuffling is that the period of the generator is greatly extended. Determine the period of the shuffled random number generator above and compare with the periods of the two constituent generators.

4. Prove: *If X is random variable uniform on the integers $\{0, \ldots, m-1\}$ and if Y is any integer-valued random variable independent of X, then the random variable $W = (X + Y)(mod\ m)$ is uniform on the integers* $\{0, \ldots, m-1\}$.

5. Consider the above quadratic residue generator $x_{n+1} = x_n^2 \bmod(m)$ with $m = 4783 \times 4027$. What is the period of the generator starting with seed $x_0 = 196$, or with seed $x_0 = 400$?

6. Verify that for the serial correlation statistic C_j, $var(C_j) = \frac{13}{144n}$, $j \geq 1$, $var(C_0) = 4/45n$.

7. Consider the turbo-pascal generator $x_{n+1} = (134775813x_n + 1) \bmod 2^{32}$. Generate a sequence of length 5000 and apply the serial correlation test. Is there evidence of dependence?

8. Generate 1000 daily "returns" $X_i, i = 1, 2, ..., 1000$ from each of the two distributions, the Cauchy and the logistic. In each case, assume that $a = 0, b = 0.1$. Graph the total return over an n day period versus n. Is there a qualitative difference in the two graphs? Repeat with a graph of the average daily return.

9. Briefly indicate an efficient algorithm for generating one random variable from each of the following distributions and generate such a random variable using one or more of the uniform[0,1] random numbers.

U_i:	*0.794*	*0.603*	*0.412*	*0.874*	*0.268*	*0.990*	*0.059*	*0.112*	*0.395*

   (a) X$\sim U[-1, 2]$.

   (b) a random variable X with probability density function $f(x) = \frac{3}{16}x^{1/2}$, $0 < x < 4$

   (c) A discrete random number X having probability function $P[X = x] = (1-p)^x p$, $x = 0, 1, 2, ..., p = 0.3$.

   (d) A random variable X with the normal distribution, mean 1 and *variance* 4.

(e) A random variable X with probability density function

$$f(x) = cx^2e^{-x}, \quad 0 \leq x < 1$$

for constant $c = 1/(2 - 5e^{-1})$.

(f) A random variable X with the following probability function:

x	0	1	2	3
$P[X = x]$	0.1	0.2	0.3	0.4

10. Consider the multiplicative pseudo-random number generator

$$x_{n+1} = ax_n \bmod(150)$$

starting with seed $x_0 = 7$. Try various values of the multiplier a and determine for which values the period of the generator appears to be maximal.

11. Consider the linear congruential generator

$$x_{n+1} = (ax_n + c) \bmod(2^8)$$

What is the maximal period that this generator can achieve when $c = 1$ and for what values of a does this seem to be achieved? Repeat when $c = 0$.

12. Evaluate the following integral by simulation:

$$\int_0^1 (1 - x^2)^{3/2} dx.$$

13. Let U be a uniform random variable on the interval [0,1]. Find a function of U which is uniformly distributed on the interval [0,2]. The interval $[a, b]$?

14. Evaluate the following integral by simulation:

$$\int_0^2 x^{3/2}(4 - x)^{1/2} dx.$$

15. Evaluate the following integral by simulation:

$$\int_{-\infty}^{\infty} e^{-x^2} dx.$$

(Hint: Rewrite this integral in the form $2\int_0^{\infty} e^{-x^2} dx$ and then change variables to $y = x/(1 + x)$)