Stochastic Processes and the Mathematics of Finance

Jonathan Block

April 1, 2008

Information for the class

Office: DRL3E2-A
Telephone: 215-898-8468
Office Hours: Tuesday 1:30-2:30, Thursday, 1:30-2:30.
Email: blockj@math.upenn.edu
References:

1. Financial Calculus, an introduction to derivative pricing, by Martin Baxter and Andrew Rennie.

2. The Mathematics of Financial Derivatives-A Student Introduction, by Wilmott, Howison and Dewynne.

3. A Random Walk Down Wall Street, Malkiel.

4. Options, Futures and Other Derivatives, Hull.

5. Black-Scholes and Beyond, Option Pricing Models, Chriss

6. Dynamic Asset Pricing Theory, Duffie

I prefer to use my own lecture notes, which cover exactly the topics that I want. I like very much each of the books above. I list below a little about each book.

1. Does a great job of explaining things, especially in discrete time.

2. Hull—More a book in straight finance, which is what it is intended to be. Not much math. Explains financial aspects very well. Go here for details about financial matters.

3. Duffie— This is a full fledged introduction into continuous time finance for those with a background in measure theoretic probability theory. Too advanced. But you might want to see how our course compares to a PhD level course in this material.

4. Wilmott, Howison and Dewynne—Immediately translates the issues into PDE. It is really a book in PDE. Doesn't really touch much on the probabilistic underpinnings of the subject.

Class grade will be based on homework and a take-home final. I will try to give homework every week and you will have a week to hand it in.

Syllabus

1. Probability theory. The following material will not be covered in class. I am assuming familiarity with this material (from Stat 430). I will hand out notes regarding this material for those of you who are rusty, or for those of you who have not taken a probability course but think that you can become comfortable with this material.
   (a) Probability spaces and random variables.
   (b) Basic probability distributions.
   (c) Expectation and variance, moments.
   (d) Bivariate distributions.
   (e) Conditional probability.
2. Derivatives.
   (a) What is a derivative security?
   (b) Types of derivatives.
   (c) The basic problem: How much should I pay for an option? Fair price.
   (d) Expectation pricing. Einstein and Bachelier, what they knew about derivative pricing. And what they didn't know.
   (e) Arbitrage and no arbitrage. The simple case of futures. Arbitrage arguments.
   (f) The arbitrage theorem.
   (g) Arbitrage pricing and hedging.
3. Discrete time stochastic processes and pricing models.
   (a) Binomial methods without much math. Arbitrage and reassigning probabilities.

  (b) A first look at martingales.
  (c) Stochastic processes, discrete in time.
  (d) Conditional expectations.
  (e) Random walks.
  (f) Change of probabilities.
  (g) Martingales.
  (h) Martingale representation theorem.
  (i) Pricing a derivative and hedging portfolios.
  (j) Martingale approach to dynamic asset allocation.

4. Continuous time processes. Their connection to PDE.
  (a) Wiener processes.
  (b) Stochastic integration..
  (c) Stochastic differential equations and Ito's lemma.
  (d) Black-Scholes model.
  (e) Derivation of the Black-Scholes Partial Differential Equation.
  (f) Solving the Black Scholes equation. Comparison with martingale method.
  (g) Optimal portfolio selection.

5. Finer structure of financial time series.

Chapter 1

Basic Probability

The basic concept in probability theory is that of a *random variable*. A random variable is a function of the basic outcomes in a probability space. To define a probability space one needs three ingredients:

1. A sample space, that is a set S of "outcomes" for some experiment. This is the set of all "basic" things that can happen. This set can be a discrete set (such as the set of 5-card poker hands, or the possible outcomes of rolling two dice) or it can be a continuous set (such as an interval of the real number line for measuring temperature, etc).

2. A sigma-algebra (or σ-algebra) of subsets of S. This means a set Ω of subsets of S (so Ω itself is a subset of the power set $\mathcal{P}(S)$ of all subsets of S) that contains the empty set, contains S itself, and is closed under countable intersections and countable unions. That is when $E_j \in \Omega$, for $j = 1, 2, 3, \ldots$ is a sequence of subsets in Ω, then

$$\bigcup_{j=1}^{\infty} E_j \in \Omega$$

and

$$\bigcap_{j=1}^{\infty} E_j \in \Omega$$

In addition, if $E \in \Omega$ then its complement is too, that is we assume $E^c \in \Omega$

The above definition is not central to our approach of probability but it is an essential part of the measure theoretic foundations of probability theory. So we include the definition for completeness and in case you come across the term in the future.

When the basic set S is finite (or countably infinite), then Ω is often taken to be all subsets of S. As a technical point (that you should probably ignore), when S is a continuous subset of the real line, this is not possible, and one usually restricts attention to the set of subsets that can be obtained by starting with all open intervals and taking intersections and unions (countable) of them – the so-called Borel sets [or more generally, Lebesgue measurable subsets of S). Ω is the collection of sets to which we will assign probabilities and you should think of them as the possible events. In fact, we will most often call them events.

3. A function P from Ω to the real numbers that assigns probabilities to events. The function P must have the properties that:

   (a) $P(S) = 1, P(\emptyset) = 0$.

   (b) $0 \leq P(A) \leq 1$ for every $A \in \Omega$

   (c) If $A_i, i = 1, 2, 3, \ldots$ is a countable (finite or infinite) collection of disjoint sets (i.e., $A_i \cap A_j = \emptyset$ for all i different from j), then $P(\cup A_i)) = \sum P(A_i)$.

These axioms imply that if A^c is the complement of A, then $P(A^c) = 1 - P(A)$, and the principle of inclusion and exclusion: $P(A \cup B) = P(A) + P(B) - P(A \cap B)$, even if A and B are not disjoint.

General probability spaces are a bit abstract and can be hard to deal with. One of the purposes of random variables, is to bring things back to something we are more familiar with. As mentioned above, a random variable is a function X on a probability space (S, Ω, P) is a function that to every outcome $s \in S$ gives a real number $X(s) \in \mathbb{R}$. For a subset $A \subset \mathbb{R}$ let's define $X^{-1}(A)$ to be the following subset of S:

$$X^{-1}(A) = \{s \in S | X(s) \in A\}$$

It may take a while to get used to what $X^{-1}(A)$ means, but do not think of X^{-1} as a function. In order for X^{-1} to be a function we would need X to be one-one, which we are not assuming. $X^{-1}(A)$ is just a subset of S as defined above. The point of this definition is that if $A \subset \mathbb{R}$ then $X^{-1}(A) \subset S$, that is $X^{-1}(A)$ is the event that X of the outcome will be in A. If for $A = (a, b) \subset \mathbb{R}$ an inverval, $X^{-1}(A) \in \Omega$ (this is a techinical assumption on X, called measurability, and should really be part of the definition of random variable.) then we can define $P_X(A) = P(X^{-1}(A))$.

Proposition 1.0.1. *P_X is a probability function on $\mathbb{R}$.*

So we have in some sense transferred the probability function to the real line. P_X is called the probability density of the random variable X.

In practice, it is the probability density that we deal with and mention of (S, Ω, P) is omitted. They are lurking somewhere in the background.

1.0.1 Discrete Random Variables

These take on only isolated (discrete) values, such as when you are counting something. Usually, but not always, the values of a discrete random variable X are (subset of the) integers, and we can assign a probability to any subset of the sample space, as soon as we know the probability of any set containing one element, i.e., $P(\{X = k\})$ for all k. It will be useful to write $p(k) = P(\{X = k\})$ — set $p(k) = 0$ for all numbers not in the sample space. $p(k)$ is called the probability density function (or pdf for short) of X. We repeat, for discrete random variables, the value $p(k)$ represents the probability that the event $\{X = k\}$ occurs. So any function from the integers to the (real) interval $[0, 1]$ that has the property that

$$\sum_{k=-\infty}^{\infty} p(k) = 1$$

defines a discrete probability distribution.

1.0.2 Finite Discrete Random Variables

1. Uniform distribution – This is often encountered, e.g., coin flips, rolls of a single die, other games of chance: S is the set of whole numbers from a to b (this is a set with $b - a + 1$ elements!), Ω is the set of all subsets of S, and P is defined by giving its values on all sets consisting of one element each (since then the rule for disjoint unions takes over

to calculate the probability on other sets). "Uniform" means that the same value is assigned to each one-element set. Since $P(S) = 1$, the value that must be assigned to each one element set is $1/(b - a + 1)$.

For example, the possible outcomes of rolling one die are $\{1\}, \{2\}, \{3\}, \{4\}, \{5\}$ and $\{6\}$. Each of these outcomes has the same probability, namely $1/6$. We can express this by making a table, or specifying a function $f(k) = 1/6$ for all $k = 1, 2, 3, 4, 5, 6$ and $f(k) = 0$ otherwise. Using the disjoint union rule, we find for example that $P(\{1, 2, 5\}) = 1/2$, $P(\{2, 3\}) = 1/3$, etc..

2. Binomial distribution – flip n fair coins, how many come up heads? i.e., what is the probability that k of them come up heads? Or do a sample in a population that favors the Democrat over the Republican 60 percent to 40 percent. What is the probability that in a sample of size 100, more than 45 will favor the Republican?

   The sample space S is $0, 1, 2, ...n$ since these are the possible outcomes (number of heads, number of people favoring the Republican [n=100 in this case]). As before, the sigma algebra Ω is the set of all subsets of S. The function P is more interesting this time:

$$P(\{k\}) = \frac{\binom{n}{k}}{2^n}$$

   where $\frac{n}{k}$ is the binomial coefficient

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}$$

   which equals the number of subsets of an n-element set that have exactly k elements.

1.0.3 Infinite Discrete Random Variables

3. Poisson distribution (with parameter α) – this arises as the number of (random) events of some kind (such as people lining up at a bank, or Geiger-counter clicks, or telephone calls arriving) per unit time. The sample space S is the set of all nonnegative integers $S = 0, 1, 2, 3, ...,$

and again Ω is the set of all subsets of S. The probability function on Ω is derived from:

$$P(\{k\}) = e^{-\alpha}\frac{\alpha^k}{k!}$$

Note that this is an honest probability function, since we will have

$$P(S) = P(\cup_{k=1}^{\infty}\{k\}) = \sum_{k=1}^{\infty} e^{-\alpha}\frac{\alpha^k}{k!} = e^{-\alpha}\sum_{k=1}^{\infty}\frac{\alpha^k}{k!} = e^{-\alpha}e^{\alpha} = 1$$

1.0.4 Continuous Random Variables

A continuous random variable can take on only any real values, such as when you are measuring something. Usually the values are (subset of the) reals, and for technical reasons, we can only assign a probability to certain subsets of the sample space (but there are a lot of them). These subsets, either the collection of Borel sets ((sets that can be obtained by taking countable unions and intersections of intervals)) or Lebesgue-measurable sets ((Borels plus a some other exotic sets)) comprise the set Ω. As soon as we know the probability of any interval, i.e., $P([a, b])$ for all a and b, we can calculate the probabililty of any Borel set. Recall that in the discrete case, the probabilities were determined once we knew the probabilities of individual outcomes. That is P was determined by $p(k)$. In the continuous case, the probability of a single outcome is always 0, $P(\{x\}) = 0$. However, it is enough to know the probabilities of "very small" intervals of the form $[x, x + dx]$, and we can calculate continuous probabilities as integrals of "probability density functions", so-called pdf's. We think of dx as a very small number and then taking its limit at 0. Thus, in the continuous case, the pdf is

$$p(x) = \lim_{dx\to 0}\frac{1}{dx}P([x, x + dx])$$

so that

$$P[a, b] = \int_a^b p(x)dx \tag{1.0.1}$$

This is really the defining equation of the continuous pdf and I can not stress too strongly how much you need to use this. Any function $p(x)$ that takes on only positive values (they don't have to be between 0 and 1 though), and whose integral over the whole sample space (we can use the whole real line if we assign the value $p(x) = 0$ for points x outside the sample space) is equal to 1 can be a pdf. In this case, we have (for small dx) that $p(x)dx$ represents (approximately) the probability of the set (interval) $[x, x + dx]$ (with error that goes to zero faster than dx does). More generally, we have the probability of the set (interval) $[a, b]$ is:

$$p([a, b]) = \int_a^b p(x)\, dx$$

So any nonnegative function on the real numbers that has the property that

$$\int_{-\infty}^{\infty} p(x)\, dx = 1$$

defines a continuous probability distribution.

Uniform distribution

As with the discrete uniform distribution, the variable takes on values in some interval $[a, b]$ and all variables are equally likely. In other words, all small intervals $[x, x + dx]$ are equally likely as long as dx is fixed and only x varies. That means that $p(x)$ should be a constant for x between a and b, and zero outside the interval $[a, b]$. What constant? Well, to have the integral of $p(x)$ come out to be 1, we need the constant to be $1/(b - a)$. It is easy to calculate that if $a < r < s < b$, then

$$p([r, s]) = \frac{s - r}{b - a}$$

The Normal Distribution (or Gaussian distribution)

This is the most important probability distribution, because the distribution of the average of the results of repeated experiments always approaches a normal distribution (this is the "central limit theorem"). The sample space for the normal distribution is always the entire real line. But to begin, we need to calculate an integral:

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

We will use a trick that goes back (at least) to Liouville: Now note that

$$\begin{aligned} I^2 &= \left(\int_{-\infty}^{\infty} e^{-x^2} dx\right) \cdot \left(\int_{-\infty}^{\infty} e^{-x^2} dx\right) && (1.0.2) \\ &= \left(\int_{-\infty}^{\infty} e^{-x^2} dx\right) \cdot \left(\int_{-\infty}^{\infty} e^{-y^2} dy\right) && (1.0.3) \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{-x^2-y^2} dxdy && (1.0.4) \end{aligned}$$

because we can certainly change the name of the variable in the second integral, and then we can convert the product of single integrals into a double integral. Now (the critical step), we'll evaluate the integral in polar coordinates (!!) – note that over the whole plane, r goes from 0 to infinity as θ goes from 0 to 2π, and $dxdy$ becomes $rdrd\theta$:

$$I^2 = \int_0^{2\pi} \int_0^{\infty} e^{-r^2} rdrd\theta = \int_0^{2\pi} \frac{1}{2} e^{-r^2}|_0^{\infty} d\theta = \int_0^{2\pi} \frac{1}{2} d\theta = \pi$$

Therefore, $I = \sqrt{\pi}$. We need to arrange things so that the integral is 1, and for reasons that will become apparent later, we arrange this as follows: define

$$N(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$$

Then $N(x)$ defines a probability distribution, called the standard normal distribution. More generally, we define the normal distribution with parameters μ and σ to be

$$p(x) = \frac{1}{\sigma} N(\frac{x-\mu}{\sigma}) = \frac{1}{\sqrt{2\pi}\sigma} e^{\frac{-(x-\mu)^2}{2\sigma^2}}$$

Exponential distribution (with parameter α)

This arises when measuring waiting times until an event, or time-to-failure in reliability studies. For this distribution, the sample space is the positive part of the real line $[0, \infty)$ (or we can just let $p(x) = 0$ for $x < 0$). The probability density function is given by $p(x) = \alpha e^{-\alpha x}$. It is easy to check that the integral of $p(x)$ from 0 to infinity is equal to 1, so $p(x)$ defines a bona fide probability density function.

1.1 Expectation of a Random Variable

The expectation of a random variable is essentially the average value it is expected to take on. Therefore, it is calculated as the weighted average of the possible outcomes of the random variable, where the weights are just the probabilities of the outcomes. As a trivial example, consider the (discrete) random variable X whose sample space is the set $\{1, 2, 3\}$ with probability function given by $p(1) = 0.3, p(2) = 0.1$ and $p(3) = 0.6$. If we repeated this experiment 100 times, we would expect to get about 30 occurrences of $X = 1$, 10 of $X = 2$ and 60 of $X = 3$. The average X would then be $((30)(1) + (10)(2) + (60)(3))/100 = 2.3$. In other words, $(1)(0.3) + (2)(0.1) + (3)(0.6)$. This reasoning leads to the defining formula:

$$E(X) = \sum_{x \in S} x p(x)$$

for any discrete random variable. The notation $E(X)$ for the expectation of X is standard, also in use is the notation $\langle X \rangle$.

For continuous random variables, the situation is similar, except the sum is replaced by an integral (think of summing up the average values of x by dividing the sample space into small intervals $[x, x + dx]$ and calculating the probability $p(x)dx$ that X fall s into the interval. By reasoning similar to the previous paragraph, the expectation should be

$$E(X) = \lim_{dx \to 0} \sum_{x \in S} x p(x) dx = \int_S x p(x) dx.$$

This is the formula for the expectation of a continuous random variable.

Example 1.1.1. :

1. Uniform discrete: We'll need to use the formula for $\sum k$ that you

learned in freshman calculus when you evaluated Riemann sums:

$$\begin{aligned}
E(X) &= \sum_{k=a}^{b} k \cdot \frac{1}{b-a+1} && (1.1.1)\\
&= \frac{1}{b-a+1}\left(\sum_{k=0}^{b} k - \sum_{k=0}^{a-1} k\right) && (1.1.2)\\
&= \frac{1}{b-a+1}\left(\frac{b(b+1)}{2} - \frac{(a-1)a}{2}\right) && (1.1.3)\\
&= \frac{1}{2}\frac{1}{b-a+1}(b^2+b-a^2+a) && (1.1.4)\\
&= \frac{1}{2}\frac{1}{b-a+1}(b-a+1)(b+a) && (1.1.5)\\
&= \frac{b+a}{2} && (1.1.6)
\end{aligned}$$

which is what we should have expected from the uniform distribution.

2. Uniform continuous: (We expect to get $(b+a)/2$ again, right?). This is easier:

$$\begin{aligned}
E(X) &= \int_a^b x\frac{1}{b-a}dx && (1.1.7)\\
&= \frac{1}{b-a}\frac{b^2-a^2}{2} = \frac{b+a}{2} && (1.1.8)
\end{aligned}$$

3. Poisson distribution with parameter α: Before we do this, recall the Taylor series formula for the exponential function:

$$e^x = \sum_{k=0}^{\infty} \frac{x^k}{k!}$$

Note that we can take the derivative of both sides to get the formula:

$$e^x = \sum_{k=0}^{\infty} \frac{kx^{k-1}}{k!}$$

If we multiply both sides of this formula by x we get

$$xe^x = \sum_{k=0}^{\infty} \frac{kx^k}{k!}$$

We will use this formula with x replaced by α.

If X is a discrete random variable with a Poisson distribution, then its expectation is:

$$E(X) = \sum_{k=0}^{\infty} k \frac{e^{-\alpha}\alpha^k}{k!} = e^{-\alpha}\left(\sum_{k=0}^{\infty} \frac{k\alpha^k}{k!}\right) = e^{-\alpha}\left(\alpha e^{\alpha}\right) = \alpha.$$

4. Exponential distribution with parameter α. This is a little like the Poisson calculation (with improper integrals instead of series), and we will have to integrate by parts (we'll use $u = x$ so $du = dx$, and $dv = e^{-\alpha x}dx$ so that v will be $-\frac{1}{\alpha}e^{-\alpha x}$):

$$E(X) = \int_0^{\infty} x\alpha e^{-\alpha x}dx = \alpha\left(-\frac{1}{\alpha}xe^{-\alpha x} - \frac{1}{\alpha^2}e^{-\alpha x}\right)\Big|_0^{\infty} = \frac{1}{\alpha}.$$

Note the difference between the expectations of the Poisson and exponential distributions!!

5. By the symmetry of the respective distributions around their "centers", it is pretty easy to conclude that the expectation of the binomial distribution (with parameter n) is n/2, and the expectation of the normal distribution (with parameters μ and σ) is μ.

6. Geometric random variable (with parameter r): It is defined on the sample space $\{0, 1, 2, 3, ...\}$ of non-negative integers and, given a fixed value of r between 0 and 1, has probability function given by

$$p(k) = (1-r)r^k$$

The geometric series formula:

$$\sum_{k=0}^{\infty} ar^k = \frac{a}{1-r}$$

shows that $p(k)$ is a probability function. We can compute the expectation of a random variable X having a geometric distribution with parameter r as follows (the trick is reminiscent of what we used on the Poisson distribution):

$$E(X) = \sum_{k=0}^{\infty} k(1-r)r^k = (1-r)r\sum_{k=0}^{\infty} kr^{k-1}$$

$$= (1-r)r\frac{d}{dr}\sum_{k=0}^{\infty} r^k = (1-r)r\frac{d}{dr}\frac{1}{1-r} = \frac{r}{1-r}$$

7. Another distribution defined on the positive integers $\{1, 2, 3, ...\}$ by the function $p(k) = \frac{6}{\pi^2 k^2}$. This is a probability function based on the famous formula due to Euler :

$$\sum_{k=1}^{\infty} \frac{1}{k^2} = \frac{\pi^2}{6}$$

But an interesting thing about this distribution concerns its expectation: Because the harmonic series diverges to infinity, we have that

$$E(X) = \frac{6}{\pi^2}\sum_{k=1}^{\infty}\frac{k}{k^2} = \frac{6}{\pi^2}\sum_{k=1}^{\infty}\frac{1}{k} = \infty$$

So the expectation of this random variable is infinite (Can you interpret this in terms of an experiment whose outcome is a random variable with this distribution?)

If X is a random variable (discrete or continuous), it is possible to define related random variables by taking various functions of X, say X^2 or $\sin(X)$ or whatever.

If $Y = f(X)$ for some function f, then the probability of Y being in some set A is defined to be the probability of X being in the set $f^{-1}(A)$.

As an example, consider an exponentially distributed random variable X with parameter α=1. Let $Y = X^2$. Since X can only be positive, the probability that Y is in the interval $[a, b]$ is the same as the probability that X is in the interval $[\sqrt{a}, \sqrt{b}]$.

We can calculate the probability density function $p(y)$ of Y by recalling that the probability that Y is in the interval $[0, y]$ (actually $(-\infty, y]$) is the integral of $p(y)$ from 0 to y. In other words, $p(y)$ is the integral of the function $h(y)$, where $h(y)$ = the probability that Y is in the interval $[0, y]$. But $h(y)$ is the same as the probability that X is in the interval $[0, \sqrt{y}]$. We calculate:

$$p(y) = \frac{d}{dy}\int_0^{\sqrt{y}} e^{-x}dx = \frac{d}{dy}\left(-e^{-x}|_0^{\sqrt{y}}\right) = \frac{d}{dy}(1-e^{-\sqrt{y}}) = \frac{1}{2\sqrt{y}}e^{-\sqrt{y}}$$

There are two ways to calculate the expectation of Y. The first is obvious: we can integrate $yp(y)$. The other is to make the change of variables $y = x^2$ in this integral, which will yield (Check this!) that the expectation of $Y = X^2$ is

$$E(Y) = E(X^2) = \int_0^\infty x^2 e^{-x} dx$$

Theorem 1.1.2. (Law of the unconscious statistician) If $f(x)$ is any function, then the expectation of the function $f(X)$ of the random variable X is

$$E(f(X)) = \int f(x)p(x)dx \qquad \text{or} \qquad \sum_x f(x)p(x)$$

where $p(x)$ is the probability density function of X if X is continuous, or the probability density function of X if X is discrete.

1.2 Variance and higher moments

We are now lead more or less naturally to a discussion of the "moments" of a random variable.

The r^{th} *moment* of X is defined to be the expected value of X^r. In particular the first moment of X is its expectation. If $s > r$, then having a finite s^{th} moment is a more restrictive condition than having and r^{th} one (this is a convergence issue as x approaches infinity, since $x^r > x^s$ for large values of x.

A more useful set of moments is called the set of *central moments*. These are defined to be the rth moments of the variable $X - E(X)$. In particular, the second moment of $X - E(X)$ is called the variance of X and is denoted $\sigma^2(X)$. (Its square root is the standard deviation.) It is a crude measure of the extent to which the distribution of X is spread out from its expectation value. It is a useful exercise to work out that

$$\sigma^2(X) = E((X - E(X))^2) = E(X^2) - (E(X))^2.$$

As an example, we compute the variance of the uniform and exponential distributions:

1. Uniform discrete: If X has a discrete uniform distribution on the interval $[a, b]$, then recall that $E(X) = (b + a)/2$. We calculate $\mathrm{Var}(X)$ as follows:

$$\mathrm{Var}(X) = E(X^2) - \left(\frac{b+a}{2}\right)^2 = \sum_{k=a}^{b} \frac{k^2}{b-a+1} - \left(\frac{b+a}{2}\right)^2 = \frac{1}{12}(b-a+2)(b-a).$$

2. Uniform continuous: If X has a continuous uniform distribution on the interval $[a, b]$, then its variance is calculated as follows:

$$\sigma^2(X) = \int_a^b \frac{x^2}{b-a}dx - \left(\frac{b+a}{2}\right)^2 = \frac{1}{12}(b-a)^2.$$

3. Exponential with parameter α: If X is exponentially distributed with parameter α, recall that $E(X) = 1/\alpha$. Thus:

$$\text{Var}(x) = \int_0^\infty x^2 \alpha e^{-\alpha x} dx - \frac{1}{\alpha^2} = \frac{1}{\alpha^2}$$

The variance decreases as α increases; this agrees with intuition gained from the graphs of exponential distributions shown last time.

1.3 Bivariate distributions

It is often the case that a random experiment yields several measurements (we can think of the collection of measurements as a random vector) – one simple example would be the numbers on the top of two thrown dice. When there are several numbers like this, it is common to consider each as a random variable in its own right, and to form the joint probability density (or joint probability) function $p(x, y, z, ...)$. For example, in the case of two dice, X and Y are discrete random variables each with sample space $S = \{1, 2, 3, 4, 5, 6\}$, and $p(x, y) = 1/36$ for each (x, y) in the cartesian product $S\times$S. More generally, we can consider discrete probability functions $p(x, y)$ on sets of the form $S \times T$, where X ranges over S and Y over T. Any function $p(x, y)$ such that $p(x, y)$ is between 0 and 1 for all (x, y) and such that

$$\sum_{x\in S}\sum_{y\in T} p(x, y) = 1$$

defines a (joint) probability function on $S \times T$.

For continuous functions, the idea is the same. $p(x, y)$ is the joint pdf of X and Y if $p(x, y)$ is non-negative and satisfies

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} p(x, y)dxdy = 1$$

We will use the following example of a joint pdf throughout the rest of this section:

X and Y will be random variables that can take on values (x, y) in the triangle with vertices $(0, 0)$, $(2, 0)$ and $(2, 2)$. The joint pdf of X and Y will be given by $p(x, y) = 1/(2x)$ if (x, y) is in the triangle and 0 otherwise. To see that this is a probability density function, we need to integrate $p(x, y)$ over the triangle and get 1:

$$\int_0^2 \int_0^x \frac{1}{2x} dy dx = \int_0^2 \left(\frac{y}{2x} \right) |_0^x dx = \int_0^2 \frac{1}{2} dx = 1.$$

For discrete variables, we let $p(i, j)$ be the probability that $X = i$ and $Y = j$, $P(X = i \cap Y = j)$. This give a function p, the joint probability function, of X and Y that is defined on (some subset of) the set of pairs of integers and such that $0 \le p(i, j) \le 1$ for all i and j and

$$\sum_{i=-\infty}^{\infty} \sum_{j=-\infty}^{\infty} p(i, j) = 1$$

When we find it convenient to do so, we will set $p(i, j) = 0$ for all i and j outside the domain we are considering.

For continuous variables, we define the joint probability density function $p(x, y)$ on (some subset of) the plane of pairs of real numbers. We interpret the function as follows: $p(x, y)dxdy$ is (approximately) the probability that X is between x and $x + dx$ and Y is between y and $y + dy$ (with error that goes to zero faster than dx and dy as they both go to zero). Thus, $p(x, y)$ must be a non-negative valued function with the property that

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} p(x, y) \, dxdy = 1$$

As with discrete variables, if our random variables always lie in some subset of the plane, we will define $p(x, y)$ to be 0 for all (x, y) outside that subset.

We take one simple example of each kind of random variable. For the discrete random variable, we consider the roll of a pair of dice. We assume that we can tell the dice apart, so there are thirty-six possible outcomes and each is equally likely. Thus our joint probability function will be

$$p(i, j) = 1/36 \quad \text{if} \quad 1 \le i \le 6, \ 1 \le j \le 6$$

and $p(i,j) = 0$ otherwise.

For our continuous example, we take the example mentioned at the end of the last lecture:

$$p(x,y) = \frac{1}{2x}$$

for (x,y) in the triangle with vertices $(0,0)$, $(2,0)$ and $(2,2)$, and $p(x,y) = 0$ otherwise. We checked last time that this is a probability density function (its integral is 1).

1.4 Marginal distributions

Often when confronted with the joint probability of two random variables, we wish to restrict our attention to the value of just one or the other. We can calculate the probability distribution of each variable separately in a straightforward way, if we simply remember how to interpret probability functions. These separated probability distributions are called the marginal distributions of the respective individual random variables.

Given the joint probability function $p(i,j)$ of the discrete variables X and Y, we will show how to calculate the marginal distributions $p_X(i)$ of X and $p_Y(j)$ of Y. To calculate $p_X(i)$, we recall that $p_X(i)$ is the probability that $X = i$. It is certainly equal to the probability that $X = i$ and $Y = 0$, or $X = i$ and $Y = 1$, or .... In other words the event $X = i$ is the union of the events $X = i$ and $Y = j$ as j runs over all possible values. Since these events are disjoint, the probability of their union is the sum of the probabilities of the events (namely, the sum of $p(i,j)$). Thus:

$$p_X(i) = \sum_{j=-\infty}^{\infty} p(i,j)$$

. Likewise,

$$p_Y(j) = \sum_{i=-\infty}^{\infty} p(i,j)$$

Make sure you understand the reasoning behind these two formulas!

An example of the use of this formula is provided by the roll of two dice discussed above. Each of the 36 possible rolls has probability $1/36$ of occurring, so we have probability function $p(i, j)$ as indicated in the following table:

$i \backslash j$	1	2	3	4	5	6	$p_X(i)$
1	1/36	1/36	1/36	1/36	1/36	1/36	1/6
2	1/36	1/36	1/36	1/36	1/36	1/36	1/6
3	1/36	1/36	1/36	1/36	1/36	1/36	1/6
4	1/36	1/36	1/36	1/36	1/36	1/36	1/6
5	1/36	1/36	1/36	1/36	1/36	1/36	1/6
6	1/36	1/36	1/36	1/36	1/36	1/36	1/6
$p_Y(j)$	1/6	1/6	1/6	1/6	1/6	1/6	

The marginal probability distributions are given in the last column and last row of the table. They are the probabilities for the outcomes of the first (resp second) of the dice, and are obtained either by common sense or by adding across the rows (resp down the columns).

For continuous random variables, the situation is similar. Given the joint probability density function $p(x, y)$ of a bivariate distribution of the two random variables X and Y (where $p(x, y)$ is positive on the actual sample space subset of the plane, and zero outside it), we wish to calculate the marginal probability density functions of X and Y. To do this, recall that $p_X(x)dx$ is (approximately) the probability that X is between x and $x + dx$. So to calculate this probability, we should sum all the probabilities that both X is in $[x, x + dx]$ and Y is in $[y, y + dy]$ over all possible values of Y. In the limit as dy approaches zero,this becomes an integral:

$$p_X(x)dx = \left(\int_{-\infty}^{\infty} p(x, y)\, dy \right) dx$$

In other words,

$$p_X(x) = \int_{-\infty}^{\infty} p(x, y)\, dy$$

Similarly,

$$p_Y(y) = \int_{-\infty}^{\infty} p(x, y)\, dx$$

Again, you should make sure you understand the intuition and the reasoning behind these important formulas.

We return to our example:

$$p(x, y) = \frac{1}{2x}$$

for (x, y) in the triangle with vertices $(0,0), (2,0)$ and $(2,2)$, and $p(x,y) = 0$ otherwise, and compute its marginal density functions. The easy one is $p_X(x)$ so we do that one first. Note that for a given value of x between 0 and 2, y ranges from 0 to x inside the triangle:

$$p_X(x) = \int_{-\infty}^{\infty} p(x,y)\,dy = \int_0^x \frac{1}{2x}\,dy$$

$$= \left(\frac{y}{2x}\right)\Big|_{y=0}^{y=x} = \frac{1}{2}$$

if $0 \le x \le 2$, and $p_X(x) = 0$ otherwise. This indicates that the values of X are uniformly distributed over the interval from 0 to 2 (this agrees with the intuition that the random points occur with greater density toward the left side of the triangle but there is more area on the right side to balance this out).

To calculate $p_Y(y)$, we begin with the observation that for each value of y between 0 and 2, x ranges from y to 2 inside the triangle:

$$p_Y(y) = \int_{-\infty}^{\infty} p(x,y)\,dx = \int_x^2 \frac{1}{2x}\,dx$$

$$= \left(\frac{1}{2}\ln(x)\right)\Big|_{x=y}^{x=2} = \frac{1}{2}(\ln(2) - \ln(y))$$

if $0 \le y \le 2$ and $p_Y(y) = 0$ otherwise. Note that $p_Y(y)$ approaches infinity as y approaches 0 from above, and $p_Y(y)$ approaches 0 as y approaches 2. You should check that this function is actually a probability density function on the interval $[0,2]$, i.e., that its integral is 1.

1.5 Functions of two random variables

Frequently, it is necessary to calculate the probability (density) function of a function of two random variables, given the joint probability (density) function. By far, the most common such function is the sum of two random variables, but the idea of the calculation applies in principle to any function of two (or more!) random variables.

The principle we will follow for discrete random variables is as follows: to calculate the probability function for $F(X,Y)$, we consider the events

$\{F(X,Y) = f\}$ for each value of f that can result from evaluating F at points of the sample space of (X,Y). Since there are only countably many points in the sample space, the random variable F that results is discrete. Then the probability function $p_F(f)$ is

$$p_F(f) = \sum_{\{(x,y)\in S | F(x,y)=f\}} p(x,y)$$

This seems like a pretty weak principle, but it is surprisingly useful when combined with a little insight (and cleverness).

As an example, we calculate the distribution of the sum of the two dice. Since the outcome of each of the dice is a number between 1 and 6, the outcome of the sum must be a number between 2 and 12. So for each f between 2 and 12:

$$p_F(f) = \sum_{x=-\infty}^{\infty} p(x, f-x) = \sum_{\max(1,f-6)}^{\min(6,f-1)} \frac{1}{36} = \frac{1}{36}(6 - |f-7|)$$

A table of the probabilities of various sums is as follows:

f	2	3	4	5	6	7	8	9	10	11	12
$p_F(f)$	1/36	2/36	3/36	4/36	5/36	6/36	5/36	4/36	3/36	2/36	1/36

The "tent-shaped" distribution that results is typical of the sum of (independent) uniformly distributed random variables.

For continuous distributions, our principle will be a little more complicated, but more powerful as well. To enunciate it, we recall that to calculate the probability of the event $F < f$, we integrate the pdf of F from $-\infty$ to f:

$$P(F < f) = \int_{-\infty}^{f} p_F(g)\, dg$$

Conversely, to recover the pdf of F, we can differentiate the resulting function:

$$p_F(f) = \frac{d}{df} \int_{-\infty}^{f} p_F(g)\, dg$$

(this is simply the first fundamental theorem of calculus). Our principle for calculating the pdf of a function of two random variables $F(X,Y)$ will be to calculate the probabilities of the events $\{F(X,Y) < f\}$ (by integrating the joint pdf over the region of the plane defined by this inequality), and then to differentiate with respect to f to get the pdf.

We apply this principle to calculate the pdf of the sum of the random variables X and Y in our example:

$$p(x,y) = \frac{1}{2x}$$

for (x,y) in the triangle T with vertices $(0,0),(2,0)$ and $(2,2)$, and $p(x,y)=0$ otherwise. Let $Z = X+Y$. To calculate the pdf $p_Z(z)$, we first note that for any fixed number z, the region of the plane where $Z < z$ is the half plane below and to the left of the line $y = z - x$. To calculate the probability $P(Z < z)$, we must integrate the joint pdf $p(x,y)$ over this region. Of course, for $z \leq 0$, we get zero since the half plane $z < 0$ has no points in common with the triangle where the pdf is supported. Likewise, since both X and Y are always between 0 and 2 the biggest the sum can be is 4. Therefore $P(Z < z) = 1$ for all $z \geq 4$.

For z between 0 and 4, we need to integrate $1/2x$ over the intersection of the half-plane $x+y<z$ and the triangle T. The shape of this intersection is different, depending upon whether z is greater than or less than 2: If $0 \leq z \leq 2$, the intersection is a triangle with vertices at the points $(0,0),(z/2,z/2)$ and $(z,0)$. In this case, it is easier to integrate first with respect to x and then with respect to y, and we can calculate:

$$P(Z<z) = \int_0^{z/2}\int_y^{z-y} \frac{1}{2x}\,dxdy = \frac{1}{2}\int_0^{z/2}\left(\ln(x)|_{x=y}^{x=z-y}\right)dy$$

$$= \frac{1}{2}\int_0^{z/2}(\ln(z-y)-\ln(y))\,dy = \frac{1}{2}(z-(z-y)\ln(z-y)-y\ln(y))|_{y=0}^{y=z/2}$$

$$= \frac{1}{2}z\ln(2)$$

And since the (cumulative) probability that $Z<z$ is $\frac{1}{2}z\ln(2)$ for $0<z<2$, the pdf over this range is $p_Z(z) = \frac{\ln(2)}{2}$.

The calculation of the pdf for $2 \leq z \leq 4$ is somewhat trickier because the intersection of the half-plane $x+y<z$ and the triangle T is more complicated. The intersection in this case is a quadrilateral with vertices at the points $(0,0)$, $(z/2,z/2)$, $(2,z-2)$ and $(2,0)$. We could calculate $P(Z<z)$ by integrating p(x,y) over this quadrilateral. But we will be a

little more clever: Note that the quadrilateral is the "difference" of two sets. It consists of points inside the triangle with vertices $(0,0)$, $(z/2, z/2)$, $(z,0)$ that are to the left of the line $x = 2$. In other words it is points inside this large triangle (and note that we already have computed the integral of $1/2x$ over this large triangle to be $\frac{1}{2}z\ln(2)$) that are NOT inside the triangle with vertices $(2,0)$, $(2, z-2)$ and $(z,0)$. Thus, for $2 \leq z \leq 4$, we can calculate $P(Z < z)$ as

$$\begin{aligned} P(Z < z) &= \frac{1}{2}z\ln(2) - \int_2^z \int_0^{z-x} \frac{1}{2x}\, dydx \\ &= \frac{1}{2}z\ln(2) - \int_2^z \left(\frac{y}{2x}|_{y=0}^{y=z-x} \right) dx \\ &= \frac{1}{2}z\ln(2) - \int_2^z \frac{z-x}{2x}\, dx = \frac{1}{2}\ln(2) + \frac{1}{2}(z\ln(x) - x)|_{x=2}^{x=z} \\ &= \frac{1}{2}z\ln(2) - \frac{1}{2}(z\ln(z) - z - z\ln(2) + 2) \\ &= z\ln(2) - \frac{1}{2}z\ln(z) + \frac{1}{2}z - 1 \end{aligned}$$

To get the pdf for $2 < z < 4$, we need only differentiate this quantity, to get

$$p_Z(z) = \frac{d}{dz}(z\ln(2) - \frac{1}{2}z\ln(z) + \frac{1}{2}z - 1) = \ln(2) - \frac{1}{2}\ln(z)$$

Now we have the pdf of $Z = X + Y$ for all values of z. It is $p_Z(z) = \ln(2)/2$ for $0 < z < 2$, it is $p_Z(z) = \ln(2) - \frac{1}{2}\ln(z)$ for $2 < z < 4$ and it is 0 otherwise. It would be good practice to check that the integral of $p_Z(z)$ is 1.

The following fact is extremely useful.

Proposition 1.5.1. *If X and Y are two random variables, then $E(X+Y) = E(X) + E(Y)$.*

1.6 Conditional Probability

In our study of stochastic processes, we will often be presented with situations where we have some knowledge that will affect the probability of whether some event will occur. For example, in the roll of two dice, suppose we already know that the sum will be greater than 7. This changes the probabilities from those that we computed above. The event $F \geq 8$ has probability

$(5+4+3+2+1)/36 = 15/36$. So we are restricted to less than half of the original sample space. We might wish to calculate the probability of getting a 9 under these conditions. The quantity we wish to calculate is denoted $P(F=9|F\geq 8)$, read "the probability that F=9 given that $F \geq 8$".

In general to calculate $P(A|B)$ for two events A and B (it is not necessary that A is a subset of B), we need only realize that we need to compute the fraction of the time that the event B is true, it is also the case the A is true. In symbols, we have

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

For our dice example (noting that the event $F = 9$ is a subset of the event $F \geq 8$), we get

$$P(F=9|F\geq 8) = \frac{P(F=9)}{P(F\geq 8)} = \frac{4/36}{15/36} = \frac{4}{15}.$$

As another example (with continuous probability this time), we calculate for our $1/2x$ on the triangle example the conditional probabilities: $P(X > 1|Y > 1)$ as well as $P(Y > 1|X > 1)$ (just to show that the probabilities of A given B and B given A are usually different).

First $P(X > 1|Y > 1)$. This one is easy! Note that in the triangle with vertices $(0,0)$, $(2,0)$ and $(2,2)$ it is true that $Y > 1$ implies that $X > 1$. Therefore the events $(Y > 1) \cap (X > 1)$ and $Y > 1$ are the same, so the fraction we need to compute will have the same numerator and denominator. Thus $P(X > 1|Y > 1) = 1$.

For $P(Y > 1|X > 1)$ we actually need to compute something. But note that $Y > 1$ is a subset of the event $X > 1$ in the triangle, so we get:

$$P(Y>1|X>1) = \frac{P(Y>1)}{P(X>1)} = \frac{\int_1^2 \int_1^x \frac{1}{2x}\,dydx}{\int_1^2 \int_0^x \frac{1}{2x}\,dydx} = \frac{\frac{1}{2}(1-\ln(2))}{\frac{1}{2}} = 1 - \ln(2)$$

Two events A and B are called independent if the probability of A given B is the same as the probability of A (with no knowledge of B) and vice versa. The assumption of independence of certain events is essential to many probabilistic arguments. Independence of two random variables is expressed by the equations:

$$P(A|B) = P(A) \qquad P(B|A) = P(B)$$

and especially

$$P(A \cap B) = P(A) \cdot P(B)$$

Two random variables X and Y are independent if the probability that $a < X < b$ remains unaffected by knowledge of the value of Y and vice versa. This reduces to the fact that the joint probability (or probability density) function of X and Y "splits" as a product:

$$p(x, y) = p_X(x)p_Y(y)$$

of the marginal probabilities (or probability densities). This formula is a straightforward consequence of the definition of independence and is left as an exercise.

Proposition 1.6.1. *1. If X is a random variable, then $\sigma^2(aX) = a^2\sigma^2(X)$ for any real number a.*

2. If X and Y are two independent random variables, then $\sigma^2(X + Y) = \sigma^2(X) + \sigma^2(Y)$.

1.7 Law of large numbers

The point of probability theory is to organize the random world. For many centuries, people have noticed that, though certain measurements or processes seem random, one can give a quantitative account of this randomness. Probability gives a mathematical framework for understanding statements like the chance of hitting 13 on a roulette wheel is 1/38. This belief about the way the random world works is embodied in the empirical law of averages. This law is not a mathematical theorem. It states that if a random experiment is performed repeatedly under identical and independent conditions, then the proportion of trials that result in a given outcome converges to a limit as the number of trials goes to infinity. We now state for future reference the strong law of large numbers. This theorem is the main rigorous result giving credence to the empirical law of averages.

Theorem 1.7.1. (The strong law of large numbers) Let $X_1, X_2, \ldots$ be a sequence of independent random variables, all with the same distribution, whose expectation is μ. Then

$$P(\lim_{n\to\infty} \frac{\sum_{i=1}^n X_i}{n} \to \mu) = 1$$

Exercises

1. Make your own example of a probability space that is finite and discrete. Calculate the expectation of the underlying random variable X.

2. Make your own example of a probability space that is infinite and discrete. Calculate the expectation of the underlying random variable X.

3. Make your own example of a continuous random variable. Calculate its expectation.

4. Prove that the normal distribution function

$$\phi(x; \mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-(x-\mu)^2}{2\sigma^2}}$$

is really a probability density function on the real line (i.e., that it is positive and that its integral from $-\infty$ to ∞ is 1). Calculate the expectation of this random variable.

5. What is the relationship among the following partial derivatives? $\frac{\partial\phi}{\partial\sigma}$, $\frac{\partial^2\phi}{\partial x^2}$, and $\frac{\partial\phi}{\partial t}$, where $t = \sigma^2$ (for the last one, rewrite ϕ as a function of x, μ and t).

6. Consider the experiment of picking a point at random from a uniform distribution on the disk of radius R centered at the origin in the plane ("uniform" here means if two regions of the disk have the same area, then the random point is equally likely to be in either one). Calculate the probability density function and the expectation of the random variable D, defined to be the distance of the random point from the origin.

7. Prove the principle of inclusion and exclusion, $p(A \cup B) = p(A) + p(B) - p(A \cap B)$ and $p(A^c) = 1 - p(A)$.

8. If S is a finite set with n elements, how many elements are in $\mathcal{P}(S)$ the power set of S. (Hint, try it for the first few values of n, and don't forget to count the empty set.)

9. Show that for a function $f : S \to \mathbb{R}$ that for $A_1, A_2, A_3, \ldots$, a collection of subsets of the real numbers, that

$$f^{-1}(\cup_{j=1}^{\infty} A_j) = \cup_{j=1}^{\infty} f^{-1}(A_j)$$

Use this to prove that if f is a random variable on a probability space $(S, \Omega, P$ then its distribution P_f as defined above is also a probability.

10. For each of the examples of random variables you gave in problems 1, 2 and 3 calculate the expectation and the variance, if they exist.

11. Calculate the variance of the binomial, Poisson and normal distributions. The answer for the normal distribution is σ^2.

12. Let X be any random variable with finite second moment. Consider the function $f(a)$ defined as follows: $f(a) = E((X - a)^2)$. Show that the minimum value of f occurs when $a = E(X)$.

13. Fill in the details of the calculation of the variance of the uniform and exponential distributions. Also, prove that

$$\sigma^2(X) = E((X - E(X))^2) = E(X^2) - (E(X))^2.$$

14. Let X be a random variable with the standard normal distribution. find the mean and variance of each of the following:$|X|$, X^2 and e^{tX}.

15. Let X be the sine of an angle in radians chosen uniformly from the interval $(-\pi/2, \pi/2)$. Find the pdf of X, its mean and its variance.

16. Calculate the variance of the binomial, Poisson and normal distributions. The answer for the normal distribution is σ^2.

17. Let $p(x, y)$ be the uniform joint probability density on the unit disk, i.e.,

$$p(x, y) = \frac{1}{\pi} \quad \text{if} \;\; x^2 + y^2 < 1$$

and $p(x, y) = 0$ otherwise. Calculate the pdf of $X + Y$.

18. Suppose X and Y are independent random variables, each distributed according to the exponential distribution with parameter α. Find the joint pdf of X and Y (easy). Find the pdf of $X + Y$. Also find the mean and variance of $X + Y$.

Chapter 2

Aspects of finance and derivatives

The point of this book is to develop some mathematical techniques that can be used in finance to model movements of assets and then to use this to solve some interesting problems that occur in concrete practice.

Two of the problems that we will encounter is asset allocation and derivative valuation. The first, asset allocation, asks how do we allocate our resources in various assets to achieve our goals? These goals are usually phrased in terms of how much risk we are willing to acquire, how much income we want to derive from the assets and what level of return we are looking for. The second, derivative valuation, asks to calculate an intrinsic value of a derivative security (for example, an option on a stock). Though it might seem more esoteric, the second question is actually more basic than the first (at least in our approach to the problem). Thus we will attack derivative valuation first and then apply it to asset allocation.

2.1 What is a derivative security?

What is a derivative security? By this we mean any financial instrument whose value depends on (i.e. is derived from) an *underlying* security. The underlying in theory could be practically anything: a stock, a stock index, a bond or much more exotic things. Our most fundamental example is an *option* on a *stock*. An options contract gives the right but not the obligation to the purchaser of the contract to buy (or sell) a specified (in our case) stock at a specific price at a specific future time. For example, I could buy a (call) option to buy 100 shares of IBM at \$90 a share on January 19. Currently, this contract is selling for $\$19\frac{1}{2}$. The seller of this option earns the $\$19\frac{1}{2}$ (the price (premium) of the options contract) and incurs the obligation to sell me 100 shares of IBM stock on January 31 for \$90 if I should so choose. Of course, when January 19 rolls around, I would exercise my option only if the price of IBM is more than \$90.

Another well known and simpler example is a *forward*. A forward contract is an agreement/contract between two parties that gives the right and the obligation to one of the parties to purchase (or sell) a specified commodity (or financial instrument) at a specified price at a specific future time. No money changes hands at the outset of the agreement.
For example, I could buy a June forward contract for 50,000 gallons of frozen orange juice concentrate for $\$.25/gallon$. This would obligate me to take delivery of the frozen orange juice concentrate in June at that price. No money changes hands at the establishment of the contract.

A similar notion is that of a *futures contract*. The difference between a forward and a future contract is that futures are generally traded on exchanges and require other technical aspects, such as margin accounts.

2.2 Securities

The two most common types of securities are an ownership stake or a debt stake in some entity, for example a company or a government. You become a part owner in a company by buying shares of the company's stock. A debt security is usually acquired by buying a company's or a government's bonds. A debt security like a bond is a loan to a company in exchange for interest income and the promise to repay the loan at a future maturity date.

Securities can either be sold in a market (an exchange) or over the counter. When sold in an exchange the price of the security is subject to market forces. When sold over the counter, the price is arrived at through negotiation by the parties involved.

Securities sold on an exchange are generally handled by market makers or specialists. These are traders whose job it is to always make sure that the security can always be bought or sold at some price. Thus they provide liquidity for the security involved. Market makers must quote *bid* prices (the amount they are willing to buy the security for)and *ask* prices (the amount they are willing to sell the security for) of the securities. Of course the bid price is usually greater than the ask and this is one of the main ways market makers make their money. The *spot* price is the amount that the security is currently selling for. That is, we will usually assume that the bid and ask prices are the same. This is one of many idealizations that we will make as we analyze derivatives.

2.3 Types of derivatives

A *European call option* (on a stock) is a contract that gives the buyer the right but not the obligation to buy a specified stock for a specified price K, the *strike price*, at a specified date, T, the *expiration date* from the seller of the option. If the owner of the option can exercise the option any time up to T, then it is called an *American call*.

A *put* is just the opposite. It gives the owner of the contract the right to sell. It also comes in the European and American varieties.

Some terminology: If one owns an asset, one is said to have a *long position* in that asset. Thus is I own a share of IBM, I am long one share in IBM.

The opposite position is called *short*. Thus if I owe a share of IBM, I am short one share of IBM. Short selling is a basic financial maneuver. To short sell, one borrows the stock (from your broker, for example), sells it in the market. In the future, one must return the stock (together with any dividend income that the stock earned). Short selling allows negative amounts of a stock in a portfolio and is essential to our ability to value derivatives.

One can have long or short positions in any asset, in particular in options. Being long an option means owning it. Be short an option means that you are the writer of the option.

These basic call and put options are sold on exchanges and their prices are determined by market forces.

There are also many types of exotic options.

Asian options: these are options allowing one to buy a share of stock for the average value of the stock over the lifetime of the option. Different versions of these are based on calculating the average in different ways. An interesting feature of these is that their pay off depends not only on the value of the stock at expiration, but on the entire history of the stock's movements. We call this feature path dependence.

Barrier options: These are options that either go into effect or out of effect if

the stock value passes some threshold. For example, a down and out option is an option that expires worthless if the barrier X is reached from above before expiration. There are also down and in, up and in, up and out, and you are encouraged to think what these mean. These are path dependent.
Lookback options: Options whose pay-off depends on the the maximum or minimum value of the stock over the lifetime of the option.

2.4 The basic problem: How much should I pay for an option? Fair value

A basic notion of modern finance is that of risk. Contrary to common parlance, risk is not only a bad thing. Risk is what one gets rewarded for in financial investments. Quantifying risk and the level of reward that goes along with a given level of risk is a major activity in both theoretical and practical circles in finance. It is essential that every individual or corporation be able to select and manipulate the level of risk for a given transaction. This redistribution of risk towards those who are willing and able to assume it is what modern financial markets are about.

Options can be used to reduce or increase risk. Markets for options and other derivatives are important in the sense that agents who anticipate future revenues or payments can ensure a profit above a certain level or insure themselves against a loss above a certain level.

It is pretty clear that a prerequisite for efficient management of risk is that these derivative products are correctly valued, and priced. Thus our basic question: How much should I pay for an option?

We make a formal distinction between value and price. The price is what is actually charged while the value is what the thing is actually worth.

When a developer wants to figure out how much to charge for a new house say, he calculates his costs: land, material and labor, insurance, cost of financing etc.. This is the houses value. The developer, wanting to make a profit, adds to this a markup. The total we call the price. Our problem is to determine the value of an option.

In 1997, the Nobel Prize in economics was awarded to Robert Merton and Myron Scholes, who along with Fischer Black, wrote down a solution to this problem. Black died in his mid-fifties in 1995 and was therefore not awarded the Nobel prize. Black and Scholes wrote down there formula, the Black-Scholes Formula in 1973. They had a difficult time getting their paper published. However, within a year of its publication, Hewlett Packard had a calculator with this formula hardwired into it. Today this formula is an indispensable tool for thousands of traders and investors to value stock options in markets throughout the world.

Their method is quite general and can be used and modified to handle may kinds of exotic options as well as to value insurance contracts, and even has applications outside of finance.

2.5 Expectation pricing. Einstein and Bachelier, what they knew about derivative pricing. And what they didn't know

As opposed to the case of pricing a house, an option seems more like pricing bets in a casino. That is, the amount the casino charges for a given bet is based on models for the payoffs of the bets. Such models are of course statistical/probabilistic.

As an example, suppose we play a (rather silly) game of tossing a die and the casino pays the face value on the upside of the die in dollars. How much should the casino charge to play this game. Well, the first thing is to figure out the fair value. The casino of course is figuring that people will play this game many many times. Assuming that the die is fair, (i.e. $P(X = i) = 1/6$, for $i = 1, 2, \ldots, 6$ and that each roll of the die doesn't affect (that is independent of) any other roll, the law of large numbers says that if there are n plays of the game, that the average of the face values on the die will be with probability close to 1

$$E(X) = 1 \cdot 1/6 + 2 \cdot 1/6 + \cdots + 6 \cdot 1/6 = 21/6$$

Thus after n plays of the game, the casino feels reasonably confident that they will be paying off something close to $\$21/6 \cdot n$. So 21/6 is the fair valuer per game. The casino desiring to make a profit sets the actual cost to play at 4. Not too much higher than the fair value or no would want to play, they would lose their money too fast.

This is an example of expectation pricing. That is, the payoff of some experiment is random. One models this payoff as a random variable and uses the strong law of large numbers to argue that the fair value is the expectation of the random variable. This works much of the time.

Some times it doesn't. There are a number of reasons why the actual value might not be the expectation price. In order for the law of large numbers to kick in, the experiment has to be performed many times, in independent trials. If you are offered to play a game once, it might not be a very good value. What's the fair value to play Russian Roulette? As we will see below, there are more subtle and interesting reason why the fair value is not enforced in the market.

Thinking back to stocks and options, what would be the obvious way to use expectation pricing to value an option. The first order of business would be to model the movement of the underlying stock. Unfortunately, the best we can do here as well is to give a probabilistic model.

The French mathematician Louis Bachelier in his doctoral dissertation around 1900 was one of the earliest to attempt to model stock movements. He did so by what we now call a Random Walk (discrete version) or a Wiener process (continuous version). This actually predated the use of these to model Brownian motion, (for which Einstein got his Nobel Prize). Bachelier's thesis advisor, Jacques Hadamard did not have such high regard for Bachelier's work.

2.5.1 Payoffs of derivatives

We will have a lot to say in the future about models for stock movements. But for now, let's see how might we use this to value an option. Let us consider a European call option C to buy a share of XYZ stock with strike price K at time T. Let S_t denote the value of the stock at time t.The one time we know the value of the option is at expiration. At expiration there are two possibilities. If the value of the stock at expiration is less than the strike price $S_T \leq K$, we don't exercise our option and expires worthless. If $S_T > K$ then it makes sense to exercise the option, paying K for the stock, and selling it on the market for S_T, the whole transaction netting \$$S_T - K$. Thus, the option is worth $\max(S_t - K, 0)$. This is what we call the payoff of the derivative. It is what the derivative pays at expiration, which in this case only depends on the value of the stock at expiration.

On your own, calculate the payoff of all the derivatives defined in section 2.3.

Assuming a probabilistic model for S_t we would arrive at a random variable S_T describing in a probabilistic way the value of XYZ at time T. Thus the value of the option would be the random variable $\max(S_T - K, 0)$. And expectation pricing would say that the fair price should be $E(\max(S_T - K, 0))$. (I am ignoring the time value of money, that is, I am assuming interest rates are 0.)

Bachelier attempted to value options along these lines, unfortunately, it turns out that expectation pricing (naively understood at least) does not work. There is another wrinkle to option pricing. It's called arbitrage.

2.6 The simple case of forwards. Arbitrage arguments

I will illustrate the failure of expectation pricing with the example of a forward contract. This time I will also keep track of the time value of money. The time value of money is expressed by interest rates. If one has $\$B$ at time t , then at time $T > t$ it will be worth $\$\exp(r(T-t))B$. Bonds are the market instruments that express the time value of money. A zero-coupon bond is a promise of a specific amount of money (the face or par value) at some specified time in the future, the expiration date. So if a bond promises to pay $\$B_T$ at time T, then its current value B_t can be expressed as B_T discounted by some interest rate r, that is $B_t = \exp(-r(T-t))B_T$. For the meanwhile, we will assume that the interest rate r is constant. This makes some sense. Most options are rather short lived, 3 months to a year or so. Over this period, the interest rates do no vary so much.

Now suppose that we enter into a forward contract with another person to buy one share of XYZ stock at time T. We want to calculate what the forward price $\$F$ is. Thus at time T, I must buy a share of XYZ for $\$F$ from him and he must sell it to me for $\$F$. So at time T, the contract is worth $\$S_T - F$. According to expectation pricing, and adjusting for the time value of money, this contract should at time t be worth $\$\exp(r(T-t))E(S_T - F)$. Since the contract costs nothing to set up, if this is not zero, it would lead to profits or losses over the long run, by the law of large numbers. So expectation pricing says $F = E(S_T)$.

On the other hand, consider the following strategy for the party who must deliver the stock at time T. If the stock is now, at time t worth $\$S_t$, he borrows the money at the risk free interest rate r, buys the stock now, and holds the stock until time T. At time T he can just hand over the stock, collect $\$F$ and repay the loan $\exp(r(T-t))S_t$. After the transaction is over he has $\$(F - \exp(r(T-t))S_t$. At the beginning of the contract, it is worth nothing; no money changes hands. So he starts with a contract costing no money to set up, and in the end has $\$(F-\exp(r(T-t))S_t$. If this is positive, then we have an arbitrage opportunity. That is a risk free profit. No matter how small the profit is on any single transaction, one can parley this into unlimited profits by performing the transaction in large multiples. On the other hand, what will happen in practice if this is positive. Someone else will come along and offer the same contract but with a slightly smaller F still making a positive (and thus a potentially unlimited) positive. Therefore, as long as $\$(F - \exp(r(T-t))S_t > 0$ someone can come along and undercut him and still make unlimited money. So $\$(F - \exp(r(T-t))S_t \leq 0$

If $\$(F - \exp(r(T-t))S_t < 0$ you can work it from the other side. That is, the party to the contract who must buy the stock for F at time T, can at time t short the stock (i.e. borrow it) and sell it in the market for $\$S_t$. Put the money in the risk free instrument. At time T, buy the stock for $\$F$, return the stock, and now has $\exp(r(T-t))S_t$ in the risk free instrument. So again, at time t the contract costs nothing to set up, and at time T she has $\$(\exp(r(T-t))S_t - F)$ which is positive. Again, an arbitrage opportunity. A similar argument as in the paragraph above shows that $\$(\exp(r(T-t))S_t - F) \leq 0$. So we arrive at $F = \exp(r(T-t))S_t$. So we see something very interesting. The forward price has nothing to do with the expected value of the stock at time S_T. It only has to do with the value of the stock at the time of the beginning of the contract. This is an example of the basic arbitrage argument. In practice, the price arrived at by expectation pricing $F = E(S_T)$ would be much higher than the price arrived at by the arbitrage argument.

Let's give another example of an arbitrage argument. This is also a basic result expressing the relationship between the value of a call and of a put.

Proposition 2.6.1. *(Put-Call parity) If S_t denotes the value of a stock XYZ at time t, and C and P denote the value of a call and put on XYZ with the same strike price K and expiration T, then at time t, $P + S_t = C + K\exp(-r(T-t))$.*

Proof: By now, you should have calculated that the payoff on a put is $\max(K - S_T, 0)$. Now form the portfolio $P + S - C$. That is, long a share of stock, and a put, and short a call. Then at time T we calculate.

Case 1: $S_T \leq K$. Then $P+S-C = \max(K-S_T, 0)+S_T-\max(S_T-K, 0) = K - S_T + S_T - 0 = K$

Case 2: $S_T > K$. Then $P+S-C = \max(K-S_T, 0)+S_T-\max(S_T-K, 0) = 0 + S_T - (S_T - K) = K$
So the payoff at time T from this portfolio is K no matter what. That is it's risk free. Regardless of what S_T is, the portfolio is worth K at time T. What other financial entity has this property? A zero-coupon bond with par value K. Hence the value of this portfolio at time t is $K\exp(-r(T-t))$. Giving the theorem.
Q.E.D.

2.7 Arbitrage and no-arbitrage

Already we have made arguments based on the fact that it is natural for competing players to undercut prices as long as they are guaranteed a positive risk free profit, and thus able to generate large profits by carrying out multiple transactions. This is formalized by the no arbitrage hypothesis. (We formalize this even more in the next section.) An *arbitrage* is a portfolio that guarantees an (instantaneous) risk free profit. It is important to realize that a zero-coupon bond is not an arbitrage. It is risk free to be sure, but the growth of the risk free bond is just an expression of the time value of money. Thus the word instantaneous in the definition is equivalent to saying that the portfolio is risk free and grows faster than the risk-free interest rate.

Definition 2.7.1. The no-arbitrage hypothesis is the assumption that arbitrage opportunities do not exist.

What is the sense of this? Of course arbitrage opportunities exist. There exist people called arbitrageurs whose job it is to find these opportunities. (And some of them make a lot of money doing it.)

Actually, a good discussion of this is in Malkiel's book, along with *the efficient market hypothesis*. But Malkiel and others would insist that in actual fact, arbitrage opportunities don't really exist. But even if you don't buy this, it seems to be a reasonable assumption (especially in view of the reasonable arguments we made above.) as a starting point.

In fact, the existence of arbitrageurs makes the no-arbitrage argument more reasonable. The existence of many people seeking arbitrage opportunities out means that when such an opportunity arises, it closes all the more quickly. On publicly traded stocks the existence of arbitrage is increasingly rare. So as a basic principle in valuing derivatives it is a good starting point. Also, if one sees derivatives mispriced, it might mean the existence of arbitrage opportunities. So you have a dichotomy; you can be right (about the no-arbitrage hypothesis), in which case your evaluation of derivative prices is better founded, or you can make money (having found an arbitrage opportunity).

2.8 The arbitrage theorem

In this section we establish the basic connection between finance and mathematics. Our first goal is to figure out what arbitrage means mathematically and for this we model financial markets in the following way. We consider the following situation. Consider N possible securities, labeled $\{1, \ldots, N\}$. These may be stocks, bonds, commodities, options, etc. Suppose that now, at time 0, these securities are selling at $S_1, S_2, S_3, \ldots, S_N$ units of account (say dollars if you need closure). Note that this notation is different than notation in the rest of the notes in that subscript does not denote time but indexes the different securities. Consider the column vector S whose components are the S_j's. At time 1, the random nature of financial markets are represented by the possible states of the world, which are labeled $1, 2, 3, 4, \ldots, M$. That is, from time 0 to time 1 the world has changed from its current state (reflected in the securities prices S), to one of M possible states. Each security, say i has a payoff d_{ij} in state j. This payoff reflects any change in the security's price, any dividends that are paid or any possible change to the over all worth of the security. We can organize these payoffs into an $N \times M$ matrix,

$$D = \begin{pmatrix} d_{1,1} & d_{1,2} & \ldots & d_{1,M} \\ d_{2,1} & d_{2,2} & \ldots & d_{2,M} \\ \ldots & \ldots & \ldots & \ldots \\ d_{N,1} & d_{N,2} & \ldots & d_{N,M} \end{pmatrix} \tag{2.8.1}$$

Our portfolio is represented by a $N \times 1$ column matrix

$$\theta = \begin{pmatrix} \theta_1 \\ \theta_2 \\ \cdots \\ \theta_N \end{pmatrix} \tag{2.8.2}$$

where θ_i represents the number of copies of security i in the portfolio. This can be a negative number, indicating short selling.

Now, the market value of the portfolio is $\theta \cdot S = \theta^T S$. The pay-off of the portfolio can be represented by the $M \times 1$ matrix $D^T \theta$

$$= \begin{pmatrix} \sum_{j=1}^N d_{j,1}\theta_j \\ \sum_{j=1}^N d_{j,2}\theta_j \\ \cdots \\ \sum_{j=1}^N d_{j,M}\theta_j \end{pmatrix} \tag{2.8.3}$$

where the ith row of the matrix, which we write as $(D^T\theta)_i$ represents the payoff of the portfolio in state i. The portfolio is *riskless* if the payoff of the portfolio is independent of the state of the world, i.e. $(D^T\theta)_i = (D^T\theta)_j$ for any states of the world i and j.

Let

$$\mathbb{R}^n_+ = \{(x_1, \ldots, x_n)|x_i \geq 0 \text{ for all } i\}$$

and

$$\mathbb{R}^n_{++} = \{(x_1, \ldots, x_n)|x_i > 0 \text{ for all } i\}$$

. We write $v \geq 0$ if $v \in \mathbb{R}^n_+$ and $v > 0$ if $v \in \mathbb{R}^n_{++}$.

Definition 2.8.1. An *arbitrage* is a portfolio θ such that either

1. $\theta \cdot S \leq 0$ and $D^T\theta > 0$ or

2. $\theta \cdot S < 0$ and $D^T\theta \geq 0$

These conditions express that the portfolio goes from being worthless (respectively, less that worthless) to being worth something (resp. at least worth nothing), regardless of the state at time 1. In common parlance, an arbitrage is an (instantaneous) risk free profit, in other words, a free-lunch.

Definition 2.8.2. A *state price vector* is a $M \times 1$ column vector ψ such that $\psi > 0$ and $S = D\psi$.

Theorem 2.8.1. (Arbitrage theorem) There are no arbitrage portfolios if and only if there is a state price vector.

Before we prove this theorem, we need an abstract theorem from linear algebra. Let's recall that a subset $C \subset \mathbb{R}^n$ is called convex if the line between any two points of C is completely contained in C. That is, for any two points $x, y \in C$ and any $t \in [0,1]$ one has

$$x + t(y - x) \in C$$

Theorem 2.8.2. (The separating hyperplane theorem) Let C be a compact convex subset of $\mathbb{R}^n$ and $V \subset \mathbb{R}^N$ a vector subspace such that $V \cap C = \emptyset$. Then there exists a point $\xi_0 \in \mathbb{R}^n$ such that $\xi_0 \cdot c > 0$ for all $c \in C$ and $\xi_0 \cdot v = 0$ for all $v \in V$.

First we need a lemma.

Lemma 2.8.3. Let C be a closed and convex subset of $\mathbb{R}^n$ not containing 0. Then there exists $\xi_0 \in C$ such that $\xi_0 \cdot c \geq ||\xi_0||^2$ for all $c \in C$.

Proof: Let $r > 0$ be such that $B_r(0) \cap C \neq \emptyset$. Here $B_r(x)$ denotes all those points a distance r or less from x. Let $\xi_0 \in C \cap B_r(0)$ be an element which minimizes the distance from the origin for elements in $C \cap B_r(0)$. That is,

$$||\xi_0|| \leq ||c|| \tag{2.8.4}$$

for all $c \in B_r(0) \cap C$. Such a ξ_0 exists by the compactness of $C \cap B_r(0)$. Then it follows that (2.8.4) holds for all $c \in C$ since if $c \in C \cap B_r(0)$ then $||\xi_0|| \leq r \leq ||c||$.

Now since C is convex, it follows that for any $c \in C$, we have $\xi_0 + t(c - \xi_0) \in C$ so that

$$||\xi_0||^2 \leq ||\xi_0 + t(c - \xi_0)||^2 = ||\xi_0||^2 + 2t\xi_0 \cdot (c - \xi_0) + t^2||c - \xi_0||^2$$

Hence

$$2||\xi_0||^2 - t||c - \xi_0|| \leq 2\xi_0 \cdot c$$

for all $t \in [0,1]$. So $\xi_0 \cdot c \geq ||\xi_0||^2$
QED.

Proof (of the separating hyperplane theorem) Let C be compact and convex, $V \subset \mathbb{R}^n$ as in the statement of the theorem. Then consider

$$C - V = \{c - v \mid c \in C, v \in V\}$$

This set is closed and convex. (Here is where you need C be compact.) Then since $C \cap V = \emptyset$ we have that $C - V$ doesn't contain 0. Hence by the lemma there exists $\xi_0 \in C - V$ such that $\xi_0 \cdot x \geq ||\xi_0||^2$ for all $x \in C - V$. But this implies that for all $c \in C, v \in V$ that $\xi_0 \cdot c - \xi_0 \cdot v \geq ||\xi_0||^2$. Or that

$$\xi_0 \cdot c > \xi_0 \cdot v$$

for all $c \in C$ and $v \in V$. Now note that since V is a linear subspace, that for any $v \in V$, $av \in V$ for all $a \in \mathbb{R}$. Hence $\xi_0 \cdot c > \xi_0 \cdot av = a\xi_0 \cdot v$ for all a. But this can't happen unless $\xi_0 \cdot v = 0$. So we have $\xi_0 \cdot v = 0$ for all $v \in V$ and also that $\xi_0 \cdot c > 0$.
QED.
Proof (of the arbitrage theorem): Suppose $S = D\psi$, with $\psi > 0$. Let θ be a portfolio with $S \cdot \theta \leq 0$. Then

$$0 \geq S \cdot \theta = D\psi \cdot \theta = \psi \cdot D^T\theta$$

and since $\psi > 0$ this implies $D^T\theta \leq 0$. The same argument works to handle the case that $S \cdot \theta < 0$.

To prove the opposite implication, let

$$V = \{(-S \cdot \theta, D^T\theta) \mid \theta \in \mathbb{R}^N\} \subset \mathbb{R}^{M+1}$$

Now consider $C = \mathbb{R}_+ \times \mathbb{R}_+^M$. C is closed and convex. Now by definition, there exists an arbitrage exactly when there is a non-zero point of intersection of V and C. Now consider

$$C_1 = \{(x_0, x_1, \ldots, x_M) \in C \mid x_0 + \cdots + x_M = 1\}$$

C_1 is compact and convex. Hence there exists according to (2.8.2) $x \in C_1$ with

1. $x \cdot v = 0$ for all $v \in V$ and

2. $x \cdot c > 0$ for all $c \in C_1$.

Then writing $x = (x_0, x_1, \ldots, x_M)$ we have that $x \cdot v = 0$ implies that for all $\theta \in \mathbb{R}^N$, $(-S\cdot\theta)x_0 + D^T\theta\cdot(x_1, \ldots, x_M)^T = 0$. So $D^T\theta\cdot(x_1, \ldots, x_M)^T = x_0 S\cdot\theta$ or $\theta \cdot D(x_1, \ldots, x_M)^T = x_0 S \cdot \theta$. Hence letting

$$\psi = \frac{(x_1, \ldots, x_M)^T}{x_0}$$

we have $\theta \cdot D\psi = S \cdot \theta$ for all θ. Thus $D\psi = S$ and ψ is a state price vector. Q.E.D.

Let's consider the following example. A market consisting of two stocks 1, 2. And two states 1, 2. Suppose that $S_1 = 1$ and $S_2 = 1$. And

$$D = \begin{pmatrix} 1 & 1 \\ \frac{1}{2} & 2 \end{pmatrix} \tag{2.8.5}$$

So security 1 pays off 1 regardless of state and security 2 pays of 1/2 in state 1 and 2 in state 2. Hence if

$$\theta = \begin{pmatrix} \theta_1 \\ \theta_2 \end{pmatrix} \tag{2.8.6}$$

is a portfolio its payoff will be $D^T\theta$

$$\begin{pmatrix} \theta_1 + \theta_2/2 \\ \theta_1 + 2\theta_2 \end{pmatrix} \tag{2.8.7}$$

Is there a ψ with $S = D\psi$. One can find this matrix easily by matrix inversion so that

$$\begin{pmatrix} 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ \frac{1}{2} & 2 \end{pmatrix} \begin{pmatrix} \psi_1 \\ \psi_2 \end{pmatrix} \tag{2.8.8}$$

So that

$$\begin{pmatrix} \psi_1 \\ \psi_2 \end{pmatrix} = \begin{pmatrix} 4/3 & -2/3 \\ -1/3 & 2/3 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 2/3 \\ 1/3 \end{pmatrix} \tag{2.8.9}$$

Now we address two natural and important questions.

What is the meaning of the state price vector?

How do we use this to value derivatives?

We formalize the notion of a derivative.

Definition 2.8.3. A *derivative* (or a *contingency claim*) is a function on the set of states.

We can write this function as an $M \times 1$ vector. So if h is a contingency claim, then h will payoff h_j if state j is realized. To see how this compares with what we have been talking about up until now, let us model a call on security 2 in the example above. Consider a call with strike price $1\frac{1}{2}$ on security 2 with expiration at time 1. Then $h_j = \max(d_{2,j} - 1\frac{1}{2}, 0)$. Or

$$h = \begin{pmatrix} h_1 \\ h_2 \end{pmatrix} = \begin{pmatrix} 0 \\ \frac{1}{2} \end{pmatrix} \tag{2.8.10}$$

So our derivatives are really nothing new here, except that it allows the possibility of a payoff depending on a combination of securities, which is certainly a degree of generality that we might want.

To understand what the state price vector means, we introduce some important special contingency claims. Let ϵ^j be a contingency claim that pays 1 if state change occurs, and 0 if any other state occurs. So

$$\epsilon^j = \begin{pmatrix} 0 \\ 0 \\ \cdots \\ 1 \text{ in the } j\text{th row} \\ \cdots \\ 0 \end{pmatrix} \tag{2.8.11}$$

These are often called binary options. It is just a bet paying off 1 that state j occurs. How much is this bet worth? Suppose that we can find a portfolio θ^j such that $D^T\theta^j = \epsilon^j$. Let ψ be a state price vector. Then the market value of the portfolio θ^j is

$$S \cdot \theta^j = D\psi \cdot \theta^j = \psi \cdot D^T\theta^j = \psi \cdot \epsilon^j = \psi_j$$

Hence we have shown, that the market value of a portfolio whose payoff is ϵ^j is ψ_j. So the state price vector is the market value of a bet that state j will be realized. This is the meaning of ψ.

Now consider

$$B = \sum_{j=1}^{M} \epsilon^j = \begin{pmatrix} 1 \\ 1 \\ \cdots \\ 1 \end{pmatrix} \tag{2.8.12}$$

This is a contingency claim that pays 1 regardless of state. That is, B is a riskless zero-coupon bond with par value 1 and expiration at time 1. Its market value will then be $R = \sum_{j=1}^{M} \psi_j$. So this R is the discount factor for riskless borrowing. (you might want to think of R as $(1+r)^{-1}$ where r is the interest rate.) Now we can form a vector $Q = \frac{1}{R}\psi$. Note that $Q \in \mathbb{R}_{++}$ and $\sum_{j=1}^{M} q_j = 1$. So Q can be interpreted as a probability function on the state space. **It is important to remark that these probabilities are not the real world probabilities that might be estimated for the states j. They are probabilities that are derived only from the no-arbitrage condition. They are therefore often called *pseudo-probabilities.***

Coming back to our problem of valuing derivatives, consider a derivative h. If θ is a portfolio whose payoff is h, i.e. $D^T\theta = h$, we say that θ *replicates* h. Then the market value of θ is

$$S \cdot \theta = D\psi \cdot \theta = \psi \cdot D^T\theta$$

$$= \psi \cdot h = RQ \cdot h = D\sum_{j=1}^{M} q_j h_j = RE_Q(h)$$

That is, the market value of θ (and therefore the derivative) is the discounted expectation of h with respect to the pseudo-probability Q. **Hence, we have regained our idea that the value of a contingency claim should be an expectation, but with respect to probabilities derived from only from the no-arbitrage hypothesis, not from any actual probabilities of the states occurring.**

Let us go back to the example above. Let's go ahead and value the option on security number 2. $R = 1$ in this example, so the value of h is

$$RE_Q(h) = 1 \cdot (0\frac{2}{3} + \frac{1}{2}\frac{1}{3}) = 1/6$$

Coming back to general issues, there is one further thing we need to discuss. When is there a portfolio θ whose payoff $D^T\theta$ is equal to a contingency claim h. For this we make the following

Definition 2.8.4. A market is *complete* if for each of the ϵ^j there is a portfolio θ^j with $D^T\theta^j = \epsilon^j$.

Now we see that if a market is complete, then for any contingency claim h, we have that for $\theta = \sum_{j=1}^{M} h_j\theta^j$ that $D^T\theta = h$. That is, a market is complete if and only if every claim can be replicated.

We state this and an additional fact as a

Theorem 2.8.4. 1. A market is complete if and only if every contingency claim can be replicated.

2. An arbitrage free market is complete if and only if the state price vector is unique.

Proof: Having proved the first statement above we prove the second. It is clear from the first statement that if a market is complete then $D^T : \mathbb{R}^N \to \mathbb{R}^M$ is onto, that is has rank M. Thus if ψ^1 and ψ^2 are two state price vectors, then $D(\psi^1 - \psi^2) \cdot \theta = (S - S) \cdot \theta = 0$ for all θ. But then $0 = D(\psi^1 - \psi^2) \cdot \theta = (\psi^1 - \psi^2) \cdot D^T\theta$. Since D^T is onto, this means that $(\psi^1 - \psi^2) \cdot v = 0$ for all $v \in \mathbb{R}^N$ which means that $\psi^1 - \psi^2 = 0$.

We prove the other direction by contradiction. So assume that the market is not complete, and ψ^1 is a state price vector. Then D^T is not onto. So we can find a vector τ orthogonal to the image, i.e. $D^T\theta \cdot \tau = 0$ for all θ. Then $\theta \cdot D\tau = 0$ for all θ so $D\tau = 0$ and $\psi^2 = \psi^1 + \tau$ is another state price vector. QED.

Exercises

1. Consider a variant of the game described above, tossing a die and the casino pays the face value on the upside of the die in dollars. Except this time, if the player doesn't like her first toss, she is allowed one additional toss and the casino then pays the face value on the second toss. What is the fair value of this game?.

2. Suppose a stock is worth $\$S_-$ just before it pays a dividend, $\$D$. What is the value of the stock S_+ just after the dividend? Use an arbitrage argument.

3. Suppose that stock XYZ pays no dividends, and C is an American call on XYZ with strike price K and expiration T. Show, using an arbitrage argument, that it is never advantageous to exercise early, that is, for any intermediate time $t < T$, $C_t > S_t - K$. (Hint: Calculate the payoff of the option as seen from the same time T in two cases: 1) If you exercise early at time $t < T$, 2) If you hold it until expiration. To finance the first case, you borrow K at time t to buy the stock, buy the stock, and repay the loan at time T. Calculate this portfolio at time T. The other case you know the payoff. Now compare and set up an arbitrage if $S_t - K \geq C_t$.)

4. Suppose that C is a call on a non-dividend paying stock S, with strike price K and expiration T. Show that for any time $t < T$ that $C_t > S_t - \exp(-r(T-t))K$.

5. Suppose that C and D are two calls on the same stock with the same expiration T, but with strike prices $K < L$ respectively. Show by an arbitrage argument that $C_t \geq D_t$ for all $t < T$.

6. Suppose that C and D are two calls on the same stock with the same strike price K, but with expirations $T_1 < T_2$, respectively. Show by an arbitrage argument that $D_t \geq C_t$ for all $t < T_1$.

7. Using the arbitrage theorem, show that there can only be one bond price. That is, if there are 2 securities paying \$1, regardless of state, then their market value is the same.

8. Consider the example of the market described in section (2.7). Let h be a put on security 2 with strike price $\frac{1}{2}$. Calculate its market value.

9. Make up your own market, and show that there is a state price vector. Consider a specific derivative (of your choice) and calculate its market value.

Chapter 3

Discrete time

In this chapter, we proceed from the one period model of the last chapter to the mutli-period model.

3.1 A first look at martingales

In this section I would like to give a very intuitive sense of what a martingale is, why they come up in option pricing, and then give an application to calculating expected stopping times.

The concept of a martingale is one of primary importance in the modern theory of probability and in theoretical finance. Let me begin by stating the slogan that a martingale is a mathematical model for a fair game. In fact, the term martingale derives from the well known betting strategy (in roulette) of doubling ones bet after losing. First let's think about what we mean by a fair game.

Giorolamo Cardano, an Italian, was one of the first mathematicians to study the theory of gambling. In his treatise, *The Book of Games of Chance* he wrote: "The most fundamental principle of all in gambling is simply equal conditions, for example, of opponents, of bystanders, of money, of situation, of the dice box, and of the die itself. To the extent to which you depart from that equality, if it is in your opponent's favor, you are a fool, and if in your own, you are unjust." (From A. Hald, A History of Probability and Statistics and their applications before 1750, John Wiley and Sons, New York.)

How do we mathematically model this notion of a fair game, that is, of equal conditions? Consider a number of gamblers, say N, playing a gambling game. We will assume that this game is played in discrete steps called hands. We also assume that the gamblers make equal initial wagers and the winner wins the whole pot. It is reasonable to model each gambler as the gamblers' current fortunes. That is, this is the only information that we really need to talk about. Noting the gamblers' fortunes at a single time is not so useful, so we keep track of their fortunes at each moment of time. So let F_n denote fortune of one of the gamblers after the nth hand. Because of the random nature of gambling we let F_n be a random variable for each n. A sequence of random variables, F_n is what is called a *stochastic process*.

Now for the equal conditions. A reasonable way to insist on the equal conditions of the gamblers is to assume that each of the gamblers' fortune processes are distributed (as random variables) in exactly the same way. That is, if $p_n(x)$ denotes the probability density function of F_n, and G_n is any of the other gamblers' fortunes, with pdf $q_n(x)$, then we have $p_n(x) = q_n(x)$ for all n and x.

What are the consequences of this condition? The amount that the gambler wins in hand n is the random variable $\xi_n = F_n - F_{n-1}$ and the equal distribution of the gamblers' fortunes implies that the gamblers' wins in the nth hand are also equally distributed. From the conditions on the gambling game described above, we must have $E(\xi_n) = 0$. Think about this. Hence we have that

$$E(F_n) = E(F_{n-1} + \xi_n) = E(F_{n-1}) + E(\xi_n) = E(F_{n-1})$$

So the expected fortunes of the gambler will not change over time. This is close to the Martingale condition. The same analysis says that if the gambler has F_{n-1} after the $n-1$ hand, then he can expect to have F_{n-1} after the nth hand. That is, $E(F_n|F_{n-1}) = F_{n-1}$. (Here $E(F_n|F_{n-1}) = F_{n-1}$ denotes a conditional expectation, which we explain below.) This is closer to the Martingale condition which says that $E(F_n|F_{n-1}, F_{n-1}, \cdots, F_0) = F_{n-1}$.

O Romeo, Romeo! wherefore art thou Romeo?
Deny thy father and refuse thy name;
Or, if thou wilt not, be but sworn my love,
And I'll no longer be a Capulet.

-Juliet

Now let me describe a very pretty application of the notion of martingale to a simple problem. Everyone has come accross the theorem in a basic probability course that a monkey typing randomly at a typewriter with a probability of 1/26 of typing any given letter (we will always ignore spaces, capitals and punctuation) and typing one letter per second will eventually type the complete works of Shakespeare with probability 1. But how long will it take. More particularly, let's consider the question of how long it will take just to type: "O Romeo, Romeo". Everyone recognizes this as Juliet's line in Act 2 scene 2 of Shakespeare's Romeo and Juliet. Benvloio also says this in Act 3 scene 1. So the question more mathematically is what is the expected time for a monkey to type "O Romeo Romeo", ignoring spaces and commas. This seems perhaps pretty hard. So let's consider a couple of similar but simpler questions and also get our mathematical juices going.

1. Flipping a fair coin over and over, what is the expected number of times it takes to flip before getting a Heads and then a Tails: HT ?

2. Flipping a fair coin over and over, what is the expected number of times it takes to flip two Heads in a row: HH ?

We can actually calculate these numbers by hand. Let T_1 denote the random variable of expected time until HT and T_2 the expected time until HH. Then we want to calculate $E(T_1)$ and $E(T_2)$. If T denotes one of these RV's (random variables) then

$$E(T) = \sum_{n=1}^{\infty} nP(T = n)$$

We then have

$$\begin{aligned} P(T_1 = 1) = &\quad 0 \\ P(T_1 = 2) = &\quad P(\{HT\}) = 1/4 \\ P(T_1 = 3) = &\quad P(\{HHT, THT\}) = 2/8 \\ P(T_1 = 4) = &\quad P(\{HHHT, THHT, TTHT\}) = 3/16 \\ P(T_1 = 5) = &\quad P(\{HHHHT, THHHT, \\ &\quad TTHHT, TTTHT\}) = 4/32 \\ P(T_1 = 6) = &\quad P(\{HHHHHT, THHHHT, \\ &\quad TTHHHT, TTTHHT, TTTTHT\}) = 5/64 \end{aligned} \tag{3.1.1}$$

We can see the pattern $P(T_1 = n) = \frac{n-1}{2^n}$ and so

$$E(T_1) = \sum_{n=2}^{\infty} n(n-1)/2^n$$

. To sum this, let's write down the "generating function"

$$f(t) = \sum_{n=1}^{\infty} n(n-1)t^n$$

so that $E(T_1) = f(1/2)$ To be able to calculate with such generating functions is a useful skill. First not that the term $n(n-1)$ is what would occur in the general term after differentiating a power series twice. On the other hand, this term sits in front of t^n instead of t^{n-2} so we need to multiply a second derivative by t^2. At this point, we make some guesses and then adjustments to see that

$$f(t) = t^2(\sum_{n=0}^{\infty} t^n)''$$

so it is easy to see that

$$f(t) = \frac{2t^2}{(1-t)^3}$$

and that $E(T_1) = f(1/2) = 4$.

So far pretty easy and no surprises. Let's do $E(T_2)$. Here

$$\begin{aligned}
P(T_2 = 1) = &\ 0 \\
P(T_2 = 2) = &\ P(\{HH\}) = 1/4 \\
P(T_2 = 3) = &\ P(\{THH\}) = 1/8 \\
P(T_2 = 4) = &\ P(\{T-THH, H-THH\} = 2/16 \\
P(T_2 = 5) = &\ P(\{TT-THH, HT-THH, TH-THH\}) = 3/32 \\
P(T_2 = 6) = &\ P(\{TTT-THH, TTH-THH, THT-THH, \\
&\ HTT-THH, HTH-THH\}) = 5/64 \\
P(T_2 = 7) = &\ P(\{TTTT-THH, TTTH-THH, TTHT-THH, \\
&\ THTT-THH, HTTT-THH, THTH-THH, \\
&\ HTHT-THH, HTTH-THH\} = 8/128
\end{aligned} \tag{3.1.2}$$

Again the pattern is easy to see: $P(T_2 = n) = f_{n-1}/2^n$ where f_n denotes the nth Fibonacci number (Prove it!).(Recall that the Fibonacci numbers are defined by the recursive relationship, $f_1 = 1$, $f_2 = 1$ and $f_n = f_{n-1} + f_{n-2}$.) So

$$E(T_2) = \sum_{n=2}^{\infty} n f_{n-1}/2^n$$

Writing $g(t) = \sum_{n=1}^{\infty} n f_{n-1} t^n$ we calculate in a similar manner (but more complicated) that

$$g(t) = \frac{2t^2 - t^3}{(1 - t - t^2)^2}$$

and that $E(T_2) = g(1/2) = 6$. A little surprising perhaps.

What is accounting for the fact that HH takes longer? A simple answer is that if we are awaiting the target string (HT or HH) then if we succeed to get the first letter at time n but then fail at time $n+1$ to get the second letter, the earliest we can succeed after that is $n+2$ in the case of HT but is $n+3$ in the case of HH.

The same kind of analysis could in theory be done for "OROMEOROMEO" but it would probably be hopelessly complicated.

To approach "OROMEOROMEO" let's turn it into a game. Let T denote the random variable giving me the expected time until "OROMEOROMEO". I walk into a casino and sitting at a table dressed in the casino's standard uniform is a monkey in front of a typewriter. People are standing around the table making bets on the letters of the alphabet. Then the monkey strikes a key at random (it is a fair monkey) and the people win or lose according to whether their bet was on the letter struck. If their letter came up, they are paid off at a rate of 26 to 1. The game then continues. For simplicity, assume that a round of monkey typing takes exactly one minute. Clearly the odds and the pay-off are set up to yield a "fair game": this is an example of a martingale. I know in my mathematical bones that I can not expect to do better on average than to break even. That is:

Theorem 3.1.1. No matter what strategy I adopt, on average, I can never do better than break even.

Eventually this will be a theorem in martingale theory (and not a very hard one at that). Together a group of gamblers conspire together on the following betting scheme. On the first play, the first bettor walks up to the table and bets one Italian Lira (monkey typing is illegal in the United States) ITL 1 that the first letter is an "O". If he wins, he gets ITL 26 and bets all ITL 26 on the second play that the second letter is an "R". If he wins he bets all ITL 26^2 that on the third play the letter is an "O". Of course if at some point he loses, all the money is confiscated and he walks away with a net loss of the initial ITL 1. Now in addition to starting the above betting sequence on the first play, the second bettor begins the same betting sequence starting with the second play, that is, regardless of what has happened on the first play, she also bets ITL 1 that the second play will be an O, etc. One after another, the bettors (there are a lot of them) initiate a new betting sequence at each play. If at some point one of the bettors gets through "OROMEOROMEO", the game ends, and the conspirators walk away with their winnings and losses which they share. Then they calculate their total wins and losses.

Again, because the game is fair, the conspirators expected wins is going to be 0. On the other hand, let's calculate the amount they've won once "OROMEOROMEO" has been achieved. Because they got through, the betting sequence beginning "OROMEOROMEO" has yielded ITL 26^{11}. In addition to this though, the sequence starting with the final "O" in the first ROMEO and continuing through the end has gone undefeated and yields ITL 26^6. And finally the final "O" earns an additional ITL 26. On the other hand, at each play, ITL 1 the betting sequence initiated at that play has been defeated, except for those enumerated above. Each of these then counts as a ITL 1 loss. How many such losses are there? Well, one for each play of monkeytyping. Therefore, their net earnings will be

$$26^{11} + 26^6 + 26 - T$$

Since the expected value of this is 0, it means that the expected value $E(T) = 26^{11} + 26^6 + 26$. A fair amount of money even in Lira.

Try this for the "HH" problem above to see more clearly how this works.

3.2 Introducing stochastic processes

A *stochastic process* which is discrete in both space and time is a sequence of random variables, defined on the same sample space (that is they are jointly distrubuted) $\{X_0, X_1, X_2, \ldots\}$ such that the totality of values of all the X_n's combined form a discrete set. The different values that the X_n's can take are called the *states* of the stochastic process. The subscript n in X_n is to be thought of as time.

We are interested in understanding the evolution of such a process. One way to describe such a thing is to consider how X_n depends on $X_0, X_1, \ldots, X_{n-1}$.

Example 3.2.1. Let W_n denote the position of a random walker. That is, at time 0, $W_0 = 0$ say and at each moment in time, the random walker has a probability p of taking a step to the left and a probability $1-p$ of taking a step to the right. Thus, if at time n, $W_n = k$ then at time $n+1$ there is probability p of being at $k-1$ and a probability $1-p$ of being at $k+1$ at time $n+1$. We express this mathematically by saying

$$P(W_{n+1} = j|W_n = k) \quad = \quad \begin{cases} p & \text{if } j = k-1 \\ 1-p & \text{if } j = k+1 \\ 0 & \text{otherwise} \end{cases} \tag{3.2.1}$$

In general, for a stochastic process X_n we might specify

$$P(X_{n+1} = j|X_n = k_n, X_{n-1} = k_{n-1}, \cdots, X_0 = k_0)$$

If as in the case of a random walk

$$P(X_{n+1} = j|X_n = k_n, X_{n-1} = k_{n-1}, \cdots, X_0 = k_0) = P(X_{n+1} = j|X_n = k_n)$$

the process is called a *Markov process*. We write

$$p_{jk} = P(X_{n+1} = k|X_n = j)$$

The numbers p_{jk} form a matrix (an infinite by infinite matrix if the random variables X_n are infinite), called the *matrix of transition probabilities*. They satisfy (merely by virtue of being probabilities)

1. $p_{jk} \geq 0$
2. $\sum_k p_{jk} = 1$ for each j.

The matrix p_{jk} as we've defined it tells us how the probabilities change from the nth to the $n+1$st step. So p_{jk} a priori depends on n. If it doesn't, then the process is called a *stationary* Markov process. Let's reiterate: p_{jk} represents the probability of going from state j to state k in one time step. Thus, if X_n is a stationary Markov process, once we know the p_{jk} the only information left to know is the initial probability distribution of X_0, that is $p_0(k) = P(X_0 = k)$. The basic question is: Given p_0 and p_{jk} what is the probability distribution of X_n? Towards this end we introduce the "higher" transition probablilities

$$p_{jk}^{(n)} = P(X_n = k|X_0 = j)$$

which by stationarity also equals $P(X_{m+n} = k|X_m = j)$ the probability of going from j to k in n steps.

Proposition 3.2.1.

$$p_{jk}^{(n)} = \sum_h p_{jh}^{(n-1)} p_{hk}$$

Proof:

$$p_{jk}^{(n)} = P(X_n = k|X_0 = j)$$

Now a process that goes from state j to state k in n steps, must land somewhere on the $n-1$st step. Suppose it happens to land on the state h. Then the probability

$$P(X_n = k|X_{n-1} = h) = P(X_1 = k|X_0 = h) = p_{hk}$$

by stationarity. Now the probability of going from state j to state h in $n-1$ steps and then going from state h to state k in one step is

$$P(X_{n-1} = h|X_0 = j)P(X_n = k|X_{n-1} = h) = p_{jh}^{(n-1)} p_{hk}$$

Now the proposition follows by summing up over all the states one could have landed upon on the $n-1$st step. Q.E.D.

It follows by repeated application of the preceding proposition that

$$p_{jk}^{(n+m)} = \sum_h p_{jh}^{(n)} p_{hk}^{(m)}$$

This is called the Chapman-Kolmogorov equation. Expressed in terms of matrix mulitplication, it just says that if $P = p_{jk}$ denotes the matrix of transition probabilities from one step to the next, then P^n the matrix raised to the nth power represents the higher transition probabilities and of course $P^{n+m} = P^n P^m$.

Now we can answer the basic question posed above. If $P(X_0 = j) = p_0(j)$ is the initial probability distribution of a stationary Markov process with transition matrix $P = p_{jk}$ then

$$P(X_n = k) = \sum_j p_0(j) p_{jk}^{(n)}$$

3.2.1 Generating functions

Before getting down to some examples, I want to introduce a technique for doing calculations with random variables. Let X be a discrete random variable whose states are any integer j and has distribution $p(k)$. We introduce its generating function

$$F_X(t) = \sum_k p(k)t^k$$

At this point we do not really pay attention to the convergence of F_X, we will do formal manipulations with F_X and therefore F_X is called a formal power series (or more properly a formal Laurent series since it can have negative powers of t). Here are some of the basic facts about the generating function of X;

1. $F_X(1) = 1$ since this is just the sum over all the states of the probabilities, which must be 1.

2. $E(X) = F_X'(1)$. Since $F_X'(t) = \sum_k kp(k)t^{k-1}$. Now evaluate at 1.

3. $\text{Var}(X) = F_X''(1) + F_X'(1) - (F_X'(1))^2$. This is an easy calculation as well.

4. If X and Y are independent random variables, then $F_{X+Y}(t) = F_X(t)F_Y(t)$.

Before going on, let's use these facts to do some computations.

Example 3.2.2. Let X be a Random variable representing a Bernoulli trial, that is $P(X = -1) = p$, $P(X = 1) = 1 - p$ and $P(X = k) = 0$ otherwise. Thus $F_X = pt^{-1} + (1 - p)t$. Now a Random walk can be described by

$$W_n = X_1 + \cdots + X_n$$

where each X_k is an independent copy of X: X_k represents taking a step to the left or right at the kth step. Now according to fact 4) above, since the X_k are mutually independent we have

$$F_{W_n} = F_{X_1}F_{X_2}\cdots F_{X_n} = (F_X)^n = (pt^{-1} + (1-p)t)^n$$

$$= \sum_{j=0}^{n} \binom{n}{j}(pt^{-1})^j((1-p)t)^{n-1}$$

$$= \sum_{j=0}^{n} \binom{n}{j} p^j (1-p)^{n-j} t^{n-2j}$$

Hence the distribution for the random walk W_n can be read off as the coefficients of the generating function. So $P(W_n = n - 2j) = \binom{n}{j} p^j (1-p)^{n-j}$ for $j = 0, 1, 2, \ldots, n$ and is zero for all other values. (Note: This is our third time calculating this distribution.) Also, $E(W_n) = F'_{W_n}(1)$. $F'_{W_n}(t) = \frac{d}{dt}(pt^{-1} + (1-p)t)^n = n(pt^{-1} + (1-p)t)^{n-1}((1-p) - pt^{-2})$ and evaluating at 1 gives $n(p + (1-p))((1-p) - p) = n((1-p) - p)$. Similarly, F''_{W_n}

$$= n(n-1)(pt^{-1} + (1-p)t)^{n-2}((1-p) - pt^{-2})^2 + n(pt^{-1} + (1-p)t)^{n-1}(2pt^{-3})$$

and evaluating at 1 gives $n(n-1)((1-p) - p)^2 + n2p$. So $\mathrm{Var}(W_n)$

$$= F''_{W_n}(1) + F'_{W_n}(1) - F'_{W_n}(1)^2$$

$$= n(n-1)((1-p) - p)^2 + 2np + n((1-p) - p) - n^2((1-p) - p)^2$$

$$= -n((1-p) - p)^2 + 2np + n((1-p) - p)$$

$$= n((1-p) - p)(1 - ((1-p) - p)) + 2np = 2pn((1-p) - p) + 2np$$

$$= 2np(1-p) - 2np^2 + 2np = 2np(1-p) + 2np(1-p) = 4np(1-p)$$

Consider the following experiment. Roll two dice and read off the sum on them. Then flip a coin that number of times and count the number of heads. This describes a random variable X. What is its distribution? We can describe X mathematically as follows. Let N denote the random variable of the sum on two dice. So according to a calculation in a previous set of notes, $P(N = n) = \frac{1}{36}(6 - |7 - n|)$ for $n = 2, 3, \ldots 12$ and is 0 otherwise. Also, let Y count the number of heads in one toss, so that $P(Y = k) = \frac{1}{2}$ if $k = 0$ or 1 and is 0 otherwise. Hence our random variable X can be written as

$$X = \sum_{j=1}^{N} Y_j \tag{3.2.2}$$

where Y_j are independent copies of Y and note that the upper limit of the sum is the random variable N. This is an example of a composite random variable, that is, in general a composite random variable is one that can be expressed by (3.2.2)where N and Y are general random variables. The generating function of such a composite is easy to calculate.

Proposition 3.2.2. *Let X be the composite of two random variables N and Y. Then $F_X(t) = F_N(F_Y(t)) = F_N \circ F_Y$.*

Proof:

$$P(X = k) = \sum_{n=0}^{\infty} P(N = n)P(\sum_{j=1}^{n} Y_j = k)$$

Now, $F_{\sum_{j=1}^{n} Y_j} = (F_Y(t))^n$ since the Y_j's are mutually independent. Therefore

$$F_X = \sum_{n=1}^{\infty} P(N = n)(F_Y(t))^n = F_N(F_Y(t))$$

Q.E.D.

So in our example above of rolling dice and then flipping coins, $F_N = \sum_{n=2}^{1} 2\frac{1}{36}(6-|7-n|)t^n$ and $F_Y(t) = 1/2+1/2t$. Therefore $F_X(t) = \sum_{n=2}^{12} \frac{1}{36}(6-|7-n|)(1/2 + 1/2t)^n$ from which one can read off the distribution of X.

The above sort of compound random variables appear naturally in the context of *branching processes*. Suppose we try to model a population (of amoebae, of humans, of subatomic particles) in the following way. We denote the number of members in the nth generation by X_n. We start with a single individual in the zeroth generation. So $X_0 = 1$. This individual splits into (or becomes) X_1 individuals (offspring) in the first generation. The number of offspring X_1 is a random variable with a certain distribution $P(X_1 = k) = p(k)$. $p(0)$ is the probability that X_0 dies. $p(1)$ is the probability that X_0 has one descendent, $p(2)$, 2 descendants etc. Now in the first generation we have X_1 individuals, each of which will produce offspring according to the same probability distribution as its parent. Thus in the second generation,

$$X_2 = \sum_{n=1}^{X_1} X_{1,n}$$

where $X_{1,n}$ is an independent copy of X_1. That is, X_2 is the compound random variable of X_1 and itself. Continuing on, we ask that in the nth generation, each individual will produce offspring according to the same distribution as its ancestor X_1. Thus

$$X_n = \sum_{n=1}^{X_{n-1}} X_{1,n}$$

Using generating functions we can calculate the distribution of X_n. So we start with $F_{X_1}(t)$. Then by the formula for the generating function of a compound random variable,

$$F_{X_2}(t) = F_{X_1}(F_{X_1}(t))$$

and more generally,

$$F_{X_n} = F_{X_1} \circ \cdots \text{n- times} \circ F_{X_1}$$

Without generating functions, this expression would be quite difficult.

3.3 Conditional expectation

We now introduce a very subtle concept of utmost importance to us: conditional expectation. We will begin by defining it in its simplest incarnation: expectation of a random variable conditioned on an event. Given an event A, define its *indicator* to be the random variable

$$1_A = \left\{ \begin{matrix} 1 & \text{if} & \omega \in A, \\ 0 & \text{if} & \omega \notin A. \end{matrix} \right\}$$

Define

$$E(X|A) = \frac{E(X \cdot 1_A)}{P(A)}.$$

In the discrete case, this equals

$$\sum_x xP(X = x|A)$$

and in the continous case

$$\int xp(x|A)\,dx$$

Some basic properties of conditional expectation are

Proposition 3.3.1. *(Properties of conditional expectation)*

1. $E(X + Y|A) = E(X|A) + E(Y|A)$.
2. *If X is a constant c on A, then $E(XY|A) = cE(Y|A)$.*

3. *If* $B = \coprod_{i=1}^{n} A_i$, *then*

$$E(X|B) = \sum_{i=1}^{n} E(X|A_i) \frac{P(A_i)}{P(B)}$$

(Jensen's inequality) A function f is convex *if for all* $t \in [0,1]$ *and* $x < y$ f *satisfies the inequality*

$$f((1-t)x + ty) \leq (1-t)f(x) + tf(y)$$

Then for f convex, $E(f(X)) \geq f(E(X))$ *and* $E(f(X)|A) \geq f(E(X|A))$.

If X and Y are two random variables defined on the same sample space, then we can talk of the particular kind of event $\{Y = j\}$. Hence it makes sense to talk of $E(X|Y = j)$. There is nothing really new here. But a more useful way to think of the conditional expectation of X conditioned on Y is as a random variable. As j runs over the values of the variable Y, $E(X|Y = j)$ takes on different values. These are the values of a new random variable. Thus we make the following

Definition 3.3.1. Let X and Y be two random variables defined on the same sample space. Define a random variable called $E(X|Y)$ and called the *expectation of X conditioned on Y*. The values of this random variable are $E(X|Y = j)$ as j runs over the values of Y. Finally

$$P[E(X|Y) = e] = \sum_{\{j|E(X|Y=j)=e\}} P[Y = j]$$

More generally, if $Y_0, Y_1, \ldots, Y_n$ are a collection of random variables, we may define $E(X|Y_0 = y_0, Y_1 = y_1, \ldots, Y_n = y_n)$ and $E(X|Y_0,, Y_1, \ldots, Y_n)$.

Some basic properties of the random variable $E(X|Y)$ are listed in the following proposition.

Proposition 3.3.2. *Let X, Y be two random variables defined on the same sample space. Then*

1. *(Tower property)* $E(E(X|Y)) = E(X)$.

2. *If X, Y are independent, then* $E(X|Y) = E(X)$.

3. *Letting* $Z = f(Y)$, *then* $E(ZX|Y) = ZE(X|Y)$

Proof:
1.

$$P(E(X|Y) = e) = \sum_{\{j|E(X|Y=j)=e\}} P(Y = j)$$

Then

$$\begin{aligned} E(E(X|Y)) &= \sum_e eP(E(X|Y) = e) \\ &= \sum_j E(X|Y = j)P(Y = j) \\ &= \sum_j \left(\sum_k kP(X = k|Y = j)\right) P(Y = j) \\ &= \sum_j \sum_k k(P((X = k) \cap (Y = j))/P(Y = j))P(Y = j) \\ &= \sum_j \sum_k kP((X = k) \cap (Y = j)) = \sum_k \sum_j kP((X = k) \cap (Y = j)) \\ &= \sum_k kP((X = k) \cap \left(\bigcup_j Y = j\right)) = \sum_k kP(X = k) = E(X) \end{aligned} \tag{3.3.1}$$

2. If X, Y are independent, then $E(X|Y) = E(X)$. So

$$\begin{aligned} E(X|Y = j) &= \sum_k kP(X = k|Y = j) \\ &= \sum_k kP(X = k) = E(X). \end{aligned} \tag{3.3.2}$$

This is true for all j so the conditional expectation if a constant.
3. We have

$$\begin{aligned} E(ZX|Y = j) &= \sum_{k,l} klP((Z = k) \cap (X = l)|Y = j) \\ &= \sum_{k,l} klP((f(Y) = k) \cap (X = l)|Y = j) \\ &= f(j) \sum_l lP(X = l|Y = j) = ZE(X|Y = j). \end{aligned} \tag{3.3.3}$$

Q.E.D.

We also have an analogous statement for conditioning on a set of variables.

Proposition 3.3.3. *Let X, $Y_0, \ldots, Y_n$ be random variables defined on the same sample space. Then*

1. *(Tower property) $E(E(X|Y_0, \ldots, Y_n)) = E(X)$.*

2. *If X is independent of $Y_0, \ldots, Y_n$, then $E(X|Y_0, \ldots, Y_n) = E(X)$.*

3. *Letting $Z = f(Y_0, \ldots, Y_n)$, then $E(ZX|Y_0, \ldots, Y_n) = ZE(X|Y_0, \ldots, Y_n)$*

Proof: See exercises.

3.4 Discrete time finance

Now let us return to the market. If S_n denotes the value of a security at time n, is S_n a Martingale? There are two reasons why not. The first reason is the time value of money. If S_n were a Martingale, no one would buy it. You might as well just put your money under your mattress. What about the discounted security process? This would turn a bank account into a Martingale. But any other security would not be. If the discounted stock value were a Martingale, you would expect to make as much money as in your bank account and people would just stick their money in the bank. The second reason why the discounted stock price process is not a Martingale is the non-existence of arbitrage opportunities.

Let's go back to our model of the market from the last chapter. Thus we have a vector of securities S and a pay-off matrix $D = (d_{ij})$ where security i pays off d_{ij} in state j. In fact, let's consider a very simple market consisting of two securities, a bond S_1 and a stock S_2. Then the value of the stock at time 1 is d_{2j} if state j occurs. Under the no-arbitrage condition, there is a state price vector ψ so that $S = D\psi$. Setting $R = \sum_j \psi_j$ and $q_j = \psi_j/R$ as in the last chapter, we have $\sum_j q_j = 1$ and the q_j's determine a probability Q on the set of states. Now the value of the stock at time 1 is a random variable, which we will call d_2 with respect to the probabilities Q, thus, $P(d_2 = d_{2j}) = q_j$. The expected value with respect to Q of the stock at time 1 is

$$E_Q(d_2) = \sum_j d_{2,j} q_j = (D(\psi/R))_2 = (S/R)_2 = S_2/R$$

In other words, $E(Rd_2) = S_2$. That is, the expected discounted stock price at time 1 is the same as the value at time 0, with respect to the probabilities Q which were arrived at by the no-arbitrage assumption. This is the first step to finding Q such that the discounted stock prices are martingales. This is the basic phenomena that we will take advantage of in the rest of this chapter to value derivatives.

We've seen that the no-arbitrage hypothesis led to the existence of a state price vector, or to what is more or less the same thing, to a probability on the set of states such that the expected value of the discounted price of the stock at time 1 is the same as the stock price at time 0. The arbitrage theorem of last chapter only gives us an existence theorem. It does not tell us how to find Q in actual practice. For this we will simplify our market drastically and specify a more definite model for stock movements.

Our market will consist of a single stock S and a bond B.

Let us change notation a little. Now, at time 0, the stock is worth S_0 and the bond is worth B_0. At time 1, there will be only two states of nature, u (up) or d (down). That is, at time $1 \cdot \delta t$, the stock will either go up to some value S_1^u or down to S_1^d. The bond will go to $e^{r\delta t}B_0$.

$$
\begin{array}{ccc} & & S_1^u \\ & \nearrow & \\ S_0 & & \\ & \searrow & \\ & & S_1^d \end{array} \quad , \qquad B_0 \;\rightarrow\; e^{r\delta t}B_0 \tag{3.4.1}
$$

According to our picture up there we need to find a probability Q such that $E_Q(e^{-r\delta t}S_1) = S_0$. Let q be the probability that the stock goes up. Then we want

$$S_0 = E_Q(e^{-r\delta t}S_1) = qe^{-r\delta t}S_1^u + (1-q)e^{-r\delta t}S_1^d$$

Thus we need

$$q(e^{-r\delta t}S_1^u - e^{-r\delta t}S_1^d) + e^{-r\delta t}S_1^d = S_0$$

So

$$
\begin{aligned}
q &= \frac{e^{r\delta t}S_0 - S_1^d}{S_1^u - S_1^d} \\
1-q &= \frac{S_1^u - e^{r\delta t}S_0}{S_1^u - S_1^d}
\end{aligned} \tag{3.4.2}
$$

If h is a contingency claim on S, then according to our formula of last chapter, the value the derivative is

$$
\begin{aligned}
V(h) &= RE_Q(h(S_1)) \\
&= e^{-r\delta t}(qh(S_1^u) + (1-q)h(S_1^d)) \\
&= e^{-r\delta t}\left\{\left(\frac{e^{r\delta t}S_0 - S_1^d}{S_1^u - S_1^d}\right)h(S_1^u) + \left(\frac{S_1^u - e^{r\delta t}S_0}{S_1^u - S_1^d}\right)h(S_1^d)\right\} \\
&= e^{-r\delta t}\left\{h(S_1^u)\left(\frac{e^{r\delta t}S_0 - S_1^d}{S_1^u - S_1^d}\right) + h(S_1^d)\left(\frac{S_1^u - e^{r\delta t}S_0}{S_1^u - S_1^d}\right)\right\}.
\end{aligned}
$$

Let's rederive this based on an ab initio arbitrage argument for h.

Let's try to construct a replicating portfolio for h, that is, a portfolio consisting of the stock and bond so that no matter whether the stock goes up or down, the portfolio is worth the same as the derivative. Thus, suppose at time 0 we set up a portfolio

$$\theta = \begin{pmatrix} \phi \\ \psi \end{pmatrix} \tag{3.4.3}$$

ϕ units of stock S_0 and ψ units of bond B_0.

At time δt, the value of this portfolio is worth either $\phi S_1^u + \psi e^{r\delta t} B_0$ if the state is u, or $\phi S_1^d + \psi e^{r\delta t} B_0$ if the state is d. We want this value to be the same as $h(S_1)$. So we solve for ϕ and ψ:

$$\begin{aligned} \phi S_1^u + \psi e^{r\delta t} B_0 &= h(S_1^u) \\ \phi S_1^d + \psi e^{r\delta t} B_0 &= h(S_1^d) \end{aligned} \tag{3.4.4}$$

Subtracting,

$$\phi(S_1^u - S_1^d) = h(S_1^u) - h(S_1^d).$$

So we find

$$\phi = \frac{h(S_1^u) - h(S_1^d)}{S_1^u - S_1^d}$$

while

$$\begin{aligned} \psi &= \frac{h(S_1^u) - \phi S_1^u}{e^{r\delta t} B_0} \\ &= \frac{h(S_1^u) - \frac{h(S_1^u) - h(S_1^d)}{S_1^u - S_1^d} S_1^u}{e^{r\delta t} B_0} \\ &= \frac{h(S_1^u)S_1^u - h(S_1^u)S_1^d - h(S_1^u)S_1^u + h(S_1^d)S_1^u}{e^{r\delta t} B_0 (S_1^u - S_1^d)} \\ &= \frac{h(S_1^u)S_1^u - h(S_1^u)S_1^d}{e^{r\delta t} B_0 (S_1^u - S_1^d)} \end{aligned}$$

Thus, the value of the portfolio at time 0 is

$$\begin{aligned} \phi S_0 + \psi B_0 &= \frac{h(S_1^u) - h(S_1^d)}{S_1^u - S_1^d} \cdot S_0 + \frac{h(S_1^d)S_1^u - h(S_1^u)S_1^d}{e^{r\delta t} B_0 (S_1^u - S_1^d)} \cdot B_0 \\ &= \left(\frac{e^{r\delta t}(h(S_1^u) - h(S_1^d))}{e^{r\delta t}(S_1^u - S_1^d)} \right) \cdot S_0 + \frac{h(S_1^d)S_1^u - h(S_1^u)S_1^d}{e^{r\delta t}(S_1^u - S_1^d)} \\ &= e^{-r\delta t} \left(h(S_1^u) \left(\frac{e^{r\delta t} S_0 - S_1^d}{S_1^u - S_1^d} \right) + h(S_1^d) \left(\frac{S_1^u - e^{r\delta t} S_0}{S_1^u - S_1^d} \right) \right). \end{aligned} \tag{3.4.5}$$

Since the value of this portfolio and the derivative are the same at time 1, by the no-arbitrage hypothesis, they must be equal at time 0. So we've arrived the value of the derivative by a somewhat different argument.
Now we move on to a multi-period model. We are interested in defining a reasonably flexible and concrete model for the movement of stocks.
We now want to divide the time period from now, time 0 to some horizon T, which will often be the expiration time for a derivative, into N periods, each of length δt so that $N\delta t = T$.
One Step Redo with δt .

$$\begin{array}{ccc} & & S^u_{\delta t} \\ & \overset{q}{\nearrow} & \\ S_0 & & \\ & \overset{1-q}{\searrow} & \\ & & S^d_{\delta t} \end{array} \tag{3.4.6}$$

$$q = \frac{S_0 e^{r\delta t} - S^d_{\delta t}}{S^u_{\delta t} - S^d_{\delta t}}, \quad 1 - q = \frac{S^u_{\delta t} - S_0 e^{r\delta t}}{S^u_{\delta t} - S^d_{\delta t}}$$

Then for a contigency claim h we have $V(h) = e^{-r\delta t} E_Q(h)$.
Two Steps

$$\begin{array}{ccccc} & & & & \bullet_{uu} \\ & & & \overset{q^u_1}{\nearrow} & \\ & & \bullet_u & \overset{1-q^u_1}{\longrightarrow} & \bullet_{ud} \\ & \overset{q_0}{\nearrow} & & & \\ \bullet & & & & \\ & \overset{1-q_0}{\searrow} & & & \\ & & \bullet_d & \overset{1-q^d_1}{\longrightarrow} & \bullet_{du} \\ & & & \overset{q^d_1}{\searrow} & \\ & & & & \bullet_{dd} \end{array} \tag{3.4.7}$$

$$q^d_1 = \frac{e^{r\delta t} S^d_{1\delta t} - S^{dd}_{2\delta t}}{S^{du}_{2\delta t} - S^{dd}_{2\delta t}}, \qquad 1 - q^d_1 = \frac{S^{du}_{2\delta t} - e^{r\delta t} S^d_{1\delta t}}{S^{du}_{2\delta t} - S^{dd}_{2\delta t}}$$
$$q^u_1 = \frac{e^{r\delta t} S^u_{1\delta t} - S^{ud}_{2\delta t}}{S^{uu}_{2\delta t} - S^{ud}_{2\delta t}}, \qquad 1 - q^u_1 = \frac{S^{uu}_{2\delta t} - e^{r\delta t} S^u_{1\delta t}}{S^{uu}_{2\delta t} - S^{ud}_{2\delta t}}$$

$$q_0 = \frac{e^{r\delta t}S_0 - S^d_{1\delta t}}{S^u_{1\delta t} - S^d_{1\delta t}}, \qquad 1 - q_0 = \frac{S^u_{1\delta t} - e^{r\delta t}S_0}{S^u_{1\delta t} - S^d_{1\delta t}}$$

Suppose h is a contingency claim, and that we know $h(S_{2\delta t})$. Then we have

$$h(S^u_{1\delta t}) = e^{-r\delta t}(q^u_1 h(S^{uu}_{2\delta t}) + (1 - q^u_1)h(S^{ud}_{2\delta t})),$$
$$h(S^d_{1\delta t}) = e^{-r\delta t}(q^d_1 h(S^{du}_{2\delta t}) + (1 - q^d_1)h(S^{dd}_{2\delta t})).$$

Finally,

$$h(S_0) = e^{-r\delta t}(q_0 h(S^u_{1\delta t}) + (1 - q_0)h(S^d_{1\delta t}))$$
$$= e^{-r\delta t}(q_0(e^{-r\delta t}(q^u_1 h(S^{uu}_{2\delta t})+(1-q^u_1)h(S^{ud}_{2\delta t})))+(1-q_0)(e^{-r\delta t}(q^d_1 h(S^{du}_{2\delta t})+(1-q^d_1)h(S^{dd}_{2\delta t})))$$
$$= e^{-2r\delta t}(q_0 q^u_1 h(S^{uu}_{2\delta t})+q_0(1-q^u_1)h(S^{ud}_{2\delta t})+(1-q_0)q^d_1 h(S^{du}_{2\delta t})+(1-q_0)(1-q^d_1)h(S^{dd}_{2\delta t})).$$

Let's try to organize this into an N-step tree.

First, let's notice what the coefficients of $h(S^{\bullet\bullet}_{2\delta t})$ are. $q_0 q^u_1$ represents going up twice. $q_0(1 - q^u_1)$ represents going up, then down, etc.. That is, these coefficients are the probability of ending up at a particular node of the tree. Thus, if we define a sample space

$$\Psi = \{uu, ud, du, dd\}$$

$P(uu) = q_0 q^u_1, P(ud) = q_0(1-q^u_1), P(du) = (1-q_0)q^d_1, P(dd) = (1-q_0)(1-q^d_1).$

We check that this is a probability on the final nodes of the two step tree:

$$P(uu) + P(ud) + P(du) + P(dd)$$

$$= q_0 q^u_1 + q_0(1 - q^u_1) + (1 - q_0)q^d_1 + (1 - q_0)(1 - q^d_1)$$
$$= q_0(q^u_1 + 1 - q^u_1) + (1 - q_0)(q^d_1 + 1 - q^d_1) = q_0 + 1 - q_0 = 1.$$

Now $h(S_2)$ is a random variable depending on $w_0 w_1 \in \Psi$. Hence we have

$$V(h) = e^{-2r\delta t}E_Q(h(S_2)).$$

So what's the general case?

For N steps, the sample space is $\Psi = \{w_0 w_1 \dots w_{N-1} | w_i = u \text{ or } d\}$.

$$P(w_0 w_1 \dots w_{N-1}) = \prod_{j=0}^{N-1} P(w_j)$$

where

$$P(w_0) = \left\{ \begin{array}{lll} q_0 & \text{if} & w_0 = u \\ 1 - q_0 & \text{if} & w_0 = d \end{array} \right\}$$

and more generally

$$P(w_j) = (\begin{array}{lll} q_1^{w_0 \cdots w_{j-1}} & \text{if} & w_j = u \\ 1 - q_1^{w_0 \cdots w_{j-1}} & \text{if} w_j = d & \end{array}$$

$$V(h) = e^{-rN\delta t} E_Q(h(S_N))$$

3.5 Basic martingale theory

We will now begin our formal study of martingales.

Definition 3.5.1. Let M_n be a stochastic process. We say M_n is a *Martingale* if

1. $E(|M_n|) < \infty$, and
2. $E(M_{n+1}|M_n = m_n, M_{n-1} = m_{n-1}, \ldots, M_0 = m_0) = m_n$.

Noting that $E(M_{n+1}|M_n = m_n, M_{n-1} = m_{n-1}, \ldots, M_0 = m_0)$ are the values of a random variable which we called $E(M_{n+1}|M_n, M_{n-1}, \ldots, M_0)$ we may write the condition 2. Above as $E(M_{n+1}|M_n, M_{n-1}, \ldots, M_0) = M_n$.

We also define a stochastic process M_n to be a *supermartingale* if

1. $E(|M_n|) < \infty$, and
2. $E(M_{n+1}|M_n = m_n, M_{n-1} = m_{n-1}, \ldots, M_0 = m_0) \leq m_n$ (or equivalently $E(M_{n+1}|M_n, M_{n-1}, \ldots, M_0) \leq M_n$).

We say M_n is a *submartingale* if

1. $E(|M_n|) < \infty$, and
2. $E(M_{n+1}|M_n = m_n, M_{n-1} = m_{n-1}, \ldots, M_0 = m_0) \geq m_n$. (or equivalently $E(M_{n+1}|M_n, M_{n-1}, \ldots, M_0) \geq M_n$.

The following theorem is one expression of the fair game aspect of martingales.

Theorem 3.5.1. If Y_n is a

1. supermartingale, then $E(Y_n) \leq E(Y_m)$ for $n \geq m$.
2. submartingale, then $E(Y_n) \geq E(Y_m)$ for $n \geq m$.
3. martingale $E(Y_n) = E(Y_m)$ for n, m.

Proof

1. For $n \geq m$, the tower property of conditional expectation implies $E(Y_n) = E(E(Y_n|Y_m, Y_{m-1}, \ldots, Y_0)) \leq E(Y_m)$.
2. is similar.
3. is implied by 1. and 2. together.

Q.E.D.

It will often be useful to consider several stochastic processes at once, or to express a stochastic process in terms of other ones. But we will usually assume that there is some basic or driving stochastic process and that others are expressed in terms of this one and defined on the same sample space. Thus we make the following more general definition of a martingale.

Definition 3.5.2. (Generalization) Let X_n be a stochastic process. A stochastic process $M_0, M_1, \ldots$ is a *martingale* relative to $X_0, \ldots, X_n$ if

1. $E(|M_n|) < \infty$ and
2. $E(M_{n+1} - M_n|X_0 = x_0, \ldots, X_n = x_n) = 0$.

Examples

1. Let $X_n = a + \xi_1 + \cdots \xi_n$ where ξ_i are independent random variables with $E(|\xi_i|) < \infty$ and $E(\xi_i) = 0$. Then X_n is a martingale. For $E(X_{n+1}|X_0 = x_0, \ldots, X_n = x_n)$

$$= E(X_n + \xi_{n+1}|X_0 = x_0, \ldots, X_n = x_n)$$

$$= E(X_n|X_0 = x_0, \ldots, X_n = x_n) + E(\xi_{n+1}|X_0 = x_0, \ldots, X_n = x_n)$$

Now by property 1 of conditional expectation for the first summand and property 2 for the second this is

$$= X_n + E(\xi_{n+1}) = X_n$$

and thus a martingale.

A special case of this is when ξ_i are IID with $P(\xi_i = +1) = 1/2$, and $P(\xi_i = -1) = 1/2$, the symmetric random walk.

2. There is a multiplicative analogue of the previous example. Let $X_n = a\xi_1 \cdot \xi_2 \cdots \xi_n$ where ξ_i are independent random variables with $E(|\xi_i|) < \infty$ and $E(\xi_i) = 1$. Then X_n is a martingale. We leave the verification of this to you.

3. Let $S_n = S_0 + \xi_1 + \cdots + \xi_n$, S_0 constant, ξ_i independent random variables, $E(\xi_i) = 0$, $\sigma^2(\xi_i) = \sigma_i^2$. Set $v_n = \sum_{i=1}^n \sigma_i^2$. Then $M_n = S_n^2 - v_n$ is a martingale relative to $S_0, \ldots, S_n$.

   For
$$\begin{aligned} M_{n+1} &= (S_{n+1}^2 - v_{n+1}) - (S_n^2 - v_n) \\ &= (S_n + \xi_{n+1})^2 - S_n^2 - (v_{n+1} - v_n) \\ &= 2S_n\xi_{n+1} + \xi_{n+1}^2 - \sigma_{n+1}^2, \end{aligned} \tag{3.5.1}$$
   and thus $E(M_{n+1} - M_n | S_0, S_1, \ldots, S_n)$
$$= E(2S_n\xi_{n+1} + \xi_{n+1}^2 - \sigma_{n+1}^2 | S_0, \ldots, S_n)$$
$$= 2S_n E(\xi_{n+1}) + \sigma_{n+1}^2 - \sigma_{n+1}^2 = 0$$

4. There is a very general way to get martingales usually ascribed to Levy. This will be very important to us. Let X_n be a driving stochastic process and let Y be a random variable defined on the same sample space. Then we may define $Y_n = E(Y|X_n, \ldots, X_0)$. Then Y_n is a martingale. For by the tower property of conditional expectation, we have $E(Y_{n+1}|X_n, \ldots, X_0)$ is by definition
$$E(E(Y|X_{n+1}, \ldots, X_0)|X_n, \ldots, X_0)) = E(Y|X_n, \ldots, X_0) = Y_n$$

We now introduce a very useful notation. Let $\mathcal{F}(X_0, \ldots, X_n)$ be the collection of random variables that can be written as $g(X_0, \ldots, X_n)$ for some function g of $n+1$ variables. More compactly, given a stochastic process X_k, we let $\mathcal{F}_n = \mathcal{F}(X_0, \ldots, X_n)$. $\mathcal{F}_n$ represents all the information that is knowable about the stochastic process X_k at time n and we call it the *filtration* induced by X_n. We write $E(Y|\mathcal{F}_n)$ for $E(Y|X_0, \ldots, X_n)$. We restate Proposition (3.3.3) in this notation. We also have an analogous statement for conditioning on a set of variables.

Proposition 3.5.1. *Let X, $X_0, \ldots, X_n, \ldots$ be a stochastic process with filtration $\mathcal{F}_n$, and X a random variable defined on the same sample space. Then*

1. $E(E(X|\mathcal{F}_n)|\mathcal{F}_m) = E(X|\mathcal{F}_m)$ *if* $m \leq n$.
2. *If* X, $X_0, \ldots, X_n$ *are independent, then* $E(X|\mathcal{F}_n) = E(X)$.
3. *Letting* $X \in \mathcal{F}_n$. *So* $X = f(X_0, \ldots, X_n)$, *then* $E(XY|\mathcal{F}_n) = XE(Y|\mathcal{F}_n)$

Definition 3.5.3. Let X_n be a stochastic process and $\mathcal{F}_n$ its filtration. We say a process ϕ_n is

1. *adapted* to X_n (or to $\mathcal{F}_n$) if $\phi_n \in \mathcal{F}_n$ for all n.
2. *previsible* relative to X_n (or $\mathcal{F}_n$) if $\phi \in \mathcal{F}_{n-1}$.

We motivate the definitions above and the following construction by thinking of a gambling game. Let S_n be a stochastic process representing the price of a stock at the nth tick. Suppose we are playing a game and that our winnings per unit bet in the nth game is $X_n - X_n - 1$. Let ϕ_n denote the amount of our bet on the nth hand. It is a completely natural assumption that ϕ_n should only depend on the history of the game up to time $n-1$. This is the previsible hypothesis. Given a bet of ϕ_n on the nth hand our winnings playing this game up til time n will be

$$\phi_1(X_1 - X_0) + \phi_2(X_2 - X_1) + \cdots \phi_n(X_n - X_{n-1}$$

$$= \sum_{j=1}^{n} \phi_j(X_j - X_{j-1}$$

$$= \sum_{j=1}^{n} \phi_j \Delta X_j$$

This is called the martingale transform of ϕ with respect to X_n. We denote it more compactly by $(\sum \phi\Delta X)_n$. The name comes from the second theorem below, which gives us yet another way of forming new martingales from old.

Theorem 3.5.2. Let Y_n be a supermartingale relative to $X_0, X_1, \ldots$, $\mathcal{F}_n$ the filtration associated to X_n. Let ϕ_n be a previsible process and $0 \leq \phi_n \leq c_n$. Then $G_n = G_0 + \sum_{m=1}^{n} \phi_m(Y_m - Y_{m-1})$ is an $\mathcal{F}_n$-supermartingale.

Proof $G_{n+1} = G_n + \phi_{n+1}(Y_{n+1} - Y_n)$. Thus

$$
\begin{aligned}
E(G_{n+1} - G_n)|X_0, \ldots, X_n) &= E(\phi_{n+1}(Y_{n+1} - Y_n)|X_0, \ldots, X_n) \\
&= \phi_{n+1} E(Y_{n+1} - Y_n|X_0, \ldots, X_n) \\
&\leq \phi_{n+1} \cdot 0 \leq 0.
\end{aligned} \tag{3.5.2}
$$

Q.E.D.

Theorem 3.5.3. If $\phi_{n+1} \in \mathcal{F}_n$, $|\phi_n| \leq c_n$, and Y_n is a martingale relative to X_i, then $G_n = G_0 + \sum_{m=1}^n \phi_m(Y_m - Y_{m-1})$ is a martingale.

Definition 3.5.4. T, a random variable, is called a stopping time relative to X_i, if the event $\{T = n\}$ is determined by $X_0, \ldots, X_n$.

Example: In this example, we define a simple betting strategy of betting 1 until the stopping time tells us to stop. Thus, set

$$
\phi_n(\omega) = \left\{ \begin{array}{lll} 1 & \text{if} & T(\omega) \geq n, \\ 0 & \text{if} & T(\omega) < n. \end{array} \right\} \tag{3.5.3}
$$

Let's check that $\phi_n \in \mathcal{F}(X_0, \ldots, X_n)$. By definition, $\{T(\omega) = n\}$ is determined by $X_0, \ldots, X_n$. Now $\{\phi_n(\omega) = 0\} = \{T \leq n-1\} = \bigcup_{k=0}^{n-1}\{T = k\}$ is determined by $X_0, \ldots, X_{n-1}$.

Let $T \wedge n$ be the minimum of T and n. Thus

$$
(T \wedge n)(\omega) = \left\{ \begin{array}{lll} T(\omega) & \text{if} & T(\omega) < n, \\ = n & \text{if} & T(\omega) \geq n. \end{array} \right\} \tag{3.5.4}
$$

Theorem 3.5.2. *1. Let Y_n be a supermartingale relative to X_i and T a stopping time. Then $Y_{T \wedge n}$ is a supermartingale.*

2. Let Y_n be a submartingale relative to X_i and T a stopping time. Then $Y_{T \wedge n}$ is a submartingale.

3. Let Y_n be a martingale relative to X_i and T a stopping time. Then $Y_{T \wedge n}$ is a martingale.

Proof: Let ϕ_n be the previsible process defined above, i.e.,

$$
\phi_n = \left\{ \begin{array}{lll} 1 & \text{if} & n \leq T, \\ 0 & \text{if} & n > T. \end{array} \right\} \tag{3.5.5}
$$

Then

$$
\begin{aligned}
G_n &= Y_0 + \sum_{m=1}^n \phi_m(Y_m - Y_{m-1}) \\
&= Y_0 + \phi_1(Y_1 - Y_0) + \phi_2(Y_2 - Y_1) + \cdots + \phi_n(Y_n - Y_{n-1}) \\
&= \left\{ \begin{array}{lll} Y_n & \text{if} & n \le T \\ Y_T & \text{if} & n > T. \end{array} \right\} \\
&= Y_{T\wedge n}
\end{aligned} \tag{3.5.6}
$$

So it follows by the above.
Q.E.D.
Example: The following betting strategy of doubling after each loss is known to most children and is in fact THE classic martingale, the object that gives martingales their name.

Let $P(\xi = \pm 1) = \frac{1}{2}$, ξ_i are independent, identically distributed random variables, $X_n = \sum_{i=1}^n \xi_i$, which is a martingale. Now ϕ_n is the previsible process defined in the following way: $\phi_1 = 1$. To define ϕ_n we set inductively $\phi_n = 2\phi_{n-1}$ if $\xi_{n-1} = -1$ and $\phi_n = 1$ if $\xi_{n-1} = 1$. Then ϕ_n is previsible. Now define and $T = \min\{n|\xi_n = 1\}$. Now consider the process $G_n = (\sum \phi \Delta X)_n$. Then G_n is a martingale by a previous theorem. On the other hand, $E(G_T) = 1$, that is, when you finely win according to this betting strategy, you always win 1. This seems like a sure fire way to win. But it requires an arbitrarily pot to make this work. Under more realistic hypotheses, we find

Theorem 3.5.4. (Doob's optional stopping theorem) Let M_n be a martingale relative to X_n, T a stopping time relative to X_n with $P(T_\xi < \infty) = 1$. If there is a K such that $|M_{T\wedge n}| \le K$ for all n, then $E(M_T) = E(M_0)$.

Proof We know that $E(M_{T\wedge n}) = E(M_0)$ since by the previous theorem $M_{T\wedge n}$ is a martingale. Then

$$|E(M_0) - E(M_T)| = |E(M_{T\wedge n}) - E(M_T)| \le 2KP(T > n) \to 0$$

as $n \to \infty$.
Q.E.D.

An alternative version is

Theorem 3.5.5. If M_n is a martingale relative to X_n, T a stopping time, $|M_{n+1} - M_n| \le K$ and $E(T) < \infty$, then $E(M_T) = E(M_0)$.

We now discuss one last theorem before getting back to the finance. This is the martingale representation theorem. According to what we proved above, if X_n is a martingale and ϕ_n is previsible then the martingale transform

$(\sum \phi \Delta X)_n$ is also a martingale. The martingale representation theorem is a converse of this. It gives conditions under which a martingale can be expressed as a martingale transform of a given martingale. For this, we need a condition on our driving martingale. Thus we define

Definition 3.5.5. A *generalized random walk or binomial process* is a stochastic process X_n in which at each time n, the value of $X_{n+1} - X_n$ can only be one of two possible values.

This like an ordinary random walk in that at every step the random walker can take only 2 possible steps, but what the steps she can take depend on the path up to time n. Thus

$$X_{n+1} = X_n + \xi_{n+1}(X_1, X_2, \ldots, X_n)$$

Hence assume that

$$P(\xi_{n+1}(X_1, \ldots, X_n) = a) = p$$

and

$$P(\xi_{n+1}(X_1, \ldots, X_n) = b) = 1 - p$$

where I am being a little sloppy in that a, b, and p are all dependent on $X_1, \ldots, X_n$. Then

$$\begin{aligned} E(X_{n+1}|\mathcal{F}_n) &= E(X_n + \xi_n(X_0, \ldots, X_n)|\mathcal{F}_n) \\ &= X_n + E(\xi_n(X_0, \ldots, X_n)|\mathcal{F}_n) = X_n + pa + (1-p)b \end{aligned} \tag{3.5.7}$$

So in order that X_n be a martingale we need

$$pa + (1-p)b = 0 \tag{3.5.8}$$

Theorem 3.5.6. Let $X_0, X_1, \ldots$ be a generalized random walk with associated filtration $\mathcal{F}_n$. Let M_n be a Martingale relative to $\mathcal{F}_n$. Then there exists a previsible process ϕ_n such that

$$M_n = \sum_{i=1}^{n} \phi_i(X_i - X_{i-1}) = (\sum \phi \Delta X)_n$$

Proof Since M_n is adapted to X_n, there exist functions f_n such that $M_n = f_n(X_0, \ldots, X_n)$. By induction, we need to show there exists $\phi_{n+1} \in \mathcal{F}_n$ with $M_{n+1} = \phi_{n+1}(X_{n+1} - X_n) + M_n$. Now

$$\begin{aligned} 0 &= E(M_{n+1} - M_n|\mathcal{F}_n) \\ &= E(f_{n+1}(X_0, \ldots, X_{n+1}) - f_n(X_0, \ldots, X_n)|\mathcal{F}_n) \\ &= E(f_{n+1}(X_0, \ldots, X_{n+1}|\mathcal{F}_n) - f_n(X_0, \ldots, X_n). \end{aligned} \tag{3.5.9}$$

Now by assumption, X_{n+1} can take on only 2 possible values: $X_n + a$, with probability p, and X_n+b, with probability $1-p$. So $E(f_{n+1}(X_0, \ldots, X_{n+1})|\mathcal{F}_n)$

$$= f_{n+1}(X_0, \ldots, X_n, X_n + a) \cdot p + f_{n+1}(X_0, \ldots, X_n, X_n + b) \cdot (1 - p).$$

So continuing the earlier calculation,

$$\begin{aligned} 0 &= E(f_{n+1}(X_0, \ldots, X_{n+1}|\mathcal{F}_n) - f_n(X_0, \ldots, X_n) \\ &= (f_{n+1}(X_0, \ldots, X_n, X_n + a) - f_n(X_0, \ldots, X_n)) \cdot p \\ &+ (f_{n+1}(X_0, \ldots, X_n, X_n + b) - f_n(X_0, \ldots, X_n)) \cdot (1 - p) \\ &= \tfrac{(f_{n+1}(X_0,\ldots,X_n,X_n+a) - f_n(X_0,\ldots,X_n))}{a} ap \\ &+ \tfrac{(f_{n+1}(X_0,\ldots,X_n,X_n+b) - f_n(X_0,\ldots,X_n))}{b} b(1-p) \\ &= \tfrac{(f_{n+1}(X_0,\ldots,X_n,X_n+a) - f_n(X_0,\ldots,X_n))}{a} ap \\ &+ \tfrac{(f_{n+1}(X_0,\ldots,X_n,X_n+b) - f_n(X_0,\ldots,X_n))}{b} (-ap). \end{aligned} \tag{3.5.10}$$

since $b(1 - p) = -ap$ by (3.5.8). Or equivalently, (assuming neither a nor p are zero,

$$\frac{(f_{n+1}(X_0, \ldots, X_n, X_n + a) - f_n(X_0, \ldots, X_n))}{a}$$
$$= \frac{(f_{n+1}(X_0, \ldots, X_n, X_n + b) - f_n(X_0, \ldots, X_n))}{b}$$

Set ϕ_{n+1} to be equal to this common value. Clearly, ϕ_{n+1} only depends on $X_0, \ldots, X_n$ and is therefore previsible. Now we have to cases to consider: if $X_{n+1} = X_n + a$ or $X_{n+1} = X_n + b$. Both calculations are the same. So if $X_{n+1} = X_n + a$ then

$$\phi_{n+1}(X_{n+1} - X_n) + M_n = \phi_{n+1}a + M_n = M_{n+1}$$

Q.E.D.

3.6 Pricing a derivative

We now discuss the problem we were originally out to solve: to value a derivative. So suppose we have a stock whose price process we denote by S_n, where each tick of the clock represents some time interval δt. We will assume, to be concrete, that our riskless bond is given by the standard formula, $B_n = e^{nr\delta t}B_0$. We use the following notation. If V_n represents some price process, then $\tilde{V}_n = e^{-nr\delta t}V_n$ is the discounted price process. Suppose h denotes the payoff function from a derivative on S_n.
There are three steps to pricing h.
Step 1: Find Q such that $\tilde{S}_n = e^{-rn\delta t}S_n$ is a Q-martingale. Let $\mathcal{F}_n$ denote the filtration corresponding to $\tilde{S}$.
Step 2: Set $\tilde{V}_n = E(e^{-rN\delta t}h(S_N)|\mathcal{F}_n)$, then $\tilde{V}_n$ is a martingale as well. By the martingale representation theorem there exists a previsible ϕ_i such that

$$\tilde{V}_n = V_0 + \sum_{i=1}^{n} \phi_i(\tilde{S}_i - \tilde{S}_{i-1}).$$

Step 3: Construct a self-financing hedging portfolio. Just as in the one period case, we will form a portfolio Π consisting of the stock and bond. Thus $\Pi_n = \phi_n S_n + \psi_n B_n$, where $B_n = e^{rn\delta t}B_0$. So $\tilde{\Pi}_n = \phi_n \tilde{S}_n + \psi_n B_0$. We require two conditions on Π.

1. ϕ_n and ψ_n should be previsible. From a financial point of view this is a completely reasonable assumption, since decisions about what to hold during the nth tick must be made prior to the nth tick.

2. The portfolio Π should be self financing. This too is financially justified. In order for the portfolio to be viewed as equivalent to the claim h, it must act like the claim. The claim needs no infusion of money and provides no income until expiration. And this is also what we require of the portfolio.

The self-financing condition is derived from the following considerations. After tick $n-1$ the protfolio is worth $\Pi_{n-1} = \phi_{n-1}S_{n-1} + \psi_{n-1}B_{n-1}$ Between tick $n-1$ and n we readjust the portfolio in anticipation of the nth tick. But we only want to use the resources that are available to us via the portfolio itself. Also, between the $n-1$st and nth tick, the stock is still worth S_{n-1}. Therefore we require

$$(SFC1) \qquad \phi_n S_{n-1} + \psi_n B_{n-1} = \phi_{n-1}S_{n-1} + \psi_{n-1}B_{n-1}.$$

Another equivalent formulation of the self-financing condition is arrived at by

$$\begin{aligned}(\Delta\Pi)_n &= \Pi_n - \Pi_{n-1} = \phi_n S_n + \psi_n B_n - (\phi_{n-1} S_{n-1} + \psi_{n-1} B_{n-1}) \\ &= \phi_n S_n + \psi_n B_n - (\phi_n S_{n-1} + \psi_n B_{n-1}) \\ &= \phi_n (S_n - S_{n-1}) - \psi_n (B_n - B_{n-1}) \\ &= \phi_n (\Delta S)_n + \psi_n (\Delta B)_n\end{aligned} \tag{3.6.1}$$

So we have

$$(SCF2) \qquad (\Delta\Pi)_n = \phi_n (\Delta S)_n + \psi_n (\Delta B)_n.$$

Now we have $\tilde{V}_n = E_Q(e^{-rN\delta t} h(S_N)|\mathcal{F}_n)$ and we can write

$$\tilde{V}_n = V_0 + \sum_{i=1}^{n} \phi_i \Delta \tilde{S}_i,$$

for some previsible ϕ_i. Now we want $\tilde{\Pi}_n = \phi_n \tilde{S}_n + \psi_n B_0$ to equal $\tilde{V}_n$. So set

$$\psi_n = (\tilde{V}_n - \phi_n \tilde{S}_n)/B_0$$

or equivalently

$$\psi_n = (V_n - \phi_n S_n)/B_n$$

Clearly, $\tilde{\Pi}_n = \phi_n \tilde{S}_n + \psi_n B_0 = \tilde{V}_n$.
We now need to check two things:

1. previsibility of ϕ_n and ψ_n;
2. self-financing of the portfolio.

1. Previsibility: ϕ_n is previsible by the martingale representation theorem. As for ψ_n, we have

$$\begin{aligned}\psi_n &= \frac{\tilde{V}_n - \phi_n \tilde{S}_n}{B_0} \\ &= \frac{\phi_n(\tilde{S}_n - \tilde{S}_{n-1}) + \sum_{j=1}^{n-1} \phi_j (\tilde{S}_j - \tilde{S}_{j-1}) - \phi_n \tilde{S}_n}{B_0} \\ &= \frac{-\phi_n \tilde{S}_{n-1} + \sum_{j=1}^{n-1} \phi_j (\tilde{S}_j - \tilde{S}_{j-1})}{B_0}\end{aligned}$$

which is manifestly independent of $\tilde{S}_n$, hence ψ_n is previsible.
2. Self-financing: We need to verify

$$\phi_n S_{n-1} + \psi_n B_{n-1} = \phi_{n-1} S_{n-1} + \psi_{n-1} B_{n-1}$$

or what is the same thing

$$(SFC3) \qquad (\phi_n - \phi_{n-1})\tilde{S}_{n-1} + (\psi_n - \psi_{n-1})B_0 = 0$$

We calculate the right term $(\psi_n - \psi_{n-1})B_0$

$$= \left(\frac{\tilde{V}_n - \phi_n \tilde{S}_n}{B_0} - \frac{\tilde{V}_{n-1} - \phi_{n-1}\tilde{S}_{n-1}}{B_0}\right)B_0$$

$$\begin{aligned}
&= \tilde{V}_n - \tilde{V}_{n-1} + \phi_{n-1}\tilde{S}_{n-1} - \phi_n \tilde{S}_n \\
&= \phi_n(\tilde{S}_n - \tilde{S}_{n-1}) + \phi_{n-1}\tilde{S}_{n-1} - \phi_n \tilde{S}_n \\
&= -(\phi_n - \phi_{n-1})\tilde{S}_{n-1}
\end{aligned}$$

which clearly implies (SFC3).
So finally we have constructed a self-financing hedging strategy for the derivative h. By the no-arbitrage hypothesis, we must have that the value of h at time n (including time 0 or now) is $V(h)_n = \Pi_n = V_n = e^{rn\delta t}\tilde{V}_n = e^{rn\delta t}E_Q(e^{-rN\delta t}h(S_N))$

$$= e^{-r(N-n)\delta t}E_Q(h(S_N))$$

3.7 Dynamic portfolio selection

In this section we will use our martingale theory and derivative pricing theory to solve a problem in portfolio selection. We will carry this out in a rather simplified situation, but the techniques are valid in much more generality.

Thus suppose that we want to invest in a stock, S_n and the bond B_n. In our first go round, we just want to find a portfolio $\Pi_n = (\phi_n, \psi_n)$ which maximizes something. To maximize the expected value of Π_N where N represents the time horizon, will just result in investing everything in the stock, because it has a higher expected return. This does not take into account our aversion to risk. It is better to maximize $E(U(\Pi_n))$ where U is a (differentiable) utility function that encodes our risk aversion. U is required to satisfy two conditions:

1. U should be increasing, which just represents that more money is better than less, and means $U'(x) > 0$
2. U' should be decreasing, that is $U'' < 0$ which represents that the utility of the absolute gain is decreasing, that is the utility of going from a billion dollars to a billion and one dollars is less than the utility of going from one hundred dollars to one hundred and one dollars.

Some common utility functions are

$$\begin{aligned} U_0(x) &= \ln(x) \\ U_p(x) &= \frac{x^p}{p} \text{ for } 0 \neq p < 1 \end{aligned} \tag{3.7.1}$$

We will carry out the problem for $U(x) = U_0(x) = \ln(x)$ and leave the case of $U_p(x)$ as a problem.

So our problem is to now maximize $E_P(\ln(\Pi_n))$ but we have constraints. Here P represents the real world probabilities, since we want to maximize in the real world, not the risk neutral one. First, want that the value of our portfolio at time 0 is the amount that we are willing to invest in it. So we want $\Pi_0 = W_0$ for some initial wealth W_0. Second, if we want to let this portfolio go, with no other additions or withdrawals, then we want it to be self financing. Thus Π_n should satisfy SFC. And finally, for Π_n to be realistically tradeable, it also must satisfy that it is previsible.

In chapter 3 we saw that there was an equivalence between two sets of objects:

1. Self-financing previsible portfolio processes Π_n and
2. derivatives h on the stock S.

If in addition we want the value of the portfolio to satisfy $\Pi_0 = W_0$ then our problem

$$\max(E_P(\ln(\Pi_N))) \text{ subject to } \Pi_0 = 0$$

can be reformulated as

$$\max(E_P(\ln(h))) \text{ subject to the value of } h \text{ at time } 0 \text{ is } W_0$$

And now we may use our derivative pricing formula to look at this condition, since the value of a derivative h at time 0 is

$$E_Q(\exp(-rT)h)$$

where Q is the risk neutral probabilities. Thus the above problem becomes

$$\max(E_P(\ln(h))) \text{ subject to } E_Q(\exp(-rT)h)) = W_0$$

Now we are in a position where we are really using two probabilities at the same time. We need the following theorem which allows us to compare them.

Theorem 3.7.1. *(Radon-Nikodym) Given two probabilities P and Q assume that if $P(E) = 0$ then $Q(E) = 0$ for any event E. Then there exists a random variable ξ such that*

$$E_Q(X) = E_P(\xi X)$$

for any random variable X.

Proof. In our discrete setting this theorem is easy. For each outcome ω, set $\xi(\omega) = Q(\omega)/P(\omega)$, if $P(\omega) \neq 0$ and 0 otherwise. Then

$$E_Q(X) = \sum_{\omega} X(\omega)Q(\omega)$$

$$= \sum_{\omega} X(\omega)P(\omega)\frac{Q(\omega)}{P(\omega)} = E_P(\xi X)$$

□

So now our problem is

$$\max(E_P(\ln(h))) \text{ subject to } E_P(\xi \exp(-rT)h)) = W_0$$

We can now approach this extremal problem by the method of Lagrange multipliers.

According to this method, to find an extremum of the function

$$\max(E_P(\ln(h))) \text{ subject to } E_P(\xi \exp(-rT)h)) = W_0$$

we introduce the mutliplier λ. And we solve

$$\frac{d}{d\epsilon}|_{\epsilon=0}(E_P(\ln(h+\epsilon v))) - \lambda E_P(\xi \exp(-rT)(h+\epsilon v))) = 0$$

for all random variables v. Now the left hand side is

$$\begin{aligned} E_P(\frac{d}{d\epsilon}|_{\epsilon=0} \ln(h+\epsilon v)) - \lambda E_P(\frac{d}{d\epsilon}|_{\epsilon=0} \xi \exp(-rT)(h+\epsilon v)) = \\ E_P(\frac{1}{h}v - \lambda\xi \exp(-rT)v) = \\ E_P([\frac{1}{h} - \lambda\xi\exp(-rT)]v) = \end{aligned} \tag{3.7.2}$$

and we want this to be 0 for all random variables v. In order for this expectation to be 0 for all v, it must be the case that

$$\frac{1}{h} - \lambda\xi\exp(-rT) = 0$$

almost surely. Hence $h = \exp(rT)/(\lambda\xi)$. To calculate λ we put h back into the constraint: $W_0 = E_P(\xi\exp(-rT)h)$

$$\begin{aligned} &= E_P(\xi\exp(-rT)\cdot\exp(rT)/(\lambda\xi)) \\ &= E_P(1/\lambda) \end{aligned} \tag{3.7.3}$$

So $\lambda = 1/W_0$ and thus finally we have

$$h = W_0\exp(rT)/\xi$$

To go further we need to write down a specific model for the stock. Thus let $X_1, X_2, \ldots,$ be IID with law $P(X_j = \pm 1) = 1/2$. Set S_0 be the initial value of the stock and inductively

$$S_n = S_{n-1}(1 + \mu\cdot\delta t + \sigma\sqrt{\delta t}X_n)$$

In other words, S_n is just a binomial stock model with

$$u = (1 + \mu\cdot\delta t + \sigma\sqrt{\delta t})$$

and

$$d = (1 + \mu\cdot\delta t - \sigma\sqrt{\delta t})$$

The sample space for this process can be taken to be sequences $x_1x_2x_3\cdots x_N$ where x_i is u or d representing either an up move or a down move.

Then by our formula for the risk neutral probability

$$q = \frac{\exp r\delta t - d}{u-d} = \frac{r\delta t - (\mu\delta t - \sigma\sqrt{\delta t})}{(\mu\delta t + \sigma\sqrt{\delta t}) - (\mu\delta t - \sigma\sqrt{\delta t})}$$

up to first order in δt. So

$$q = \frac{1}{2}\left(1 + (\frac{r-\mu}{\sigma})\sqrt{\delta t}\right)$$

and

$$1 - q = \frac{1}{2}\left(1 - (\frac{r-\mu}{\sigma})\sqrt{\delta t}\right)$$

This then determines the random variable ξ in the Radon-Nikodym theorem, 3.7.1:

$$\xi(x_1 x_2 x_3 \cdots x_N) = \xi_1(x_1)\xi_2(x_2)\cdots\xi_N(x_N)$$

where x_i denotes either u or d and

$$\xi_n(u) = \frac{q}{p} = \left(1 + (\frac{r-\mu}{\sigma})\sqrt{\delta t}\right) \tag{3.7.4}$$

and

$$\xi_n(d) = \frac{1-q}{1-p} = \left(1 - (\frac{r-\mu}{\sigma})\sqrt{\delta t}\right)$$

We now know which derivative has as its replicating portfolio the portfolio that solves our optimization problem. We now calculate what the replicating portfolio actually is. For this we use our theory from chapter 3 again. This tells us that we take

$$\widetilde{V}_n = E_Q(\exp(-rT)h|\mathcal{F}_n)$$

Then by the martingale representation theorem there is a previsible process ϕ_n such that

$$\widetilde{V}_n = \widetilde{V}_0 + \sum_{k=1}^{n} \phi_k(\widetilde{S}_k - \widetilde{S}_{k-1})$$

Moreover, it tells us how pick ϕ_n,

$$\phi_n = \frac{\widetilde{V}_n - \widetilde{V}_{n-1}}{\widetilde{S}_n - \widetilde{S}_{n-1}}$$

Now let's recall that by the fact that $\widetilde{S}_n$ is a martingale with respect to Q

$$\widetilde{S}_{n-1} = E_Q(\widetilde{S}_n|\mathcal{F}_{n-1}) = q\widetilde{S}_n^u + (1-q)\widetilde{S}_n^d$$

where $\widetilde{S}_n^u$ denotes the value of $\widetilde{S}_n$ if S takes an uptick at time n. Similarly, we have

$$\widetilde{V}_{n-1} = E_Q(\widetilde{V}_n|\mathcal{F}_{n-1}) = q\widetilde{V}_n^u + (1-q)\widetilde{V}_n^d$$

From this, we see that if S takes an uptick at time n that

$$\phi_n = \frac{\widetilde{V}_n - \widetilde{V}_{n-1}}{\widetilde{S}_n - \widetilde{S}_{n-1}} = \frac{\widetilde{V}_n^u - (q\widetilde{V}_n^u + (1-q)\widetilde{V}_n^d)}{\widetilde{S}_n^u - (q\widetilde{S}_n^u + (1-q)\widetilde{S}_n^d)} = \frac{\widetilde{V}_n^u - \widetilde{V}_n^d}{\widetilde{S}_n^u - \widetilde{S}_n^d}$$

A similar check will give the same exact answer if S took a downtick at time n. (After all, ϕ_n must be previsible and therefore independent of whether S takes an uptick or downtick at time n.) Finally let's remark that ϕ_n can be expressed in terms of the undiscounted prices

$$\phi_n = \frac{\widetilde{V}_n^u - \widetilde{V}_n^d}{\widetilde{S}_n^u - \widetilde{S}_n^d} = \frac{V_n^u - V_n^d}{S_n^u - S_n^d}$$

since both the numerator and the denominator are expressions at the same time n and therefore the discounting factor for the numerator and the denominator cancel.

Now we use the following refinement of the Radon-Nikodym theorem.

Theorem 3.7.2. *Given two probabilities P and Q assume that if $P(E) = 0$ then $Q(E) = 0$ for any event E. Let ξ be the random variable such that*

$$E_Q(X) = E_P(\xi X)$$

for any random variable X. Then for any event A, we have

$$E_Q(X|A) = \frac{E_P(\xi X|A)}{E_P(\xi|A)}$$

I will ask you to prove this.

From this theorem it follows that

$$\widetilde{V}_n = E_Q(\widetilde{h}|\mathcal{F}_n) = \frac{E_P(\xi\widetilde{h}|\mathcal{F}_n)}{E_P(\xi|\mathcal{F}_n)}$$

Therefore we see that

$$\widetilde{V}_n = \frac{\widetilde{V}_{n-1}}{\xi_n(x_n)}$$

recalling that x_n denotes the nth tick and ξ_n is given by (3.7.4).

Now we are ready to finish our calculation. $\widetilde{V}_n = \frac{\widetilde{V}_{n-1}}{\xi_n(x_n)}$ so

$$\widetilde{V}_n^u = \frac{\widetilde{V}_{n-1}}{\xi_n(u)} = \frac{\widetilde{V}_{n-1}}{\left(1 + (\frac{r-\mu}{\sigma})\sqrt{\delta t}\right)}$$

$$\widetilde{V}_n^d = \frac{\widetilde{V}_{n-1}}{\xi_n(d)} = \frac{\widetilde{V}_{n-1}}{\left(1 - (\frac{r-\mu}{\sigma})\sqrt{\delta t}\right)}$$

We also have

$$S_n^u = S_{n-1}(1 + \mu\delta t + \sigma\sqrt{\delta t})$$

so

$$\widetilde{S}_n^u = \widetilde{S}_{n-1}(1 + (\mu - r)\delta t + \sigma\sqrt{\delta t})$$

up to first order in δt. Similarly

$$\widetilde{S}_n^d = \widetilde{S}_{n-1}(1 + (\mu - r)\delta t - \sigma\sqrt{\delta t})$$

to first order in δt. So we have

$$\begin{aligned}
\phi_n &= \frac{\widetilde{V}_n^u - \widetilde{V}_n^d}{\widetilde{S}_n^u - \widetilde{S}_n^d} \\
&= \frac{\widetilde{V}_{n-1}\left(\frac{1}{\left(1+(\frac{r-\mu}{\sigma})\sqrt{\delta t}\right)} - \frac{1}{\left(1-(\frac{r-\mu}{\sigma})\sqrt{\delta t}\right)}\right)}{\widetilde{S}_{n-1}(2\sigma\sqrt{\delta t})} \\
&= \frac{\widetilde{V}_{n-1}}{\widetilde{S}_{n-1}}\left\{\frac{\left(1-(\frac{r-\mu}{\sigma})\sqrt{\delta t}\right) - \left(1+(\frac{r-\mu}{\sigma})\sqrt{\delta t}\right)}{\left(1-(\frac{r-\mu}{\sigma})^2\delta t\right)2\sigma\sqrt{\delta t}}\right\} \\
&= \frac{\widetilde{V}_{n-1}}{\widetilde{S}_{n-1}}\left\{-\frac{2(\frac{r-\mu}{\sigma})\sqrt{\delta t}}{2\sigma\sqrt{\delta t}}\right\}
\end{aligned} \tag{3.7.5}$$

up to terms of order δt. So we have

$$\phi_n = \frac{\widetilde{V}_{n-1}(\frac{\mu-r}{\sigma^2})}{\widetilde{S}_{n-1}} = \frac{V_{n-1}(\frac{\mu-r}{\sigma^2})}{S_{n-1}} \tag{3.7.6}$$

the last equality is because the quotient of the two discounting factors cancel. V_{n-1} represents the value of the portfolio at time $n-1$, and ϕ_n represents the number of shares of stock in this portfolio. Hence the amount of money invested in the stock is $\phi_n S_{n-1} = (\frac{\mu-r}{\sigma^2})V_{n-1}$. Thus formula (3.7.6) expresses the fact that at every tick, you should have the proportion $\frac{\mu-r}{\sigma^2}$ of your portfolio invested in the stock.

Over the 50 year period from 1926 to 1985, the S&P 500 has had excess return $\mu - r$ over the short term treasury bills of about .058 with a standard deviation of σ of .216. Doing the calculation with the values above yields $.058/(.216)^2 = 1.23 = 123\%$. That is, with log utility, the optimal portfolio is one where one invests all ones wealth in the S&P 500, and borrows an additional 23% of current wealth and invests it also in the S&P.

3.8 Exercises

1. Let S_0, S_1^u, S_1^d be as in class, for the one-step binomial model. For the pseudo-probabilities q of going up and $1-q$ of going down that we calculated in class, what is $E_Q(S_1)$.

2. We arrived at the following formula for discounting. If B is a bond with par value 1 at time T, then at time $t < T$ the bond is worth $\exp(-r(T-t))$ assuming constant interest rates. Now assume that interest rates are not constant but are described by a known function of time $r(t)$. Recalling the derivation of the formula above for constant interest rate, calculate the value of the same bond under the interest rate function $r(t)$. Your answer should have an integral in it.

3. A fair die is rolled twice. Let X denote the sum of the values. What is $E(X|A)$ where A is the event that the first roll was a 2.

4. Toss a nickel, a dime and a quarter. Let B_i denote the event that i of the coins landed heads up, $i = 0, 1, 2, 3$. Let X denote the sum of the amount shown, that is, a coin says how much it is worth only on the tales side, so X is the sum of the values of the coins that landed tales up. What is $E(X|B_i)$.

5. Let X be a random variable with pdf $p(x) = \alpha \exp(-\alpha x)$, (a continuous distribution.) Calculate $E(X|\{X \geq t\})$.

In the next three problems calculate $E(X|Y)$ for the two random variables X and Y. That is, describe $E(X|Y)$ as a function of the values of Y. Also, construct the pdf for this random variable.

6. A die is rolled twice. X is the sum of the values, and Y is the value on the first roll.

7. From a bag containing four balls numbered $1, 2, 3$ and 4 you draw two balls. If at least one of the balls is numbered 2 you win \$1 otherwise

you lose a \$1. Let X denote the amount won or lost. Let Y denote the number on the first ball chosen.

8. A bag contains three coins, but one of them is fake and has two heads. Two coins are drawn and tossed. Let X be the number of heads tossed and Y denote the number of honest coins drawn.

9. Let W_n denote the symmetric random walk starting at 0, that is $p = 1/2$ and $W_0 = 0$. Let X denote the number of steps to the right the random walker takes in the first 6 steps. Calculate

   (a) $E(X|W_1 = -1)$.

   (b) $E(X|W_1 = 1)$.

   (c) $E(X|W_1 = 1, W_2 = -1)$

   (d) $E(X|W_1 = 1, W_2 = -1, W_3 = -1)$

10. We've seen that if ξ_n are independent with expectations all 0, then

$$M_n = \sum_{j=1}^{n} \xi_j$$

is a Martingale. It is not necessary for the ξ_j's to be independent. In this exercise you are asked to prove that, though they may not be independent, they are uncorrelated. Thus, let M_n be a Martingale and let $\xi_j = M_j - M_{j-1}$ be its increments. Show

   (a) $E(\xi_j) = 0$ for all j.

   (b) $E(\xi_j \xi_k) = 0$ for all $j \neq k$. (This means that ξ_j and ξ_k have 0 covariance, since their expectations are 0.) (Hint: You will have to use several of the properties of conditional expectations.)

In the following three problems, you are asked to derive what is called the gambler's ruin using techniques from Martingale theory.

Suppose a gambler plays a game repeatedly of tossing a coin with probability p of getting heads, in which case he wins \$1 and $1-p$ of getting tails in which case he loses \$1. Suppose the gambler begins with \$$a$

and keeps gambling until he either goes bankrupt or gets \$$b$, $a < b$. Mathematically, let ξ_j, $j = 1, 2, 3, \ldots$ be IID random variables with distribution $P(\xi_j = 1) = p$ and $P(\xi_j = -1) = 1 - p$, where $0 < p < 1$. Let S_n denote the gamblers fortune after the nth toss, that is

$$S_n = a + \xi_1 + \cdots + \xi_n$$

Define the stopping time, $T = \min\{n | S_n = 0 \text{ or } b\}$ that is, T is the time when the gambler goes bankrupt or achieves a fortune of b. Define $A_n = S_n - n(2p - 1)$ and $M_n = (\frac{1-p}{p})^{S_n}$.

11. Show A_n and M_n are Martingales.

12. Calculate $P(S_T = 0)$, the probability that the gambler will eventually go bankrupt, in the following way. Using Doob's optional stopping theorem, calculate $E(M_T)$. Now use the definition of $E(M_T)$ in terms of $P(S_T = j)$ to express $P(S_T = 0)$ in terms of $E(M_T)$ and solve for $P(S_T = 0)$.

13. Calculate $E(S_T)$.

14. Calculate $E(T)$ the expected time till bust or fortune in the following way: calculate $E(A_T)$ by the optional stopping theorem. Express $E(A_T)$ in terms of $E(S_T)$ and $E(T)$. Solve for $E(T)$.

In the following problems, you are asked to find an optimal betting strategy using Martingale methods.

Suppose a gambler is playing a game repeatedly where the winnings of a \$1 bet pay-off \$1 and the probability of winning each round is $p > 1/2$ is in the gambler's favor. (By counting cards, blackjack can be made such a game.) What betting strategy should the gambler adopt. It is not hard to show that if the gambler wants to maximize his expected fortune after n trials, then he should bet his entire fortune each round. This seems counter intuitive and in fact doesn't take into much account the gambler's strong interest not to go bankrupt. It turns out to make more sense to maximize not ones expected fortune, but the expected value of the logarithm his fortune.

As above, let ξ_j be IID with $P(\xi_j = 1) = p$ and $P(\xi_j = -1) = 1 - p$ and assume this time that $p > 1/2$. Let F_n denote the fortune of the gambler at time n. Assume the gambler starts with \$$a$ and adopts a betting strategy ϕ_j, (a previsible process), that is, in the jth round he bets \$$\phi_j$. Then we can express his fortune

$$F_n = a + \sum_{j=1}^{n} \phi_j \xi_j$$

We assume that $0 < \phi_n < F_{n-1}$, that is the gambler must bet at least something to stay in the game, and he doesn't ever want to bet his entire fortune. Let $e = p\ln(p) + (1-p)\ln(1-p) + \ln 2$.

15. Show $L_n = \ln(F_n) - ne$ is a super Martingale and thus $E(\ln(F_n/F_0)) \leq ne$.

16. Show that for some strategy, L_n is a Martingale. What is the strategy? (Hint: You should bet a fixed percentage of your current fortune.)

17. Carry out a similar analysis to the above using the utility function $U_p(x) = \frac{x^p}{p}$, $p < 1$ not equal to 0. Reduce to terms of order δt. You will have to use Taylor's theorem. Your answer should not have a δt in it.

Chapter 4

Stochastic processes and finance in continuous time

In this chapter we begin our study of continuous time finance. We could develop it similarly to the model of chapter 3 using continuous time martingales, martingale representation theorem, change of measure, etc. But the probabilistic infra structure needed to handle this is more formidable. An approach often taken is to model things by partial differential equations, which is the approach we will take in this chapter. We will however discuss continuous time stochastic processes and their relationship to partial differential equations in more detail than most books.

We will warm up by progressing slowly from stochastic processes discrete in both time and space to processes continuous in both time and space.

4.1 Stochastic processes continuous in time, discrete in space

We segue from our completely discrete processes by now allowing the time parameter to be continuous but still demanding space to be discrete. Thus our second kind of stochastic process we will consider is a process X_t where the time t is a real-valued deterministic variable and at each time t, the sample space Ψ of the random variable X_t is a (finite or infinite) discrete set. In such a case, a state X of the entire process is described as a sequence of jump times $(\tau_1, \tau_2, \ldots)$ and a sequence of values of X (elements of the common discrete state space) $j_0, j_1, \ldots$. These describe a "path" of the system through S given by $X(t) = j_0$ for $0 < t < \tau_1$, $X_t = j_1$ for $\tau_1 < t < \tau_2$, etc... Because X moves by a sequence of jumps, these processes are often called "jump" processes.

Thus, to describe the stochastic process, we need, for each $j \in S$, a continuous random variable F_j that describes how long it will be before the process (starting at $X = j$) will jump, and a set of discrete random variables Q_j that describe the probabilities of jumping from j to other states when the jump occurs. We will let Q_{jk} be the probability that the state starting at j will jump to k (when it jumps), so

$$\sum_k Q_{jk} = 1 \qquad \text{and} \qquad Q_{jj} = 0$$

We also define the function $F_j(t)$ to be the probability that the process starting at $X = j$ has already jumped at time t, i.e.,

$$F_j(t) = P(\tau_1 < t \,|\, X_0 = j).$$

All this is defined so that

1. $$P(\tau_1 < t, X_{\tau_1} = k | X_0 = j) = F_j(t) Q_{jk} \tag{4.1.1}$$

   (this is an independence assumption: where we jump to is independent of when we jump).

   Instead of writing $P(\ldots | X_0 = j)$, we will begin to write $P_j(\ldots)$.

   We will make the usual stationarity and Markov assumptions as follows: We assume first that

2. $$\begin{aligned} &P_j(\tau_1 < s, X_{\tau_1} = k, \tau_2 - \tau_1 < t, X_{\tau_2} = l) \\ &= P(\tau_1 < s, X_{\tau_1} = k, \tau_2 - \tau_1 < t, X_{\tau_2} = l \,|\, X_0 = j) \\ &= F_j(s) Q_{jk} F_k(t) Q_{kl}. \end{aligned}$$

   It will be convenient to define the quantity $P_{jk}(t)$ to be the probability that $X_t = k$, given that $X_0 = j$, i.e.,

   $$P_{jk}(t) = P_j(X_t = k).$$

3. The Markov assumption is

$$P(X_t = k | X_{s_1} = j_1, \ldots, X_{s_n} = j_n, X_s = j) = P_{jk}(t - s).$$

The three assumptions 1, 2, 3 will enable us to derive a (possibly infinite) system of differential equations for the probability functions $P_{jk}(t)$ for j and $k \in S$. First, the stationarity assumption M1 tells us that

$$P_j(\tau_1 > t + s \,|\, \tau_1 > s) = P_j(\tau_1 > t).$$

This means that the probability of having to wait at least one minute more before the process jumps is independent of how long we have waited already, i.e., we can take any time to be $t = 0$ and the rules governing the process will not change. (From a group- theoretic point of view this means that the process is invariant under time translation). Using the formula for conditional probability ($P(A|B) = P(A \cap B)/P(B)$), the fact that the event $\tau_1 > t + s$ is a subset of the event $\tau_1 > s$, we translate this equation into

$$\frac{1 - F_j(t + s)}{1 - F_j(s)} = 1 - F_j(t)$$

This implies that the function 1-$F_j(t)$ is a multiplicative function, i.e., it must be an exponential function. It is decreasing since $F_j(t)$ represents the probability that something happens before time t, and $F_j(0)$=0, so we must have that 1-$F_j(t)$=$e^{-q_j t}$ for some positive number q_j. In other words,

F_j is exponentially distributed with probability density function

$$f_j(t) = q_j e^{-q_j t} \quad \text{for } t \geq 0$$

To check this, note that

$$P_j(\tau_1 > t) = 1 - F_j(t) = \int_t^\infty q_j e^{-q_j s} ds = e^{-q_j t}$$

so the two ways of computing $F_j(t)$ agree .

The next consequence of our stationarity and Markov assumptions are the Chapman-Kolmorgorov equations in this setting:

$$P_{jk}(t + s) = \sum_{l \in S} P_{jl}(t) P_{lk}(s).$$

They say that to calculate the probability of going from state j to state k in time $t+s$, one can sum up all the ways of going from j to k via any l at the intermediate time t. This is an obvious consistency equation, but it will prove very useful later on.

Now we can begin the calculations that will produce our system of differential equations for $P_{jk}(t)$. The first step is to express $P_{jk}(t)$ based on the time of the first jump from j. The result we intend to prove is the following:

$$P_{jk}(t) = \delta_{jk} e^{-q_j t} + \int_0^t q_j e^{q_j s} \left(\sum_{l \neq j} Q_{jl} P_{lk}(t-s) \right) ds$$

As intimidating as this formula looks, it can be understood in a fairly intuitive way. We take it a term at a time:

$\delta_{jk} e^{-q_j t}$: This term is the probability that the first jump hasn't happened yet, i.e., that the state has remained j from the start. That is why the Kronecker delta symbol is there (recall that $\delta_{jk} = 1$ if $j = k$ and $= 0$ if j is different from k). This term is zero unless $j = k$, in which case it gives the probability that $\tau_1 > t$, i.e., that no jump has occurred yet.

The integral term can be best understood as a "double sum", and then by using the Chapman Kolmorgorov equation. For any s between 0 and t, the definition of $f_j(t)$ tells us that the probability that the first jump occurs between time s and time s+ds is (approximately) $f_j(s)ds = q_j e^{-q_j s} ds$ (where the error in the approximation goes to zero faster than ds as ds goes to zero). Therefore, the probability that for l different from j, the first jump occurs between time s and time s+ds and the jump goes from j to l is (approximately) $q_j e^{-q_j s} Q_{jl} ds$. Based on this, we see that the probability that the first jump occurs between time s and time s+ds, and the jump goes from j to l, and in the intervening time (between s and t) the process goes from l to k is (approximately)

$$q_j e^{-q_j s} Q_{jl} P_{lk}(t-s) ds$$

To account for all of the ways of going from j to k in time t, we need to add up these probabilities over all possible times s of the first jump (this gives an integral with respect to ds in the limit as ds goes to 0), and over all possible results l different from j of the first jump. This reasoning yields the second term of the formula.

We can make a change of variables in the integral term of the formula: let s be replaced by t-s. Note that this will transform ds into -ds, but it will also reverse the limits of integration. The result of this change of variables is (after taking out factors from the sum and integral that depend neither upon l nor upon s) :

$$P_{jk}(t) = \delta_{jk}e^{-q_j t} + q_j e^{-q_j t}\int_0^t e^{q_j s}\left(\sum_{l\neq j} Q_{jl}P_{lk}(s)\right) ds$$

$$= e^{-q_j t}\left(\delta_{jk} + q_j\int_0^t e^{q_j s}\left(\sum_{l\neq j} Q_{jl}P_{lk}(s)\right) ds\right)$$

Now $P_{jk}(t)$ is a continuous function of t because it is defined as the product of a continuous function with a constant plus a constant times the integral of a bounded function of t (the integrand is bounded on any bounded t interval). But once P_{jk} is continuous, we can differentiate both sides of the last equation with respect to t and show that $P_{jk}(t)$ is also differentiable.

We take this derivative (and take note of the fact that when we differentiate the first factor we just get a constant times what we started with) using the fundamental theorem of calculus to arrive at:

$$P'_{jk}(t) = -q_j P_{jk}(t) + q_j\sum_{l\neq j} Q_{jl}P_{lk}(t).$$

A special case of this differential equation occurs when t=0. Note if we know that $X_0 = j$, then it must be true that $P_{jk}(0) = \delta_{jk}$. This observation enables us to write:

$$P'_{jk}(0) = -q_j P_{jk}(0) + q_j\sum_{l\neq j} Q_{jl}P_{lk}(0)$$

$$= -q_j\delta_{jk} + q_j\sum_{l\neq j} Q_{jl}\delta_{lk}$$

$$= -q_j\delta_{jk} + q_j Q_{jk}.$$

One last bit of notation! Let $q_{jk} = P'_{jk}(0)$. So $q_{jj} = -q_j$ and $q_{jk} = q_j Q_{jk}$ if $j \neq k$. (Note that this implies that

$$\sum_{k\neq j} q_{jk} = q_j = -q_{jj}$$

so the sum of the q_{jk} as k ranges over all of S is 0. the constants q_{jk} a re called the infinitesimal parameters of the stochastic process.

Given this notation, we can rewrite our equation for $P'_{jk}(t)$ given above as follows:

$$P'_{jk}(t) = \sum_{l \in S} q_{jl} P_{lk}(t)$$

for all $t > 0$. This system of differential equations is called the backward equation of the process.

Another system of differential equations we can derive is the forward equation of the process. To do this, we differentiate both sides of the Chapman Kolmogorov equation with respect to s and get

$$P'_{jk}(t+s) = \sum_{l \in S} P_{jl}(t) P'_{lk}(s)$$

If we set s=0 in this equation and use the definition of q_{jk}, we get the forward equation:

$$P'_{jk}(t) = \sum_{l \in S} q_{lk} P_{jl}(t).$$

4.2 Examples of Stochastic processes continuous in time, discrete in space

The simplest useful example of a stochastic process that is continuous in time but discrete in space is the Poisson process. The Poisson process is a special case of a set of processes called birth processes. In these processes, the sample space for each random variable X_t is the set of non-negative integers, and for each non-negative integer j, we choose a positive real parameter λ_j which will be equal to both $q_{j,j+1}$ and $-q_j$. Thus, in such a process, the value of X can either jump up by 1 or stay the same. The time until the process jumps up by one, starting from X=j, is exponentially distributed with parameter λ_j. So if we think of X as representing a population, it is one that experiences only births (one at a time) at random times, but no deaths.

The specialization in the Poisson process is that λ_j is chosen to have the same value, λ, for all j. Since we move one step to the right (on the j-axis) according to the exponential distribution, we can conclude without much thought that

$$P_{jk}(t) = 0 \quad \text{for } k < j \text{ and } t > 0$$

and that

$$P_{jj}(t) = e^{-\lambda t}$$

for all $t > 0$. To learn any more, we must consider the system of differential equations (the "forward equations") derived in the preceding lecture.

Recall that the forward equation

$$P'_{jk}(t) = \sum_{l \in S} q_{lk} P_{jl}(t)$$

gave the rate of change in the probability of being at k (given that the process started at j) in terms of the probabilities of jumping to k from any l just at time t (the terms of the sum for $l \neq k$) and (minus) the probability of already being at k an d jumping away just at time t (recall that q_{kk} is negative.

For the Poisson process, the forward equations simplify considerably, because q_{jk} is nonzero only if $j = k$ or if $k = j + 1$. And for these values we have:

$$q_{jj} = -q_j = -\lambda \qquad \text{and} \qquad q_{j,j+1} = \lambda$$

Therefore the forward equation reduces to:

$$P'_{jk}(t) = \sum_{l=0}^{\infty} q_{lk} P_{jl}(t) = q_{k-1,k} P_{j,k-1}(t) + q_{kk} P_{jk}(t)$$

$$= \lambda P_{j,k-1}(t) - \lambda P_{jk}(t).$$

Since we already know $P_{jk}(t)$ for $k \leq j$, we will use the preceding differential equation inductively to calculate $P_{jk}(t)$ for $k > j$, one k at a time.

To do this, though, we need to review a little about linear first-order differential equations. In particular, we will consider equations of the form

$$f' + \lambda f = g$$

where f and g are functions of t and λ is a constant. To solve such equations, we make the left side have the form of the derivative of a product by multiplying through by $e^{\lambda t}$:

$$e^{\lambda t} f' + \lambda e^{\lambda t} f = e^{\lambda t} g$$

We recognize the left as the derivative of the product of $e^{\lambda t}$ and f:

$$(e^{\lambda t} f)' = e^{\lambda t} g$$

We can integrate both sides of this 9with s substituted for t) from 0 to t to recover f(t):

$$\int_0^t \frac{d}{ds}(e^{\lambda s} f(s))\, ds = \int_0^t e^{\lambda s} g(s)\, ds$$

Of course, the integral of the derivative may be evaluated using the (second) fundamental theorem of calculus:

$$e^{\lambda t} f(t) - f(0) = \int_0^t e^{\lambda s} g(s)\, ds$$

Now we can solve algebraically for f(t):

$$f(t) = e^{-\lambda t} f(0) + e^{-\lambda t} \int_0^t e^{\lambda s} g(s)\, ds$$

Using this formula for the solution of the differential equation, we substitute $P_{jk}(t)$ for f(t), and $\lambda P_{j,k-1}(t)$ for g(t). We get that

$$P_{jk}(t) = e^{-\lambda t} P_{jk}(0) + \lambda \int_0^t e^{-\lambda(t-s)} P_{j,k-1}(s)\, ds$$

Since we are assuming that the process starts at j, i.e., $X_0 = j$, we know that $P_{jk}(0) = 0$ for all $k \neq j$ and $P_{jj}(0) = 1$. In other words

$$P_{jk}(0) = \delta_{jk}$$

(this is the Kronecker delta).

Now we are ready to compute. Note that we have to compute $P_{jk}(t)$ for all possible nonnegative integer values of j and k. In other words, we have to fill in all the entries in the table:

$$\begin{array}{cccccc}
P_{00}(t) & P_{01}(t) & P_{02}(t) & P_{03}(t) & P_{04}(t) & \cdots \\
P_{10}(t) & P_{11}(t) & P_{12}(t) & P_{13}(t) & P_{14}(t) & \cdots \\
P_{20}(t) & P_{21}(t) & P_{22}(t) & P_{23}(t) & P_{24}(t) & \cdots \\
P_{30}(t) & P_{31}(t) & P_{32}(t) & P_{33}(t) & P_{34}(t) & \cdots \\
\vdots & \vdots & \vdots & \vdots & \ddots & \cdots
\end{array}$$

Of course, we already know part of the table:

$$\begin{array}{cccccc} e^{-\lambda t} & P_{01}(t) & P_{02}(t) & P_{03}(t) & P_{04}(t) & \cdots \\ 0 & e^{-\lambda t} & P_{12}(t) & P_{13}(t) & P_{14}(t) & \cdots \\ 0 & 0 & e^{-\lambda t} & P_{23}(t) & P_{24}(t) & \cdots \\ 0 & 0 & 0 & e^{-\lambda t} & P_{34}(t) & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots & \cdots \end{array}$$

So we begin the induction:

We know that $P_{jj} = e^{-\lambda t}$, so we can use the differential equation

$$P'_{j,j+1}(t) + \lambda P_{j,j+1}(t) = \lambda P_{jj}(t)$$

to get that

$$P_{j,j+1}(t) = \lambda \int_0^t e^{-\lambda(t-s)} e^{-\lambda s} ds = \lambda t e^{-\lambda t}$$

Note that this fills in the next diagonal of the table.

Then we can use the differential equation

$$P'_{j,j+2}(t) + \lambda P_{j,j+2}(t) = \lambda P_{j,j+1}(t)$$

to get that

$$P_{j,j+2}(t) = \lambda \int_0^t e^{-\lambda(t-s)} \lambda s e^{-\lambda s} ds = \frac{(\lambda t)^2}{2} e^{-\lambda t}$$

Continuing by induction we get that

$$P_{jk}(t) = \frac{(\lambda t)^{k-j}}{(k-j)!} e^{-\lambda t}$$

provided $0 \leq j \leq k$ and $t > 0$. This means that the random variable X_t (this is the infinite discrete random variable that gives the value of j at time t, i.e., tells how many jumps have happened up until time t, has a Poisson distribution with parameter λt This is the important connection between the exponential distribution and the Poisson distribution.

4.3 The Wiener process

4.3.1 Review of Taylor's theorem

Taylor series are familiar to every student of calculus as a means writing functions as power series. The real force of Taylor series is what is called Taylor's theorem which tells one how well the Taylor polynomials approximate the function. We will be satisfied with the following version of Taylor's theorem. First, we introduce some notation.

Definition 4.3.1. For two functions f and g, we say that $f(h)$ is $o(g(h))$ (as $h \to 0$) (read "f is little o of g) if

$$\lim_{h \to 0} f(h)/g(h) = 0$$

It of course means that $f(h)$ is smaller than $g(h)$ (as h gets small) but in a very precise manner.

As an example, note that another way to say that $f'(a) = b$ is to say that

$$f(a+h) - (f(a) + bh) = o(h)$$

That is, that the difference between the function $h \mapsto f(a+h)$ and the linear function $h \mapsto f(a) + bh$ is small as $h \to 0$.

Taylor's theorem is just the higher order version of this statement.

Theorem 4.3.1. *Let f be a function defined on an interval (α, β) be n times continuously differentiable. Then for any $a \in (\alpha, \beta)$*

$$f(a+h) = f(a) + f'(a)h + \frac{f''(a)}{2!}h^2 + \frac{f^{(3)}(a)}{3!}h^3 + \cdots + \frac{f^{(n)}(a)}{n!}h^n + o(h^n)$$

We will in fact need a multivariable version of this, but only of low order.

Theorem 4.3.2. *Let $f(x, y)$ be a function of two variables defined on an open set of $O \subset \mathbb{R}^2$. Then for any point $(a, b) \in O$ one has*

$$f(a+k, b+k) = f(a,b) + \frac{\partial f}{\partial x}(a,b)h + \frac{\partial f}{\partial y}(a,b)k$$

$$+\frac{1}{2}(\frac{\partial^2 f}{\partial x^2}(a,b)h^2 + 2\frac{\partial^2 f}{\partial x \partial y}(a,b)kh + \frac{\partial^2 f}{\partial y^2}(a,b)k^2) + o(h^2, k^2, hk)$$

4.3.2 The Fokker-Plank equation

We will give a derivation of the Fokker-Planck equation, which governs the evolution of the probability density function of a random variable-valued function X_t that satisfies a "first-order stochastic differential equation". The variable evolves according to what in the literature is called a Wiener process, named after Norbert Wiener, who developed a great deal of the theory of stochastic processes during the first half of this century.

We shall begin with a generalized version of random walk (an "unbalanced" version where the probability of moving in one direction is not necessarily the same as that of moving in the other). Then we shall let both the space and time increments shrink in such a way that we can make sense of the limit as both Δx and Δy approach 0.

Recall from the previous sections the one-dimensional random walk (on the integers) beginning at $W_0 = 0$. We fix a parameter p, and let p be the probability of taking a step to the right, so that $1 - p$ is the probability of taking a step to the left, at any (integer-valued) time n. In symbols, this is:

$$p = P(W_{n+1} = x + 1 | W_n = x) \quad \text{and} \quad 1 - p = P(W_{n+1} = x - 1 | W_n = x).$$

Thus

$$W_n = \xi_1 + \cdots + \xi_n$$

where ξ_j are IID with $P(\xi_j = 1) = p$ and $P(\xi_j = -1) = 1 - p$. Since we are given that $W_0 = 0$, we know that the probability function of W_n will be essentially binomially distributed (with parameters n and p, except that at time n, x will go from $-n$ to $+n$, skipping every other integer as discussed earlier). We will set $v_{x,n} = P(W_n = x)$. Then

$$v_{x,n} = \binom{n}{\frac{1}{2}(n+x)} p^{\frac{1}{2}(n+x)} (1-p)^{\frac{1}{2}(n-x)}$$

for $x = -n, -(n-2), \ldots, (n-2), n$ and $v_{x,n} = 0$ otherwise.

For use later, we calculate the expected value and variance of W_n. First,

$$E(W_n) = \sum_{x=-n}^{n} x v_{x,n} = \sum_{k=0}^{n} (2k - n) v_{x,2k-n} = \sum_{k=0}^{n} (2k - n) \binom{n}{k} p^k (1-p)^{n-k}$$

$$= n(2p - 1)$$

(it's an exercise to check the algebra; but the result is intuitive). Also,

$$\mathrm{Var}(W_n) = \sum_{k=0}^{n}(2k-n)^2 v_{x,2k-n} = \sum_{k=0}^{n}(2k-n^2)\binom{n}{k}p^k(1-p)^{n-k}$$

$$= 4np(1-p)$$

(another exercise!).

Now lets form a new stochastic process W^Δ and we let the space steps Δx and time steps Δt be different from 1. W^Δ is a shorthand for $W^{\Delta t, \Delta x}$. First, the time steps. We will take time steps of length Δt. This will have little effect on our formulas, except everywhere we will have to note that the number n of time steps needed to get from time 0 to time t is

$$n = \frac{t}{\Delta t}$$

Each step to the left or right has length Δx instead of length 1. This has the effect of scaling W by the factor Δx. Since multiplying a random variable by a constant has the effect of multiplying its expected value by the same constant, and its variance by the square of the constant.

Therefore, the mean and variance of the total displacement in time t are:

$$E(W_t^\Delta) = \frac{t}{\Delta t}(2p-1)\Delta x \quad \text{and} \quad \mathrm{Var}(W_t^\Delta) = 4p(1-p)t\frac{(\Delta x)^2}{\Delta t}$$

We want both of these quantities to remain finite for all fixed t, since we don't want our process to have a positive probability of zooming off to infinity in finite time, and we also want $\mathrm{Var}(W(t))$ to be non-zero. Otherwise the process will be completely deterministic.

We need to examine what we have control over in order to achieve these ends. First off, we don't expect to let Δx and Δt to go to zero in any arbitrary manner. In fact, it seems clear that in order for the variance to be finite and non-zero, we should insist that Δt be roughly the same order of magnitude as $(\Delta x)^2$. This is because we don't seem to have control over anything else in the expression for the variance. It is reasonable that p and q vary with Δx, but we probably don't want their product to approach zero (that would force our process to move always to the left or always to the right), and it can't approach infinity (since the maximum value of $p(1-p)$ is $1/4$). So we are forced to make the assumption that $\frac{(\Delta x)^2}{\Delta t}$ remain bounded as they both go to zero. In fact, we will go one step further. We will insist that

$$\frac{(\Delta x)^2}{\Delta t} = 2D \tag{4.3.1}$$

for some (necessarily positive) constant D.

Now, we are in a bind with the expected value. The expected value will go to infinity as $\Delta x \to 0$, and the only parameter we have to control it is $2p - 1$. We need $2p - 1$ to have the same order of magnitude as Δx in order to keep $E(W_t^\Delta)$ bounded. Since $p = 1 - p$ when $p = 1/2$, we will make the assumption that

$$p = \frac{1}{2} + \frac{c}{2D}\Delta x \tag{4.3.2}$$

for some real (positive, negative or zero) constant c, which entails

$$1 - p = \frac{1}{2} - \frac{c}{2D}\Delta x$$

The reason for the choice of the parameters will become apparent. D is called the diffusion coefficient, and c is called the drift coefficient.

Note that as Δx and Δt approach zero, the values $v_{n,k} = P(W_{n\Delta t}^\Delta = k\Delta x) = p(k\Delta x, n\Delta t)$ specify values of the function p at more and more points in the plane. In fact, the points at which p is defined become more and more dense in the plane as Δx and Δt approach zero. This means that given any point in the plane, we can find a sequence of points at which such an p is defined, that approach our given point as Δx and Δt approach zero. We would like to be able to assert the existence of the limiting function, and that it expresses how the probability of a continuous random variable evolves with time. We will not quite do this, but, assuming the existence of the limiting function, we will derive a partial differential equation that it must satisfy. Then we shall use a probabilistic argument to obtain a candidate limiting function. Finally, we will check that the candidate limiting function is a solution of the partial differential equation. This isn't quite a proof of the result we would like, but it touches on the important steps in the argument.

We begin with the derivation of the partial differential equation. To do this, we note that for the original random walk, it was true that $v_{x,n}$ satisfies the following difference equation:

$$v_{k,n+1} = pv_{k-1,n} + (1-p)v_{k+1,n}$$

This is just the matrix of the (infinite discrete) Markov process that defines the random walk. When we pass into the Δx, Δt realm, and use $p(x,t)$ notation instead of $v_{k\Delta x, n\Delta t}$, this equation becomes:

$$(*) \qquad p(x, t+\Delta t) = pp(x-\Delta x, t) + (1-p)p(x+\Delta x, t)$$

Next, we need Taylor's expansion for a function of two variables. As we do the expansion, we must keep in mind that we are going to let $\Delta t \to 0$, and as we do this, we must account for all terms of order up to and including that of Δt. Thus, we must account for both Δx and $(\Delta x)^2$ terms (but $(\Delta x)^3$ and $\Delta t \Delta x$ both go to zero faster than Δt).

We expand each of the three expressions in equation (*) around the point (x,t), keeping terms up to order Δt and $(\Delta x)^2$. We get:

$$p(x, t+\Delta t) = p(x,t) + \Delta(t)\frac{\partial p}{\partial t}(x,t) + \cdots$$

$$p(x+\Delta x, t) = p(x,t) + \Delta(x)\frac{\partial p}{\partial x}(x,t) + \frac{(\Delta x)^2}{2}\frac{\partial^2 p}{\partial x^2}(x,t) + \cdots$$

$$p(x-\Delta x, t) = p(x,t) - \Delta(x)\frac{\partial p}{\partial x}(x,t) + \frac{(\Delta x)^2}{2}\frac{\partial^2 p}{\partial x^2}(x,t) + \cdots$$

We substitute these three equations into (*), recalling that $p + (1-p) = 1$, we get

$$\Delta t \frac{\partial p}{\partial t} = -(2p-1)\Delta x \frac{\partial p}{\partial x} + \frac{(\Delta x)^2}{2}\frac{\partial^2 p}{\partial x^2}$$

(where all the partial derivatives are evaluated at (x,t)). Finally, we divide through by Δt, and recall that $(\Delta x)^2/\Delta t = 2D$, and $2p-1 = \Delta x c/D$ to arrive at the equation:

$$\frac{\partial p}{\partial t} = -2c\frac{\partial p}{\partial x} + D\frac{\partial^2 p}{\partial x^2}$$

This last partial differential equation is called the Fokker-Planck equation.

To find a solution of the Fokker-Planck equation, we use a probabilistic argument to find the limit of the random walk distribution as Δx and Δt approach 0. The key step in this argument uses the fact that the binomial distribution with parameters n and p approaches the normal distribution with expected value np and variance $np(1-p)$ as n approaches infinity. In terms of our assumptions about Δx, Δt, p and $1-p$, we see that at time t, the binomial distribution has mean

$$t(2p-1)\frac{\Delta x}{\Delta t} = 2ct$$

and variance

$$4p(1-p)t\frac{(\Delta x)^2}{\Delta t}$$

which approaches $2Dt$ as $\Delta t \to 0$. Therefore, we expect W_t in the limit to be normally distributed with mean $2ct$ and variance $2Dt$. Thus, the probability that $W_t < x$ is

$$P(W_t < x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\frac{x-2ct}{\sqrt{2Dt}}} e^{-\frac{1}{2}y^2} dy$$

Therefore the pdf of W_t is the derivative of this with respect to x, which yields the pdf

$$p(x,t) = \frac{1}{2\sqrt{\pi Dt}} e^{-\frac{(x-2ct)^2}{4Dt}}$$

A special case of this is when $c = 0$ and $D = \frac{1}{2}$. The corresponding process is called the *Wiener process* (or *Brownian motion*). The following properties characterize the Wiener process, which we record as a theorem.

Theorem 4.3.3. *1. $W_0 = 0$ with probability 1.*

2. W_t is a continuous function of t with probability 1.

3. For $t > s$ $W_t - W_s$ is normally distributed with mean 0 and variance $t - s$.

4. For $T \geq s \geq t \geq u \geq v$, $W_s - W_t$ and $W_u - W_v$ are independent.

5. The probability density for W_t is

$$p(x,t) = \frac{1}{\sqrt{2\pi t}} e^{-\frac{x^2}{2t}}$$

Proof: Let's discuss these points.
1. This is clear since we defined W_t as a limit of random walks which all started at 0.
3. This we derived from the Fokker-Plank equation.
4. Again, we defined W_t as a limit of random walks, W^Δ. Now $W_t^\Delta = \Delta x(\xi_1 + \cdots \xi_n)$ where $n = t/\Delta t$. So if $t = n\Delta t$, $s = m\Delta t$, $u = k\Delta t$ and $v = l\Delta t$ where $n \geq m \geq k \geq l$ then we have

$$W_t - W_s = \Delta x(\xi_{m+1} + \xi_{m+2} + \cdots \xi_n)$$

and

$$W_u - W_v = \Delta x(\xi_{l+1} + \cdots \xi_k)$$

which are independent.
2. This fact is beyond the scope of our discussion but is intuitive.
Q.E.D.
While, according to 2. above, the paths W_t are continuous, they are very rough and we see according to the next proposition that they are not differentiable, with probability 1.

Proposition 4.3.4. *For all $t > 0$ and all $n = 1, 2, \ldots$,*

$$\lim_{h \to 0} P(|W_{t+h} - W_t| > n|h|) = 1$$

so that W_t is nowhere differentiable.

Proof: Assume without loss of generality that $h > 0$. Then

$$P(|W_{t+h} - W_t| > nh) = P(|W_h| > nh)$$

by 3) of **??**. Now this gives

$$1 - P(|W_h| \leq nh)$$

Now

$$P(|W_h| \leq nh)) = \frac{1}{\sqrt{2\pi h}} \int_{-nh}^{nh} e^{-\frac{x^2}{2t}} dx$$

$$\leq \frac{1}{\sqrt{2\pi h}} \int_{-nh}^{nh} 1 dx \leq \frac{1}{\sqrt{2\pi h}} 2hn$$

$$= \sqrt{\frac{2h}{\pi}}$$

which converges to 0 as $h \to 0$ for any n. Q.E.D.

Almost everything we use about the Wiener process will follow from **??**. We now do some sample computations to show how to manipulate the Wiener process and some of which will be useful later.
1. Calculate the expected distance of the Wiener process from 0.
Solution: This is just $E(|W_t|)$ which is

$$\int_{-\infty}^{\infty} |x| \frac{1}{\sqrt{2\pi t}} e^{-\frac{x^2}{2t}}$$

$$= 2\int_0^\infty x \frac{1}{\sqrt{2\pi t}} e^{-\frac{x^2}{2t}}$$

Changing variables, $u = x^2/(2t)$ this integral becomes

$$\frac{2t}{\sqrt{2\pi t}} \int_0^\infty e^{-u} du = \sqrt{2t/\pi}$$

2. Calculate $E(W_t W_s)$.
Solution: Assume without loss of generality that $t > s$. Then

$$E(W_t W_s) = E((W_t - W_s)W_s + W_s^2) = E((W_t - W_s)W_s) + E(W_s^2)$$

Now $W_t - W_s$ is independent of W_s since $t > s$ and W_s has mean 0, therefore we get

$$E(W_t - W_s)E(W_s) + \sigma^2(W_s) = s$$

since $W_t - W_s$ has mean 0.

The Wiener Process Starting at y, W^y

The defining properties of $W_t^y = W_t + y$, the Wiener process starting at y, are:

1. $W_0^y = y$.
2. W_t^y is continuous with probability 1.
3. W_t^y is normally distributed with $\mu = y$, $\sigma^2 = t$.
4. $W_t^y - W_s^y$ is normally distributed with $\mu = 0$, $\sigma^2 = t - s$, $t > s$.
5. $W_t^y - W_s^y$, $W_u^y - W_v^y$ are independent for $t \geq s \geq u \geq v$.
6. The pdf is given by

$$p(x, t|y) = \frac{1}{\sqrt{2\pi t}} e^{-\frac{(x-y)^2}{2t}}$$

$$P(a \leq W_t^y \leq b) = \frac{1}{\sqrt{2\pi t}} \int_a^b e^{-\frac{(x-y)^2}{2t}} dx$$

How do we use Wiener Process to model stock movements?

We would like to make sense of the idea that $\frac{dS_t}{dt} = \mu S_t + \text{noise}(\sigma, S_t)$, where if the noise term were absent then S_t is a bank account earning interest u. There are two obvious problems:
(1) THe Wiener process is differentiable **nowhere**, so why should S_t be differentiable anywhere?
(2) "Noise" is very difficult to make rigorous.

To motivate our way out of these problems let us look at deterministic motion. The most general 1st order differential equation is

$$\frac{df}{dt} = \phi(f(t), t)\,.$$

By integrating both sides we arrive at the following integral equation

$$f(t) - f(t_0) = \int_{t_0}^{t} \frac{df}{dt}\,dt = \int_{t_0}^{t} \phi(f(t), t)\,dt\,.$$

That is,

$$f(t) = f(t_0) + \int_{t_0}^{t} \phi(f(t), t)\,dt\,.$$

Theoretically integral equations are better behaved than differential equations (they tend to smooth out bumps). So we try to make the corresponding conversion from differential to integral equation in the stochastic world.

Our first goal is just to define a reasonable notion of stochastic integral. We had the **Martingale Transform**: If S_n was a stock, ϕ_n denoted the amount of stock at tick n, then the gain from tick 0 to tick N was

$$\sum_{i=1}^{N} \phi_i (S_i - S_{i-1})$$

It is these sorts of sums that become integrals in continuous time.

Recall the **Riemann integral** $\int_a^b f(x)\,dx$: Take a partition of $[a, b]$, $\mathcal{P} = \{a = t_0 < t_1 < \cdots < t_n = b\}$.

Theorem 4.3.5. *(Riemann's Theorem) Let f be a function such that the probability of choosing a point of discontinuity is 0. Let mesh $\mathcal{P} = \max\{|t_i - t_{i-1}|\,|i = 1, \ldots, n\}$, let $t_{i-1} < \xi_i < t_i$, then let*

$$RS(f, \mathcal{P}) = \sum_{i=1}^{n} f(\xi_i)(t_i - t_{i-1})\,.$$

(RS stands for Riemann sum). Then

$$\lim_{mesh_{\mathcal{P}}\to 0} RS(f, \mathcal{P})$$

exists, and is independent of the choices of ξ_i within their respective intervals. We call this limit the Riemann integral

$$\int_a^b f(x)dx$$

Note that the summation is beginning to look like a martingale transform.

The next level of integral technology is the **Riemann-Stieltjes integral**. Suppose $\phi(x)$ is a non decreasing function, $\mathcal{P}$ a partition, and ξ_i values between partition points. Let

$$\text{RSS}(f, \mathcal{P}, \phi) = \sum_{i=1}^{n} f(\xi_i)(\phi(t_i) - \phi(t_{i-1}))\,.$$

("RSS" stands for Riemann-Stieltjes sum.) If $\lim_{\text{mesh}_{\mathcal{P}}\to 0} \text{RSS}(f, \mathcal{P}, \phi)$ exists, then by definition, $\int_a^b f(t)\,d\phi(t)$ is this limit. Assuming ϕ is differentiable, the corresponding Riemann's theorem holds, namely

$$\lim_{\text{mesh}_{\mathcal{P}}\to 0} \text{RSS}(f, \mathcal{P}, \phi)$$

exists (and is independent of the choices of ξ_i).

Proposition 4.3.6. *If ϕ is differentiable, $\int_a^b f(t)\,d\phi(t) = \int_a^b f(t)\phi'(t)\,dt$.*

Proof. Let $\mathcal{P} = \{a = t_0 < t_1 < \cdots < t_n = b\}$.

$$\begin{aligned} \text{RSS}(f, \mathcal{P}, \phi) &= \textstyle\sum_{i=1}^n f(\xi_i)(\phi(t_i) - \phi(t_{i-1})) \\ &= \textstyle\sum_{i=1}^n f(\xi_i)\frac{\phi(t_i)-\phi(t_{i-1})}{t_i - t_{i-1}}(t_i - t_{i-1}) \\ &= \textstyle\sum_{i=1}^n f(\xi_i)(\phi'(t_i) + o(|t_i - t_{i-1}|^2))(t_i - t_{i-1}) \end{aligned} \tag{4.3.3}$$

and so $\lim_{\text{mesh}_{\mathcal{P}}\to 0} \sum_{i=1}^n f(\xi_i)\phi'(t_i)(t_i - t_{i-1}) = \int_a^b f(t)\phi'(t)\,dt$. □

Example. Let $\phi(x)$ be differentiable.

$$\int_a^b \phi(t)\,d\phi(t) = \int_a^b \phi(t)\phi'(t)\,dt = \frac{1}{2}\phi(b)^2 - \frac{1}{2}\phi(a)^2\,.$$

A stochastic integral is formed much like the Riemann-Stieltjes integral, with one twist.

For X_t and ϕ_t stochastic processes,

$$\int_a^b \phi_t dX_t$$

is to be a random variable. In the Riemann and Riemann-Stieltjes cases, the choices of the points ξ_i in the intervals were irrelevant, in the stochastic case it is very much relevant. As an example, take $\mathcal{P} = \{a = t_0 < t_1 < \cdots < t_n = b\}$. Consider

$$\sum_{i=1}^{n} \phi_{\xi_i}(X_{t_i} - X_{t_{i-1}})\,.$$

Let's work out the above sum for $X_t = \phi_t = W_t$, where W_t is the Wiener process, using two different choices for the ξ_i's.

First, let $\xi_i = t_{i-1}$, the left endpoint. Then

$$\begin{aligned}
\mathrm{E}\Big(\textstyle\sum_{i=1}^n W_{t_{i-1}}(W_{t_i} - W_{t_{i-1}})\Big) &= \textstyle\sum_{i=1}^n \mathrm{E}(W_{t_{i-1}}(W_{t_i} - W_{t_{i-1}})) \\
&= \textstyle\sum_{i=1}^n \mathrm{E}(W_{t_{i-1}})\mathrm{E}(W_{t_i} - W_{t_{i-1}}) = 0\,,
\end{aligned}$$

using independence.

Second, let $\xi_i = t_i$, the right endpoint. Then

$$\begin{aligned}
\mathrm{E}\Big(\textstyle\sum_{i=1}^n W_{t_i}(W_{t_i} - W_{t_{i-1}})\Big) &= \textstyle\sum_{i=1}^n \mathrm{E}(W_{t_i}(W_{t_i} - W_{t_{i-1}})) = \sum_{i=1}^n \mathrm{E}({W_{t_i}}^2 - W_{t_i}W_{t_{i-1}}) \\
&= \textstyle\sum_{i=1}^n \mathrm{E}({W_{t_i}}^2) - \mathrm{E}(W_{t_i}W_{t_{i-1}}) \\
&= \textstyle\sum_{i=1}^n t_i - t_{i-1} \\
&= (t_1 - t_0) + (t_2 - t_1) + \cdots + (t_n - t_{n-1}) \\
&= t_n - t_0 = b - a\,.
\end{aligned}$$

Which endpoint to choose? Following Itô, one uses the left-hand endpoint. One defines the **Itô sum**

$$\mathrm{IS}(\phi_t, \mathcal{P}, X_t) = \sum_{i=1}^{n} \phi_{t_{i-1}}(X_{t_i} - X_{t_{i-1}})$$

.

There are two motivations for this choice.

1. The lefthand choice looks a little like a previsibility condition. For example, if X_t is a stock-process and ϕ_t a portfolio process then the left-hand choice embodies the idea that our gains from time T_{i-1} to time t_i should be based on a choice made at time t_{i-1}.

2. It turns out, that if X_t is a Martingale, then the Ito sum will be a Martingale as well.

Following Stratonovich, one takes $\xi_i = \frac{1}{2}(t_{i-1} + t_i)$, the midpoint. As an advantage, the resulting Stratonovich calculus looks like traditional calculus. As a disadvantage, the derived stochastic process is not a martingale.

Let X_t be a stochastic process. Let ϕ_t be another stochastic process defined on the same sample space as X_t.

Definition 4.3.2. ϕ_t is **adapted** to X_t if ϕ_t is independent of $X_s - X_t$ for $s > t$.

Then we define the **Itô integral**

$$\int_a^t \phi_s \, dX_s \equiv \lim_{\text{mesh}\mathcal{P} \to 0} \text{IS}(\phi, \mathcal{P}, X) = \lim_{\text{mesh}\mathcal{P} \to 0} \sum_{i=1}^{n} \phi_{t_{i-1}}(X_{t_i} - X_{t_{i-1}}) \,.$$

Let us call a random variable with finite first and second moments *sqaure integrable*. One knows that for X, Y square integrable random variables defined on the same sample space, $P(X = Y) = 1$ iff $\text{E}(X) = \text{E}(Y)$ and $\text{Var}(X - Y) = 0$. Now $\text{Var}(X - Y) = \text{E}((X - Y)^2) - (\text{E}(X - Y))^2$, so

$$P(X = Y) = 1 \text{ iff } \text{E}((X - Y)^2) = 0 \,.$$

This is a criterion for checking whether two square integrable random variables are equal (almost surely).

Definition 4.3.3. A sequence X^α of square integrable random variables converges in mean-square to X if $\text{E}((X - X^\alpha)^2) \to 0$ as $\alpha \to 0$. In symbols, $\lim X^\alpha = X$. This is equivalent to $\text{E}(X^\alpha) \to \text{E}(X)$ and $\text{Var}(X - X^\alpha) \to 0$.

This is the sense of the limit in the definition of the stochastic integral.

Here is a sample computation of a stochastic integral. Let's evaluate

$$\int_a^b W_t \, dW_t$$

We need the following facts: $\mathrm{E}(W_t) = 0$, $\mathrm{E}({W_t}^2) = t$, $\mathrm{E}({W_t}^3) = 0$, and $\mathrm{E}({W_t}^4) = 3t^2$.

$$\begin{aligned}
\mathrm{IS}(W_t, \mathcal{P}, W_t) &= \textstyle\sum_{i=1}^n W_{t_{i-1}} \cdot \Delta W_{t_i} = \sum_{i=1}^n W_{t_{i-1}}(W_{t_i} - W_{t_{i-1}}) \\
&= \tfrac{1}{2} \textstyle\sum_{i=1}^n (W_{t_{i-1}} + \Delta W_{t_i})^2 - {W_{t_{i-1}}}^2 - (\Delta W_{t_i})^2 \\
&= \tfrac{1}{2} \textstyle\sum_{i=1}^n {W_{t_i}}^2 - {W_{t_{i-1}}}^2 - \tfrac{1}{2} \sum_{i=1}^n (\Delta W_{t_i})^2 \\
&= \tfrac{1}{2}(({W_{t_1}}^2 - {W_{t_0}}^2) + ({W_{t_2}}^2 - {W_{t_1}}^2) + \cdots \\
&\quad + ({W_{t_n}}^2 - {W_{t_{n-1}}}^2)) - \tfrac{1}{2} \textstyle\sum_{i=1}^n (\Delta W_{t_i})^2 \\
&= \tfrac{1}{2}({W_b}^2 - {W_a}^2) - \tfrac{1}{2} \textstyle\sum_{i=1}^n (\Delta W_{t_i})^2 \,.
\end{aligned}$$

The first part is recognizable as what the Riemann-Stieltjes integral gives formally. But what is $\sum_{i=1}^n (\Delta W_{t_i})^2$ as $\mathrm{mesh}\mathcal{P} \to 0$? The simple calculation

$$\mathrm{E}\Big(\sum_{i=1}^n (\Delta W_{t_i})^2\Big) = \sum_{i=1}^n \mathrm{E}((\Delta W_{t_i})^2)$$

$$= \sum_{i=1}^n \mathrm{E}((W_{t_i} - W_{t_{i-1}})^2) = \sum_{i=1}^n t_i - t_{i-1} = b - a$$

leads us to the

Claim:

$$\lim_{\mathrm{mesh}\mathcal{P} \to 0} E(\sum_{i=1}^n (\Delta W_{t_i})^2) = b - a$$

and therefore

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

Proof.

$$\mathrm{E}\Big((\sum_{i=1}^n (\Delta W_{t_i})^2 - (b-a))^2\Big)$$

$$\begin{aligned}
&= \mathrm{E}\Big((\textstyle\sum_{i=1}^n (\Delta {W_{t_i}}^2))^2 - 2(b-a) \sum_{i=1}^n (\Delta W_{t_i})^2 + (b-a)^2\Big) \\
&= \mathrm{E}\Big(\textstyle\sum_{i=1}^n (\Delta W_{t_i})^4 + 2 \sum_{i<j} (\Delta W_{t_i})^2 (\Delta W_{t_j})^2 - 2(b-a) \sum_{i=1}^n (\Delta W_{t_i})^2 + (b-a)^2\Big) \\
&= \textstyle\sum_{i=1}^n 3(\Delta t_i)^2 + 2 \sum_{i<j} \mathrm{E}((\Delta W_{t_i})^2) \mathrm{E}((\Delta W_{t_j})^2) - \mathrm{E}(2(b-a) \sum_{i=1}^n (\Delta W_{t_i})^2 + (b-a)^2) \\
&= 3 \textstyle\sum_{i=1}^n (\Delta t_i)^2 + 2 \sum_{i<j} (\Delta t_i)(\Delta t_j) - 2(b-a) \sum_{i=1}^n \Delta t_i + (b-a)^2 \\
&= 2 \textstyle\sum_{i=1}^n (\Delta t_i)^2 + \Big[\sum_{i=1}^n (\Delta t_i)^2 + 2 \sum_{i<j} (\Delta t_i)(\Delta t_j) \Big] - 2(b-a) \sum_{i=1}^n \Delta t_i + (b-a)^2 \\
&= 2 \textstyle\sum_{i=1}^n (\Delta t_i)^2 + \Big(\sum_{i=1}^n (\Delta t_i) \Big)^2 - 2(b-a) \sum_{i=1}^n \Delta t_i + (b-a)^2 \\
&= 2 \textstyle\sum_{i=1}^n (\Delta t_i)^2 + \Big(\sum_{i=1}^n \Delta t_i - (b-a) \Big)^2 = 2 \sum_{i=1}^n (\Delta t_i)^2
\end{aligned}$$

Now for each i, $\Delta t_i \leq \operatorname{mesh}\mathcal{P}$, so $(\Delta t_i)^2 \leq \Delta t_i \cdot \operatorname{mesh}\mathcal{P}$. (Note that only one Δt_i is being rounded up.) Summing $i = 1$ to n, $0 \leq \sum_{i=1}^{n}(\Delta t_i)^2 \leq \big(\sum_{i=1}^{n}\Delta t_i\big)\cdot \operatorname{mesh}\mathcal{P} = (b-a)\cdot\operatorname{mesh}\mathcal{P} \to 0$ as $\operatorname{mesh}\mathcal{P} \to 0$. So

$$\lim_{\operatorname{mesh}\mathcal{P}\to 0}\sum_{i=1}^{n}(\Delta t_i)^2 = 0\,.$$

Intuitively, $\Delta t_i \approx \operatorname{mesh}\mathcal{P} \approx \frac{1}{n}$, so $\lim_{\operatorname{mesh}\mathcal{P}\to 0}\sum_{i=1}^{n}(\Delta t_i)^2 \approx \lim_{n\to\infty} n\cdot(\frac{1}{n})^2 \to 0$.

The underlying reason for the difference of the stochastic integral and the Stieltjes integral is that ΔW_{t_i} is, with high probability, of order $\sqrt{\Delta t_i}$.

Slogan of the Itô Calculus: $d{W_t}^2 = dt$.

In deriving the Wiener Process, $\dfrac{(\Delta X)^2}{\Delta t} = 1$, so $(\Delta X)^2 = (\Delta W)^2 = \Delta t$.

$dW_t^2 = dt$ is only given precise sense through its integral form:

Theorem 4.3.7. . *If ϕ_s is adapted to W_s, then $\int_a^t \phi_s (dW_s)^2 = \int_a^t \phi_s\, ds$. Moreover, for $n > 2$, $\int_a^t \phi_s (dW_s)^n = 0$.*

Proving this will be in the exercises.

Properties of the Itô Integral. If ϕ_s, ψ_s are adapted to the given process (X_s or W_s), then

(1) $$\int_a^t dX_s = X_t - X_a\,,$$

(2) $$\int_a^t (\phi_s + \psi_s)\, dX_s = \int_a^t \phi_s\, dX_s + \int_a^t \psi_s\, dX_s\,,$$

(3) $$\mathrm{E}\Big(\int_a^t \phi_s\, dW_s\Big) = 0 \text{ and}$$

(4) $$\mathrm{E}\Big(\big(\int_a^t \phi_s\, dW_s\big)^2\Big) = \mathrm{E}\Big(\int_a^t {\phi_s}^2\, ds\Big)\,.$$

Proof.

1. For any partition $\mathcal{P} = \{a = s_0 < s_1 < \cdots < s_n = t\}$, $\mathrm{IS}(1, \mathcal{P}, X_s) = \sum_{i=1}^n 1 \cdot (X_{s_i} - X_{s_{i-1}})$, which telescopes to $X_{s_n} - X_{s_0} = X_t - X_a$. So the limit as mesh$\mathcal{P} \to 0$ is $X_t - X_a$.

2. It is obvious for any partition $\mathcal{P}$ that $\mathrm{IS}(\phi_s + \psi_s, \mathcal{P}, X_s) = \mathrm{IS}(\phi_s, \mathcal{P}, X_s) + \mathrm{IS}(\psi_s, \mathcal{P}, X_s)$. Now take the limit as mesh$\mathcal{P} \to 0$.

3. Given a partition $\mathcal{P}$, $\mathrm{IS}(\phi_s, \mathcal{P}, W_s) = \sum_{i=1}^n \phi_{s_{i-1}}(W_{s_i} - W_{s_{i-1}})$. Taking expected values gives

$$\begin{aligned} \mathrm{E}(\mathrm{IS}(\phi_s, \mathcal{P}, W_s)) &= \mathrm{E}\Big(\sum_{i=1}^n \phi_{s_{i-1}}(W_{s_i} - W_{s_{i-1}}) \Big) = \sum_{i=1}^n \mathrm{E}(\phi_{s_{i-1}}(W_{s_i} - W_{s_{i-1}})) \\ &= \sum_{i=1}^n \mathrm{E}(\phi_{s_{i-1}}) \mathrm{E}(W_{s_i} - W_{s_{i-1}}) = 0\,. \end{aligned}$$

Taking the limit as mesh$\mathcal{P} \to 0$ gives the result.

4. Given a partition $\mathcal{P}$,

$$\begin{aligned} &\mathrm{E}(\mathrm{IS}(\phi_s, \mathcal{P}, W_s)^2) \\ = \;& \mathrm{E}\Big((\textstyle\sum_{i=1}^n \phi_{s_{i-1}}(W_{s_i} - W_{s_{i-1}}))^2 \Big) \\ = \;& \mathrm{E}\Big(\textstyle\sum_{i=1}^n {\phi_{s_{i-1}}}^2 (W_{s_i} - W_{s_{i-1}})^2 + 2 \sum_{i<j} \phi_{s_{i-1}} \phi_{s_{j-1}} (W_{s_i} - W_{s_{i-1}})(W_{s_j} - W_{s_{j-1}}) \Big) \\ = \;& \mathrm{E}\Big(\textstyle\sum_{i=1}^n {\phi_{s_{i-1}}}^2 (\Delta W_{s_i})^2 \Big) + 2 \sum_{i<j} \mathrm{E}(\phi_{s_{i-1}} \phi_{s_{j-1}} (W_{s_i} - W_{s_{i-1}})(W_{s_j} - W_{s_{j-1}})) \\ = \;& \mathrm{E}\Big(\textstyle\sum_{i=1}^n {\phi_{s_{i-1}}}^2 (\Delta W_{s_i})^2 \Big) + 2 \sum_{i<j} \mathrm{E}(\phi_{s_{i-1}} \phi_{s_{j-1}} (W_{s_i} - W_{s_{i-1}})) \cdot \mathrm{E}(W_{s_j} - W_{s_{j-1}}) \\ = \;& \textstyle\sum_{i=1}^n \mathrm{E}({\phi_{s_{i-1}}}^2) \Delta s_i + 2 \sum_{i<j} \mathrm{E}(\phi_{s_{i-1}} \phi_{s_{j-1}} (W_{s_i} - W_{s_{i-1}})) \cdot 0 \\ = \;& \textstyle\sum_{i=1}^n \mathrm{E}({\phi_{s_{i-1}}}^2) \Delta s_i = \mathrm{IS}({\phi_s}^2, \mathcal{P}, s)\,. \end{aligned}$$

Now take the limit as mesh$\mathcal{P} \to 0$.

4.3.3 Stochastic Differential Equations.

A first-order **stochastic differential equation** is an expression of the form

$$dX_t = a(X_t, t)\, dt + b(X_t, t)\, dW_t\,,$$

where $a(x,t)$, $b(x,t)$ are functions of two variables.

A stochastic process X_t is a solution to this SDE if

$$X_t - X_a = \int_a^t dX_s = \int_a^t a(X_s, s)\, ds + b(X_s, s)\, dW_s\,.$$

Lemma 4.3.8. *(Itô's Lemma). Suppose we are given a twice differentiable function $f(x)$, and a solution X_t to the SDE $dX_t = a\,dt + b\,dW_t$. Then $f(X_t)$ is a stochastic process, and it satisfies the SDE*

$$df(X_t) = [f'(X_t)a + \frac{1}{2}f''(X_t)b^2]\,dt + f'(X_t)b\,dW_t\,.$$

More generally, if $f(x,t)$ a function of two variables, twice differentiable in x, once differentiable in t, one has $df(X_t,t)$

$$= [\frac{\partial f}{\partial t}(X_t,t) + \frac{\partial f}{dx}(X_t,t)a + \frac{1}{2}\frac{\partial^2 f}{\partial x^2}(X_t,t)b^2]\,dt + \frac{\partial f}{\partial x}(X_t,t)b\,dW_t\,.$$

Proof. This is just a calculation using two terms of the Taylor's series for f, and the rules ${dW_t}^2 = dt$ and ${dW_t}^n = 0$ for $n > 2$. (Note that $dt\,dW_t = {dW_t}^3 = 0$ and $dt^2 = {dW_t}^4 = 0$.)

$$\begin{aligned} df(X_t) &= f(X_{t+dt}) - f(X_t) = [f(X_t + dX_t)] - f(X_t) \\ &= [f(X_t) + f'(X_t)\,dX_t + \tfrac{1}{2}f''(X_t)(dX_t)^2 + o((dX_t)^3)] - f(X_t) \\ &= f'(X_t)\,dX_t + \tfrac{1}{2}f''(X_t)(dX_t)^2 \\ &= f'(X_t)[a(X_t,t)\,dt + b(X_t,t)\,dW_t] + \tfrac{1}{2}f''(X_t)[a^2\,dt^2 + 2ab\,dt\,dW_t + b^2\,{dW_t}^2] \\ &= [f'(X_t)a(X_t,t) + \tfrac{1}{2}f''(X_t)(b(X_t,t))^2]\,dt + f'(X_t)b(X_t,t)\,dW_t \end{aligned}$$

Example. To calculate $\int_a^t W_t\,dW_t$, show that is equal to $\int_a^t dX_t = X_t - X_a$ for some X_t. That is, find $f(x)$ such that $W_t\,dW_t = d(f(W_t))$. Since $\int x\,dx = \frac{1}{2}x^2$, consider $f(x) = \frac{1}{2}x^2$. Then $f'(x) = x$ and $f''(x) = 1$, so Itô's lemma, applied to $dW_t = 0 \cdot dt + 1 \cdot dW_t$, gives

$$d(f(W_t)) = f'(W_t)\,dW_t + \frac{1}{2}f''(W_t)\,dt = W_t\,dW_t + \frac{1}{2}\cdot 1\,dt = W_t\,dW_t + \frac{1}{2}\,dt\,.$$

Equivalently, $W_t\,dW_t = df(W_t) - \frac{1}{2}\,dt$. Therefore, $\int_a^t W_s\,dW_s = \int_a^t d(f(W_s)) - \int_a^t \frac{1}{2}\,dt = f(W_t) - f(W_a) - \frac{1}{2}(t-a) = \frac{1}{2}{W_t}^2 - \frac{1}{2}{W_a}^2 - \frac{1}{2}(t-a)$.

Example. Calculate $\int_a^t {W_s}^2\,dW_s$. Since $\int x^2\,dx = \frac{1}{3}x^3$, consider $f(x) = \frac{1}{3}x^3$. Then $f'(x) = x^2$ and $f''(x) = 2x$. Itô's lemma implies

$$d(\frac{1}{3}x^3) = {W_s}^2\,dW_s + \frac{1}{2}\cdot 2W_s\,ds = {W_s}^2\,dW_s + W_s\,ds\,,$$

or ${W_s}^2\,dW_s = d(\frac{1}{3}\,{W_s}^3) - W_s\,ds$, so

$$\int_a^t {W_s}^2\,dW_s = \frac{1}{3}\,{W_t}^2 - \frac{1}{3}\,{W_a}^2 - \int_a^t W_s\,ds\,.$$

(The last integrand does not really simplify.)

Example. Solve $dX_s = \frac{1}{2}X_s\,ds + X_s\,dW_s$. Consider $f'(X_s)\cdot X_s = 1$. With ordinary variables, this is $f'(x)\cdot x = 1$, or $f'(x) = \frac{1}{x}$. So we are led to consider $f(x) = \ln x$, with $f'(x) = \frac{1}{x}$, and $f''(x) = -\frac{1}{x^2}$.

Then, using $dX_s = \frac{1}{2}X_s\,ds + X_s\,dW_s$ and ${dX_s}^2 = {X_s}^2\,ds$,

$$\begin{aligned} d(\ln X_s) &= \tfrac{1}{X_s}\,dX_s + \tfrac{1}{2}(-\tfrac{1}{{X_s}^2})(dX_s)^2 \\ &= \tfrac{1}{2}\,ds + dW_s - \tfrac{1}{2}\,ds = dW_s\,. \end{aligned}$$

Now integrate both sides, getting

$$\begin{aligned} \textstyle\int_a^t d(\ln X_s) &= \textstyle\int_a^t dW_s \\ &= W_t - W_a \text{ or} \\ \ln\left(\tfrac{X_t}{X_a}\right) &= W_t - W_a \text{ and so} \\ X_t &= X_a e^{W_t - W_a}\,. \end{aligned}$$

Asset Models

The basic model for a stock is the SDE

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

where μ is the short-term expected return, and σ is the standard deviation of the the short-term returns, called **volatility**.

Note that if $\sigma = 0$, the model simplifies to $dS_t = \mu S_t\,dt$, implying $S_t = S_a e^{\mu t}$, which is just the value of an interest-bearing, risk-free security.

Compared with the generic SDE

$$dX_t = a(X_t, t)\,dt + b(X_t, t)\,dW_t$$

the stock model uses $a(x,t) = \mu x$ and $b(x,t) = \sigma x$. We have ${dS_t}^2 = \sigma^2 {S_t}^2\,dt$. The model SDE is similar to the last example, and the same change of variables will work. Using $f(x) = \ln x$, $f'(x) = \frac{1}{x}$, and $f''(x) = -\frac{1}{x^2}$, Itô's lemma gives

$$\begin{aligned} d(\ln S_t) &= \tfrac{1}{S_t}\,dS_t + \tfrac{1}{2}(-\tfrac{1}{{S_t}^2})(dS_t)^2 \\ &= \mu\,dt + \sigma\,dW_t - \tfrac{1}{2}\sigma^2\,dt \\ &= (\mu - \tfrac{1}{2}\sigma^2)\,dt + \sigma\,dW_t\,. \end{aligned}$$

Integrating both sides gives

$$\begin{aligned} \textstyle\int_0^t d(\ln S_s) &= \textstyle\int_0^t (\mu - \tfrac{1}{2}\sigma^2)\,ds + \sigma\,dW_s \\ \ln S_t - \ln S_0 &= (\mu - \tfrac{1}{2}\sigma^2)t + \sigma W_t \\ \tfrac{S_t}{S_0} &= e^{(\mu - \frac{1}{2}\sigma^2)t + \sigma W_t} \\ S_t &= S_0 e^{(\mu - \frac{1}{2}\sigma^2)t + \sigma W_t}\,. \end{aligned}$$

The expected return of S_t is

$$\mathrm{E}(\ln(S_t/S_0)) = \mathrm{E}((\mu - \frac{1}{2}\sigma^2)t + \sigma W_t) = (\mu - \frac{1}{2}\sigma^2)t\,.$$

Vasicek's interest rate model is $dr_t = a(b - r_t)\,dt + \sigma\,dW_t$. It satisfies "mean reversion", meaning that if $r_t < b$, then $dr_t > 0$, pulling r_t upwards, while if $r_t > b$, then $dr_t < 0$, pushing r_t downwards.

The Cox-Ingersoll-Ross model is $dr_t = a(b - r_t)\,dt + \sigma\sqrt{r_t}\,dW_t$. It has mean reversion, and a volatility that grows with r_t.

Ornstein-Uhlenbeck Process. An O-U process is a solution to

$$dX_t = \mu X_t\,dt + \sigma\,dW_t$$

with an initial condition $X_0 = x_0$.

This SDE is solved by using the integrating factor $e^{-\mu t}$. First,

$$e^{-\mu t}\,dX_t = \mu e^{-\mu t} X_t\,dt + \sigma e^{-\mu t}\,dW_t$$

holds just by multiplying the given SDE. Letting $f(x,t) = xe^{-\mu t}$, then

$$\frac{\partial f}{\partial t} = -\mu e^{-\mu t} X_t\,dt\,, \quad \frac{\partial f}{\partial x} = e^{-\mu t}\,, \quad \text{and} \quad \frac{\partial^2 f}{\partial x^2} = 0\,.$$

Itô's lemma gives

$$\begin{array}{rcl} df(X_t,t) & = & \frac{\partial f}{\partial t}(X_t,t)\,dt + \frac{\partial f}{\partial x}(X_t,t)\,dX_t + \frac{\partial^2 f}{\partial x^2}(X_t,t)\,dX_t{}^2 \\ d(e^{-\mu t}X_t) & = & -\mu e^{-\mu t} X_t\,dt + e^{-\mu t}\,dX_t \quad \text{or} \\ e^{-\mu t}\,dX_t & = & d(e^{-\mu t}X_t) + \mu e^{-\mu t} X_t\,dt\,. \end{array}$$

So the SDE, with integrating factor, rewrites as

$$\begin{array}{rcl} d(e^{-\mu t}X_t) + \mu e^{-\mu t} X_t\,dt & = & \mu e^{-\mu t} X_t\,dt + \sigma e^{-\mu t}\,dW_t \\ d(e^{-\mu t}X_t) & = & \sigma e^{-\mu t}\,dW_t\,. \end{array}$$

Now integrate both sides from 0 to t:

$$\begin{array}{rcl} \int_0^t d(e^{-\mu s}X_s) & = & \int_0^t \sigma e^{-\mu s}\,dW_s \\ e^{-\mu t}X_t - X_0 & = & \int_0^t \sigma e^{-\mu s}\,dW_s \\ X_t & = & e^{\mu t}X_0 + e^{\mu t}\int_0^t \sigma e^{-\mu s}\,dW_s\,. \end{array}$$

Let's go back to the stock model

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

with solution

$$S_t = S_0 e^{(\mu - \frac{1}{2}\sigma^2)t + \sigma W_t}\,.$$

We calculate the pdf of S_t as follows. First,

$$\begin{aligned} P(S_t \le S) &= P(S_0 e^{(\mu - \frac{1}{2}\sigma^2)t + \sigma W_t} \le S) \\ &= P(e^{(\mu - \frac{1}{2}\sigma^2)t + \sigma W_t} \le S/S_0) \\ &= P((\mu - \tfrac{1}{2}\sigma^2)t + \sigma W_t \le \ln(S/S_0)) \\ &= \tfrac{1}{\sqrt{2\pi\sigma^2 t}} \int_{-\infty}^{\ln(S/S_0)} e^{-\frac{(x-(\mu-\frac{1}{2}\sigma^2)t)^2}{2\sigma^2 t}}\, dx\,. \end{aligned}$$

Second, the pdf is the derivative of the above with respect to S. The fundamental theorem of calculus gives us derivatives of the form $\frac{d}{du}\int_a^u \phi(x)\,dx = \phi(u)$. The chain rule gives us, where $u = g(S)$, a derivative

$$\frac{d}{dS}\int_a^{g(S)} \phi(x)\,dx = (\frac{d}{du}\int_a^u \phi(x)\,dx) \cdot \frac{du}{dS} = \phi(u)\cdot g'(S) = \phi(g(S))\cdot g'(S)\,.$$

In the case at hand, $g(S) = \ln(S/S_0) = \ln S - \ln S_0$, so $g'(S) = 1/S$, and the pdf is

$$\begin{aligned} \tfrac{d}{dS}P(S_t \le S) &= \tfrac{d}{dS}\tfrac{1}{\sqrt{2\pi\sigma^2 t}} \int_{-\infty}^{\ln(S/S_0)} e^{-\frac{(x-(\mu-\frac{1}{2}\sigma^2)t)^2}{2\sigma^2 t}}\, dx \\ &= \tfrac{1}{S\sqrt{2\pi\sigma^2 t}} e^{-\frac{(\ln(S/S_0)-(\mu-\frac{1}{2}\sigma^2)t)^2}{2\sigma^2 t}} \end{aligned}$$

for $S > 0$ and identically 0 for $S \le 0$.

Pricing Derivatives in Continuous Time

Our situation concerns a stock with price stochastic variable S_t, and a derivative security, assumed to have expiration T, and further assumed that its (gross) payoff depends only on T and S_T. Let $V(S_T, T)$ denote this payoff. More generally, let $V(S_t, t)$ denote the payoff for this derivative if it expires at time t with stock price S_t. In particular, $V(S_t, t)$ depends only on t and S_t. That is, we have a deterministic function $V(s, t)$, and are plugging in a stochastic process S_t and getting a stochastic process $V(S_t, t)$.

To price V, construct a self-financing, hedging portfolio for V. That is, set up a portfolio consisting of ϕ_t units of stock S_t and ψ_t units of bond B_t

$$\Pi_t = \phi_t S_t + \psi_t B_t\,,$$

such that the terminal value of the portfolio equals the derivative payoff

$$\Pi_T = V_T \,,$$

subject to the **self-financing condition**

$$SFC1 \qquad \Pi_t = \Pi_0 + \int_0^t \phi_s \, dS_s + \psi_s \, dB_s \,.$$

The pricing of the portfolio is straightforward, and the no-arbitrage assumption implies $\Pi_t = V_t$ for all $t \le T$.

Recall, also, the self-financing condition in the discrete case:

$$(\Delta\Pi)_n = \phi_n \Delta S_n + \psi_n \Delta B_n \,.$$

The continuous limit is

$$SFC2 \qquad d\Pi_t = \phi_t \, dS_t + \psi_t \, dB_t \,.$$

We now model the stock and bond. The stock is $dS_t = \mu S_t \, dt + \sigma S_t \, dW_t$, while the bond is $dB_t = rB_t \, dt$.

Theorem 4.3.9. *(Black-Scholes). Given a derivative as above, there exists a self-financing, hedging portfolio* $\Pi_t = \phi_t S_t + \psi_t B_t$ *with* $\Pi_t = V(S_t, t)$ *if and only if*

$$(BSE) \qquad \frac{\partial V}{\partial t} + \frac{1}{2}\sigma^2 S^2 \frac{\partial^2 V}{\partial S^2} + rS\frac{\partial V}{\partial S} - rV = 0,$$

and furthermore, $\phi_t = \frac{\partial V}{\partial S}$.

Proof. Let's calculate $dV(S_t, t) = d\Pi_t$ two ways. First, by Ito's lemma

$$dV(S_t,t) = \frac{\partial V}{\partial t}(S_t,t)\,dt + \frac{\partial V}{\partial S}(S_t,t)\,dS_t + \frac{1}{2}\frac{\partial^2 V}{\partial S^2}(S_t,t)(dS_t)^2 \,,$$

or, suppressing the S_t, t arguments,

$$\begin{aligned} dV_t &= \tfrac{\partial V}{\partial t}\,dt + \tfrac{\partial V}{\partial S}\,dS_t + \tfrac{1}{2}\tfrac{\partial^2 V}{\partial S^2}(dS_t)^2 \\ &= (\tfrac{\partial V}{\partial t} + \mu S_t \tfrac{\partial V}{\partial S} + \tfrac{1}{2}\sigma^2 {S_t}^2 \tfrac{\partial^2 V}{\partial S^2})\,dt + \tfrac{\partial V}{\partial S}\sigma S_t \, dW_t \,. \end{aligned}$$

On the other hand, using our stock and bond model on a portfolio, the self financing condition (SFC2) says that $V_t = \Pi_t = \phi_t S_t + \psi_t B_t$,

$$\begin{aligned} dV_t = d\Pi_t &= \phi_t \, dS_t + \psi_t \, dB_t \\ &= (\phi_t \mu S_t + \psi_t r B_t)\,dt + \phi_t \sigma S_t \, dW_t \,. \end{aligned}$$

If these are to be equal, the coefficients of dt and of dW_t must be equal. Equating the coefficients of dW_t gives

$$\phi_t \sigma S_t = \frac{\partial V}{\partial S} \sigma S_t \,, \qquad \text{hence } \phi_t = \frac{\partial V}{\partial S} \,,$$

and equating the coefficients of dt gives

$$\phi_t \mu S_t + \psi_t r B_t = \frac{\partial V}{\partial t} + \mu S_t \frac{\partial V}{\partial S} + \frac{1}{2} \sigma^2 {S_t}^2 \frac{\partial^2 V}{\partial S^2} \,,$$

hence

$$\psi_t r B_t = \frac{\partial V}{\partial t} + \frac{1}{2} \sigma^2 {S_t}^2 \frac{\partial^2 V}{\partial S^2} \,.$$

Since $\psi_t B_t = \Pi_t - \phi_t S_t = V_t - \frac{\partial V}{\partial S} S_t$ we have

$$r(V_t - \frac{\partial V}{\partial S} S_t) = \frac{\partial V}{\partial t} + \frac{1}{2} \sigma^2 {S_t}^2 \frac{\partial^2 V}{\partial S^2} \,,$$

or in other works (BSE), completing the proof.

In summary, to construct a self-financing, replicating portfolio for a derivative, you need the value of the derivative $V_t = V(S_t, t)$ to satisfy the Black-Scholes equation and for the stock holding to be $\phi_t = \frac{\partial V}{\partial S}$.

The theory of partial differential equations tells us that in order to get unique solutions, we need boundary and temporal conditions. Since the BSE is first order in t, we need one time condition. And since the BSE is second order in S, we need two boundary conditions.

The time condition is simply the expiration payoff $V(S_T, T) = h(S_T)$. (Usually, one has an "initial" condition, corresponding to time 0. Here, one has a "final" or "terminal" condition.) The boundary conditions are usually based on $S = 0$ and $S \to \infty$.

For example, consider a call. $h(S) = \max(S - K, 0)$ is, by definition, the payoff of a call at expiry, so $V(S_T, T) = \max(S_T - K, 0)$ is the final condition. If the stock price is zero, a call is worthless, hence $V(0, t) = 0$ is one boundary condition. If the stock price grows arbitrarily large, the call is essentially worth the same as the stock, that is $V(S, t) \sim S$.

The One-Dimensional Heat Equation

Let the temperature of a homogeneous one-dimensional object at position x, time t be given by $u(x, t)$. The heat equation is

$$\frac{\partial u}{\partial t} = k \frac{\partial^2 u}{\partial x^2} \,.$$

where k is a constant, the thermal conductivity of the object's material.

$k = \frac{1}{2}$ gives the Fokker-Planck equation for a Wiener path, with solution

$$u(x,t) = \frac{1}{\sqrt{2\pi t}}\, e^{-\frac{x^2}{2t}}\,.$$

If we let $k = 1$, the solution is

$$u(x,t) = \frac{1}{\sqrt{4\pi t}}\, e^{-\frac{x^2}{4t}}\,.$$

What is $\lim_{t\to 0} u(x,t)$, the initial temperature distribution? To answer this, we use the following notion.

Definition. A **Dirac δ-function** is a one-parameter family δ_t, indexed by $t \geq 0$, of piecewise smooth functions such that

$$\begin{array}{lc} (1) & \delta_t(x) \geq 0 \text{ for all } t, x, \\ (2) & \int_{-\infty}^{\infty} \delta_t(x)\, dx = 1 \text{ for all } t, \\ (3) & \lim_{t\to 0} \delta_t(x) = 0 \text{ for all } x \neq 0. \end{array}$$

Example. Consider

$$\delta_t = \{ \begin{array}{cl} 0 & , \text{if } x < -t \text{ or } x > t\,, \\ \frac{1}{2t} & , \text{if } \; -t \leq x \leq t\,. \end{array}$$

Important Example. The solutions to the heat equation

$$\delta_t(x) = u(x,t) = \frac{1}{\sqrt{4\pi t}}\, e^{-\frac{x^2}{4t}}$$

are a Dirac δ-function.

Proposition. If f is any bounded smooth function, δ_t any Dirac δ-function, then

$$\lim_{t\to 0} \int_{-\infty}^{\infty} \delta_t(x) f(x)\, dx = f(0)\,,$$

more generally,

$$\lim_{t\to 0} \int_{-\infty}^{\infty} \delta_t(x-y) f(x)\, dx = f(y)\,.$$

Intuitively, this can be seen as follows. Every Dirac δ-function is something like the first example, and on a small interval, continuous functions are approximately constant, so $\lim_{t\to 0} \int_{-\infty}^{\infty} \delta_t(x-y) f(x)\, dx \approx \lim_{t\to 0} \int_{-t-y}^{t-y} \frac{1}{2t} f(x)\, dx \approx \lim_{t\to 0} 2t \cdot \frac{1}{2t} \cdot f(y) \approx f(y)$.

Proceeding, the Heat Equation

$$\frac{\partial u}{\partial t} = \frac{\partial^2 u}{\partial x^2},$$

with initial condition $u(x,0) = \phi(x)$ and boundary conditions $u(\infty,t) = u(-\infty,t) = 0$, will be solved using

$$\delta_t = \frac{1}{\sqrt{4\pi t}} e^{-\frac{x^2}{4t}},$$

which is both a Dirac δ-function and a solution to the Heat Equation for $t > 0$.

Let

$$v(y,t) = \frac{1}{\sqrt{4\pi t}} \int_{-\infty}^{\infty} e^{-\frac{(x-y)^2}{4t}} \cdot \phi(x)\, dx \text{ for } t > 0\,.$$

Claim. $\lim_{t\to 0} v(x,t) = \phi(x)$. This is just an example of the previous proposition:

$$\lim_{t\to 0} \frac{1}{\sqrt{4\pi t}} \int_{-\infty}^{\infty} e^{-\frac{(x-y)^2}{4t}} \cdot \phi(x)\, dx = \lim_{t\to 0} \int_{-\infty}^{\infty} \delta_t(x-y)\phi(x)\, dx = \phi(y)\,.$$

Claim. v satisfies the Heat Equation.

$$\begin{aligned}
\tfrac{\partial v}{\partial t} &= \tfrac{\partial}{\partial t} \int_{-\infty}^{\infty} e^{-\frac{(x-y)^2}{4t}} \cdot \phi(x)\, dx \\
&= \int_{-\infty}^{\infty} \tfrac{\partial}{\partial t} (e^{-\frac{(x-y)^2}{4t}}) \cdot \phi(x)\, dx \\
&= \int_{-\infty}^{\infty} \tfrac{\partial^2}{\partial y^2} (e^{-\frac{(x-y)^2}{4t}}) \cdot \phi(x)\, dx \\
&= \tfrac{\partial^2}{\partial y^2} \int_{-\infty}^{\infty} e^{-\frac{(x-y)^2}{4t}} \cdot \phi(x)\, dx = \tfrac{\partial^2 v}{\partial y^2}\,.
\end{aligned}$$

Back to the Black-Scholes equation

$$\frac{\partial V}{\partial t} + \frac{1}{2}\sigma^2 S^2 \frac{\partial^2 V}{\partial S^2} + rS\frac{\partial V}{\partial S} - rV = 0\,.$$

Note that μ, the expected return of S, does not appear explicitly in the equation. This is called **risk neutrality**. The particulars of the derivative are encoded in the boundary conditions. For a call $C(S,t)$ one has

$$\frac{\partial C}{\partial t} + \frac{1}{2}\sigma^2 S^2 \frac{\partial^2 C}{\partial S^2} + rS\frac{\partial C}{\partial S} - rV = 0\,,$$

$$C(S,T) = \max(S-K,0)\,,$$
$$C(0,t) = 0 \text{ for all } t,\ C(S,t) \sim S \text{ as } S \to \infty\,.$$

And for a put $P(S,t)$ one has

$$\frac{\partial P}{\partial t} + \frac{1}{2}\sigma^2 S^2 \frac{\partial^2 P}{\partial S^2} + rS\frac{\partial P}{\partial S} - rV = 0\,,$$
$$P(S,T) = \max(K-S,0)\,,$$
$$P(0,t) = K \text{ for all } t,\ P(S,t) \to 0 \text{ as } S \to \infty\,.$$

Black-Scholes Formula

The solution to the the call PDE above is

$$C(S,t) = SN(d_+) - Ke^{-r(T-t)}N(d_-)\,,$$

where

$$\begin{aligned} d_+ &= \tfrac{\ln(\frac{S}{K}) + (r+\frac{1}{2}\sigma^2)(T-t)}{\sigma\sqrt{T-t}}\,, \\ d_- &= \tfrac{\ln(\frac{S}{K}) + (r-\frac{1}{2}\sigma^2)(T-t)}{\sigma\sqrt{T-t}}\,, \end{aligned}$$

and

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

is the cumulative normal distribution.

If you are given data, including the market price for derivatives, and numerically solve the Black-Scholes formula for σ, the resulting value is called the **implied volatility**.

We derive the Black-Scholes formula from the Black-Scholes equation, by using a change of variables to obtain the heat equation, whose solutions we have identified, at least in a formal way.

The first difference is the heat equation has an initial condition at $t = 0$, while the Black-Scholes equation has a final condition at $t = T$. This can be compensated for by letting $\tau = T - t$. Then in terms of τ, the final condition at $t = T$ is an initial condition at $\tau = 0$.

The second difference is that the Black-Scholes equation has $S^2 \frac{\partial^2 V}{\partial S^2}$ while the heat equation has $\dfrac{\partial^2 u}{\partial x^2}$. If we let $S = e^x$, or equivalently, $x = \ln S$, then

$$\begin{aligned} \tfrac{\partial}{\partial x} &= \tfrac{\partial S}{\partial x}\tfrac{\partial}{\partial S} = e^x \tfrac{\partial}{\partial S} = S\tfrac{\partial}{\partial S} \\ \tfrac{\partial^2}{\partial x^2} &= S^2 \tfrac{\partial^2}{\partial S^2} + \cdots \end{aligned}$$

The third difference is that the Black-Scholes equation has $\frac{\partial V}{\partial S}$ and V terms, but the heat equation has no $\frac{\partial u}{\partial x}$ or u terms. An integrating factor can eliminate such terms, as follows. Suppose we are given

$$\frac{\partial u}{\partial t} = \frac{\partial^2 u}{\partial x^2} + a\frac{\partial u}{\partial x} + bu\,.$$

Let

$$u(x,t) = e^{\alpha x+\beta t}v(x,t)\,,$$

with α, β to be chosen later. Then

$$\frac{\partial u}{\partial t} = \beta e^{\alpha x+\beta t}v + e^{\alpha x+\beta t}\frac{\partial v}{\partial t}\,,$$

$$\frac{\partial u}{\partial x} = \alpha e^{\alpha x+\beta t}v + e^{\alpha x+\beta t}\frac{\partial v}{\partial x}\,,$$

$$\frac{\partial^2 u}{\partial x^2} = \alpha^2 e^{\alpha x+\beta t}v + 2\alpha e^{\alpha x+\beta t}\frac{\partial v}{\partial x} + e^{\alpha x+\beta t}\frac{\partial^2 v}{\partial x^2}\,.$$

So

$$\frac{\partial u}{\partial t} = \frac{\partial^2 u}{\partial x^2} + a\frac{\partial u}{\partial x} + bu$$

becomes

$$e^{\alpha x+\beta t}(\beta v+\frac{\partial v}{\partial t}) = e^{\alpha x+\beta t}(\alpha^2 v+2\alpha\frac{\partial v}{\partial x}+\frac{\partial^2 v}{\partial x^2})+a\cdot e^{\alpha x+\beta t}(\alpha v+\frac{\partial v}{\partial x})+b\cdot e^{\alpha x+\beta t}v\,,$$

or

$$\beta v + \frac{\partial v}{\partial t} = \alpha^2 v + 2\alpha\frac{\partial v}{\partial x} + \frac{\partial^2 v}{\partial x^2} + a\cdot(\alpha v + \frac{\partial v}{\partial x}) + b\cdot v\,,$$

hence,

$$\frac{\partial v}{\partial t} = \frac{\partial^2 v}{\partial x^2} + (a+2\alpha)\frac{\partial v}{\partial x} + (\alpha^2 + a\alpha + b - \beta)v\,.$$

With a and b given, one can set $a + 2\alpha = 0$ and $\alpha^2 + a\alpha + b - \beta = 0$ and solve for α and β. One gets $\alpha = -\frac{1}{2}a$ and $\beta = b - \alpha^2 = b - \frac{1}{4}a^2$. In other words, if $v(x,t)$ is a solution to

$$\frac{\partial v}{\partial t} = \frac{\partial^2 v}{\partial x^2}\,,$$

then $u(x,t) = e^{-\frac{1}{2}ax + (b - \frac{1}{4}a^2)t} v(x,t)$ satisfies

$$\frac{\partial u}{\partial t} = \frac{\partial^2 u}{\partial x^2} + a\frac{\partial u}{\partial x} + bu\,.$$

With these three differences accounted for, we will now solve the Black-Scholes equation for a European call. We are considering

$$\frac{\partial C}{\partial t} + \frac{1}{2}\sigma^2 S^2 \frac{\partial^2 C}{\partial S^2} + rS\frac{\partial C}{\partial S} - rV = 0$$

with final condition

$$C(S,T) = \max(S-K,0)\,.$$

First make the change of variables $S = Ke^x$, or equivalently, $x = \ln(S/K) = \ln S - \ln K$, and also let $\tau = \frac{1}{2}\sigma^2(T-t)$, so $C(S,t) = Ku(x,\tau)$. Then

1. $$\frac{\partial C}{\partial t} = \frac{\partial C}{\partial \tau}\cdot\frac{d\tau}{dt} = (K\frac{\partial u}{\partial \tau})\cdot(-\frac{1}{2}\sigma^2) = -\frac{1}{2}\sigma^2 K\frac{\partial u}{\partial \tau}\,,$$

2. $$\frac{\partial C}{\partial S} = \frac{\partial C}{\partial x}\cdot\frac{dx}{dS} = (K\frac{\partial u}{\partial x})\cdot\frac{1}{S} = \frac{K}{S}\frac{\partial u}{\partial x}\,,$$

3. $$\frac{\partial^2 C}{\partial S^2} = \frac{\partial}{\partial S}(\frac{K}{S}\frac{\partial u}{\partial x}) = -\frac{K}{S^2}\frac{\partial u}{\partial x} + \frac{K}{S}\frac{\partial}{\partial S}(\frac{\partial u}{\partial x})$$
$$= -\frac{K}{S^2}\frac{\partial u}{\partial x} + \frac{K}{S}\frac{\partial}{\partial x}(\frac{\partial u}{\partial x})\frac{\partial x}{\partial S} = -\frac{K}{S^2}\frac{\partial u}{\partial x} + \frac{K}{S^2}\frac{\partial^2 u}{\partial x^2}\,.$$

Substituting this into the Black-Scholes equation, we have

$$-\frac{1}{2}\sigma^2 K\frac{\partial u}{\partial \tau} + \frac{1}{2}\sigma^2 S^2(-\frac{K}{S^2}\frac{\partial u}{\partial x} + \frac{K}{S^2}\frac{\partial^2 u}{\partial x^2}) + rS(\frac{K}{S}\frac{\partial u}{\partial x}) - rKu = 0$$

or

$$\frac{1}{2}\sigma^2\frac{\partial u}{\partial \tau} = \frac{1}{2}\sigma^2\frac{\partial^2 u}{\partial x^2} - \frac{1}{2}\sigma^2\frac{\partial u}{\partial x} + r\frac{\partial u}{\partial x} - ru$$

or

$$\frac{\partial u}{\partial \tau} = \frac{\partial^2 u}{\partial x^2} + (\frac{2r}{\sigma^2} - 1)\frac{\partial u}{\partial x} - \frac{2r}{\sigma^2}u\,.$$

Set $k = \frac{2r}{\sigma^2}$. The equation becomes

$$\frac{\partial u}{\partial \tau} = \frac{\partial^2 u}{\partial x^2} + (k-1)\frac{\partial u}{\partial x} - ku\,.$$

And the final condition $C(S,T) = \max(S-K,0)$ becomes the initial condition $Ku(x,0) = \max(Ke^x - K, 0)$, which becomes $u(x,0) = \max(e^x - 1, 0)$.

As we saw, this reduces to the heat equation if we let $u(x,\tau) = e^{\alpha x + \beta\tau} v(x,\tau)$, with $\alpha = -\frac{1}{2}(k-1)$ and $\beta = -k - \frac{1}{4}(k-1)^2 = -\frac{1}{4}(k+1)^2$. And the initial condition $u(x,0) = \max(e^x - 1, 0)$ becomes

$$e^{-\frac{1}{2}(k-1)x} v(x,0) = \max(e^x - 1, 0)$$

or equivalently,

$$v(x,0) = \max(e^x \cdot e^{\frac{1}{2}(k-1)x} - e^{\frac{1}{2}(k-1)x}, 0) = \max(e^{\frac{1}{2}(k+1)x} - e^{\frac{1}{2}(k-1)x}, 0)\,.$$

Recall that the heat equation

$$\frac{\partial v}{\partial \tau} = \frac{\partial^2 v}{\partial x^2} \qquad v(x,0) = \phi(x)$$

is solved by

$$v(x,\tau) = \frac{1}{\sqrt{4\pi\tau}} \int_{-\infty}^{\infty} e^{-\frac{(y-x)^2}{4\tau}} \phi(y)\,dy\,,$$

which in this case has $\phi(y) = \max(e^{\frac{1}{2}(k+1)y} - e^{\frac{1}{2}(k-1)y}, 0)$.

Now, $e^{\frac{1}{2}(k+1)y} > e^{\frac{1}{2}(k-1)y}$ if and only if $e^{\frac{1}{2}(k+1)y}/e^{\frac{1}{2}(k-1)y} > 1$ if and only if $e^y > 1$ if and only if $y > 0$. So in our situation, we can write

$$\phi(y) = \begin{cases} e^{\frac{1}{2}(k+1)y} - e^{\frac{1}{2}(k-1)y} & \text{if } y > 0 \\ 0 & \text{if } y \le 0 \end{cases}$$

Now make the change of variables $s = (y-x)/\sqrt{2\tau}$. Then $ds = dy/\sqrt{2\tau}$, so $dy = \sqrt{2\tau}\,ds$. Then since $\phi(\sqrt{2\tau}s + x) = 0$ when $\sqrt{2\tau}s + x < 0$, that is, when $s < \frac{-x}{\sqrt{2\tau}}$,

$$\begin{aligned} v(x,\tau) &= \tfrac{\sqrt{2\tau}}{\sqrt{4\pi\tau}} \int_{-\infty}^{\infty} e^{-\frac{(x-(\sqrt{2\tau}s+x))^2}{4\tau}} \phi(\sqrt{2\tau}s + x)\,ds \\ &= \tfrac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{2\tau s^2}{4\tau}} \phi(\sqrt{2\tau}s + x)\,ds \\ &= \tfrac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{s^2}{2}} \phi(\sqrt{2\tau}s + x)\,ds \\ &= \tfrac{1}{\sqrt{2\pi}} \int_{-\frac{x}{\sqrt{2\tau}}}^{\infty} e^{-\frac{s^2}{2}} \left(e^{\frac{1}{2}(k+1)(\sqrt{2\tau}s+x)} - e^{\frac{1}{2}(k-1)(\sqrt{2\tau}s+x)}\right) ds \\ &= I_+ - I_-, \end{aligned}$$

where

$$I_\pm = \frac{1}{\sqrt{2\pi}} \int_{-\frac{x}{\sqrt{2\tau}}}^{\infty} e^{-\frac{1}{2}s^2+\frac{1}{2}(k\pm1)(\sqrt{2\tau}s+x)} ds\,.$$

We calculate I_+ by completing the square inside the exponent.

$$\begin{aligned}
I_+ &= \tfrac{1}{\sqrt{2\pi}}\int_{-\frac{x}{\sqrt{2\tau}}}^{\infty} e^{-\frac{1}{2}s^2+\frac{1}{2}(k+1)(\sqrt{2\tau}s+x)} ds \\
&= \tfrac{1}{\sqrt{2\pi}} \cdot e^{\frac{1}{2}(k+1)x}\int_{-\frac{x}{\sqrt{2\tau}}}^{\infty} e^{-\frac{1}{2}s^2+\frac{1}{2}(k+1)\sqrt{2\tau}s} ds \\
&= \tfrac{e^{\frac{1}{2}(k+1)x}}{\sqrt{2\pi}}\int_{-\frac{x}{\sqrt{2\tau}}}^{\infty} e^{-\frac{1}{2}(s-\frac{1}{2}(k+1)\sqrt{2\tau})^2+\frac{1}{4}(k+1)^2\tau} ds \\
&= \tfrac{e^{\frac{1}{2}(k+1)x+\frac{1}{4}(k+1)^2\tau}}{\sqrt{2\pi}}\int_{-\frac{x}{\sqrt{2\tau}}}^{\infty} e^{-\frac{1}{2}(s-\frac{1}{2}(k+1)\sqrt{2\tau})^2} ds \\
&= \tfrac{e^{\frac{1}{2}(k+1)x+\frac{1}{4}(k+1)^2\tau}}{\sqrt{2\pi}}\int_{-\frac{x}{\sqrt{2\tau}}-\frac{1}{2}(k+1)\sqrt{2\tau}}^{\infty} e^{-\frac{1}{2}t^2} dt \\
&= \tfrac{e^{\frac{1}{2}(k+1)x+\frac{1}{4}(k+1)^2\tau}}{\sqrt{2\pi}}\int_{-\infty}^{\frac{x}{\sqrt{2\tau}}+\frac{1}{2}(k+1)\sqrt{2\tau}} e^{-\frac{1}{2}t^2} dt \\
&= e^{\frac{1}{2}(k+1)x+\frac{1}{4}(k+1)^2\tau} \cdot N(d_+) = e^{x-\alpha x-\beta\tau}N(d_+)\,.
\end{aligned}$$

using $\alpha = -\frac{1}{2}(k-1)$ and $\beta = -\frac{1}{4}(k+1)^2$ and where $d_+ = \frac{x}{\sqrt{2\tau}} + \frac{1}{2}(k+1)\sqrt{2\tau}$. The second to last line was arrived at since

$$\int_a^\infty e^{-t^2/2} dt = \int_{-\infty}^{-a} e^{-t^2/2} dt$$

Now $S = Ke^x$ and $2\tau = \sigma^2(T-t)$, so

$$\begin{aligned}
d_+ &= \tfrac{x}{\sqrt{2\tau}} + \tfrac{1}{2}(k+1)\sqrt{2\tau} \\
&= \tfrac{\ln(S/K)}{\sigma\sqrt{T-t}} + \tfrac{1}{2}(\tfrac{2r}{\sigma^2}+1)\sigma\sqrt{T-t} \\
&= \tfrac{\ln(\frac{S}{K})+(r+\frac{1}{2}\sigma^2)(T-t)}{\sigma\sqrt{T-t}}\,.
\end{aligned}$$

The integral I_- is treated in the same manner and gives

$$I_- = e^{\frac{1}{2}(k-1)+\frac{1}{4}(k-1)^2\tau}N(d_-)$$

$$= e^{-\alpha x-\beta\tau-k\tau}N(d_-)$$

where $d_- = \frac{x}{\sqrt{2\tau}} + \frac{1}{2}(k-1)\sqrt{2\tau} = \frac{\ln(\frac{S}{K})+(r-\frac{1}{2}\sigma^2)(T-t)}{\sigma\sqrt{T-t}}$. So with $v(x,\tau) = I_+ - I_-$ established, we have substituting everything back in: $x = \ln(S/K)$,

$\tau = \frac{\sigma^2}{2}(T-t)$ and $k = \frac{2r}{\sigma^2}$.

$$\begin{aligned} C(S,t) &= Ku(x,\tau) = Ke^{\alpha x+\beta\tau}v(x,\tau) \\ &= Ke^{\alpha x+\beta\tau}(I_+ - I_-) \\ &= Ke^{\alpha x+\beta\tau}e^{x-\alpha x-\beta\tau}\cdot N(d_+) - Ke^{\alpha x+\beta\tau}e^{-\alpha x-\beta\tau-k\tau}N(d_-) \\ &= Ke^{x}N(d_+) - Ke^{-k\tau}N(d_-) \\ &= SN(d_+) - Ke^{-r(T-t)}N(d_-) \end{aligned}$$

4.4 Exercises

1. Draw some graphs of the distribution of a Poisson process X_t for various values of t (and λ). What happens as t increases?

2. Since for a Poisson process $P_{jk}(t) = P_{0,k-j}(t)$ for all positive t, show that if $X_0 = j$, then the random variable $X_t - j$ has a Poisson distribution with parameter λt. Then, use the Markov assumption to prove that $X_t - X_s$ has Poisson distribution with parameter $\lambda(t-s)$.

3. Let T_m be the continuous random variable that gives the time to the mth jump of a Poisson process with parameter λ. Find the pdf of T_m.

4. For a Poisson process with parameter λ, find $P(X_s = m|X_t = n)$, both for the case
$$0 \le n \le m \qquad \text{and} \qquad 0 \le t \le s$$
and the case
$$0 \le m \le n \qquad \text{and} \qquad 0 \le s \le t$$

5. Verify that the expected value and variance of W_t^{Δ} are as claimed in the text.

6. Memorize the properties of the Wiener process and close your notes and books and write them on them down.

7. Verify that the function $p(x,t)$ obtained at the end of the section satisfies the Fokker-Planck equation.

8. Calculate $E(e^{\lambda W_t})$.

9. Suppose a particle is moving along the real line following a Wiener process. At time 10 you observe the position of the particle is 3. What is the probability that at time 12 the particle will be found in the interval $[5, 7]$. (Write down the expression and then use a computer, or a table if you must, to estimate this value.)

10. On a computer (or a calculator), find the smallest value of $n = 1, 2, 3, \cdots$ such that $P(|W_1| \leq n)$ is calculated as 0.

11. What is $E(W_s^2 W_t)$. You must break up the cases $s < t$ and $s > t$.

The following three problems provide justification for the Ito calculus rule $(dW_s)^2 = ds$. This problems is very similar to the calculation of the Ito integral $\int W_t dW_t$. Recall that for a process ϕ_t adapted to X_t how the Ito integral

$$\int_{t_0}^{t} \phi_s dX_s$$

is defined. First, for a partition $\mathcal{P} = \{t_0 < t_1 < t_2 < \cdots t_n = t\}$ we define the Ito sum

$$I(\phi_s, \mathcal{P}, dX_s) = \sum_{j=1}^{n} \phi_{t_{j-1}} \Delta X_{t_j}$$

where $\Delta X_{t_j} = X_{t_j} - X_{t_{j-1}}$. Then one forms the limit as the mesh size of the partitions go to zero,

$$\int_{t_0}^{t} \phi_s dX_s = \lim_{\text{mesh } \mathcal{P} \to 0} I(\phi_s, \mathcal{P}, dX_s)$$

12. Let W_t be the Wiener process. Show
    (a) Show $E((\Delta W_{t_i})^2) = \Delta t_i$.
    (b) Show $E(((\Delta W_{t_i})^2 - \Delta t_i)^2) = 2(\Delta t_i)^2$.

13. For a partition $\mathcal{P}$ as above, calculate the mean square difference which is by definition

$$E((I(\phi_s, \mathcal{P}, (dW_s)^2) - I(\phi_s, \mathcal{P}, dt))^2)$$

$$= E((\sum_{j=1}^{n} \phi_{t_{j-1}}(\Delta W_{t_j})^2 - \sum_{j=1}^{n} \phi_{t_{j-1}}\Delta t_j)^2)$$

and show that it equals

$$2\sum_{j=1}^{n} E(\phi_{t_{j-1}}^2)(\Delta t_j)^2$$

14. Show that for an adapted process ϕ_s such that $E(\phi_s^2) \leq M$ for some $M > 0$ and for all s, that the limit as the mesh $\mathcal{P} \to 0$ of this expression is 0 and thus that for any adapted ϕ_s

$$\int_{t_0}^{t} \phi_s(dW_s)^2 = \int_{t_0}^{t} \phi_s ds$$

15. Calculate

$$\int_0^t W_s^n dW_s$$

16. (a) Solve the stock model SDE

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

   (b) Find the pdf of S_t.

   (c) If $\mu = .1$ and $\sigma = .25$ and $S_0 = 100$ what is

$$P(S_1 \geq 110)$$

.

17. What is $d(sin(W_t))$

18. Solve the Vasicek model for interest rates

$$dr_t = a(b - r_t)dt + \sigma dW_t$$

subject to the initial condition $r_0 = c$. (Hint: You will need to use a multiplier similar to the one for the Ornstein-Uhlenbeck process.)

19. This problem is long and has many parts. If you can not do one part, go onto the next.

    In this problem, you will show how to synthesize any derivative (whose pay-off only depends on the final value of the stock) in terms of calls with different strikes. This allows us to express the value of any derivative in terms of the Black-Scholes formula.

    (a) Consider the following functions: Fix a number $h > 0$. Define

    $$\delta_h(x) = \begin{Bmatrix} 0 \text{ if } |x| \geq h \\ (x+h)/h^2 \text{ if } x \in [-h, 0] \\ (h-x)/h^2 \text{ if } x \in [0, h] \end{Bmatrix} \tag{4.4.1}$$

    Show that collection of functions $\delta_h(x)$ for $h > 0$ forms a δ-function.

    Recall that this implies that for any continuous function f that

    $$f(y) = \lim_{h \to 0} \int_{-\infty}^{\infty} f(x)\delta_h(x-y)dx$$

    (b) Consider a non-divdend paying stock S. Suppose European calls on S with expiration T are available with arbitrary strike prices K. Let $C(S, t, K)$ denote the value at time t of the call with strike K. Thus, for example, $C(S, T, K) = \max(S - K, 0)$.

    Show

    $$\delta_h(S-K) = \frac{1}{h^2}\left(C(S, T, K+h) - 2C(S, T, K) + C(S, T, K-h)\right)$$

    (c) Let $h(S)$ denote the payoff function of some derivative. Then we can synthesize the derivative by

    $$h(S) = \lim_{h \to 0} \int_0^{\infty} h(K)\delta_h(S-K)dK$$

    If $V(S, t)$ denotes the value at time t of the derivative with payoff function h if stock price is S (at time t) then why is $V(S, t) =$

    $$\lim_{h \to 0} \int_0^{\infty} h(K)\frac{1}{h^2}\left(C(S, T, K+h) - 2C(S, T, K) + C(S, T, K-h)\right)dK$$

(d) Using Taylor's theorem, show that

$$\lim_{h\to 0} \frac{1}{h^2}\left(C(S,T,K+h) - 2C(S,T,K) + C(S,T,K-h)\right)$$

$$= \frac{\partial^2 C(S,t,K)}{\partial K^2}$$

(e) Show that $V(S,t)$ is equal to

$$h(K)\frac{\partial C(S,t,K)}{\partial K}\Big|_{K=0}^{\infty} - h'(K)C(S,t,K)\Big|_{K=0}^{\infty} + \int_0^{\infty} h''(K)C(S,t,K)dK$$

(Hint: Integrate by parts.)