Notes on Stochastic Processes

Kiyoshi Igusa

December 17, 2006

These are lecture notes from a an undergraduate course given at Brandeis University in Fall 2006 using the second edition of Gregory Lawler's book "Introduction to Stochastic Processes".

Contents

Math 56a: Introduction to Stochastic Processes and Models

Kiyoshi Igusa, Mathematics

August 31, 2006

A *stochastic process* is a random process which evolves with time. The basic model is the *Markov chain*. This is a set of "states" together with transition probabilities from one state to another. For example, in simple epidemic models there are only two states: S = "susceptible" and I = "infected." The probability of going from S to I increases with the size of I. In the simplest model The $S \to I$ probability is proportional to I, the $I \to S$ probability is constant and time is discrete (for example, events happen only once per day). In the corresponding deterministic model we would have a first order recursion.

In a continuous time Markov chain, transition events can occur at any time with a certain probability density. The corresponding deterministic model is a first order differential equation. This includes the "general stochastic epidemic."

The number of states in a Markov chain is either finite or countably infinite. When the collection of states becomes a continuum, e.g., the price of a stock option, we no longer have a "Markov chain." We have a more general stochastic process. Under very general conditions we obtain a Wiener process, also known as Brownian motion. The mathematics of hedging implies that stock options should be priced as if they are exactly given by this process. Ito's formula explains how to calculate (or try to calculate) stochastic integrals which give the long term expected values for a Wiener process.

This course will be a theoretical mathematics course. The goal is to give rigorous mathematical definitions and derive the consequences of simple stochastic models, most of which will be Markov chains. I will not explain the principles of biology, economics and physics behind the models, although I would invite more qualified guest lecturers to explain the background for the models. There will be many examples, not just the ones outlined above.

The prerequisite for the course will be Math 36a (probability using calculus), linear algebra (Math 15) and multivariable calculus (Math 20). Basic linear differential equations will be reviewed at the beginning of the course. Probability will not be reviewed. This is an advanced course for students already familiar with probability. Linear algebra is also heavily used. Statistics is not required.

Outline of course

1. Review of linear recursion and linear differential equations.
2. Definitions and general notions about stochastic processes
3. Finite Markov chains
4. Renewal processes
5. Continuous time Markov chains
6. Martingales
7. Wiener processes (Brownian motion)
8. Stochastic integration and Ito's formula
9. Applications to population biology and epidemiology
10. Application to financial security analysis

Applications will vary according to the interests of students and teacher.

Required text *Introduction to Stochastic Processes*, Gregory Lawler, Chapman & Hall

Recommended books:

- *Markov Chains*, J.R. Norris, Cambridge University Press. (This is an excellent book which develops Markov chains in a more leisurely way but does not have stochastic integrals.)
- *Epidemic Modelling, An Introduction*, D.J. Daley & J.Gani, Cambridge University Press
- *Financial Derivatives, Pricing, Applications and Mathematics*, Jamil Baz & George Chacko, Cambridge University Press
- *The Mathematics of Financial Derivatives, A Student Introduction*, Paul Wilmott, Sam Howison, Jeff Dewynne, Cambridge University Press

Grading 50% homework, 50% in-class performance. Expected grade: A-/B+

There will be weekly homework. The first HW might have the problem: Find a formula for the n-th Fibonacci number by solving the linear recurrence. Students are encouraged to work on their homework in groups and to access all forms of aid including expert advice, internet and other resources. The work you hand in should, however, be in your own words and in your own handwriting. And you should understand what you have written.

In-class activities: "quizzes" will be given every week or other week. Students should form groups of 3 or 4 to work on these problems in class, solve them and help other students in the group to

understand them. Each group should hand in their answers signed by all members of the group. Every student is required to give at least one short oral presentation in front of the class. Attendance is required and counts as part of the grade.

Students with disability If you are a student with a documented disability at Brandeis University and if you wish to request a reasonable accommodation for this class, please see the instructor immediately.

Academic integrity All members of the Brandeis academic community are expected to maintain the highest standard of academic integrity as outlined in "Rights and Responsibilities." Any violations will be treated seriously.

MATH 56A: STOCHASTIC PROCESSES
CHAPTER 0

0. Chapter 0

I "reviewed" basic properties of linear differential equations in one variable. I still need to do the theory for several variables.

0.1. **linear differential equations in one variable.** In the first lecture I discussed linear differential equations in one variable. The problem in degree $n = 2$ is to find a function $y = f(x)$ so that:

$$y'' + ay' + by + c = 0 \tag{0.1}$$

where a, b, c are constants.

The general n-th order linear diff eq in one variable is:

$$y^{(n)} + a_1 y^{(n-1)} + a_2 y^{(n-2)} + \cdots + a_n y + a_{n+1} = 0$$

where $a_1, \cdots, a_{n+1}$ are constant.

0.1.1. *associated homogeneous equation.* The standard procedure is to first take the *associated homogeneous equation* which is given by setting $a_{n+1} = 0$. For the 2nd order equation we get:

$$y'' + ay' + by = 0 \tag{0.2}$$

Lemma 0.1. *If $y = f(x)$ is a solution of the homogeneous equation then so is $y = \alpha f(x)$ for any scalar α.*

Lemma 0.2. *If $y = f_0(x)$ is a solution of the original equation (0.1) and $y = f_1(x)$ is a solution of the homogeneous equation (0.2) then $y = f_0(x) + f_1(x)$ is a solution of the original equation (0.1). Similarly, if f_0, f_1 are solutions of the homogeneous equation then so is $f_0 + f_1$.*

Theorem 0.3. *The set of solutions of the homogenous equation is a vector space of dimension equal to n (the order of the equation).*

This means that if we can find n linearly independent solutions $f_1, f_2, \cdots, f_n$ of the homogeneous equation then the general solution (of the homogenous equation) is a linear combination:

$$y = \alpha_1 f_1 + \alpha_2 f_2 + \cdots + \alpha_n f_n$$

Date: September 28, 2006.

0.1.2. *complete solution of linear diffeqs.* A general solution of the linear diffeq is given by adding a particular solution f_0 to the general solution of the homogeneous equation:

$$y = f_0 + \alpha_1 f_1 + \alpha_2 f_2 + \cdots + \alpha_n f_n$$

The particular solution is easy to find:

$$f_0(x) = -\frac{a_{n+1}}{a_n}$$

if $a_n \neq 0$

$$f_0(x) = - - \frac{a_{n+1}}{a_{n-1}} x$$

if $a_n = 0$ but $a_{n-1} \neq 0$.

The solutions of the homogeneous equations are given by guessing. We just need to find n linearly independent solutions. We guess that $y = e^{\lambda x}$. Then

$$y^k = \lambda^k e^{\lambda x}$$

So, the homogenous equation is:

$$\lambda^n e^{\lambda x} + a_1 \lambda^{n-1} e^{\lambda x} + \cdots + a_n e^{\lambda x} = 0$$

Since $e^{\lambda x}$ is never zero we can divide to get:

$$\lambda^n + a_1 \lambda^{n-1} + \cdots + a_n = 0$$

This is a monic (coefficient of λ^n is 1) polynomial of degree n. So, it has n complex roots $\lambda_1, \lambda_2, \cdots, \lambda_n$. If the roots are distinct then the solutions

$$f_1(x) = e^{\lambda_1 x}, f_2(x) = e^{\lambda_2 x}, \cdots$$

are linearly independent and span the solution space.

If roots repeat, e.g., if $\lambda_1 = \lambda_2 = \lambda_3$ then the functions f_2, f_3 are given by

$$f_2(x) = x e^{\lambda_1 x}, \ f_3(x) = x^2 e^{\lambda_1 x},$$

0.1.3. *separation of variables.* Finally, I talked about *separation of variables*. This just means put all the x's on one side of the equation and all the y's on the other side. For example:

$$\frac{dy}{dx} = xy$$

This is a nonlinear diffeq. Separating variables we get:

$$\frac{dy}{y} = x dx$$

Now integrate both sides:

$$\int \frac{dy}{y} = \int x\,dx$$

$$\ln y = \frac{x^2}{2} + C$$

Note that there is only one constant. (You get a constant on both sides and C is the difference between the two constants.) The final solution is:

$$y = y_0 exp\left(\frac{x^2}{2}\right)$$

where $y_0 = e^C$.

0.2. **Kermack-McKendrick.** This is from the book *Epidemic Modelling, An Introduction*, D.J. Daley & J.Gani, Cambridge University Press. Kermack-McKendrick is the most common model for the general epidemic. I made two simplifying assumptions:

- the population is homogeneous and
- no births or deaths by other means

Since there are no births, the size of the population N is fixed.

In this model there are three states:

S: = susceptible
I: = infected
R: = removed (immune)

Let $x = \#S, y = \#I, z = \#R$. So

$$N = x + y + z$$

I assume that $z_0 = 0$ (If there are any "removed" people at $t = 0$ we ignore them.)

As time passes, susceptible people become infected and infected recover and become immune. So the size of S decreases and the size of R increases. People move as shown by the arrows:

$$S \longrightarrow I \longrightarrow R$$

The infection rate is given by the *Law of mass action* which says:

The rate of interaction between two different subsets of the population is proportional to the product of the number of elements in each subset.

So,

$$\frac{dx}{dt} = -\beta xy$$

where β is a positive constant.

Infected people recover at a constant rate:

$$\frac{dz}{dt} = \gamma y$$

where γ is a positive constant.

This is a system of nonlinear equations. To solve it we make it linear by dividing:

$$\frac{dx}{dz} = \frac{dx/dt}{dz/dt} = \frac{-\beta xy}{\gamma y} = \frac{-\beta x}{\gamma}$$

This is a linear differential equation with solution

$$x = x_0 exp\left(\frac{-\beta}{\gamma} z\right) = x_0 e^{-z/\rho}$$

where $\rho = \gamma/\beta$ is the *threshold* population size.

Since $x + y + z$ is fixed we can find y:

$$y = N - x - z = N - x_0 e^{-z/\rho} - z$$

Differentiating gives:

$$\frac{dy}{dz} = \frac{x_0}{\rho} e^{-z/\rho} - 1$$

$$\frac{d^2y}{dz^2} = -\frac{x_0}{\rho^2} e^{-z/\rho} < 0$$

So, the function is concave down with initial slope

$$\frac{dy}{dz} = \frac{x_0}{\rho} - 1$$

which is positive if and only if $x_0 > \rho$. So, according to this model, the number of infected will initially increase if and only if the number of susceptibles is greater than the threshold value $\rho = \gamma/\beta$.

By plotting the graphs of the functions we also see that the infection will taper off with a certain number of susceptibles x_∞ who never get infected.

0.3. **systems of first order equations.** I explained that differential equations involving second and higher order derivatives can be reduced to a system of first order equations by introducing more variables. Then I did the following example.

$$y' = z$$

$$z' = 6y - z$$

In matrix form this is:

$$\frac{d}{dt}\begin{pmatrix} y \\ z \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ 6 & -1 \end{pmatrix}\begin{pmatrix} y \\ z \end{pmatrix}$$

Which we can write as $Y = AY$ with

$$Y = \begin{pmatrix} y \\ z \end{pmatrix}, \qquad A = \begin{pmatrix} 0 & 1 \\ 6 & -1 \end{pmatrix}$$

0.3.1. *exponential of a matrix.* The solution of this equations is

$$Y = e^{tA}Y_0$$

where $Y_0 = \begin{pmatrix} y_0 \\ z_0 \end{pmatrix}$ and

$$e^{tA} := I_2 + tA + \frac{t^2A^2}{2} + \frac{t^3A^3}{3!} + \cdots$$

This works because the derivative of each term is A times the previous term:

$$\frac{d}{dt}\frac{t^kA^k}{k!} = \frac{kt^{k-1}A^k}{k!} = \frac{t^{k-1}A^k}{(k-1)!} = A\frac{t^{k-1}A^{k-1}}{(k-1)!}$$

So,

$$\frac{d}{dt}e^{tA} = Ae^{tA}$$

0.3.2. *diagonalization (corrected).* Then I explained how to compute r^{tA}. You have to *diagonalize* A. This means

$$A = QDQ^{-1}$$

where D is a diagonal matrix $D = \begin{pmatrix} d_1 & 0 \\ 0 & d_2 \end{pmatrix}$.

I should have explained this formula so that I get it right: If X_1, X_2 are eigenvalues of A with eigenvalues d_1, d_2 then $AX_1 = X_1d_1, AX_2 = X_2d_2$ and

$$A(X_1X_2) = (X_1d_1X_2d_2) = (X_1X_2)\begin{pmatrix} d_1 & 0 \\ 0 & d_2 \end{pmatrix}$$

Solve for A gives

$$A = (X_1 X_2) \begin{pmatrix} d_1 & 0 \\ 0 & d_2 \end{pmatrix} (X_1 X_2)^{-1} = QDQ^{-1}$$

where $Q = (X_1, X_2)$.

This is a good idea because $A^2 = QDQ^{-1}QDQ^{-1} = QD^2Q^{-1}$ and more generally, $t^k A^k = t^k Q D^k Q^{-1}$. Divide by $k!$ and sum over k to get:

$$e^{tA} = Qe^{tD}Q^{-1} = Q \begin{pmatrix} e^{td_1} & 0 \\ 0 & e^{td_2} \end{pmatrix} Q^{-1}$$

0.3.3. *eigenvectors and eigenvalues.* The diagonal entries d_1, d_2 are the eigenvalues of the matrix A and $Q = (X_1 X_2)$ where X_i is the eigenvector corresponding to d_i. This works if the eigenvalues of A are distinct. The eigenvalues are defined to be the solutions of the equation

$$\det(A - \lambda I) = 0$$

but there is a trick to use for 2×2 matrices. The determinant of a matrix is always the product of its eigenvalues:

$$\det A = d_1 d_2 = -6$$

The trace (sum of diagonal entries) is equal to the sum of the eigenvalues:

$$trA = d_1 + d_2 = -1$$

So, $d_1 = 2, d_2 = -3$. The eigenvalues are $X_1 = \begin{pmatrix} 1 \\ 2 \end{pmatrix}$ and $X_2 = \begin{pmatrix} 1 \\ -3 \end{pmatrix}$. So

$$Q = (X_1 X_2) = \begin{pmatrix} 1 & 1 \\ 2 & -3 \end{pmatrix}$$

$$Q^{-1} = \frac{1}{\det Q} \begin{pmatrix} -3 & -1 \\ -2 & 1 \end{pmatrix} = \begin{pmatrix} 3/5 & 1/5 \\ 2/5 & -1/5 \end{pmatrix}$$

The solution to the original equation is

$$\begin{pmatrix} y \\ z \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 2 & -3 \end{pmatrix} \begin{pmatrix} e^{2t} & 0 \\ 0 & e^{-3t} \end{pmatrix} \begin{pmatrix} 3/5 & 1/5 \\ 2/5 & -1/5 \end{pmatrix} \begin{pmatrix} y_0 \\ z_0 \end{pmatrix}$$

0.4. **Linear difference equations.** We are looking for a sequence of numbers $f(n)$ where n ranges over all the integers from K to N ($K \leq n \leq N$) so that

$$f(n) = af(n-1) + bf(n+1) \tag{0.3}$$

I pointed out that the solution set is a vector space of dimension 2. So we just have to find two linearly independent solutions. Then I followed the book.

The solution has the form $f(n) = c^n$ where you have to solve for c:

$$c^n = ac^{n-1} + bc^{n+1}$$

$$bc^2 - c + a = 0 \tag{0.4}$$

$$c = \frac{1 \pm \sqrt{1-4ab}}{2b}$$

There were two cases.

Case 1: ($4ab \neq 1$) When the quadratic equation (0.4) has two roots C_1, c_2 then the linear combinations of c_1^n and c_2^n give all the solutions of the homogeneous linear recursion (0.3).

Case 2: ($4ab = 1$) In this case there is only one root $c = \frac{1}{2b}$ and the two independent solutions are $f(n) = c^n$ and nc^n. The reason we get a factor of n is because when a linear equation has a double root then this root will also be a root of the derivative. This gives $f(n) = nc^{n-1}$ as a solution. But then you can multiply by the constant c since the equation is homogeneous.

Example 0.4. *(Fibonacci numbers) These are given by $f(0) = 1, f(1) = 1$ and $f(n+1) = f(n) + f(n-1)$ or:*

$$f(n) = f(n+1) - f(n-1)$$

This is $a = -1, b = 1$. The roots of the quadratic equation are $c = \frac{1 \pm \sqrt{5}}{2}$. So,

$$f(n) = \frac{1}{\sqrt{5}} \left(\frac{1+\sqrt{5}}{2} \right)^n - \frac{1}{\sqrt{5}} \left(\frac{1-\sqrt{5}}{2} \right)^n$$

This is a rational number since it is Galois invariant (does not change if you switch the sign of $\sqrt{5}$). However, it is not clear from the formula why it is an integer.

MATH 56A: STOCHASTIC PROCESSES
CHAPTER 1

1. Finite Markov chains

For the sake of completeness of these notes I decided to write a summary of the basic concepts of finite Markov chains. The topics in this chapter are:

(1) definition of a Markov chain
(2) communication classes
(3) classification of states
(4) periodic/aperiodic
(5) invariant probability distribution
(6) return time
(7) substochastic matrix

1.1. **definition.** A *stochastic process* is a random process which evolves with time. In other words, we have random variables X_t, Y_t, etc. which depend on time t. For example X_t could be your location at time t (where t is in the future).

A *finite Markov chain* is a special kind of stochastic process with the following properties

- There are only a finite number of states. If we are talking about your location, this means we are only interested in which building you are in and not your exact position in the building.
- Time is discrete: For example, things change only at 1pm, 2pm, etc. and never at 1:12pm or any time in between. X_n is the state at time n where $n = 0, 1, 2$, etc.
- No memory. Your (random) movement at time n depends only on X_n and is independent of X_t for $t < n$ (You forget the past and your decision making process is based only on the present state.)
- Time homogeneous: Your rules of movement are the same at all times.

To summarize: You have a finite number of building that you can move around in. You can only move on the hour. Your decision making process is random and depends only on your present location and not

Date: December 15, 2006.

on past locations and does not take into account what time it is. (You move at midnight in the same way that you do at noon.)

Definition 1.1. *A* finite Markov chain *is a sequence of random variables $X_0, X_1, \cdots$ which take values in a finite set S called the* state space *and for any n and any values of $x_0, x_1, \cdots, x_n$ we have:*

$$\mathbb{P}(X_{n+1} = x \,|\, X_0 = x_0, X_1 = x_1, \cdots, X_n = x_n) = \mathbb{P}(X_1 = x \,|\, X_0 = x_n)$$

The $S \times S$ matrix P with entries

$$p(x, y) := \mathbb{P}(X_1 = y \,|\, X_0 = x)$$

is called the transition matrix*.*

For example, suppose that you have 4 points: 0,1,2,3 and at each step you either increase one or decrease one with equal probability except at the end points. Suppose that, when you reach 0, you cannot leave. (0 is called an *absorbing state*.) Suppose that when you reach 3 you always go to 2. (3 is called a *reflecting wall*.) Then the transition matrix is

$$P = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 1/2 & 0 & 1/2 & 0 \\ 0 & 1/2 & 0 & 1/2 \\ 0 & 0 & 1 & 0 \end{pmatrix}$$

Notice that the numbers are all nonnegative and the numbers in each row add up to 1. This characterizes all transition matrices.

In the discussion of Markov chains, there are qualitative non-numerical concepts and quantitative, computational concepts. The qualitative concepts are: communication classes, their classification and periodicity.

1.2. **communication classes.** Two states *communicate with each other* if they are equal or if it is possible (with positive probability) to get from either one to the other in a finite amount of time. We write $x \leftrightarrow y$. This is an equivalence relation and the equivalence classes are called *communication classes*.

In the example above, $\{0\}$ and $\{1, 2, 3\}$ are the communication classes.

A Markov chain is *irreducible* if it has only one communication class, i.e., if it is possible to get from any state to any other state.

1.3. **classification of states: transient and recurrent.** There are two types of communication classes: recurrent and transient. At this point, I allowed the state space S to be infinite so that you don't get the wrong idea.

A communication class C is *recurrent* if for any state $x \in C$, you will keep returning to x an infinite number of times with probability one. A communication class C is *transient* if, starting at any $x \in C$, you will return to x only a finite number of times with probability one.

The theorem is that these are the only two possibilities. I proved this in class:

Lemma 1.2. *If $p = p(i,j) > 0$ and i is recurrent then so is j. In fact, if you start in state i you will go to state j an infinite number of times with probability one.*

Proof. This is the same as saying that the probability of going to state j only a finite number of times is zero. To prove this suppose that the Markov chain goes to state j only a finite number of times. Then there is a last time, say $X_m = j$. Then you can never return to j.

But i is recurrent. So, with probability one, you will go to i an infinite number of times after time m. Say at times $n_1 < n_2 < n_3 < \cdots$ (all $> m$). But

$$\mathbb{P}(X_{n_1+1} \neq j \,|\, X_{n_1} = i) = 1 - p$$

$$\mathbb{P}(X_{n_2+1} \neq j \,|\, X_{n_2} = i) = 1 - p$$

The product is $(1-p)^2, (1-p)^3$, etc. which converges to 0. So, the probability is zero that in all of these infinite times that you visit i you will never go to j. This is what I was claiming. □

Theorem 1.3. *Once you leave a communication class you can never return.*

Theorem 1.4. *Recurrent communication classes are absorbing.*

The lemma shows that, if you could leave a recurrent communication class, you will with probability one. This would be a contradiction to the definition of recurrent. So, you cannot leave a recurrent class.

The lemma that I proved can be rephrased as follows:

Lemma 1.5. *If you make an infinite number of attempts and you have a fixed positive probability p of success then, with probability one, you will succeed an infinite number of times.*

The *strong law of large numbers* says that, with probability one, the proportion of trials which are successful will converge to p as the number of trials goes to infinity. I.e., the *experimental value* of the probability p will converge to the *theoretical value* with probability one.

1.4. **periodic chains.** We are interested in the time it takes to return to a state i. The *return time* T_i to state i is the smallest positive integer n so that $X_n = i$ given that $X_0 = i$. In other words, you start at state i and count how many turns it takes to return to the same state i. This number is T_i. It is random. For example, in the random walk example given above, $\mathbb{P}(T_3 = 2) = 1/2, \mathbb{P}(T_2 = 2) = 3/4$.

The *period* of a state i is the greatest common divisor of all possible return times to state i. For the random walk on an infinite straight line (or on a finite line with reflecting walls), the period of every state is 2 because it always takes an even number of steps (the same number right as left) to get back to the same point.

A state i is *aperiodic* if the period is 1.

Theorem 1.6. *States in the same communication class have the same period.*

1.5. **Invariant probability distribution.**

Definition 1.7. *A probability distribution π is called* invariant *if*

$$\pi P = \pi$$

Remember that a *probability distribution* is a vector with nonnegative coordinates adding up to 1:

$$\sum_{i=1}^{n} \pi(i) = 1, \quad \pi(i) \geq 0$$

where n is the number of states. As an application of the Perron-Frobenius theorem we get:

Theorem 1.8. *If the Markov chain is irreducible and aperiodic then P^n converges to a matrix in which every row is equal to the invariant distribution π which is unique:*

$$P^\infty = \begin{pmatrix} \pi \\ \pi \\ \cdots \\ \pi \end{pmatrix}$$

If the Markov chain is periodic then P^n depends on n modulo the period d. However, the average value of P^n will still converge:

Theorem 1.9. *For any finite Markov chain the average value of P^n converges to a matrix in which every row is equal to the unique invariant distribution π:*

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} P^k = \begin{pmatrix} \pi \\ \cdots \\ \pi \end{pmatrix}$$

Proof. Let

$$S_n = \frac{1}{n}\sum_{k=1}^{n} P^k$$

Then each row of S_n adds to 1 and

$$S_n P = S_n + \frac{1}{n}(P^{n+1} - P)$$

So, $S_n P \approx S_n$ and $S_\infty P = S_\infty$. Since each row of S_∞ adds to 1, each row is equal to the invariant distribution π (which is unique by Perron-Frobenius). □

1.6. **Return time.** I explained in class the relation between the return time T_i to state i and the value of the invariant distribution $\pi(i)$:

$$\mathbb{E}(T_i) = \frac{1}{\pi(i)}$$

Proof. Begin with the last theorem:

$$\pi(j) = \lim_{n\to\infty} \frac{1}{n}\sum_{m=1}^{n} \mathbb{P}(X_m = j)$$

Now use the fact that probability is the expected value of the indicator function: $\mathbb{P}(A) = \mathbb{E}(I(A))$ and expected value is linear: $\sum \mathbb{E} = \mathbb{E}\sum$

$$\pi(j) = \lim_{n\to\infty} \frac{1}{n}\mathbb{E}\left(\sum_{m=1}^{n} I(X_m = j)\right)$$

This is the average expected number of visits to state j. If, in n steps, you visit a state k times then the average time from one visit to the next is n/k and the average number of visits is k/n. So they are inverse to each other:

$$\pi(j) = \frac{k}{n} = \frac{1}{n/k} = \frac{1}{\mathbb{E}(T_j)}$$

□

1.7. **substochastic matrix.** If there are both absorbing classes and transient classes in the Markov chain then you get a *substochastic matrix* Q which is the transient-to-transient transition matrix: $Q = (p(x, y))$ where we take x, y only from transient states. (Actually, you can take any subset of the set of states.) Since this is only part of the transition matrix P, the rows may not add up to 1. But we know that the entries are all nonnegative and the rows add up to at most 1.

Definition 1.10. *A* substochastic matrix *is a square matrix Q whose entries are all nonnegative with rows adding up to at most 1.*

I used a model from economics to explain what was the point of doing this.

1.7.1. *Leontief model.* In this model we have r factories which produce goods. For every dollar worth of goods that factor i produces, it needs q_{ij} dollars worth of the output of factory j. In order to be profitable or to at least break even we need the production cost to be less than or equal to one dollar:

$$q_{i1} + q_{i2} + \cdots + q_{ir} \leq 1$$

In other words, each row of the matrix $Q = (q_{ij})$ must add up to at most 1. So, Q is a substochastic matrix.

To analyze how this works we follow the dollar. q_{ij} represents goods going from j to i and it represents money going from i to j.

Now look at the total amount of money and what happens to it. Out of each dollar that factory i gets, it must give $\sum q_{ij}$ to other factories. What remains:

$$1 - \sum q_{ij}$$

is profit. Let's say it puts this money in the bank. When we add the bank to the system we get a Markov chain with $r + 1$ states. The transition matrix is $P = (p(i, j))$ where

$$p(i,j) = \begin{cases} q_{ij} & \text{if } i, j \neq 0 \\ 1 - \sum_k q_{ik} & \text{if } i \neq 0, j = 0 \\ 0 & \text{if } i = 0, j \neq 0 \\ 1 & \text{if } i = j = 0 \end{cases}$$

Note that this formula can be used to convert any substochastic matrix into a Markov chain by adding one absorbing state.

Problem The problem is to figure out how much each factory needs to produce in order for the net production (not counting inter-industry consumption) to be equal to a fixed vector γ. This is the *consumption vector*. The consumer wants γ_i dollars worth of stuff from factory i. To find the answer we just follow the money after the consumer buys the goods.

I explained it in class like this: The consumer orders γ_i worth of goods from factory i. On day zero, each factory i produces the requested goods using its inventory of supplies. Then it orders supplies from the other factories to replenish its inventory. On day one, each factory produces goods to fill the orders from the other factories using its inventory. And so on. Eventually, (in the limit as $n \to \infty$), the inventories are all back to normal and all the money is in the bank *assuming that all the factories are transient states*. The total production

is given by adding up the production on each day. Factory i produces γ_i on day 0,

$$\sum_j \gamma_j q_{ji} = (\gamma Q)_i$$

on day 1,

$$\sum_j \gamma_j q_{jk} \sum_k q_{ki} = (\gamma Q^2)_i$$

on day 2, $(\gamma Q^3)_i$ on day 3, etc. So, the total production of factory i is

$$(\gamma(I + Q + Q^2 + Q^3 + \cdots))_i = (\gamma(I - Q)^{-1})_i$$

Lemma 1.11. *The factories are all transient states if and only if $I-Q$ is invertible.*

Proof. If the factories are all transient then the money will all eventually end up in the bank. Equivalently, the matrices Q^n converge to zero. So $I - Q$ is invertible. Conversely, if there is a recurrent class, it must consist of nonprofit organizations which require only the output from other nonprofits. Then these nonprofit factories give a Markov process with equilibrium distribution π. This will be a vector with $\pi Q = \pi$. So, it is a null vector of $I - Q$ showing that $I - Q$ is not invertible. □

1.7.2. *avoiding states.* Substochastic matrices are used to calculate the probability of reaching one set of states A before reaching another set B (assuming that they are disjoint subsets of S). To do this you first combine them into two recurrent states. You also need to assume there are no other recurrent states (otherwise the Markov chain could get stuck in a third state and never reach A or B).

Suppose that there are n transient states and two absorbing states A and B. Let Q be the $n \times n$ transient-to-transient transition matrix and let S be the $n \times 2$ transient-to-recurrent matrix. Then the (i, A) entry of $Q^k S$ is the probability of getting from state i to state A in exactly $k + 1$ steps. So, the total probability of ending up in state A is

$$\mathbb{P}_i(A) = ((I + Q + Q^2 + Q^3 + \cdots)S)_{iA} = ((I - Q)^{-1}S)_{iA}$$

MATH 56A: STOCHASTIC PROCESSES
CHAPTER 2

2. Countable Markov Chains

I started Chapter 2 which talks about Markov chains with a countably infinite number of states. I did my favorite example which is on page 53 of the book.

2.1. **Extinction probability.** In this example we consider a population of one cell creatures which reproduce asexually. At each time interval, each creature produces X offspring and then dies. Here X is a random variable equal to $0, 1, 2, 3$ etc with probabilities $p_k = \mathbb{P}(X = k)$.

You can read the complete analysis in the book, but what I explained in class was the most striking part: if the average number of offspring is equal to 1 (and it is somewhat random, i.e., not always equal to 1) then the probability of extinction of the population is one.

In this model the state space is the set of nonnegative integers. If the state is i then there are i creature in the population. At each time interval each of these creatures dies and produces a random number of offspring. The state 0 is an absorbing state. We want to know the probability that we eventually land in that state. We start with the definition:

$$a := \mathbb{P}(X_n = 0 \text{ eventually} | X_0 = 1)$$

So, a is the probability that the species eventually becomes extinct if it starts out with exactly one creature. The theorem is that $a = 1$. This implies that population dies out if we start with any number of creatures. The reason is that, because of asexual reproduction, the descendants of each individual will not mix and so, *assuming independence of the probabilities of extinction* for each "family," we get that

$$a^k = \mathbb{P}(X_n = 0 \text{ eventually} | X_0 = k)$$

The point is that, if $a = 1$ then $a^k = 1$ for any $k \geq 0$.

To calculate a we look at what happens after one time interval.

$$a = \sum_k \mathbb{P}(X_n = 0 \text{ eventually} | X_1 = k)\mathbb{P}(X_1 = k) = \sum_k a^k p_k$$

Date: September 28, 2006.

But this is the *generating function* for p_k which is defined by

$$\phi(s) := \mathbb{E}(s^X) = \sum_k s^k \mathbb{P}(X = k) = \sum_k s^k p_k$$

The extinction probability is equal to the generating function!!

$$a = \phi(a)$$

The generating function has the property that $\phi(1) = 1$. Here is the proof:

$$\phi(1) = \sum_k 1^k p_k = \sum_k p_k = 1$$

The derivative of $\phi(s)$ is

$$\phi'(s) = \sum_k k s^{k-1} p_k \tag{2.1}$$

If we put $s = 1$ we get the expected value of X (the number of offspring) which we are assuming is equal to 1.

$$\phi'(1) = \sum_k k p_k = \mathbb{E}(X) = 1$$

The second derivative is

$$\phi''(s) = \sum_{k \geq 2} k(k-1) s^{k-2} p_k > 0$$

This is greater than zero (for all s) if $p_2 > 0$ or $p_k > 0$ for some $k \geq 2$. But, if X has average 1 and is not always equal to 1 then it must be sometimes more and sometimes less than 1. So, there is a positive probability that $X \geq 2$. So, $\phi''(s) > 0$.

Now, graph the function $y = \phi(s)$. Since $\phi(1) = 1$ and $\phi'(1) = 1$, the graph goes through the point $(1, 1)$ with slope 1. Since it is concave up, it has to curve away from the line $y = s$ on both sides of that point. So, the only solution to the equation $a = \phi(a)$ is $a = 1$.

For the general analysis we need the following lemma.

Lemma 2.1. *a is the smallest nonnegative solution to the equation $a = \phi(a)$.*

Proof. The eventual extinction probability is a limit of finite extinction probabilities:

$$a = \lim_{n \to \infty} a_n$$

where

$$a_n = \mathbb{P}(X_n = 0 | X_0 = 1)$$

These finite extinction probabilities are calculated recursively as follows:

$$a_n = \sum_k \mathbb{P}(X_n = 0|X_1 = k)\mathbb{P}(X_1 = k|X_0 = 1)$$

$$= \sum_k \mathbb{P}(X_{n-1} = 0|X_0 = k)\mathbb{P}(X_1 = k|X_0 = 1) = \sum_k a_{n-1}^k p_k = \phi(a_{n-1})$$

Let $\hat{a}$ be the smallest nonnegative real number so that $\hat{a} = \phi(\hat{a})$. Then we just have to show that $a_n \leq \hat{a}$ for every $n \geq 0$. This is true for $n = 0$ since

$$a_0 = \mathbb{P}(X_0 = 0|X_0 = 1) = 0 \leq \hat{a}$$

Suppose by induction that $a_{n-1} \leq \hat{a}$. Then we have to show that $a_n \leq \hat{a}$. But, if you look at the equation (2.1) you see that $\phi'(s) \geq 0$. So,

$$a_n = \phi(a_{n-1}) \leq \phi(\hat{a}) = \hat{a}$$

Therefore, $a_n \leq \hat{a}$ for all $n \geq 0$. So, $a \leq \hat{a}$. So, $a = \hat{a}$. □

Theorem 2.2. *If $\mathbb{E}(X) > 1$ then the probability of extinction is less than 1. If $\mathbb{E}(X) \leq 1$ then the probability of extinction is one, except in the case when the population is constant (i.e., when $p_1 = 1$).*

Proof. By the lemma, the extinction probability a is the first point of intersection of the graph of $y = \phi(s)$ with the graph of $y = s$. But $\phi(s)$ goes through the point $(1, 1)$ with slope $\mathbb{E}(X)$ and is always concave up. A drawing of the graphs proves the theorem. □

2.2. **Random walk in an integer lattice.** Today I want to explain the proof that the simple random walk in the lattice $\mathbb{Z}^d$ is recurrent if $d \leq 2$ and transient if $d \geq 3$.

2.2.1. *One dimensional case.* First look at the case $d = 1$. Then the state space is $\mathbb{Z}$ the set of all integers. At each step we go either left or right with probability $\frac{1}{2}$. This is periodic with period 2. We need an even number of steps to get back to the starting point. Let's say the starting point is 0.

$$p_{2n}(0,0) = C(2n,n)\left(\frac{1}{2}\right)^{2n} = \frac{(2n)!}{n!n!}\left(\frac{1}{2}\right)^{2n}$$

Use Stirling's formula:

$$n! \sim e^{-n}n^n\sqrt{2\pi n}$$

where $\sim$ means the ratio of the two sides converges to 1 as $n \to \infty$. So,

$$(2n)! \sim e^{2n}(2n)^{2n}\sqrt{4\pi n}$$

and

$$\frac{(2n)!}{n!n!} \sim \frac{e^{2n}(2n)^{2n}\sqrt{4\pi n}}{e^{2n}n^{2n}2\pi n} = \frac{2^{2n}}{\sqrt{\pi n}}$$

which means that

$$p_{2n}(0,0) \sim \frac{1}{\sqrt{\pi n}}$$

The expected number of return visits to 0 is

$$\sum_{n=1}^{\infty} p_{2n}(0,0) \approx \sum \frac{1}{\sqrt{\pi n}} = \infty$$

So, 0 is recurrent in $\mathbb{Z}$. Since there is only one communication class, all states are recurrent.

2.2.2. *Higher dimensions.* In the planar lattice $\mathbb{Z}^2$, both coordinates must be 0 at the same time in order for the particle to return to the origin. Therefore,

$$p_{2n}(0,0) \sim \frac{1}{\pi n}$$

and

$$\sum_{n=1}^{\infty} p_{2n}(0,0) \approx \sum \frac{1}{\pi n} = \infty$$

and $\mathbb{Z}^2$ is recurrent.

When $d > 2$ we get

$$p_{2n}(0,0) \sim \frac{1}{(\pi n)^{d/2}}$$

and

$$\sum_{n=1}^{\infty} p_{2n}(0,0) \approx \sum \frac{1}{(\pi n)^{d/2}}$$

which converges by the integral test. So, the expected number of visits is finite and $\mathbb{Z}^3, \mathbb{Z}^4$, etc are transient.

2.2.3. *Stirling's formula.* Exercise 2.18 on page 62 gives a rigorous proof of Stirling's formula. I will give a simpler proof which misquotes (!) the central limit theorem. It starts with Y_n the Poisson random variable with mean n. It has probability distribution

$$\mathbb{P}(Y_n = k) = e^{-n}\frac{n^k}{k!}$$

The variance of a Poisson variable is equal to its mean. So, the standard deviation is $\sqrt{n}$. The central limit theorem says that, as $n \to \infty$ the Poisson distribution is approximated by the normal distribution $N(n, \sqrt{n})$ with $\mu = n, \sigma = \sqrt{n}$. But Poisson is discrete so you have to take integer steps:

$$\mathbb{P}(Y_n = n) \sim \int_{n-\frac{1}{2}}^{n+\frac{1}{2}} N(n, \sqrt{n}) = \int_{-\frac{1}{2}}^{\frac{1}{2}} N(0, \sqrt{n})$$

$$e^{-n}\frac{n^n}{n!} \sim \frac{1}{\sqrt{2\pi n}} \int_{-\frac{1}{2}}^{\frac{1}{2}} e^{-x^2/n}\, dx$$

But,

$$\lim_{n\to\infty} \int_0^1 e^{-x^2/n}\, dx = \int_0^1 e^0\, dx = 1$$

So,

$$e^{-n}\frac{n^n}{n!} \sim \frac{1}{\sqrt{2\pi n}}$$

This is Stirling's formula when you solve for $n!$.

Note: the central limit theorem does not actually give approximations for single values of discrete probability distributions, it only approximates *sums* of values over a range of values which is a fixed multiple of the standard deviation. However, the book points out that the Poisson distribution is fairly uniform in its values. So, the sum over a range is approximated by a single value times the size of the range. (This is the point of part (b) of the exercise.)

2.3. **Transient-recurrent.** If X_n is an irreducible aperiodic Markov chain then there are 3 possibilities:

(1) X_n is transient
(2) X_n is null recurrent
(3) X_n is positive recurrent

What do these mean and how can you tell which category you are in?

First of all, in the finite case, you always have positive recurrence. Null recurrence is something that happens only in the infinite case.

To tell whether X_n is transient or recurrent we look at the function $\alpha(x)$ defined as follows.

Fix a state $z \in S$. (S is the set of all states. $|S| = \infty$.)

Take $\alpha(x)$ to be the probability that you go from x to z. I.e., you start at x and see whether you are ever in state z:

$$\alpha(x) := \mathbb{P}(X_n = z \text{ for some } n \geq 0 \,|\, X_0 = x)$$

This function satisfies three equations:

(1) $0 \leq \alpha(x) \leq 1$ (since $\alpha(x)$ is a probability)
(2) $\alpha(z) = 1$ (since you start at $x = z$)
(3) If $x \neq z$ then

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

(To get from $x \neq z$ to z you take one step to y and then go from y to z.)

The above three equations are always true. The next equation tells us whether the chain is transient or recurrent.

(4) $\inf_x \alpha(x) = 0$ iff X_n is transient
(5) $\alpha(x) = 1$ for all $x \in S$ iff X_n is recurrent.

Theorem 2.3. *$\alpha(x)$ is the smallest solution of equations (1),(2),(3). I.e., if $\hat{\alpha}(x)$ is another solution then*

$$\alpha(x) \leq \hat{\alpha}(x)$$

for all $x \in S$.

Remark 2.4. *I pointed out in class that $\alpha(x) = 1$ is always a solution of equations (1),(2),(3) since (3), in matrix form is*

$$A = PA$$

I.e., $A = (\alpha(x))$ is a right eigenvector of P with eigenvalue 1.

Proof. $\alpha(x) = \lim_{n\to\infty} \alpha_n(x)$ where $\alpha_n(x)$ is the probability that you get from x to z in n steps or less:

$$\alpha_n(x) := \mathbb{P}(X_k = z \text{ for some } 0 \le k \le n \,|\, X_0 = x)$$

Then, I claim that

$$\alpha_n(x) \le \hat{\alpha}(x) \tag{2.2}$$

for all n and all x. This is true for $n = 0$ since $\alpha_0(x) = 0$ for all $x \neq z$. Suppose that (2.2) holds for $n-1$. Then

$$\alpha_n(x) = \sum_y p(x,y)\alpha_{n-1}(y) \le \sum_y p(x,y)\hat{\alpha}(y) = \hat{\alpha}(x)$$

So, (2.2) holds for n. By induction it holds for all n. So, $\alpha(x) = \lim \alpha_n(x) \le \hat{\alpha}(x)$. □

Corollary 2.5. *Given that X_n is irreducible and aperiodic, X_n is transient iff there is a solution $\hat{\alpha}(x)$ of equations (1)-(4).*

Proof. The real probability $\alpha(x)$ is $\le \hat{\alpha}(x)$. So, $0 \le \alpha(x) \le \hat{\alpha}(x) \to 0$ implies that $\alpha(x) \to 0$. So, X_n is transient. □

2.4. **example.** Take a random walk on $S = \{0, 1, 2, 3, \cdots\}$ with partially reflecting wall at 0. So, the probability of going left (or standing still at 0) is $p > 0$ and the probability of going right is $q = 1 - p > 0$:

$$p(n, n+1) = q, \quad p(n, n-1) = p, \quad p(0,0) = p$$

Let $z = 0$. We want to find the smallest solution $\alpha(n)$ of (1),(2),(3). But we already know how to do this. Equation (3) says:

$$\alpha(n) = p\alpha(n-1) + q\alpha(n+1)$$

The solution is $\alpha(x) = r^n$ where

$$r = p + qr^2$$

So,

$$r = \frac{1 \pm \sqrt{1-4pq}}{2q} = \frac{1 - \sqrt{1-4pq}}{2q}$$

(We want the smaller solution.)

2.4.1. $p = 1/2$. In this case $r = 1$. So, $\alpha(n) = r^n = 1$ and the Markov chain is recurrent. In fact we will see that it is null recurrent.

2.4.2. $p < 1/2$. In this case we can write

$$p = \frac{1}{2} - \epsilon, \quad pq = \frac{1}{4} - \epsilon^2$$
$$q = \frac{1}{2} + \epsilon, \quad 4pq = 1 - 4\epsilon^2$$
$$2q = 1 + 2\epsilon, \quad 1 - 4pq = 4\epsilon^2$$
$$r = \frac{1 - 2\epsilon}{1 + 2\epsilon} < 1$$

So, $\alpha(n) = r^n \to 0$ and X_n is transient.

2.4.3. $p > 1/2$. This is the part I skipped in class.

$$p = \frac{1}{2} + \epsilon, \quad pq = \frac{1}{4} - \epsilon^2$$
$$q = \frac{1}{2} - \epsilon, \quad 4pq = 1 - 4\epsilon^2$$
$$2q = 1 - 2\epsilon, \quad 1 - 4pq = 4\epsilon^2$$
$$r = \frac{1 - 2\epsilon}{1 - 2\epsilon} = 1$$

So, $\alpha(n) = r^n = 1$ for all n and X_n is recurrent.

2.5. **Null recurrence-positive recurrence.**

Definition 2.6. *An irreducible aperiodic Markov chain is called* null recurrent *if it is recurrent but*

$$\lim_{n\to\infty} p_n(x, y) = 0$$

for all states x, y. It is called positive recurrent *if it is recurrent but not null recurrent.*

Theorem 2.7. *If a Markov chain is positive recurrent then*

$$\lim_{n\to\infty} p_n(x, y) = \pi(y) > 0$$

is an invariant probability distribution. Also,

$$\mathbb{E}(T) = \frac{1}{\pi(y)}$$

where T is the first return time to y:

$$T = \text{ smallest } n > 0 \text{ so that } X_n = y \text{ given that } X_0 = y$$

Remember that an invariant distribution is a left eigenvector of P:

$$\pi P = \pi$$

with $\sum \pi(x) = 1$.

Corollary 2.8. *There is an invariant probability distribution π if and only if the Markov chain is positive recurrent.*

Proof. Since π is invariant,

$$\sum_x \pi(x)p_n(x,y) = \pi(y)$$

But $\pi(y)$ is positive and constant (does not change as we increase n). Therefore, the probabilities $p_n(x,y)$ cannot all go to zero and X_n cannot be null recurrent. □

2.6. **example, continued.** Going back to the random walk on $S = \{0,1,2,3,\cdots\}$ with partially reflecting wall at 0, the definition of invariant distribution says:

$$\pi(y) = \sum_x \pi(x)p(x,y)$$

In the random walk this is:

$$\pi(n) = q\pi(n-1) + p\pi(n+1) \tag{2.3}$$

which has solution

$$\pi(n) = \frac{r^n}{1-r}$$

(We have to divide by $1 - r = 1 + r + r^2 + \cdots$ so that $\sum \pi(n) = 1$.)

$$r = \frac{1-\sqrt{1-4pq}}{2p}$$

If $p = \frac{1}{2} + \epsilon$ then $q = \frac{1}{2} - \epsilon$ and

$$r = \frac{1-2\epsilon}{1+2\epsilon}, \quad 1 - r = \frac{4\epsilon}{1+2\epsilon}$$

So, we get an invariant distribution:

$$\pi(n) = \frac{r^n}{1-r} = \frac{1-2\epsilon}{4\epsilon}$$

Therefore, the chain is positive recurrent if $p > 1/2$.

If $p = 1/2$ then $r = 1$ and $\pi(n) = 1$ or $\pi(n) = n$ and neither solution can be normalized (scaled so that the sum is 1). Therefore, X_n is null recurrent if $p = 1/2$.

2.7. **Chart.** An irreducible aperiodic Markov chain has three possible types of behavior:

	$\lim_{n\to\infty} p_n(x,y)$	$\mathbb{E}(T)$	$\mathbb{P}(T<\infty)$	$\alpha(x)$
transient	$=0$	$=\infty$	<1	$\inf \alpha(x)=0$
0 *recurrent*	$=0$	$=\infty$	$=1$	$\alpha(x)=1\ \forall x$
+ *recurrent*	$=\pi(y)>0$	$=\frac{1}{\pi(y)}<\infty$	$=1$	$\alpha(x)=1\ \forall x$

In the transient case, $\mathbb{P}(T<\infty)<1$ is the same as $\mathbb{P}(T=\infty)>0$. This implies transient because it says that there is a chance that you never return. If a guy keeps leaving and there is a chance that he doesn't return each time then eventually, with probability one, he will not return. So, the number of visits if finite. If $\mathbb{P}(T=\infty)=0$ then he always returns and so he will keep coming back an infinite number of times. Since $\mathbb{P}(T=\infty)$ is either 0 or positive, *this proves that we either have transience or recurrence!!*

MATH 56A: STOCHASTIC PROCESSES
CHAPTER 3

Plan for rest of semester

(1) st week (8/31, 9/6, 9/7) Chap 0: Diff eq's and linear recursion
(2) nd week (9/11...) Chap 1: Finite Markov chains
(3) rd week (9/18...) Chap 1: Finite Markov chains
(4) th week (9/25...) Chap 2: Countable Markov chains
(5) th week (oct 3,4,5) Chap 3: Continuous time Markov chains
(6) th week (oct 9,11,12) Ch 4: Stopping time
(7) th week (oct 16,18,19) Ch 5: Martingales
(8) th week (oct 23,25,26) Ch 6: Renewal processes
(9) th week (oct 30,1,2) Ch 7: Reversible Markov chains
(10) th week (nov 6,8,9) Ch 8: Weiner process
(11) th week (nov 13,15,16) Ch 8: more
(12) th week (nov 20,22) (short week) Ch 9: Stochastic integrals
(13) th week (nov 27,29,30,4) (extra day) Ch 9: more

3. Continuous Markov Chains

The idea of continuous Markov chains is to make time continuous instead of discrete. This idea only works when the system is not jumping back and forth at each step but rather moves gradually in a certain direction.

3.1. **making time continuous.** On the first day I discussed the problem of converting to continuous time. In the discrete Markov chain we have the transition matrix P with entries $p(i,j)$ giving the probability of going from i to j in one unit of time. The n-th power, say P^5 has entries

$$p_5(i,j) = \mathbb{P}(X_5 = j \mid X_0 = i)$$

We want to interpolate and figure out what happened for all positive time t. (Negative time is discussed in Chapter 7.) We already know how to do that. You write:

$$P = QDQ^{-1}$$

Date: October 9, 2006.

where D is a diagonal matrix whose diagonal entries are the eigenvalues of P and Q is the matrix of right eigenvectors of P. The first eigenvector of P is 1 and the first right eigenvector is the column vector having all 1's.

If the eigenvalues are all positive then we can raise them to arbitrary values:

$$P^t = QD^tQ^{-1}$$

Usually you take logarithms. For example, if there are 3 states:

$$D = \begin{pmatrix} 1 & 0 & 0 \\ 0 & d_2 & 0 \\ 0 & 0 & d_3 \end{pmatrix} = \begin{pmatrix} e^0 & 0 & 0 \\ 0 & e^{\ln d_2} & 0 \\ 0 & 0 & e^{\ln d_3} \end{pmatrix}$$

Then $P^t = e^{tA}$ where

$$A = Q(\ln D)Q^{-1} = Q \begin{pmatrix} 0 & 0 & 0 \\ 0 & \ln d_2 & 0 \\ 0 & 0 & \ln d_3 \end{pmatrix} Q^{-1}$$

This uses:

Theorem 3.1. $P = QDQ^{-1} = Qe^{\ln D}Q^{-1} = e^{Q \ln DQ^{-1}}$

Proof. Let $L = \ln D$ then

$$D = e^L := I + L + \frac{L^2}{2} + \frac{L^3}{3!} + \cdots$$

Conjugate by Q:

$$Qe^LQ^{-1} = QQ^{-1} + QLQ^{-1} + \frac{QL^2Q^{-1}}{2} + \frac{QL^3Q^{-1}}{3!} + \cdots$$

This is equal to $e^{QLQ^{-1}}$ since $QL^nQ^{-1} = (QLQ^{-1})^n$. □

The other theorem I pointed out was:

Theorem 3.2.

$$\frac{d}{dt}P^t = P^tA = AP^t$$

Proof. This is just term by term differentiation.

$$\frac{d}{dt}P^t = \sum \frac{d}{dt}\frac{Qt^nL^nQ^{-1}}{n!} = \sum \frac{Qnt^{n-1}L^nQ^{-1}}{n(n-1)!}$$

$$= QLQ^{-1} \sum \frac{Qt^nL^nQ^{-1}}{n!} = AP^t$$

□

3.2. **Poisson processes.** On the second day I explained continuous Markov chains as generalizations of Poisson processes.

A *Poisson process* is

- an event which occurs from time to time
- is time homogeneous (i.e., the probability that it will occur tomorrow is the same as the probability that it will occur today)
- and the occurrences are independent

The independence of occurrences of a Poisson event means that the probability of future occurrence is independent of both past and present. Markov processes are independent of the past. They depend only on the present. We will transform a Poisson processes so that it looks more like a Markov process.

Here is an example where a Poisson event occurs three times in a time interval $\Delta t = t_1 - t_0$. (We put $t_0 = 0$ in class so that $\Delta t = t_1$.)

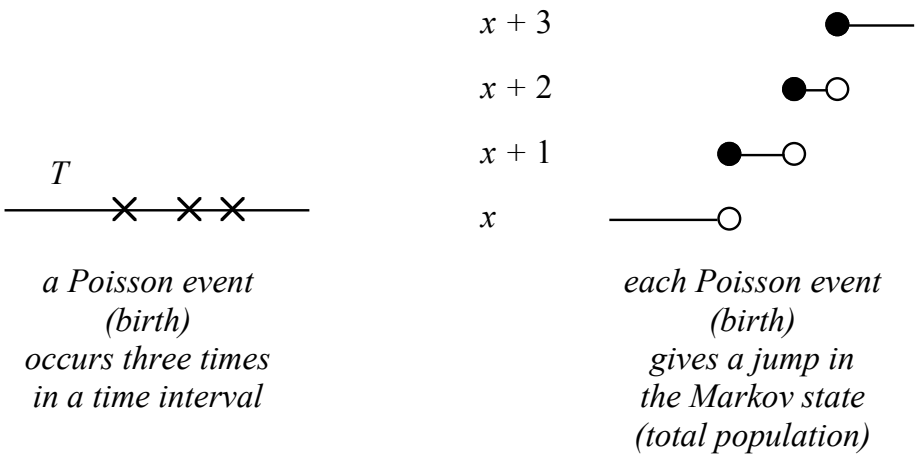

FIGURE 1. Poisson to Markov

The Poisson process has one parameter λ called the *rate*. This is measured in inverse time units (number of occurrences per unit time). Thus $\lambda\Delta t$ is the expected number of occurrences in any time interval of length Δt.

3.2.1. *variables associated to a Poisson process.* There are two random variables associated with a Poisson process:

Poisson variable (nonnegative integer)	Exponential variable (positive real)
$X =$ number of occurrences in Δt $\mathbb{P}(X = k) = e^{\lambda t_1}\dfrac{\lambda^k t_1^k}{k!}$	$T =$ time until 1st occurrence
$\mathbb{P}$(event does not occur) = $\mathbb{P}(X = 0) = e^{\lambda t_1} = 1 - \lambda t_1 + \dfrac{\lambda^2 t_1^2}{2} - \cdots$	$\mathbb{P}$(event does not occur) = $\mathbb{P}(T > t_1) = e^{\lambda t_1}$
$\mathbb{P}$(event occurs) = $\mathbb{P}(X \geq 1) = 1 - e^{\lambda \Delta t} = \lambda \Delta t + o(\Delta t)$	$\mathbb{P}$(event occurs in time Δt) = $\mathbb{P}(T \leq \Delta t) = 1 - e^{\lambda \Delta t} \approx \lambda \Delta t$

Here the book uses the "little oh" notation $o(\Delta t)$ to denote anything which vanishes faster than Δt:

$$\frac{o(\Delta t)}{\Delta t} \to 0 \quad \text{as} \quad \Delta t \to 0$$

3.2.2. *Poisson to Markov.* There are two changes we have to make to transform a Poisson process into a continuous time Markov process.

a) Every time an event occurs, you need to move to a new state in the Markov process. Figure 1 shows an example where the state is the total population:

$$X_t := \text{ population at time } t = X_0 + \#\text{births} - \#\text{deaths}$$

b) The rate $\alpha(x)$ depends on the state x. For example, the rate at which population grows is proportional to the size of the population:

$$\alpha(x) = \lambda x, \quad \lambda : \text{constant}$$

Notice that, when the rate increases, the events will occur more frequently and the waiting time will decrease. So, there is the possibility of *explosion*, i.e., an infinite number of jumps can occur in a finite amount of time.

3.3. **Definition of continuous Markov chain.** This is from Lawler, p. 69.

We need to start with an *infinitesmal generator* A which is a matrix with entries $\alpha(x, y)$ for all states $x, y \in S$ so that $\alpha(x, y) \geq 0$ for $x \neq y$ and $\alpha(x, x) \leq 0$ and so that the sum of the rows is zero:

$$\sum_{y \in S} \alpha(x, y) = 0$$

We use the notation

$$\alpha(x) = -\alpha(x,x) = \sum_{y \neq x} \alpha(x,y)$$

Definition 3.3. *A continuous time Markov chain with infinitesmal generator $A = (\alpha(x,y))$ is a function $X : [0,\infty) \to S$ so that*

1. *X is right continuous. I.e., X_t is equal to the limit of $X_{t+\Delta t}$ as Δt goes to zero from the right (the positive side).*
2. *$\mathbb{P}(X_{t+\Delta t} = x \,|\, X_t = x) = 1 - \alpha(x)\Delta t + o(\Delta t)$*
3. *$\mathbb{P}(X_{t+\Delta t} = y \,|\, X_t = x) = \alpha(x,y)\Delta t + o(\Delta t)$ for $y \neq x$.*
4. *X_t is time homogeneous*
5. *X_t is Markovian ($X_{\Delta t}$ depends on X_t but is independent of the state before time t.)*

I pointed out that the numbers $\alpha(x), \alpha(x,y)$ were necessarily ≥ 0 and that

$$\alpha(x) = \sum_{y \neq x} \alpha(x,y)$$

since $X_{t+\Delta t}$ must be in some state. The (x,x) entry of the matrix A is $\alpha(x,x) = -\alpha(x)$. So, the rows of the matrix A add up to zero and all negative numbers lie on the diagonal.

3.4. **probability distribution vector.** At any time t we have a *probability distribution vector* telling what is the probability of being in each state.

$$p_x(t) := \mathbb{P}(X_t = x)$$

This should not be confused with the time dependent probability transition matrix:

$$p_t(x,y) := \mathbb{P}(X_t = y \,|\, X_0 = x)$$

Theorem 3.4. *The time derivative of the probability distribution function $p_x(t)$ is given by*

$$\frac{d}{dt} p_x(t) = \sum_{y \in S} p_y(t)\alpha(y,x)$$

In matrix notation this is

$$\frac{d}{dt} p(t) = p(t)A$$

The unique solution of this differential equation is:

$$p(t) = p(0)e^{tA}$$

This implies that $P_t := e^{tA}$ is the time t probability transition matrix.

Proof. The difference $p_x(t+\Delta t)-p_x(t)$ is equal to the probability that the state moves into x minus the probability that it will move out of x in the time period from t to $t+\Delta t$. So,

$$p_x(t+\Delta t)-p_x(t) = \mathbb{P}(X_{t+\Delta t}=x, X_t=y\neq x)-\mathbb{P}(X_{t+\Delta t}=y\neq x, X_t=x)$$

$$= \sum_{y\neq x}\mathbb{P}(X_t=y)\mathbb{P}(X_{t+\Delta t}=x\,|\,X_t=y)-\sum_{y\neq x}\mathbb{P}(X_t=x)\mathbb{P}(X_{t+\Delta t}=y\,|\,X_t=x)$$

$$\approx \sum_{y\neq x} p_y(t)\alpha(y,x)\Delta t - \sum_{y\neq x} p_x(t)\alpha(x,y)\Delta t$$

$$= \sum_{y\neq x} p_y(t)\alpha(y,x)\Delta t - p_x(t)\alpha(x)\Delta t$$

$$= \sum_{y} p_y(t)\alpha(y,x)\Delta t$$

So,

$$\frac{p_x(t+\Delta t)-p_x(t)}{\Delta t} \approx \sum_y p_y(t)\alpha(y,x)$$

Take the limit as $\Delta t\to 0$ to get the theorem. □

3.5. **example.** What is the probability that $X_4=1$ given that $X_0=0$ if the infinitesmal generator is

$$A=\begin{pmatrix}-1 & 1\\ 2 & -2\end{pmatrix}?$$

The answer is given by the $(0,1)$ entry of the matrix e^{4A}. The given information is that $p(0)=(1,0)$ and the question is: What is $p_1(4)$? The solution in matrix terms is the second coordinate of

$$p(4)=(p_0(4),p_1(4))=(1,0)e^{4A}$$

We worked out the example: Since the trace of A is $-1+-2=-3=d_1+d_2$ and $d_1=0$ we must have $d_2=-3$. So, $A=QDQ^{-1}$ where

$$D=\begin{pmatrix}0 & 0\\ 0 & -3\end{pmatrix}$$

and Q is the matrix of right eigenvectors of A:

$$Q=\begin{pmatrix}1 & 1\\ 1 & -2\end{pmatrix},\quad Q^{-1}=\begin{pmatrix}2/3 & 1/3\\ 1/3 & -1/3\end{pmatrix}$$

The time 4 transition matrix is

$$e^{4A}=Qe^{4D}Q^{-1}=\begin{pmatrix}1 & 1\\ 1 & -2\end{pmatrix}\begin{pmatrix}1 & 0\\ 0 & e^{-12}\end{pmatrix}\begin{pmatrix}2/3 & 1/3\\ 1/3 & -1/3\end{pmatrix}$$

$$= P_4 = \begin{pmatrix} \frac{2+e^{-12}}{3} & \frac{1-e^{-12}}{3} \\ \frac{2-2e^{-12}}{3} & \frac{1+2e^{-12}}{3} \end{pmatrix}$$

So, the answer is

$$p_1(4) = \frac{1-e^{-12}}{3}$$

3.6. **equilibrium distribution, positive recurrence.** An equilibrium distribution does not change with time. In other words, the time derivative is zero:

$$\frac{d}{dt}\pi(t) = \pi(t)A = 0$$

So, $\pi(t) = \pi(0)$ is the left eigenvector of A normalized by:

$$\sum_{x \in S} \pi(x) = 1$$

Since π does not change with time, we forget the t and write $\pi(x)$ for $\pi_x(t)$. Recall that irreducible Markov chains are *positive recurrent* if and only if there is an equilibrium distribution.

The example above is irreducible and finite, therefore positive recurrent. The equilibrium distribution is $\pi = (2/3, 1/3)$.

3.7. **birth-death chain.**

Definition 3.5. *A* birth-death chain *is a continuous Markov chain with state space $S = \{0, 1, 2, 3, \cdots\}$ (representing population size) and transition rates:*

$$\alpha(n, n+1) = \lambda_n, \quad \alpha(n, n-1) = \mu_n, \quad \alpha(n) = \lambda_n + \mu_n$$

representing births and deaths which occur one at a time.

Notice that the total flow between the set of states $\{0, 1, 2, \cdots, n\}$ to the states $\{n+1, n+2, \cdots\}$ is given by the birth rate λ_n and the death rate μ_{n+1}. So, $\pi(n)$ is an equilibrium if and only if

$$\pi(n)\lambda_n = \pi(n+1)\mu_{n+1}$$

Solving for $\pi(n+1)$ gives:

$$\pi(n+1) = \frac{\lambda_n}{\mu_{n+1}}\pi(n) = \frac{\lambda_n\lambda_{n-1}}{\mu_{n+1}\mu_n}\pi(n-1) = \cdots = \frac{\lambda_n\lambda_{n-1}\cdots\lambda_0}{\mu_{n+1}\mu_n\cdots\mu_1}\pi(0)$$

If we can normalize these numbers we get an equilibrium distribution. So,

Theorem 3.6. *The birth-death chain is positive recurrent if and only if*

$$\sum_{n=0}^{\infty} \frac{\lambda_n \lambda_{n-1} \cdots \lambda_0}{\mu_{n+1}\mu_n \cdots \mu_1} < \infty$$

3.8. **birth and explosion.** If there is no death, the birth-death chain is obviously transient. The population is going to infinity but how fast? Suppose that T_n is the time that the population stays in state n. Then (when $\mu_n = 0$) T_n is exponential with rate λ_n. So,

$$\mathbb{E}(T_n) = \frac{1}{\lambda_n}$$

Theorem 3.7. *a) If* $\sum_{n=0}^{\infty} \frac{1}{\lambda_n} < \infty$ *then explosion occurs with probability one.*

b) If $\sum_{n=0}^{\infty} \frac{1}{\lambda_n} = \infty$ *then the probability of explosion is zero.*

For example, in the Yule process with $\lambda_n = n\lambda$, explosion will not occur since

$$\sum \frac{1}{\lambda_n} = \sum \frac{1}{\lambda n} = \frac{1}{\lambda} \sum \frac{1}{n} = \infty$$

3.9. **transient birth-death chains.** Recall that an irreducible Markov chain is transient if and only if there is a right eigenvector of P with entries converging to zero corresponding to eigenvector 1. In the continuous time case, this is the same as a right eigenvector of A corresponding to eigenvalue 0. So, we want numbers $a(n)$ such that

$$a(n-1)\mu_n + a(n)(-\lambda_n - \mu_n) + a(n+1)\lambda_n = 0$$

This equation can be rewritten as

$$[a(n+1) - a(n)]\lambda_n = [a(n) - a(n-1)]\mu_n$$

$$[a(n+1) - a(n)] = \frac{\mu_n}{\lambda_n}[a(n) - a(n-1)] = \frac{\mu_n \mu_{n-1} \cdots \mu_1}{\lambda_n \lambda_{n-1} \cdots \lambda_1}[a(1) - a(0)]$$

$a(k+1)$ is the sum of these numbers:

$$a(k+1) = a(0) + \sum_{n=0}^{k} [a(n+1) - a(n)] = a(0) + \sum_{n=0}^{k} \frac{\mu_n \mu_{n-1} \cdots \mu_1}{\lambda_n \lambda_{n-1} \cdots \lambda_1}[a(1) - a(0)]$$

Theorem 3.8. *A birth-death chain is transient if and only if*

$$\sum_{n=0}^{k} \frac{\mu_n \mu_{n-1} \cdots \mu_1}{\lambda_n \lambda_{n-1} \cdots \lambda_1} < \infty$$

Proof. Let L be this limit. Let $a(0) = 1$ and $a(1) = 1 - 1/L$. Then $a(k+1)$ given by the above formula will converge to zero. Conversely, if the $a(k+1)$ goes to zero, the infinite sum must converge. □

MATH 56A: STOCHASTIC PROCESSES
CHAPTER 4

4. Optimal stopping time

On the first day I explained the basic problem using the example in the book. On the second day I explained how the solution to the problem is given by a "minimal superharmonic" and how you could find one using an iteration algorithm. Also, a simple geometric construction gives the solution for fair random walks. On the third day I explained the variations of the game in which there is a fixed cost per move or if the payoff is discounted. I also explained the transition to continuous time.

4.1. **The basic problem.** The problem is to find a "stopping time" which optimizes the expected value of a payoff function. I think I gave the same example as in the book: You roll a die. If you get a 6 you lose and get nothing. But if you get any other number you get the value on the die (1,2,3,4 or 5 dollars). If the value is too low you can roll over. The question is: When should you stop? The answer needs to be a strategy: "Stop when you get 4 or 5." or maybe "Stop when you get 3,4 or 5." You want the best "stopping time."

4.1.1. *stopping time.*

Definition 4.1. *In a stochastic process a* stopping time *is a time T which has the property that you can tell when it arrives. I.e., whether or not T is the stopping time is determined by the information that you have at time T.*

Basically, a stopping time is a formula which, given $X_1, X_2, \cdots, X_n$ tells you whether to stop at step n. (Or in continuous time, given X_t for $t \leq T$, tells you whether T is the stopping time.)

Some examples of stopping time are:

(1) the 5th visit to state x
(2) the smallest time T at which $X_1 + X_2 + \cdots + X_T > 100$.
(3) the first visit to the set $\{3, 4, 5\}$.

Date: October 22, 2006.

If T is the first visit to state x then $T-1$ is not a stopping time. (You cannot say "stop right before you get to x." since the process is stochastic and you can't tell the future.)

4.1.2. *payoff function.* The *payoff function* assigns to each state $x \in S$ a number $f(x) \in \mathbb{R}$ which can be positive or negative. This represents what you gain (or lose) if you stop at state x. To figure out whether to stop you need to look at what you can expect to happen if you don't stop.

(1) If you stop you get $f(x)$.
(2) If, starting at x, you take one step and then stop you get

$$\sum p(x,y)f(y)$$

We assume that there is only one transient communication class and $f(x) = 0$ on all recurrent classes.

4.1.3. *value function.* The *value function* $v(x)$ is the expected payoff using the optimal strategy starting at state x.

$$v(x) = \mathbb{E}(f(X_T) \,|\, X_0 = x)$$

Here T is the optimal stopping time. If you don't know what T is then you need another equation:

$$v(x) = \max_T \mathbb{E}(f(X_T) \,|\, X_0 = x)$$

This says you take all possible stopping times T and take the one which gives the maximal expected payoff.

Theorem 4.2. *The value function $v(x)$ satisfies the equation*

$$v(x) = \max(f(x), \sum_y p(x,y)f(y))$$

The basic problem is to find the optimal stopping time T and calculate the value function $v(x)$.

4.2. **Solutions to basic problem.** On the second day I talked about solutions to the optimal stopping time problem. I started with an outline:

(1) Minimal superharmonic is optimal.
(2) Iteration algorithm converges to minimal solution.
(3) Random walks have concave solutions.
(4) Solution for continuous time.

I explained the solutions for discrete time, then converted these into solutions for continuous time.

4.2.1. *minimal superharmonic.*

Definition 4.3. *A* superharmonic *for the Markov chain X_n is a real valued function $u(x)$ for $x \in S$ so that*

$$u(x) \geq \sum_{y \in S} p(x,y)u(y)$$

In matrix form the definition is

$$u(x) \geq (Pu)(x)$$

where u is a column vector.

Example 4.4. *Roll one die and keep doing it until you get a 6. (6 is an absorbing state.) The payoff function is:*

states x	*payoff $f(x)$*	*probability $\mathbb{P}$*
1	150	1/6
2	150	1/6
3	150	1/6
4	300	1/6
5	300	1/6
6	0	1/6

The transition matrix in this example is actually 6×6. But I simplified this to 3 states: $A = 1, 2$ or 3, $B = 4$ or 5 and $C = 6$:

states x	payoff $f(x)$	probability $\mathbb{P}$
A	150	1/2
B	300	1/3
C	0	1/6

Then P is a 3×3 matrix:

$$P = \begin{pmatrix} 1/2 & 1/3 & 1/6 \\ 1/2 & 1/3 & 1/6 \\ 0 & 0 & 1 \end{pmatrix}$$

The best payoff function you can hope for is $u =$ the column matrix $(300, 300, 0)$. Then

$$Pu = \begin{pmatrix} 1/2 & 1/3 & 1/6 \\ 1/2 & 1/3 & 1/6 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 300 \\ 300 \\ 0 \end{pmatrix} = \begin{pmatrix} 250 \\ 250 \\ 0 \end{pmatrix}$$

The equation $u(x) \geq (Pu)(x)$ means the x-entry of the matrix x is $\geq$ the x-entry of the matrix Pu. So, $300 \geq 250$ makes $u = (300, 300, 0)$ superharmonic.

Theorem 4.5. *The value function $v(x)$ is the minimal superharmonic so that $v(x) \geq f(x)$ for all states x.*

This gives a theoretical solution which is useful in some cases (such as the simple random walk).

4.2.2. *iteration algorithm.* As I explained it, $u(x)$ is your estimated expected payoff. The algorithm works like this. You start with u_1 which is the most optimistic. This the payoff you get if you cheat on the next roll.

$$u_1(x) = \begin{cases} 0 & \text{if } x \text{ is absorbing} \\ \max f(y) & \text{if } x \text{ is transient} \end{cases}$$

Next, u_2 is your expected payoff if you play fair for one round and then cheat. u_n is your payoff if you wait n turns before cheating. The recursive formula for u_{n+1} given u_n is

$$u_{n+1}(x) = \max(f(x), (Pu_n)(x))$$

At each stage u_n is superharmonic and $u_n(x) \geq f(x)$ but the values get smaller and smaller and become minimal in the limit:

$$v(x) = \lim_{n \to \infty} u_n(x)$$

$v(x)$ is your expected payoff if you put off cheating indefinitely.

In the example,

$$\begin{aligned} u_1 &= (300, \quad 300, \quad 0) \\ u_2 &= (250, \quad 300, \quad 0) \\ u_3 &= (225, \quad 300, \quad 0) \\ u_4 &= (212.5, \quad 300, \quad 0) \\ v = u_\infty &= (x, \quad 300, \quad 0) \end{aligned}$$

where $x = 200$ is the solution of the equation:

$$x = \frac{x}{2} + \frac{300}{3}$$

4.2.3. *convex value function.* Suppose you have a simple random walk with absorbing walls. Then a function $u(x)$ is superharmonic if

$$u(x) \geq \frac{u(x-1) + u(x+1)}{2}$$

In other words, the point $(x, u(x))$ is above the point which is midway between $(x-1, u(x-1))$ and $(x+1, u(x+1))$. So, superharmonic is

the same as convex (concave down). So, the theorem that the value function $v(x)$ is the minimal superharmonic so that $v(x) \geq f(x)$ means that the graph of $v(x)$ is the convex hull of the graph of $f(x)$.

4.2.4. *continuous time.* In a continuous Markov chain you have an infinitesmal generator A which is a matrix with transition rates $\alpha(x, y)$ which are all nonnegative except for $\alpha(x, x) = -\alpha(x)$. Since the rows add to zero we have

$$\alpha(x) = \sum_{y \neq x} \alpha(x, y)$$

So, you get a probability matrix P with entries

$$p(x, y) := \frac{\alpha(x, y)}{\alpha(x)}$$

for $x \neq y$ (and $p(x, x) = 0$). This is the probability of first jumping to y from x:

$$p(x, y) = \mathbb{P}(X_J = y \,|\, X_0 = x)$$

where J is the first jump time:

$$J = J_1 := \inf\{t \,|\, X_t \neq X_0\}$$

Anyway, you use the discrete probability transition matrix P and transform it into continuous time by looking only at jump times: The optimal stopping time for the continuous process is J_T, the T-th jump time where T is the stopping time for the discrete process.

4.3. **Cost functions.** The *cost function* $g(x)$ gives the price you must pay to continue from state x. If T is your stopping time then you continued T times. So your total cost is

$$g(X_0) + g(X_1) + \cdots + g(X_{T-1}) = \sum_{j=0}^{T-1} g(X_j)$$

So, your net gain is

$$f(X_T) - \sum_{j=0}^{T-1} g(X_j)$$

The value function $v(x)$ is the expected net gain when using the optimal stopping time starting at state x. It satisfies the equation:

$$v(x) = \max(f(x), (Pv)(x) - g(x))$$

4.3.1. *iteration algorithm.* First, you take $u_1(x)$ to be your most optimistic estimate of expected gain. If you go one step in the Markov chain then you have to pay $g(x)$ so your best possible net gain would be

$$\max_{y \in S} f(y) - g(x)$$

If this is less than $f(x)$ you can't possibly do better by continuing. So

$$u_1(x) = \begin{cases} 0 & \text{if } x \text{ is absorbing} \\ f(x) & \text{if } f(x) \geq \max_{y \in S} f(y) - g(x) \\ \max_{y \in S} f(y) - g(x) & \text{otherwise} \end{cases}$$

u_{n+1} is given in terms of u_n by:

$$u_{n+1}(x) = \max(f(x), (Pu_n)(x) - g(x))$$

where $(Pu_n)(x) = \sum p(x,y)u_n(y)$.

4.3.2. *random walk.* For the simple random walk with absorbing walls, the value function is the smallest function $v(x) \geq f(x)$ so that

$$v(x) \geq \frac{v(x-1) + v(x+1)}{2} - g(x)$$

4.4. **Discounted payoff.** Here we assume that the payoff is losing value at a fixed rate so that after T steps it will only be worth $\alpha^T f(x)$ where α is the discount rate, say $\alpha = .90$. Then the value function satisfies the equation

$$v(x) = \max(f(x), \alpha(Pv)(x))$$

Again there is a recursive formula converging to this answer:

$$u_{n+1}(x) = \max(f(x), \alpha(Pu_n)(x))$$

where you start with

$$u_1(x) = \begin{cases} 0 & \text{if } x \text{ is absorbing} \\ f(x) & \text{if } f(x) \geq \alpha f(y) \text{ for all } y \\ \max \alpha f(y) & \text{otherwise} \end{cases}$$

MATH 56A: STOCHASTIC PROCESSES
CHAPTER 5

5. Martingales

On the first day I gave the intuitive definition of "information," conditional expectation and martingale using the fair value of your place in a game. On the second day I gave you the mathematical definition of "information." On the third day I explained the mathematical definition of conditional expected value. We also discussed the definition of "integrability" and "uniform integrability" and the two theorems: Optimal Sampling Theorem and the Martingale Convergence Theorem.

5.1. **Intuitive description of martingale.** In the previous chapter we talked about optimal stopping time in a game in which the worst thing that could happen is you don't get anything. This time we are talking about a martingale: You have the opportunity to buy a "share" in a random game that someone else is playing. The game may or may not be fair. The question is: How much should you pay? This question becomes easier if you assume that you can sell your share after one round of play. So the formula or strategy should tell you how much your share of the game will be tomorrow. If we don't have any transaction fees or discount rate then the fair price you should pay today should be exactly equal to the price that you expect to sell it for tomorrow given the information that you have today.

5.1.1. *information.* We have a stochastic process X_n in discrete time n. X_n is not necessarily Markovian.

$\mathcal{F}_n$ represents all the information that you have about X_n for time $\leq n$. This is basically just $X_0, X_1, \cdots, X_n$. Suppose that we have a function

$$Y_n = f(X_0, X_1, \cdots, X_n).$$

Then, given $\mathcal{F}_n$, Y_n is known. Given $\mathcal{F}_0$, Y_n is random but $\mathbb{E}(Y_n \,|\, \mathcal{F}_0)$ is known. As time progresses (gets closer to n), you usually have a better idea of what Y_n might be until finally,

$$\mathbb{E}(Y_n \,|\, \mathcal{F}_n) = Y_n$$

Date: November 12, 2006.

5.1.2. *example: Bernoulli.* Suppose that $X_1, X_2, \cdots$, are independent identically distributed (i.i.d.) with distribution

$$X_n = \begin{cases} 1 & \text{with probability } p \\ -1 & \text{with probability } 1-p \end{cases}$$

Let $Y_n = S_n$ be the sum:

$$Y_n = S_n = X_1 + X_2 + \cdots + X_n$$

The information at time 0 (before we flip the first coin) is $\mathcal{F}_0 : (S_0 = 0)$.

Suppose first that $p = 1/2$ Then S_n is simple random walk on $\mathbb{Z}$. The expected value of S_n changes with time. At the beginning we expect it to be zero: $\mathbb{E}(S_n \,|\, \mathcal{F}_0) = 0$. But later our estimate changes. For example,

$$\mathbb{E}(S_n \,|\, \mathcal{F}_{n-1}) = S_{n-1}$$

Why is that? Given $\mathcal{F}_{n-1}$ we know S_{n-1} but X_n is still random:

$$S_n = \underbrace{S_{n-1}}_{\text{known}} + \underbrace{X_n}_{\pm 1}$$

When $p = 1/2$ the expected value of X_n is zero: $\mathbb{E}(X_n) = 0$.

5.1.3. *expectation.* Before doing the case of general p I reviewed the definition of *expectation*:

$$\mathbb{E}(Y) := \sum_y y\mathbb{P}(Y = y) \quad \text{for discrete } Y$$

$$\mathbb{E}(Y) := \int_{-\infty}^{\infty} y f(y)\, dy \quad \text{for continuous } Y$$

So,

$$\begin{aligned} \mathbb{E}(X_n) &= 1 \cdot \mathbb{P}(X_n = 1) + (-1) \cdot \mathbb{P}(X_n = -1) \\ &= 1 \cdot p + (-1)(1-p) \\ &= p - 1 + p = 2p - 1 \end{aligned}$$

The expected value is what we use when we don't know X_n:

$$\mathbb{E}(X_n \,|\, \mathcal{F}_m) = \begin{cases} 2p-1 & \text{if } n > m \\ X_n & \text{if } n \le m \end{cases}$$

Recall that $\mathbb{E}$ is a *linear function.* So,

$$\mathbb{E}(S_n) = \mathbb{E}(X_1) + \mathbb{E}(X_2) + \cdots + \mathbb{E}(X_n) = n(2p-1)$$

$$\begin{aligned} \mathbb{E}(S_n \,|\, \mathcal{F}_{n-1}) &= \underbrace{\mathbb{E}(X_1 \,|\, \mathcal{F}_{n-1}) + \cdots + \mathbb{E}(X_{n-1} \,|\, \mathcal{F}_{n-1})}_{\text{not random}} + \underbrace{\mathbb{E}(X_n \,|\, \mathcal{F}_{n-1})}_{\text{random}} \\ &= X_1 + X_2 + \cdots + X_{n-1} + 2p - 1 \end{aligned}$$

So,

$$\mathbb{E}(S_n \,|\, \mathcal{F}_{n-1}) = S_{n-1} + 2p - 1$$

In general,

$$\mathbb{E}(S_n \,|\, \mathcal{F}_m) = \begin{cases} S_m + (n-m)(2p-1) & \text{if } n > m \\ S_n & \text{if } n \leq m \end{cases}$$

If $p \neq 1/2$ the value of S_n is expected to change in the future. If S_n is the payoff function you want to play this game if $p > 1/2$ and you don't want to play if $p < 1/2$.

5.1.4. *the martingale.* Continuing with the same example, let

$$M_n = X_1 + \cdots + X_n - n(2p-1) = S_n - n(2p-1)$$

This is the random number S_n minus its expected value. Then

$$\begin{aligned} \mathbb{E}(M_n \,|\, \mathcal{F}_m) &= \mathbb{E}(S_n \,|\, \mathcal{F}_m) - n(2p-1) \\ &= \begin{cases} S_m - m(2p-1) = M_m & \text{if } n > m \\ S_n - n(2p-1) = M_n & \text{if } n \leq m \end{cases} \end{aligned}$$

Definition 5.1. *A sequence of random variables $M_0, M_1, \cdots$ with $\mathbb{E}(|M_i|) < \infty$ is a* martingale *with respect to $\{\mathcal{F}_n\}$ if*

$$\mathbb{E}(M_n \,|\, \mathcal{F}_m) = M_m$$

It follows by induction on n that this definition is equivalent to the condition:

$$\mathbb{E}(M_n \,|\, \mathcal{F}_{n-1}) = M_{n-1}$$

For example,

$$\mathbb{E}(M_2 \,|\, \mathcal{F}_0) = \mathbb{E}(\mathbb{E}(M_2 \,|\, \mathcal{F}_1) \,|\, \mathcal{F}_0) = \mathbb{E}(M_1 \,|\, \mathcal{F}_0) = M_0$$

(using the rule of iterated expectation)

5.2. **theory: conditional expectation with respect to information.** On the second and third days I tried to explain the mathematical definition of information as a σ-subalgebra of the σ-algebra of all events. I started with a review of basic probability.

5.2.1. *basic probability.*

Definition 5.2. *A* probability space $(\Omega, \mathcal{F}, \mathbb{P})$ *consists of*

- $\Omega =$ *the* sample space*,*
- $\mathcal{F} =$ *the σ-algebra of all measurable subsets of Ω, (elements of $\mathcal{F}$ are called* events*) and*
- $\mathbb{P} =$ *the* probability measure *which assigns a measure $\mathbb{P}(A)$ for every $A \in \mathcal{F}$.*

The only condition is: $\mathbb{P}(\Omega) = 1$. *Note that*

$$A \in \mathcal{F} \iff \mathbb{P}(A) \text{ is defined}$$

This definition assumes the definition of "measure." "measurable" and "σ-algebra."

Definition 5.3. *A* σ-algebra *on a set* Ω *is a collection* $\mathcal{F}$ *of subsets* A *(called* measurable subsets *of* Ω*) satisfying the following axioms:*

(1) $\mathcal{F}$ *is closed under countable union. I.e., if* $A_1, A_2, \cdots$ *are measurable (elements of* $\mathcal{F}$*) then*

$$\bigcup_{i=1}^{\infty} A_i \in \mathcal{F}$$

(2) *If* A *is measurable then so is its complement* $\Omega - A$. *(This implies that* $\mathcal{F}$ *is closed under countable intersection.)*

(3) $\emptyset, \Omega \in \mathcal{F}$.

A measure $\mathbb{P} : \mathcal{F} \to [0, \infty)$ *is a function which assigns to each* $A \in \mathcal{F}$ *a nonnegative real number s.t.* $\mathbb{P}$ *takes countable disjoint union to sum:*

$$\mathbb{P}\left(\coprod A_i\right) = \sum \mathbb{P}(A_i).$$

(Compare with the definition: A *topology* on Ω is a collection of subsets called *open subsets* which is closed under finite intersection and arbitrary union. The complement of an open set may not be open.)

Definition 5.4. *A function* $X : \Omega \to \mathbb{R}$ *is called* measurable *with respect to* $\mathcal{F}$ *if the inverse image of every measurable subset of* $\mathbb{R}$ *is measurable, i.e., an element of* $\mathcal{F}$. *(This is the same as saying that the inverse images of open, closed and half open intervals* (a, b), $[a, b]$, $(a, b], [a, b)$ *are measurable or, equivalently, the subset of* Ω *on which* $a < X \leq b$ *is measurable and therefore the measure*

$$\mathbb{P}(a < X \leq b)$$

is defined.) Measurable functions on Ω *are called* random variables.

(Compare with the definition: A function is *continuous* if the inverse image of every open set is open.)

5.2.2. *information.* is defined to be a σ-subalgebra of the σ-algebra $\mathcal{F}$ of all events $A \subseteq \Omega$. When the book says that $\mathcal{F}_n$ is the information given by $X_0, \cdots, X_n$ it means that $\mathcal{F}_n$ is the collection of all subsets of Ω which are given by specifying the values of $X_0, X_1, \cdots, X_n$.

A random variable Y' is *$\mathcal{F}_n$-measurable* if it can be written as a function of $X_0, X_1, \cdots, X_n$:

$$Y' = f(X_0, X_1, \cdots, X_n)$$

5.2.3. *filtration.* $\{\mathcal{F}_n\}$ is called a *filtration.* I drew the following diagrams to illustrate what that means in the case when X_1 takes 3 values and X_2 takes two values:

TABLE 1. The σ-subalgebra $\mathcal{F}_0$ has only the two required elements $\mathcal{F}_0 = \{\emptyset, \Omega\}$

X_0: known $X_0 =$ one value

Ω

TABLE 2. The σ-subalgebra $\mathcal{F}_1$ has $2^3 = 8$ elements given by the values of X_0, X_1

$$X_1 = 1, 2, 3$$

$$\mathcal{F}_1 = \{\emptyset, A, B, C, A \cup B, A \cup C, B \cup C, \Omega\}$$

A	B	C
$X_1 = 1$	$X_1 = 2$	$X_1 = 3$

Ω

TABLE 3. The σ-subalgebra $\mathcal{F}_2$ has $2^6 = 64$ elements given by the values of X_0, X_1, X_2

$$X_2 = 1, 2$$

$X_1 = 1$ $X_2 = 2$	$X_1 = 2$ $X_2 = 2$	$X_1 = 3$ $X_2 = 2$
$X_1 = 1$ $X_2 = 1$	$X_1 = 2$ $X_2 = 1$	$X_1 = 3$ $X_2 = 1$

Ω

The increasing sequence of σ-algebras

$$\mathcal{F}_0 \subseteq \mathcal{F}_1 \subseteq \mathcal{F}_2 \subseteq \cdots$$

is an example of a *filtration.*

Definition 5.5. *A* filtration *is an increasing sequence of σ-subalgebras of $\mathcal{F}$.*

$$Y' \text{ is } \mathcal{F}_0 - \text{measurable} \iff Y' \text{ is constant}$$
$$Y' \text{ is } \mathcal{F}_n - \text{measurable} \iff Y' = f(X_0, X_1, \cdots, X_n)$$

5.2.4. *conditional expectation.* The definition of martingale uses conditional expectation with respect to information. This is defined mathematically by:

$\mathbb{E}(Y \mid \mathcal{F}_n) :=$ Y': the $\mathcal{F}_n$-measurable function which best approximates Y

In the example above, $Y' = \mathbb{E}(Y \mid \mathcal{F}_2)$ is a random variable which takes only 6 values, one for each of the 6 blocks in the third figure above. For example, in the lower left corner we have

$$Y' = \mathbb{E}(Y \mid X_1 = 1, X_2 = 2)$$

Theorem 5.6 (rule of iterated expectation)**.** *If $\mathcal{F}_n$ is a filtration and $n > m$ then*

$$\mathbb{E}(\mathbb{E}(Y \mid \mathcal{F}_n) \mid \mathcal{F}_m) = \mathbb{E}(Y \mid \mathcal{F}_m)$$

Assuming that $\mathbb{E}(|Y| \mid \mathcal{F}_m) < \infty$.

Proof. I gave the proof in the case when $n = m + 1$ assuming that X_m is given and $X = X_n$ is random. Then we have to show

$$\mathbb{E}(\mathbb{E}(Y \mid X)) = \mathbb{E}(Y)$$

The RHS is given by

$$\mathbb{E}(Y) = \int_{-\infty}^{\infty} y f_Y(y)\, dy$$

Substituting the formula for $f_Y(y)$ in terms of the joint distribution $f(x, y)$:

$$f_Y(y) = \int_{-\infty}^{\infty} f(x, y)\, dx$$

gives

$$\mathbb{E}(Y) = \iint_{\mathbb{R}^2} y f(x, y)\, dxdy$$

On the LHS we have

$$\mathbb{E}(Y \mid X = x) = \frac{\int_{-\infty}^{\infty} yf(x,y)\,dy}{f_X(x)}$$

$\mathbb{E}(\mathbb{E}(Y \mid X))$ is the expected value of this function:

$$\begin{aligned}\mathbb{E}(\mathbb{E}(Y \mid X)) &= \int_{-\infty}^{\infty} \mathbb{E}(Y \mid X = x) f_X(x)\,dx \\ &= \int_{-\infty}^{\infty} \frac{\int_{-\infty}^{\infty} yf(x,y)\,dy}{f_X(x)} f_X(x)\,dydx \\ &= \iint_{\mathbb{R}^2} yf(x,y)\,dydx = \mathbb{E}(Y)\end{aligned}$$

assuming that $|y|f(x,y)$ has a finite integral. (This is Foubini's Theorem. You can reverse the order of integration only when the absolute value has a finite integral.) □

5.3. **Optimal sampling theorem.** The optimal sampling theorem says that, under certain conditions,

$$\mathbb{E}(M_T \mid \mathcal{F}_0) = M_0$$

where M_n is a martingale and T is a stopping time. We know that this is not always true because of the Monte Carlo gambling strategy:

5.3.1. *Monte Carlo stopping time.* This is the strategy where you stop when you win and double your bet if you lose. You can express it as a stopping time for a martingale as follows.

Suppose that $X_1, X_2, \cdots$ are independent Bernoulli variables where X_n takes values $\pm 2^{n-1}$ with equal probability. Then

$$M_n = S_n = X_1 + \cdots + X_n$$

is a martingale with $M_0 = 0$. It represents the game where you keep doubling your bet no matter what happens.

Now, let T be the first time that you win:

$$T = \inf\{n \geq 1 \mid X_n > 0\}$$

Since the simple random walk on $\mathbb{Z}$ is (null) recurrent, your probability is 1 that $T < \infty$. And when you stop, you will have $M_T = 1$. So,

$$\mathbb{E}(M_T \mid \mathcal{F}_0) = 1 \neq M_0 = 0$$

The optimal sampling theorem does not hold for Monte Carlo. So, we had better make sure that the statement excludes this case and all "similar" cases.

One way to avoid this counterexample is to put an upper bound (a time limit) on T.

Theorem 5.7 (1st OST)**.** *Suppose that M_n is a martingale. Then $\mathbb{E}(M_T) = M_0$ if T is a bounded stopping time (i.e., $T < C$).*

5.3.2. *integrability.* Now we have a bunch of theorems that assume integrability. A random variable Y is called *integrable* (or, more precisely, L^1) if $\mathbb{E}(|Y|) < \infty$. I don't remember if I got to this in class but it is in my notes:

Theorem 5.8. *Suppose that $\mathcal{F}_n$ is a filtration and Y_n is $\mathcal{F}_n$ measurable. Suppose*

(1) *T is a stopping time and*
(2) *$\mathbb{P}(T < \infty) = 1$*

Then $M_n := \mathbb{E}(Y_T \,|\, \mathcal{F}_n)$ is a martingale assuming that each M_n is integrable.

Proof. By definition we have:

$$\mathbb{E}(M_{n+1} \,|\, \mathcal{F}_n) = \mathbb{E}(\mathbb{E}(Y_T \,|\, \mathcal{F}_{n+1}) \,|\, \mathcal{F}_n)$$

By Theorem 5.6 this is

$$= \mathbb{E}(Y_T \,|\, \mathcal{F}_n) = M_n$$

□

5.3.3. *2nd OST and uniform integrability.* The second optimal sampling theorem requires "uniform integrability."

Theorem 5.9 (2nd OST)**.** *Suppose that $M_0, M_1, \cdots$ is a martingale with respect to the filtration $\mathcal{F}_n$ and T is a stopping time. Then*

$$\mathbb{E}(M_T \,|\, \mathcal{F}_0) = M_0$$

provided that

(1) *$\mathbb{P}(T < \infty) = 1$*
(2) *$\mathbb{E}(|M_T|) < \infty$ (M_T is integrable).*
(3) *$M_0, M_1, \cdots$ are uniformly integrable*

When you say that Y is integrable, you mean that the improper integral

$$\int_{-\infty}^{\infty} y f(y)\, dy = \lim_{K \to \infty} \int_{-K}^{K} y f(y)\, dy$$

converges.

Definition 5.10. *The functions $M_0, M_1, \cdots$ are* uniformly integrable *if the corresponding improper integrals for M_n converge uniformly. In other words, for every $\epsilon > 0$ there is a $K > 0$ so that*

$$\int_{-\infty}^{-K} |y| f_n(y)\, dy + \int_{K}^{\infty} |y| f_n(y)\, dy < \epsilon$$

for all $n \geq 0$ where f_n is the density function for M_n: The tails are getting smaller at the same rate.

In the book the sum of the two tails is written as a single integral:

$$\int_{-\infty}^{-K} |y| f_n(y)\, dy + \int_{K}^{\infty} |y| f_n(y)\, dy = \int_{-\infty}^{\infty} I_{|y|\geq K} |y| f(y)\, dy$$

where $I_{|y|\geq K}$ is the *indicator function* equal to 1 when the condition ($|y| \geq K$) is true and 0 when it is false.

5.3.4. *example: random walk.* The OST can be used in reverse. If $\mathbb{E}(M_T \,|\, \mathcal{F}_0) \neq M_0$ then it must be because one of the conditions does not hold. I gave an example using simple random walk on $\mathbb{Z}$. You take $X_0 = 0$ and let T be the first visit to 1. Then M_n is a martingale, but

$$M_T = 1 \neq M_0 = 0$$

So, the (conclusion of) OST does not hold. Let's check the conditions:

(1) $\mathbb{P}(T < \infty) = 1$. This holds because the Markov chain is recurrent.
(2) $M_T = 1$ is constant and therefore integrable. $\mathbb{E}(|M_T|) = 1 < \infty$.

The conclusion is that the third condition must fail: $M_0, M_1, \cdots$ are *not* uniformly integrable. "The tails remain fat."

5.3.5. *example: optimal stopping time.* Suppose that $X_n \in \{1, 2, 3, 4, 5, 6\}$ and T is the 1st visit to the set $\{1, 3, 4, 5, 6\}$, i.e., this is the optimal stopping time in the game that we analyzed in the last chapter when the payoff is equal to X_n when it is > 1 and zero if you every reach 1.

Let

$$M_n = v(X_n) = \mathbb{E}(f(X_T) \,|\, \mathcal{F}_n)$$

Then,

(1) M_n is a martingale and
(2) Optimal sampling holds. I.e., $\mathbb{E}(M_T \,|\, \mathcal{F}_0) = M_0$.

In your homework you computed the value function v using the iteration algorithm which assumes that $v(X_n)$ is a Martingale.

5.4. **Martingale convergence theorem.** The other question we dealt with is: When does a martingale converge?

Theorem 5.11 (Martingale convergence theorem)**.** *Suppose that $\{M_n\}$ is a martingale with respect to the filtration $\{\mathcal{F}_n\}$.*

(1) *If there exists $C < \infty$ so that $\mathbb{E}(|M_n|) < C$ for all n then*

$$M_n \to M_\infty$$

where M_∞ is some integrable random variable.

(2) *If M_n are uniformly integrable and $M_n \to M_\infty$ then*

$$\mathbb{E}(M_n) \to \mathbb{E}(M_\infty)$$

5.4.1. *example: log normal distribution.* Suppose that $X_1, X_2, \cdots$ are i.i.d. where each X_i can take only two values $3/2$ and $1/2$ with equal probability:

$$\mathbb{P}(X_i = 3/2) = \frac{1}{2} = \mathbb{P}(X_i = 1/2)$$

The expected value of each X_i is

$$\mathbb{E}(X_i) = \frac{1}{2}(3/2 + 1/2) = 1$$

Let $M_0 = 1$ and $M_n = X_1 X_2 \cdots X_n$ (the product). Since these are independent we have

$$\mathbb{E}(M_n) = \mathbb{E}(X_1)\mathbb{E}(X_2)\cdots\mathbb{E}(X_n) = 1$$

$$\mathbb{E}(M_{n+1} \,|\, \mathcal{F}_n) = \underbrace{X_1 \cdots X_n}_{M_n} \underbrace{\mathbb{E}(X_{n+1} \,|\, \mathcal{F}_n)}_{1} = M_n$$

So, M_n is a martingale. Also, since $M_n > 0$ it is equal to its absolute value and

$$\mathbb{E}(|M_n|) = \mathbb{E}(M_n) = 1 \text{ is bounded}$$

Therefore, the first part of the martingale convergence theorem tells us that M_n converges to some function M_∞. But, the following calculation shows that $\mathbb{E}(M_\infty) = 0$. Therefore, the second part of the theorem tells us that M_n are not uniformly integrable.

Here is the calculation. Take the natural log of M_n:

$$\ln M_n = \sum_{i=1}^{n} \ln X_i$$

For each i,

$$\mathbb{E}(\ln X_i) = \frac{1}{2}(\ln 3/2 + \ln 1/2) \approx -.1438$$

By the strong law of large numbers we have that

$$\frac{1}{n} \ln M_n \to \mathbb{E}(\ln X_i) \approx -.1438$$

with probability one. Therefore, $\ln M_n \to -\infty$ and $M_n \to 0$ with probability one.

By the central limit theorem, $\frac{1}{n} \ln M_n$ becomes normal for large n. Then M_n becomes "log normal" which means that its logarithm is normal. For example, the size of grains of sand is distributed approximately log normally since each time it breaks the size is multiplied by a random factor.

MATH 56A: STOCHASTIC PROCESSES
CHAPTER 6

6. Renewal

Mathematically, renewal refers to a continuous time stochastic process with states $0, 1, 2, \cdots$

$$N_t \in \{0, 1, 2, 3, \cdots\}$$

so that you only have jumps from x to $x + 1$ and the probability of jumping from x to $x + 1$ depends only on how long the process was at state x. *Renewal* occurs at each jump.

$$N_t := \text{number of jumps that occur in time interval } (0, t]$$

The jumps (renewals) occur at times $Y, Y + T_1, Y + T_1 + T_2$, etc. and

$$Y + T_1 + \cdots + T_n = \inf\{t \,|\, N_t = n + 1\}$$

The interpretation is that there is an object or process which lasts for a certain amount of time which is random. When the object dies or the process stops then you replace the object with a new one or you restart the process from the beginning: You "renew" the process each time it stops. The number N_t is equal to the number of times renewal has taken place up to time t. Y is the lifetime of the initial process, T_1 is the lifetime of the second one, T_2 is the lifetime of the third one, etc. If the first process starts from the beginning then $Y = 0$ and the numbering is different. T_n becomes the lifetime of the nth process:

$$T_1 + \cdots + T_n = \inf\{t \,|\, N_t = n\}$$

I gave a light bulb as an example. There are three kinds of light bulbs:

(1) The guaranteed light bulb which will last exactly 1000 hours.
(2) The Poisson light bulb. This light bulb is as good as new as long as it is working. Assume it has an expected life of 1000 hours. ($\lambda = 1/1000$).
(3) A general light bulb which has a general probability distribution with the property that its expected life is 1000 hours.

Date: November 14, 2006.

In all three cases,

$$\mu = \mathbb{E}(T) = 1000$$

where T is the length of time that the light bulb lasts.

The first question is: Which light bulb is worth more? The answer is that they are all worth the same. They all give an expected utility of 1000 hours of light. With the general light bulb, there is another question: How long do you expect the last light bulb to last after it has been used for a certain amount of time? This depends on the light bulb. For example, if the guaranteed light bulb has been used for 500 hours then it is only worth half as much as a new one. If the Poisson light bulb lasts 500 hours then it is still worth the same as a new one. We will look at the value of a general light bulb (or a nenewal process with a general distribution.)

6.1. **Renewal theorem.** The guaranteed light bulb is an example of a *periodic* renewal process. Each renewal occurs at multiples of 1000 hours.

Definition 6.1. *A renewal process is* periodic *if renewals always occur at (random) integer multiples of a fixed time interval Δt starting with the first renewal which occurs at time Y.*

The renewal theorem says that, if renewal is not periodic, then the occurrences of the renewal will be spread out evenly around the clock. The probability that it will occur will depend only on the length of time you wait. Since the average waiting time is μ, the probability is approximately the proportion of μ that you wait: $\mathbb{P} \cong \Delta t/\mu$.

For the lightbulb, suppose you install a million lightbulbs at the same time. Then after a while the number of light bulbs that burn out each day will be constant. This (after dividing by one million) will be the *equilibrium distribution.*

Theorem 6.2 (Renewal Theorem)**.** *If a renewal process is aperiodic then, as $t \to \infty$,*

$$\mathbb{P}(\textit{renewal occurs in time } (t, t+dt]) \to \frac{dt}{\mu}$$

where $\mu = \mathbb{E}(T)$. This is equivalent to saying that

$$\lim_{t\to\infty} \mathbb{E}(\textit{number of renewals in time } (t, t+s]) = \frac{s}{\mu}$$

$$\lim_{t\to\infty} \mathbb{E}(N_{t+s} - N_t) = \frac{s}{\mu}$$

6.2. **age of current process.** At any time t, let A_t be the life of the current process. This would be the answer to the question: How long ago did you replace the light bulb? The book says that the pair (N_t, A_t) determines the future of the process. B_t denotes the remaining life of the current process. (How long will the current light bulb last?) First I needed the following lemma.

6.2.1. *picture for an expected value.*

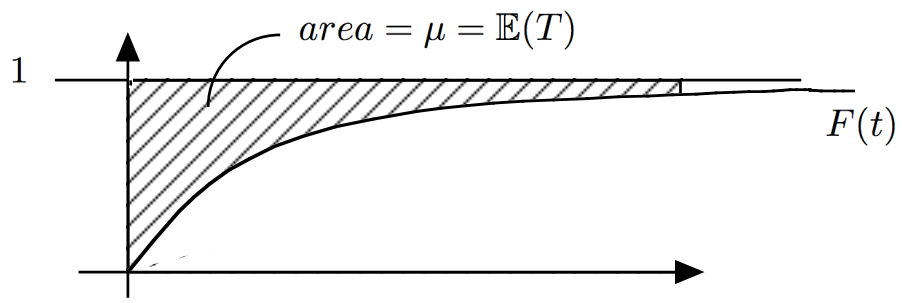

FIGURE 1. The shaded area above the distribution function $F(t)$ is equal to the expectation.

Lemma 6.3. *If $T \geq 0$ is a nonnegative random variable then the expected value of T is given by*

$$\mathbb{E}(T) = \int_0^\infty 1 - F(t)\, dt$$

Proof. The expected value of T is defined by the integral

$$\mathbb{E}(T) = \int_0^\infty tF(t)\, dt$$

Substituting the integral

$$t = \int_0^t ds = \int_{0 \leq s \leq t} ds$$

we get:

$$\mathbb{E}(T) = \iint_{0 \leq s \leq t} f(t)\, dsdt$$

On the other hand,

$$1 - F(s) = \mathbb{P}(T > s) = \int_s^\infty f(t)\, dt$$

So,

$$\int_0^\infty 1 - F(s)\, ds = \int_0^\infty \int_s^\infty f(t)\, dtds = \iint_{0\le s\le t} f(t)\, dsdt = \mathbb{E}(T)$$

□

6.2.2. *distribution of current age.* What is the density function for the current age A_t for large t? I.e., what is $\mathbb{P}(s < A_t \le s + \Delta s)$? This is given by

$$\mathbb{P}(s < A_t \le s + \Delta s) \cong_{t\to\infty} \frac{\Delta s}{\mu}(1 - F(s))$$

because: The renewal event must occur in a time interval Δs: By the renewal theorem this has probability approximately $\Delta s/\mu$. Then the next renewal event must occur at some time greater than s. That has probability $1-F(s)$ where $F(s)$ is the distribution function of the length of time that each renewal process lasts. This is an approximation for large t which depends on the Renewal Theorem. See the figure.

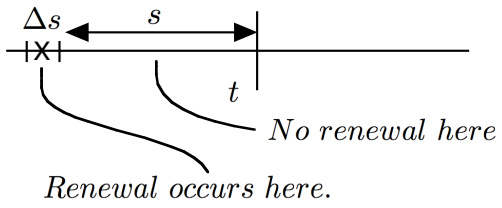

FIGURE 2. The age of the current process tells when was the last renewal.

To get the density function of the current age function A_t we have to take the limit as $\Delta s \to 0$:

$$\psi_A(s) = \lim_{\Delta s\to 0} \frac{1}{\Delta s}\mathbb{P}(s < A_t \le s + \Delta s) = \frac{1 - F(s)}{\mu}$$

The lemma says that the integral of this density function is 1 (as it should be):

$$\int_0^\infty \psi_A(s)\, ds = \int \frac{1 - F(s)}{\mu}\, ds = \frac{\mu}{\mu} = 1$$

For the case of the exponential distribution we have $1 - F(t) = e^{-\lambda t}$ and $\mu = 1/\lambda$. So

$$f(t) = \lambda e^{-\lambda t} = \lambda(1 - F(t)) = \frac{1 - F(t)}{\mu} = \psi_A(t)$$

and the age of the current process has the same distribution as the entire lifespan of the process.

6.2.3. *distribution of remaining life.* The remaining life or *residual life* of the process at time t is simply how long we have to wait for the next renewal. It is called B_t. It is a little more complicated to analyze.

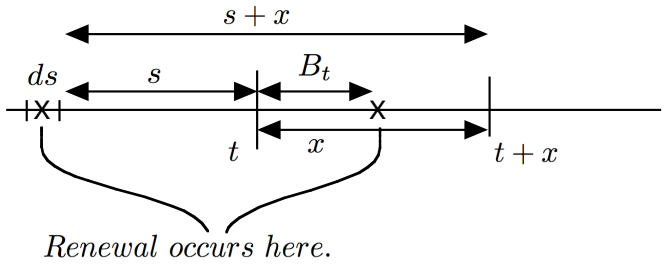

FIGURE 3. The residual life B_t seems to depend on the time of the last renewal.

We want to calculate the distribution function $\Psi_B(x) = \mathbb{P}(B_t \leq x)$. In order for this even to occur, we first need the last renewal to occur in some interval ds before t. This has probability ds/μ. Then we need the event not to occur again during the time interval s before time t but we need it to occur sometime in the time interval x after time t. This occurs with probability $F(s + x) - F(s)$. But $s \geq 0$ is arbitrary. So we sum over all s and get:

$$\Psi_B(x) = \int_0^\infty \frac{ds}{\mu}(F(s + x) - F(s))$$

(See the figure.) To get the density function we should differentiate with respect to x:

$$\psi_B(x) = \Psi'_B(x) = \int_0^\infty \frac{ds}{\mu} f(s + x)$$

If we substitute $t = s + x, dt = ds$ we get:

$$\psi_B(x) = \int_x^\infty \frac{dt}{\mu} f(t) = \frac{1 - F(x)}{\mu} = \psi_A(x)$$

In other words, the current age A_t and the residual life B_t have the same probability distribution.

6.2.4. *relativity argument.* The symmetry between past and future was the point which I wanted to explain using "relativity." Instead of having a fixed time t and looking at the renewal process as occurring at random times, you could think of the renewals as fixed and pick your current time t at random. If you pick t in some renewal period of length C then the point of time that you choose is uniformly distributed (has equal probability of being at any point in the interval). In particular, the left and right parts have the same distribution.

The sum $C_t := A_t + B_t$ is equal to the total duration of the current process. To find the distribution function for C_t you can use this relativity argument. Assume that renewal has occurred a very large number of times, say N. By the law of large number, the total amount of time this takes is very close to $N\mu$. Of these N renewals, $f(x)dx$ represents the proportion of renewal period of duration x to $x + dx$. So, $Nf(x)dx$ is the number of times this occurs. Since the renewal periods all have the same length, the total length of time for all of these renewal periods is just the product $xNf(x)dx$. If you pick a time at random then the probability that the time you pick will be in one of these intervals is

$$\mathbb{P}(x < C_t \leq x + dx) = \frac{xNf(x)dx}{N\mu} = \frac{xf(x)}{\mu}dx$$

Therefore, the density function for C_t is $\psi_C(x) = xf(x)/\mu$.

For example, for the exponential distribution with rate λ, we get:

$$\psi_C(x) = \frac{xf(x)}{\mu} = \frac{x\lambda e^{-\lambda x}}{1/\lambda} = \lambda^2 x e^{-\lambda x}$$

This is the Gamma-distribution with parameters λ and $\alpha = 2$. The reason is that, in this case, C_t is the sum of two independent exponential variables A_t, B_t with rate λ.

6.3. **Convolution.** The convolution is used to describe the density function for the sum of independent random variables. It occurs in this chapter because the lifespan of the renewal periods are independent. So, the density function for the n-th renewal is given by a convolution.

6.3.1. *density of $X + Y$.* Suppose that X, Y are independent random variables with density functions $f(x), g(y)$ respectively. Then we discussed in class that there are two ways to find the density $h(z)$ of $Z = X + Y$. The first method is intuitive and the second is rigorous.

Method 1. I assumed that X, Y are ≥ 0. But this assumption was not necessary. It just made it easier to talk about.

$h(z)dz$ is the probability that $X+Y$ will lie in the interval $[z, z+dz]$. But in order for this to happen we first need X to lie in some interval $[x, x+dx]$ where $0 \leq x \leq z$. This occurs with probability $f(x)dx$. Then we need Y to be in the interval $[z-x, z-x+dz]$. This occurs with probability $g(z-x)dz$. So,

$$h(z)dz = \int_0^z f(x)g(z-x)\,dxdz$$

Divide by dz to get

$$h(z) = \int_0^z f(x)g(z-x)\,dx$$

This is the convolution:

$$h = f * g$$

Method 2. Suppose that the distribution functions of X, Y, Z are F, G, H. Then

$$H(z) = \mathbb{P}(X+Y \leq z) = \int_{-\infty}^{\infty} G(z-x)f(x)dx$$

Differentiate both sides to get

$$h(z) = \int_{-\infty}^{\infty} g(z-x)f(x)dx$$

6.3.2. Γ *distribution.* Suppose that you have a Poisson process with rate λ. Let T be the length of time you have to wait for the αth occurrence of the event. Then T has a Γ distribution with parameters λ and α. Since the expected value of the waiting time for the Poisson event is $1/\lambda$ the expected value of T must be $\mathbb{E}(T) = \alpha/\lambda$.

To get the density function of T we take the convolution of α exponential densities:

$$f = \underbrace{\phi * \phi * \phi * \cdots * \phi}_{\alpha}$$

For example when $\alpha = 2$ we get:

$$f(t) = \int_0^t \phi(x)\phi(t-x)\,dx = \int_0^t \lambda e^{-\lambda x}\lambda e^{-\lambda(t-x)}\,dx$$

$$= \int_0^t \lambda^2 e^{-\lambda t}\,dx = \lambda^2 t e^{-\lambda t}$$

In general you get:

$$f(t) = \frac{1}{(\alpha-1)!}\lambda^\alpha t^{\alpha-1}e^{-\lambda t}$$

if α is an integer and for any α:

$$f(t) = \frac{1}{\Gamma(\alpha)}\lambda^\alpha t^{\alpha-1}e^{-\lambda t}$$

where $\Gamma(\alpha)$ is what it has to be when $\lambda = 1$:

$$\Gamma(\alpha) = \int_0^\infty t^{\alpha-1}e^{-t}\,dt$$

One example of the Γ-distribution is χ^2_r, the chi-squared distribution with r degrees of freedom. This is the Γ-distribution with $\lambda = 1/2$ and $\alpha = r/2$.

6.4. **M/G/1-queueing.** In this model, we have people lining up in a queue and one server taking care of these people one at a time. Let's assume the server is a machine.

In the notation "$M/G/1$" the "1" stands for the number of servers. The "M" means that the "customers" are entering the queue according to a Poisson process with some fixed rate λ. The "G" means that the servers does its job according to some fixed probability distribution which is "general." i.e., it could be anything. This is a renewal process where "renewal" occurs at the moment the queue is empty. At that time, the system is back in its original state with no memory of what happened.

$X_n = \#$ people who enter the line during the n-th service period.

$U_n =$ length of time to serve the n-th person.

So, $\mathbb{E}(X_n) = \lambda\mu$ where $\mu = \mathbb{E}(U_n)$. We need to assume that $\lambda\mu < 1$. Otherwise, the line gets longer and long.

$Y_n = \#$ people in queue right after the n-th person has been served.

Then

$$Y_{n+1} - Y_n = X_{n+1} - 1$$

because X_{n+1} is the number of people who enter the line and one person leaves. (Let $Y_0 = 1$ so that the equation also holds for $n = 0$.)

6.4.1. *stopping time. Busy period* is when the queue and server are active. *Rest periods* are when there is noone in the line. The queue will alternate between busy periods and rest periods. Define the *stopping*

time τ to be the number of people served during the first busy period. Then the first busy time (duration of the 1st busy period) is

$$S_1 = U_1 + U_2 + \cdots + U_\tau$$

To find a formula for τ we used exercise 5.16 on p.128:

(a) $M_n = X_1 + X_2 + \cdots + X_n - n\mathbb{E}(X)$ is a uniformly integrable martingale.
(b) $M_0 = 0$
(c) $OST \Rightarrow \mathbb{E}(M_\tau) = \mathbb{E}(M_0) = 0$. This gives us:
(d) (*Wald's equation*)

$$\mathbb{E}\left(X_1 + \cdots + X_\tau\right) = \mathbb{E}(\tau)\mathbb{E}(X)$$

But the sum of the numbers X_n gives the total number of people who entered the line after the first person. So:

$$X_1 + \cdots + X_\tau = \tau - 1$$

Put this into Wald's equation and we get:

$$\mathbb{E}(\tau) - 1 = \mathbb{E}(\tau)\mathbb{E}(X) = \mathbb{E}(\tau)\lambda\mu$$

where $\mu = \mathbb{E}(U)$. Solve for $\mathbb{E}(\tau)$ to get

$$\mathbb{E}(\tau) = \frac{1}{1 - \lambda\mu}$$

6.4.2. *equilibrium distribution.* We want to know about the equilibrium distribution of the numbers Y_n. The stopping time τ is the smallest number so that $Y_\tau = 0$. This means that τ is the time it takes for Y_n to go from state $Y_0 = 1$ to state $Y_\tau = 0$. So $\tau + 1$ is the number of steps to go from 0 to 0. (In one step it goes to 1.) Therefore, in the equilibrium distribution π of the Markov chain Y_n we have

$$\mathbb{E}(\tau) + 1 = \frac{1}{\pi_0}$$

or

$$\pi_0 = \frac{1}{\mathbb{E}(t) + 1} = \frac{1 - \lambda\mu}{2 - \lambda\mu}$$

MATH 56A: STOCHASTIC PROCESSES
CHAPTER 7

7. Reversal

This chapter talks about time reversal. A Markov process is a state X_t which changes with time. If we run time backwards what does it look like?

7.1. **Basic equation.** There is one point which is obvious. As time progresses, we know that a Markov process will converge to equilibrium. If we reverse time then it will tend to go away from the equilibrium (contrary to what we expect) unless we start in equilibrium. If a process is in equilibrium, it will stay in equilibrium (fluctuating between the various individual states which make up the equilibrium). When we run the film backwards, it will fluctuate between the same states. So, we get a theorem:

Theorem 7.1. *A Markov process with equilibrium distribution π remains a Markov process (with the same equilibrium) when time is reversed provided that*

(1) *left limits are replaced by right limits,*
(2) *the process is irreducible*
(3) *and nonexplosive.*

The time revered process has a different transition matrix

$$\widehat{P} = \Pi^{-1} P^t \Pi$$

where $P = (p(x, y))$ is the transition matrix for the original process and

$$\Pi = \begin{pmatrix} \pi(1) & 0 & \cdots \\ 0 & \pi(2) & \cdots \\ 0 & 0 & \cdots \end{pmatrix}$$

is the diagonal matrix with diagonal entries $\pi(1), \pi(2), \cdots$ given by the equilibrium distribution. In other words,

$$\hat{p}(x, y) = \pi(x)^{-1} p(y, x) \pi(y)$$

or

$$\pi(x)\hat{p}(x, y) = \pi(y)p(y, x)$$

Date: December 14, 2006.

This makes sense because $\pi(y)p(y,x)$ is the equilibrium probability that a random particle will start at y and then go to x. When we run the film backwards we will see that particle starting at x and moving to y. So, the probability of that is $\pi(x)\hat{p}(x,y)$

$$x\bullet \xleftarrow{p(y,x)} \bullet_{\pi(y)} y$$

$$x\bullet_{\pi(x)} \xrightarrow{\hat{p}(x,y)} \bullet y$$

7.1.1. *Example 1.* Take the continuous Markov chain with infinitesmal generator

$$A = \begin{pmatrix} -2 & 2 & 0 \\ 0 & -4 & 4 \\ 1 & 1 & -2 \end{pmatrix}$$

The row are required to have sum zero and terms off the diagonal must be nonnegative. The equilibrium distribution (satisfying $\pi A = 0$) is

$$\pi = (1/4, 1/4, 1/2)$$

So, the time reversed process is

$$\widehat{A} = \Pi^{-1} A^t \Pi = \begin{pmatrix} 4 & & \\ & 4 & \\ & & 2 \end{pmatrix} \begin{pmatrix} -2 & 0 & 1 \\ 2 & -4 & 1 \\ 0 & 4 & -2 \end{pmatrix} \begin{pmatrix} 1/4 & & \\ & 1/4 & \\ & & 1/2 \end{pmatrix}$$

$$\widehat{A} = \begin{pmatrix} -2 & 0 & 2 \\ 2 & -4 & 2 \\ 0 & 2 & -2 \end{pmatrix}$$

7.2. **Reversible process.**

Definition 7.2. *A Markov process is called* reversible *if $\widehat{P} = P$. This is the same as $\widehat{A} = A$. We say it is* reversible with respect to a measure π *if*

$$\pi(x)p(x,y) = \pi(y)p(y,x)$$

Example 1 is not a reversible process because $\hat{A} \neq A$.

Theorem 7.3. *If a Markov chain is reversible wrt a measure π then*

(1) *If $\sum \pi(k) < \infty$ then*

$$\lambda(j) = \frac{\pi(j)}{\sum \pi(k)}$$

is the (unique) invariant probability distribution.

(2) *If $\sum \pi(k) = \infty$ then the process is not positive recurrent.*

7.2.1. *example 2.* Take the random walk on $S = \{0, 1, 2, \cdots\}$ where the probability of going right is p. I.e., $p(k, k+1) = p, p(k+1, k) = 1 - p$.

(1) Show that this is a reversible process.
(2) Find the measure π
(3) What is the invariant distribution λ?

To answer the first two questions we have to solve the equation:

$$\pi(k)p(k, k+1) = \pi(k+1)p(k+1, k)$$

or:

$$\pi(k+1) = \frac{p}{1-p}\pi(k)$$

This has an obvious solution:

$$\pi(k) = \left(\frac{p}{1-p}\right)^k$$

Therefore, the random walk is reversible.

Now we want to find the invariant distribution λ.

If $p < 1/2$ then

$$\sum_{k=0}^{\infty} \pi(k) = \frac{1-p}{1-2p}$$

So, the equilibrium distribution is

$$\lambda(k) = \frac{p^k(1-2p}{(1-p)^{k+1}}$$

If $p \geq 1/2$ then

$$\sum_{k=0}^{\infty} \pi(k) = \infty$$

since the terms don't go to zero. So the process is not positive recurrent and there is no equilibrium.

7.3. **Symmetric process.**

Definition 7.4. *A Markov chain is called* symmetric *if* $p(x, y) = p(y, x)$. *This implies reversible with respect to the uniform measure:* $\pi(x) = 1$ *for all* x *and the process is positive recurrent if and only if there are finitely many states.*

I talked about one example which is related to the final exam. It is example 3 on page 160: Here S is the set of all N-tuples $(a_1, a_2, \cdots, a_N)$ where $a_i = 0, 1$ and the infinitesimal generator is

$$\alpha(a, b) = \begin{cases} 1 & \text{if } a, b \text{ differ in exactly one coordinate} \\ 0 & \text{otherwise} \end{cases}$$

This is symmetric: $\alpha(a,b) = \alpha(b,a)$.

We want to find the second largest eigenvalue
$_2$ of A. (The largest eigenvalue is $\lambda_1 = 0$. The second largest is negative with minimal absolute value.) The eigenvectors of A are also eigenvectors of $P = e^A$ with eigenvalue e^λ, the largest being $e^0 = 1$ and the second largest being $e^{\lambda_2} < 1$.

The first thing I said was that these eigenvectors are π-orthogonal.

Definition 7.5.

$$\langle v, w \rangle_\pi := \sum_{x \in S} v(x) w(x) \pi(x)$$

When $\pi(x) = 1$ (as is the case in this example) this is just the dot product. v, w are called *π-orthogonal* if

$$\langle v, w \rangle_\pi = 0$$

According to the book the eigenvalues of A are

$$\lambda_j = -2j/N$$

for $j = 0, 1, 2, \cdots, N$. This implies that the distance from X_t to the equilibrium distribution decreases at the rate of $-2/N$ on the average:

$$\mathbb{E}(||X_t - \pi||) \leq e^{-2t/N} ||X_0 - \pi||$$

7.4. **Statistical mechanics.** I was trying to explain the Gibbs potential in class and I gave you a crash course in statistical mechanics.

The fundamental assumption is that *All states are equally likely.* Suppose that we have two systems A, B with energy E_1, E_2. Suppose that

$$\Omega_A(E_1) = \#\text{states of } A \text{ with energy } E_1$$

$$\Omega_B(E_2) = \#\text{states of } B \text{ with energy } E_2$$

Then

$$\Omega_A(E_1)\Omega_B(E_2) = \#\text{states of } (A,B) \text{ with energy } E_1 \text{ for } A,\ E_2 \text{ for } B$$

Suppose that the two systems can exchange energy. Then they will exchange energy until the number of states is maximal. This is the same as when the log of the number of states is maximal:

$$\ln \Omega_A(E_1 + \Delta E) + \ln \Omega_B(E_2 - \Delta E) = 0$$

or:

$$\frac{\partial}{\partial E_1} \ln \Omega_A(E_1) = \frac{\partial}{\partial E_2} \ln \Omega_B(E_2) = \beta \quad \text{(constant)}$$

Define the *entropy* of the system A to be $S(E) = \ln \Omega_A(E)$. In equilibrium we have to have

$$\frac{\partial}{\partial E} S(E) = \beta$$

We think of B as an infinite reservoir whose temperature will not change if we take energy out.

Every state has equal probability. But, a state x of A with energy $E(x)$ cannot exist without taking $E(x)$ out of the environment B. Then the number of states of the environment decreases by a factor of $e^{-\beta E(x)}$. Therefore, the probability of the state is proportional to $e^{-\beta E(x)}$. So, the probability of state x is

$$\mathbb{P}(x) = \frac{e^{-\beta E(x)}}{\sum_{y \in S} e^{-\beta E(y)}}$$

The denominator is the *partition function*

$$Z(\beta) = \sum_{y \in S} e^{-\beta E(y)}$$

We looked at the *Ising model* in which there are points in a lattice and a state x is given by putting a sign $\epsilon_i(x) = \pm 1$ at each lattice point i and the energy of the state x is given by

$$E(x) = \sum_{i-j} |\epsilon_i(x) - \epsilon_j(x)| \cdot H$$

(This is $2H$ times the number of adjacent lattice point i, j so that the signs $\epsilon_i(x), \epsilon_j(x)$ are different.) Then I tried to explain the *Gibbs sampler* which is the Markov process which selects a lattice site i at random (with probability $1/\#$lattice points) and then changes $\epsilon_i(x)$ according to the probability of the new state y. So,

$$p(x, y) = \frac{1}{\#\text{lattice points}} \frac{\mathbb{P}(y)}{\mathbb{P}(y) + \mathbb{P}(y')}$$

if x, y differ at only one possible location i and y' is the other possible state which might differ from x at location i. (So, $x = y$ or $x = y'$.)

The Gibbs sampler has the effect of slowly pushing every state towards equilibrium.

MATH 56A: STOCHASTIC PROCESSES
CHAPTER 8

8. Brownian motion

We will be spending the rest of the course on Brownian motion and integration with respect to Brownian motion (stochastic integrals). The topics in the book are closely interrelated so we need to go over everything in the book plus additional material such as Lévy's remarkable theorem which I will explain today. Here is an outline of the chapter.

(0) Definition of Brownian motion
(1) Martingales and Lévy's theorem
(2) Strong Markov property and reflection principle
(3) Fractal dimension of the zero set
(4) Brownian motion and the heat equation in several dimensions
(5) Recurrence and transience
(6) Fractal dimension of the path
(7) Scaling and the Cauchy distribution
(8) Drift

8.0. **Definition of Brownian motion.** First of all this is a random process in continuous time and continuous space. We will start with dimension one: Brownian motion on the real line.

The idea is pretty simple. A particle is bouncing around and its position at time t is X_t. The process is

(1) *memoryless*: What happens for time $s > t$ depends only on its position X_t and not on how it got there.
(2) *time and space homogeneous*: The behavior of the particle remains the same if we reset both time and space coordinates. I.e., the distribution of $Y_t = X_{s+t} - X_s$ depends only on t and is independent of the time s and position X_s.
(3) *continuity*: The particle moves on a continuous path (without jumping from one point to another).

These conditions almost guarantee that we have Brownian motion. But we need a little more. Here is the definition.

Date: November 16, 2006.

8.0.1. *definition.*

Definition 8.1. *A random function* $X : [0, \infty) \to \mathbb{R}$ *written* X_t *is* Brownian motion *with variance* σ^2 *starting at 0 if:*

(1) $X_0 = 0$
(2) *For any* $s < t_1 \leq s_2 < t_2 \leq \cdots \leq s_n < t_n$ *the random variables* $X_{t_1} - X_{s_1}, X_{t_2} - X_{s_2}, \cdots, X_{t_n} - X_{s_n}$ *are independent.*
(3) *The path* X_t *is continuous*
(4) *For* $s < t$,

$$X_t - X_s \sim N(0, (t-s)\sigma^2)$$

i.e., $X_t - X_s$ *is normally distributed with mean 0 and variance* $(t-s)\sigma^2$.

Theorem 8.2. *The last condition is equivalent to the condition:*

(4') $X_t - X_s$ *and* $X_{t+c} - X_{s+c}$ *are identically distributed with mean 0 and variance* $(t-s)\sigma^2$.

Proof. (Outlined in the book.) Certainly (4) $\Rightarrow$ (4′). To prove the converse, assume (4′). Let $\Delta t = (t-s)/N$ for large N. Then

$$X_t - X_s = (X_{s+\Delta t} - X_s) + (X_{s+2\Delta t} - X_{s+\Delta t}) + \cdots + (X_t - X_{t-\Delta t})$$

This is a sum of N independent identically distributed random variables with mean 0 and variance $(\Delta t)\sigma^2$. By the central limit theorem we get

$$X_t - X_s \approx N(0, N\Delta t\, \sigma^2) = N(0, (t-s)\sigma^2)$$

Now take the limit as $N \to \infty$. (This is not rigorous because we are not using the precise statement of the CLT.) □

Recall that the *variance* of a random variable X is defined by

$$Var(X) := \mathbb{E}((X - \mathbb{E}(X))^2) = \mathbb{E}(X^2) - \mathbb{E}(X)^2$$

and it has the property that it is additive for *independent* random variables:

$$Var(X_1 + X_2 + \cdots + X_n) = Var(X_1) + Var(X_2) + \cdots + Var(X_n)$$

8.0.2. *as limit of random walk.* Brownian motion can be obtained as a limit of random walks: Take time to be integer multiples of a fixed interval Δt and we take points on the line at integer multiples of $\sigma\sqrt{\Delta t}$. For each unit time assume that position changes by $\pm\sigma\sqrt{\Delta t}$ with equal probability. This is Bernoulli with mean 0 and variance

$$(\pm\sigma\sqrt{\Delta t}) = \sigma^2 \Delta t$$

In a time interval $N\Delta t$ the change of position is given by a sum of N independent random variables. So, the mean would be 0 and the

variance would be $N\Delta t\sigma^2$. The point is that this is σ^2 times the length of the time interval. As $\Delta t \to 0$, assuming the sequence of random walks converges to a continuous function, the limit gives Brownian motion with variance σ^2 by the theorem.

8.0.3. *nowhere differentiable.* Notice that, as Δt goes to zero, the change in position is approximately $\sigma\sqrt{\Delta t}$ which is much bigger than Δt. This implies that the limit

$$\lim_{t\to 0} \frac{X_t}{t}$$

diverges. So, Brownian motion is, almost surely, nowhere differentiable. (*Almost surely* or a.s. means "with probability one.")

8.1. **Martingales and Lévy's theorem.**

Theorem 8.3. *Suppose that X_t is Brownian motion. Then*

(1) *X_t is a continuous martingale.* starting at any point!
(2) *$X_t^2 - t\sigma^2$ is a martingale*

Proof. (1) is easy: If $t > s$ then

$$\begin{array}{rcccc} \mathbb{E}(X_t \,|\, \mathcal{F}_s) & = & \mathbb{E}(X_t - X_s \,|\, \mathcal{F}_s) & + & \mathbb{E}(X_s \,|\, \mathcal{F}_s) \\ & = & 0 & + & X_s \end{array}$$

The proof does not use the assumption $X_0=0$

where $\mathcal{F}_s$ is the information contained in X_r for all $r \le s$.

For (2), we need the equation:

$$\begin{aligned} (X_t - X_s)^s &= X_t^2 - 2X_tX_s + 2X_s^2 - X_s^2 \\ &= X_t^2 - 2(X_t - X_s)X_s - X_s^2 \end{aligned}$$

Taking $\mathbb{E}(- \,|\, \mathcal{F}_s)$ of both sides gives:

$$\mathbb{E}((X_t - X_s)^2 \,|\, \mathcal{F}_s) = Var(X_t - X_s) = (t-s)\sigma^2 =$$

$$\mathbb{E}(X_t^2 \,|\, \mathcal{F}_s) - 2\mathbb{E}(X_t - X_s \,|\, \mathcal{F}_s)X_s - X_s^2 = \mathbb{E}(X_t^2 \,|\, \mathcal{F}_s) - X_s^2$$

Which gives

$$\mathbb{E}(X_t^2 - t\sigma^2 \,|\, \mathcal{F}_s) = X_s^2 - s\sigma^2$$

□

Lévy's theorem is the converse:

Theorem 8.4 (Lévy)**.** *Suppose that X_t is a continuous martingale and $X_t^2 - t\sigma^2$ is also a martingale. Then X_t is Brownian motion with variance σ^2* (with X_0 arbitrary)

This famous theorem has been proved many times. I will try to find the proof using stochastic integrals. One amazing consequence is the following.

Corollary 8.5. *Any continuous martingale M_t is Brownian motion reparametrized and starting at $C = M_0$. I.e.*

$$M_t = X_{\phi(t)} + C$$

where X_s is standard Brownian motion (with $\sigma = 1$).

Proof. (When I did this in class I forgot to "center" the martingale by subtracting M_0.) The idea is to let $\phi(t) = \mathbb{E}((M_t - C)^2)$ and apply Lévy's theorem. (I'll look for the details). □

8.2. **Strong Markov property and Reflection principle.**

Theorem 8.6 (strong Markov property)**.** *Let T be a stopping time for Brownian motion X_t. Let*

$$Y_t = X_{t+T} - X_T$$

Then Y_t is independent of $\mathcal{F}_T$ (for $t > 0$).

One consequence of this is the *reflection principle* which the book uses over and over.

8.2.1. *reflection principle.*

Corollary 8.7 (reflection principle)**.** *Suppose that $a < b$ then the probability that you will reach b from a within time t is twice the probability that at time t you will be past b. I.e.:*

$$\mathbb{P}(X_s = b \text{ for some } 0 < s < t \,|\, X_0 = a) = 2\mathbb{P}(X_t > b \,|\, X_0 = a)$$

Proof. If you reach the point b at some time before time t then half the time you will end up above b and half the time you will end up below b since the probability that $X_t = b$ is zero. So,

$$\mathbb{P}(\ X_s \text{ reaches } b \text{ sometime before } t \text{ and ends up higher} \,|\, X_0 = a)$$

$$= \frac{1}{2}\mathbb{P}(X_s = b \text{ for some } 0 < s < t \,|\, X_0 = a)$$

But the event "X_s reaches b sometime before t and ends up higher" is the same as the event "$X_t > b$" since X_t is continuous and therefore cannot get to a point $X_t > b$ starting at $a < b$ without passing through b. This proves the reflection principle.

Why is the reflection principle a corollary of the strong Markov property? The reason is that we are using the stopping time $T =$ the first time that $X_T = b$. And $Y = X_t - X_T$. For every fixed T this is normally distributed with mean 0 and variance $(t - T)\sigma^2$. So,

$$\mathbb{P}(Y > 0 \,|\, T < t) = \frac{1}{2}$$

By the formula for conditional probability, this is

$$\frac{1}{2} = \mathbb{P}(Y > 0 \,|\, T < t) = \frac{\mathbb{P}(Y > 0 \text{ and } T < t)}{\mathbb{P}(T < t)}$$

But "$Y > 0$ and $T < t$" is the same as "$X_t > b$" and "$T < t$" is the same as "X_s reaches b sometime before t." So, this gives the reflection principle again. □

8.2.2. *density function.* If $X_0 = a$, then $X_t - a \sim N(0, t\sigma^2)$. The normal distribution $N(0, t\sigma^2)$ has density function

$$\phi_t(x) = \frac{1}{\sqrt{2\pi t\sigma^2}}\, e^{-x^2/2\sigma^2 t}$$

The probability density function for X_t is given by shifting the normal distribution by $a = X_0$.

$$f_{X_t}(x) = \phi_t(x - a) = p_t(a, x) = \frac{1}{\sqrt{2\pi t\sigma^2}}\, e^{-(x-a)^2/2\sigma^2 t}$$

It is called $p_t(a, x)$ because it is the (infinitesimal) transition matrix:

$$p_t(a, x)dx = \mathbb{P}(x < X_t \leq x + dx \,|\, X_0 = a)$$

The integral of this over any interval I is equal to the probability that X_t will lie in I. E.g.,

$$\mathbb{P}(X_t > b \,|\, X_0 = a) = \int_b^\infty p_t(a, x)\, dx = \int_b^\infty \frac{1}{\sqrt{2\pi t\sigma^2}}\, e^{-(x-a)^2/2\sigma^2 t}\, dx$$

In the reflection principle we get twice this number:

$$\mathbb{P}(X_s = b \text{ for some } 0 < s < t \,|\, X_0 = a) = 2\int_b^\infty \frac{1}{\sqrt{2\pi t\sigma^2}}\, e^{-(x-a)^2/2\sigma^2 t}\, dx$$

If we make the substitution

$$y = \frac{x - a}{\sigma\sqrt{t}}, \quad dy = \frac{dx}{\sqrt{\sigma^2 t}}$$

we get the standard normal distribution:

$$2\int_{(b-a)/\sigma\sqrt{t}}^\infty \frac{1}{\sqrt{2\pi}} e^{-y^2/2}\, dy$$

This is the well-know rule: *To convert to standard normal, subtract the mean then divide by the standard deviation.* It works for integrals.

8.2.3. *The Chapman-Kolmogorov equation.* is an example of a formula which is easy to understand using Brownian motion:

$$p_{s+t}(x,y) = \int_{-\infty}^{\infty} p_s(x,z)p_t(z,y)\,dz$$

This is just a continuous version of the matrix equation

$$P_{s+t} = P_s P_t$$

and it holds for *all Markov processes*.

In the particular case of Brownian motions, the integral is a convolution and the Chapman-Kolmogorov equation can be rewritten as:

$$\phi_{s+t} = \phi_s * \phi_t$$

As I explained to you last week, convolution of density functions gives the density function for the sum of two random variables. In this case:

$$pdf(X_{s+t} - X_0) = pdf(X_s - X_0) * pdf(X_{t+s} - X_s)$$

8.2.4. *example 1.* Here we want the probability that standard Brownian motion, starting at 0, will return to 0 sometime between time 1 and time $t > 1$.

$$\mathbb{P}(X_s = 0 \text{ for some } 1 < s < t \,|\, X_0 = 0) = ?$$

We first look at where the particle is at time 1. Half the time X_1 will be positive and half the time it will be negative. So, we will assume that $X_1 > 0$ then multiply by 2. By the reflection principle, the probability of returning to 0 before time t is twice the probability that $X_t < 0$. So, the answer will be

$$4\mathbb{P}(X_1 > 0 \text{ and } X_t < 0 \,|\, X_0 = 0)$$

$$= 4\mathbb{P}(X_1 = b > 0 \text{ and } X_t - X_1 < -b \,|\, X_0 = 0)$$

The probability for fixed b (in the interval $(b, b+db]$) is

$$\phi_1(b)db\,\Phi_{t-1}(-b)$$

where Φ_{t-1} is the cumulative distribution function:

$$\Phi_{t-1}(-b) = \int_{-\infty}^{-b} \phi_{t-1}(x)\,dx = \int_{b}^{\infty} \phi_{t-1}(x)\,dx = \int_{b/\sqrt{t-1}}^{\infty} \phi_1(y)\,dy$$

where we used the "convert to standard normal" rule. The answer is now given by integrating over all $b > 0$ and multiplying by 4:

$$4\int_0^{\infty}\int_{b/\sqrt{t-1}}^{\infty} \phi_1(b)\phi_1(y)\,dydb$$

The integrand is

$$\phi_1(b)\phi_1(y) = \frac{1}{\sqrt{2\pi}}\, e^{-b^2/2} \frac{1}{\sqrt{2\pi}}\, e^{-y^2/2} = \frac{1}{2\pi}\, e^{-(b^2+y^2)/2}$$

Now convert to polar coordinates:

$$\phi_1(b)\phi_1(y)\, dy\, db = \frac{1}{2\pi}\, e^{-r^2/2}\, r dr d\theta$$

The answer is

$$4 \int_0^\infty \int_{\arctan(1/\sqrt{t-1})}^{\pi/2} \frac{1}{2\pi}\, e^{-r^2/2}\, r d\theta dr$$

$$= 4\left(\pi/2 - \arctan(1/\sqrt{t-1})\right) \frac{1}{2\pi} \int_0^\infty e^{-r^2/2}\, r dr$$

$$= 1 - \frac{2}{\pi} \arctan \frac{1}{\sqrt{t-1}}$$

8.2.5. *example 2.* In this example we have to show that, a.s.,

$$\lim_{t\to\infty} \frac{X_t}{t} = 0$$

where X_t is standard Brownian motion.

First, let $t = n$ be an integer going to infinity. Then

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

This is a sum of n iid random variables. So, by the strong law of large numbers, the average will converge to the expected value with probability one:

$$\lim_{n\to\infty} \frac{X_n}{n} = \mathbb{E}(X_1 - X_0) = 0$$

Next, we have to show that, as t goes from one integer n to the next $n+1$, X_t doesn't deviate too far from X_n. What the book shows is that, a.s., for all but a finite number of n, the difference

$$|X_t - X_n| < 2\sqrt{\ln n} = a$$

When we divide by n the difference between X_t and X_n will go to zero. Dividing by t is even better because $t > n$.

We want the probability that at some time $t \in (n, n+1)$, $X_t - X_n$ goes above a or below $-a$. By symmetry this is twice the probability that it will go above a. By the reflection principle this is 4 times the probability that it will end up above a:

$$\mathbb{P}(|X_t - X_n| > a \text{ for some } n < t < n+1) = 4\mathbb{P}(X_{n+1} - X_n > a)$$

$$= 4 \int_a^\infty \frac{1}{\sqrt{2\pi}}\, e^{-x^2/2}\, dx \le 4 \int_a^\infty \frac{1}{\sqrt{2\pi}}\, e^{-ax/2}$$

$$= \frac{8}{a\sqrt{2\pi}}\, e^{-a^2/2} = \frac{8}{2\sqrt{\ln n}\sqrt{2\pi}n^2}$$

(because $-a^2/2 = -2\ln n = \ln n^{-2}$)

This is an upper bound for the probability that X_t deviates too far from X_n for a single n. If we sum over all n we get an upper bound for the expected number of times that this will happen. But

$$\sum_{n=2}^{\infty} \frac{8}{2\sqrt{\ln n}\sqrt{2\pi}n^2} \leq C \sum \frac{1}{n^2} < \infty$$

So, the expected number is finite. So, a.s., the deviation occurs only finitely many times. (If there were a nonzero probability p that the event occurs an infinite number of times then the expected number would be at least $p \cdot \infty = \infty$.)

8.3. **Fractal dimension of the zero set.** The *zero set* Z is just the set of all times t so that $X_t = 0$. This is a subset of the positive real line: $Z \subset [0, \infty)$. The zero set is a *fractal* in the sense that it looks the same on the small scale as it does on the big scale. The "fractal dimension" of the set measures the scale at which the set is self-similar. We use the box dimension definition.

8.3.1. *self-similarity of Z.*

Theorem 8.8. *Suppose that X_t is Brownian motion with variance σ^2. Then*

1. *$Y_t = bX_{at}$ is Brownian motion with variance $b^2 a\sigma^2$ ($= \sigma^2$ if $b^2 = 1/a$).*
2. *$Y_t = tX_{1/t}$ is Brownian motion with variance σ^2.*

Proof. For (2) you need to use the fact that, a.s.,

$$\lim_{t\to 0} tX_{1/t} = \lim_{1/t\to\infty} \frac{X_{1/t}}{1/t} = 0$$

Therefore, $Y_t = tX_{1/t}$ is continuous at $t = 0$. This settles the continuity condition. The other conditions are clear: Since the functions at and $1/t$ are monotonic, Y_t is a memoryless process in both cases. We just have to calculate the variance (and see that it is constant). This is easy:

$$Var(bX_{at}) = \mathbb{E}(b^2 X_{at}^2) = b^2 at\sigma^2$$

$$Var(tX_{1/t}) = \mathbb{E}(t^2 X_{1/t}^2) = t^2(1/t)\sigma^2 = t\sigma^2$$

Here is a more careful proof (mainly for my reference).

We have to show that $Y_t - Y_s$ is normally distributed with variance proportional to $t - s$. (It obviously has zero mean.) In case (1) this is easy:

$$Y_t - Y_s = bX_{at} - bX_{as} \sim bN(0, (at - as)\sigma^2) = N(0, b^2 a(t - s)\sigma^2)$$

In case (2) we have:

$$Y_t - Y_s = tX_{1/t} - sX_{1/s} = (t - s)X_{1/t} + s(X_{1/t} - X_{1/s})$$
$$\sim N(0, (t - s)^2 \frac{1}{t}\sigma^2) + N(0, s^2 \left(\frac{1}{t} - \frac{1}{s}\right)\sigma^2)$$

Then use the fact that the sum of independent normal distributions is normal with mean the sum of the means and variance the sum of the variances. Then calculate:

$$(t - s)^2 \frac{1}{t}\sigma^2 + s^2 \left(\frac{1}{t} - \frac{1}{s}\right)\sigma^2 = (t - s)\sigma^2$$

□

What does this mean in terms of the set Z?

First of all, if we multiply X_t by a constant, the zero set is unchanged since $bX_t = 0 \Leftrightarrow X_t = 0$. Therefore, the theorem says:

(1) Z looks like aZ for any positive constant a.
(2) Z looks like $1/Z$.
(3) Z does not depend of the variance σ^2. (So, we assume $\sigma^2 = 1$.)

When I say "looks like" I mean it "has the same probability distribution as."

8.3.2. *gaps in Z.* Example 1 from 8.2 calculates the probability that Z meets the set $[1, t]$

$$\mathbb{P}(Z \cap [1, t] \neq \emptyset) = 1 - \frac{2}{\pi} \arctan\left(\frac{1}{\sqrt{t - 1}}\right)$$

or:

$$\mathbb{P}(Z \cap [1, t] = \emptyset) = \frac{2}{\pi} \arctan\left(\frac{1}{\sqrt{t - 1}}\right)$$

This is equal to $1/2$ when $t = 2$. And as $t \to \infty$ this probability goes to zero ($\arctan 0 = 0$). So,

$$\mathbb{P}(Z \cap [1, \infty) = \emptyset) = 0$$

The scaling theorem now says the same is true for any rescaling:

$$\mathbb{P}(Z \cap [t, 2t] = \emptyset) = \frac{1}{2}$$

and

$$\mathbb{P}(Z \cap [t, \infty) = \emptyset) = 0$$

for any $t > 0$.

8.3.3. *fractal dimension.* First I'll explain the simple version of the idea and do some examples. Then I'll give the mathematical definition.

Take the unit square. If we scale this by a factor of 10 then we get something which can be cut into $10^2 = 100$ squares. If we take a cube and scale by 10 we get $10^3 = 1,000$ cubes. The dimension is equal to the exponent that we need to take the scaling factor to. For the Cantor set C, if you scale by a factor of 3 then you get 2 Cantor sets. So, its dimension D is the solution of the equation

$$2 = 3^D$$

Taking the log of both sides gives

$$D = \dim C = \frac{\ln 2}{\ln 3} \approx 0.631$$

Instead of scaling the object and seeing how big it gets, you could just as well scale the units down and sees how many smaller units you need. For example take a unit square. How many little $1/10 \times 1/10$ squares do you need to cover it? It is $10^2 = (1/10)^{-2}$ just like before except that now the dimension is the negative of the exponent of the scale of the little pieces. It is the same concept.

Definition 8.9. *The* box dimension *of a bounded subset A of $\mathbb{R}^d$ is equal to the infimum of $D > 0$ so that as $\epsilon \to 0$, the number of cubes with sides ϵ needed to cover A becomes $< C\epsilon^{-D}$ where C is a constant.*

The set A needs to be bounded otherwise you need an infinite number of little cubes to cover it.

8.3.4. *dimension of Z.* Take the bounded set $Z_1 = Z \cap [0, 1]$. What is the expected number of intervals of length $\epsilon = 1/n$ needed to cover Z_1? It should be $\sim n^D$ where D is the dimension of Z_1.

The expected number of intervals needed is equal to the sum of probabilities

$$\mathbb{E}(\text{number of intervals } [k/n, (k+1)/n] \text{ that meet } Z_1)$$
$$= \sum_{k=0}^{n-1} \mathbb{P}(Z_1 \cap \left[\frac{k}{n}, \frac{k+1}{n}\right] \neq \emptyset)$$

But the scaling theorem tells us that

$$\mathbb{P}(Z_1 \cap \left[\frac{k}{n}, \frac{k+1}{n}\right] \neq \emptyset) = \mathbb{P}(Z \cap \left[1, \frac{k+1}{k}\right] \neq \emptyset)$$
$$= 1 - \frac{2}{\pi} \arctan \sqrt{k}$$

So, the expected number of intervals is

$$\sum_{k=0}^{n-1} 1 - \frac{2}{\pi} \arctan \sqrt{k}$$

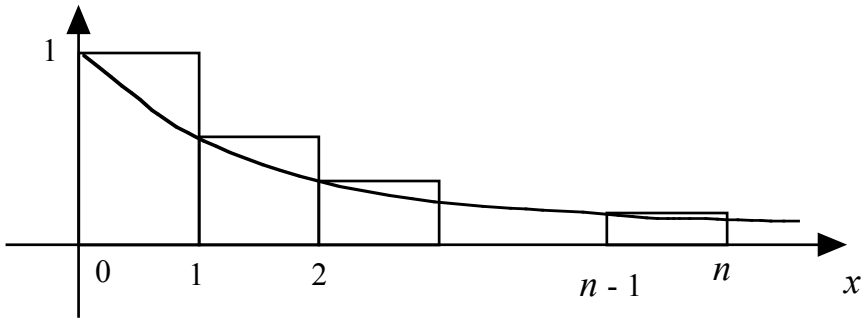

FIGURE 1. The sum is equal to the integral plus the little triangles which can be stacked up to give about 1/2

This is a Riemann sum. (See Figure 1.) So, it is approximately equal to

$$\frac{1}{2} + \int_0^n 1 - \frac{2}{\pi} \arctan \sqrt{x}\, dx$$
$$= \frac{1}{2} + n - \frac{2}{\pi}(n+1) \arctan \sqrt{n} + 2\sqrt{n}$$

Using the approximation

$$\arctan x \approx \frac{\pi}{2} - \frac{1}{x}$$

this becomes

$$\approx \frac{1}{2} + n - \left(n + 1 - \frac{2n+2}{\pi\sqrt{n}} \right) + 2\sqrt{n} \approx \sqrt{n} \left(2 - \frac{2}{\pi} \right) \approx 1.36\sqrt{n}$$

The dimension of Z_1 is the exponent of n which is $D = 1/2$.

8.4. **Brownian motion and the heat equation in several dimensions.** (When you read this section ask yourself: What do people mean when they say:

"The infinitesmal generator of Brownian motion is $\frac{1}{2}\frac{\partial^2}{\partial x^2}$" ?)

8.4.1. *definition.* With several variables, Brownian motion can be written in rectangular or polar coordinates. I prefer the version which is obviously rotationally invariant. (You can rotate the coordinates and the equation does not change.)

Definition 8.10. Standard d-dimensional Brownian motion *is a vector valued stochastic process, i.e., random function* $\mathbf{X} : [0, \infty) \to \mathbb{R}^d$ *so that*

(1) $\mathbf{X}_0 = 0$
(2) *For any* $s < t_1 \leq s_2 < t_2 \leq \cdots \leq s_n < t_n$ *the random variables* $\mathbf{X}_{t_1} - \mathbf{X}_{s_1}, \mathbf{X}_{t_2} - \mathbf{X}_{s_2}, \cdots, \mathbf{X}_{t_n} - \mathbf{X}_{s_n}$ *are independent.*
(3) *The path* $\mathbf{X}_t$ *is continuous*
(4) *For* $s < t$*, the random variable* $\mathbf{X}_t - \mathbf{X}_s$ *has density function*

$$\phi_r(\mathbf{x}) = \frac{1}{(2\pi r)^{d/2}}\, e^{-||\mathbf{x}||^2/2r}$$

where $r = t - s$ *and* $||\mathbf{x}||$ *is the length of the vector* $\mathbf{x} \in \mathbb{R}^d$*.*

The coordinates $X_t^j - X_s^j$ of the vector $\mathbf{X}_t - \mathbf{X}_s$ are independent standard 1-dimensional Brownian motions with densities

$$\frac{1}{(2\pi r)^{1/2}}\, e^{-x_j^2/2r}$$

whose product is $\phi_r(\mathbf{x})$. The elapsed time or time increment is denoted $r = t - s$ in the definition. The *covariance matrix* is the $d \times d$ matrix

$$\left[\mathbb{E}((X_t^i - X_s^i)(X_t^j - X_s^j))\right] = (t - s)\mathbf{I}$$

The *transition density* for time increment t is

$$p_t(\mathbf{x}, \mathbf{y}) = \phi_t(\mathbf{y} - \mathbf{x}) = \frac{1}{(2\pi t)^{d/2}}\, e^{-||\mathbf{y}-\mathbf{x}||^2/2t}$$

One important feature is that the transition density is *symmetric*, i.e.,

$$p_t(\mathbf{x}, \mathbf{y}) = p_t(\mathbf{y}, \mathbf{x})$$

This satisfies the *Chapman-Kolmogorov* equation just like any other transition density.

8.4.2. *diffusion.* If you have a large number of particles moving independently according to the rules of Brownian motion, then the distribution of these particles will change in a deterministic process called *diffusion.*

Let $f_t(\mathbf{x})$ denote the density of particles at time t and position $\mathbf{x}$. After an increment of time δt, the density will change to

$$f_{t+\delta t}(\mathbf{y}) = \int_{\mathbb{R}^d} f_t(\mathbf{x}) p_{\delta t}(\mathbf{x}, \mathbf{y})\, \mathbf{dx} \tag{8.1}$$

where I used the abbreviation $\mathbf{dx} = dx_1 dx_2 \cdots dx_d$. Since $p_{\delta t}(\mathbf{x}, \mathbf{y}) = p_{\delta t}(\mathbf{y}, \mathbf{x})$ we can rewrite this as

$$f_{t+\delta t}(\mathbf{y}) = \int_{\mathbb{R}^d} f_t(\mathbf{x}) p_{\delta t}(\mathbf{y}, \mathbf{x})\, \mathbf{dx}$$

Now switch "$\mathbf{x}$" and "$\mathbf{y}$":

$$f_{t+\delta t}(\mathbf{x}) = \int_{\mathbb{R}^d} f_t(\mathbf{y}) p_{\delta t}(\mathbf{x}, \mathbf{y})\, \mathbf{dy} \tag{8.2}$$

These equations look similar but they mean different things. The first equation (8.1) gives the density of particles as an sum over all places where the particles came from. Equation (8.2) says that the future density at $\mathbf{x}$ will be equal to the expected value of the present density function of the new (random) location of a single particle starting at the point $\mathbf{x}$. The first equation is deterministic and the second is probabilistic!

Equation (8.2) can be written:

$$f_{t+\delta t}(\mathbf{x}) = \mathbb{E}^{\mathbf{x}}(f_t(\mathbf{X}_{\delta t})) = \mathbb{E}(f_t(\mathbf{X}_{\delta t}) \,|\, \mathbf{X}_0 = \mathbf{x})$$

where we use the abbreviation $\mathbb{E}^{\mathbf{x}} = \mathbb{E}(- \,|\, \mathbf{X}_0 = \mathbf{x})$. I changed the equation from that in the book to clarify what is absolute time and what is relative time.

8.4.3. *the differential equation.* Now take the limit as $\delta t \to 0$:

$$\frac{\partial}{\partial t} f_t(\mathbf{x}) = \lim_{\delta t \to 0} \frac{1}{\delta t} \mathbb{E}^{\mathbf{x}}(f_t(\mathbf{X}_{\delta t}) - f_t(\mathbf{X}_0)) \tag{8.3}$$

On the RHS we are taking density at a fixed time and variable position. The book estimates the density first in the case $d = 1$:

$$f_t(X_{\delta t}) = f_t(X_0) + \frac{\partial}{\partial x} f_t(x)(X_{\delta t} - X_0) + \frac{1}{2} \frac{\partial^2}{\partial x^2} f_t(x)(X_{\delta t} - X_0)^2 + o((X_{\delta t} - X_0)^2)$$

Now take expected value

$$\mathbb{E}^x(f_t(X_{\delta t}) - f_t(X_0)) = \frac{\partial}{\partial x} f_t(x) \mathbb{E}^x(X_{\delta t} - X_0) + \frac{1}{2} \frac{\partial^2}{\partial x^2} f_t(x) \mathbb{E}^x((X_{\delta t} - X_0)^2) + o(\delta t)$$

But $\mathbb{E}^x(X_{\delta t} - X_0) = 0$ and $\mathbb{E}^x((X_{\delta t} - X_0)^2) = (\delta t - 0)\sigma^2 = \delta t$. So,

$$\mathbb{E}^x(f_t(X_{\delta t}) - f_t(X_0)) = \frac{1}{2}\frac{\partial^2}{\partial x^2} f_t(x)\delta t + o(\delta t)$$

Dividing by δt and taking the limit as $\delta t \to 0$ gives

$$\frac{\partial}{\partial t} f_t(x) = \frac{1}{2}\frac{\partial^2}{\partial x^2} f_t(x)$$

For the random vector $\mathbf{X}_{\delta t}$ we get:

$$f_t(\mathbf{X}_{\delta t}) - f_t(\mathbf{X}_0) =$$

$$\sum_i \frac{\partial}{\partial x_i} f_t(\mathbf{x})({X^i}_{\delta t} - X_0^i) + \sum_{i,j} \frac{1}{2}\frac{\partial^2}{\partial x_i \partial x_j} f_t(\mathbf{x})(({X^i}_{\delta t} - X_0^i)({X^j}_{\delta t} - X_0^j)) + o(\delta t)$$

Taking expected value, the first term gives zero. The expected value of the second term is given by the covariance matrix $\delta t\mathbf{I}$. So

$$\mathbb{E}^{\mathbf{x}}(f_t(\mathbf{X}_{\delta t}) - f_t(\mathbf{X}_0)) = \sum_i \frac{1}{2}\frac{\partial^2}{\partial x_i^2} f_t(\mathbf{x})\delta t + o(\delta t)$$

Referring back to the original differential equation (8.3) we get

$$\frac{\partial}{\partial t} f_t(\mathbf{x}) = \frac{1}{2}\Delta f_t(\mathbf{x})$$

where Δ is the *Laplacian*

$$\Delta = \sum_{i=1}^{d} \frac{1}{2}\frac{\partial^2}{\partial x_i^2}$$

If the particles are moving according to Browning motion with variance σ^2 then the density changes according to the equation

$$\frac{\partial}{\partial t} f_t(\mathbf{x}) = \frac{D}{2}\Delta f_t(\mathbf{x}) \tag{8.4}$$

where $D = \sigma^2$.

8.4.4. *heat equation.* The equation (8.4) is called the *heat equation* with *diffusion constant* D. We will see how the time reversal trick explained above can be used to solve this equation using stochastic methods.

Suppose that B is a region in $\mathbb{R}^d$ with boundary ∂B. Suppose we start with a heat density function f on B which changes according to the heat equation and a constant heat density of g on ∂B. If the temperature in the interior point $\mathbf{x}$ at time t is $u(t, \mathbf{x})$ then the probabilistic interpretation is that $u(t, \mathbf{x})$ is equal to the expected value of the density at time 0 at the position a particle starting at $\mathbf{x}$ will end up at time t.

The boundary, which in forward time is emitting heat at a constant rate, will, in backward time, act like flypaper for randomly moving particles. The particle starts at $\mathbf{X}_0 = \mathbf{x}$ and moves according to Brownian motion and stops the first time it hits the boundary ∂B. This is a stopping time. Call it $\tau = \tau_{\partial B}$.

The equation for this expected value is

$$u(t, \mathbf{x}) = \mathbb{E}^{\mathbf{x}}(g(\mathbf{X}_\tau)I(\tau < t) + f(\mathbf{X}_t)I(t \leq \tau))$$

As t goes to infinity, the temperature reaches a steady state given by

$$v(\mathbf{x}) = \mathbb{E}^{\mathbf{x}}(g(\mathbf{X}_\tau))$$

So, this is the solution of the equation:

$$\Delta v(\mathbf{x}) = 0$$

(in the interior of B) with boundary condition $v = g$ on ∂B.

8.4.5. *example 1: probabilistic method.* Let $d = 1$ and $B = (a, b) \subset \mathbb{R}$ where $0 \leq a < b < \infty$. Then the boundary is just two points $\partial B = \{a, b\}$. Suppose the function on the boundary is $g(a) = 0, g(b) = 1$. We start at some point $x \in (a, b)$ and stopping time τ is when we reach either a or b. This is the "gambler's ruin" because it describes what happens to a gambler playing a fair game who starts with $\$x$ and quits when he reaches either a or b.

$$v(x) = \mathbb{E}^x(g(X_\tau)) = \mathbb{P}^x(X_\tau = b)$$

By the strong Markov property, $\mathbb{E}^x(X_\tau) = \mathbb{E}^x(X_0) = x$. So,

$$\mathbb{E}^x(X_\tau) := a\mathbb{P}^x(X_\tau = a) + b\mathbb{P}^x(X_t = b) = a(1 - v(x)) + bv(x) = x$$

$$v(x) = \frac{x - a}{b - a}$$

8.4.6. *example 2: analytic method.* Now consider the case when $g(0) = g(2\pi) = 0$ and $B = (0, \pi)$. The heat equation

$$\frac{\partial}{\partial t}u(t, x) = \frac{1}{2}\frac{\partial^2}{\partial x^2}v(t, x)$$

is homogeneous and has a complete list of basic solutions:

$$u(t, x) = e^{-tn^2/2}\sin(nx)$$

and any solution is an infinite linear combination

$$u(t, x) = \sum_{n=1}^{\infty} C_n e^{-tn^2/2}\sin(nx)$$

Where does this come from? The idea is to write $u(t,x) = \psi(t)\phi(x)$ where

$$\frac{\partial}{\partial t}\psi(t) = -\frac{\lambda^2}{2}\psi(t) \quad (\psi(t) = e^{-\lambda^2 t/2} \text{ works})$$

$$\Delta\phi(x) = -\lambda^2\phi(x) \quad (\phi(x) = \sin(\lambda x) \text{ works as does } \cos(\lambda x)$$

Then $u(t,x) = \psi(t)\phi(x)$ is a solution of the heat equation.

Start with $f(x) = \delta_y(x)$ being a Dirac delta function at some point $y \in (0, 2\pi)$. This means that

$$\delta_y(x) = \sum C_n \sin(nx)$$

To determine the coefficients C_n we multiply by $\sin(mx)$ and integrate:

$$\int_0^\pi \delta_y(x)\sin(mx)\,dx = \sum \int_0^\pi C_n \sin(nx)\sin(mx)\,dx$$

$$\sin(my) = \frac{\pi}{2}C_m$$

So $C_n = \frac{2}{\pi}\sin(ny)$ and

$$u(t,x) = \sum_{n=1}^{\infty} \frac{2}{\pi} e^{-tn^2/2}\sin(ny)\sin(nx)$$

The book points out that one of these terms lasts longer than the others. For large values of t, the term $e^{-tn^2/2}$ tends to be really small for larger n. So, the $n = 1$ term will dominate and we get the following approximation for large t.

$$u(t,x) \approx \frac{2}{\pi} e^{-t/2}\sin(y)\sin(x)$$

8.4.7. *example 3: general solution.* We already pointed out that we should look for solutions of the form

$$e^{-t\lambda_n^2/2}\phi_n(x)$$

where

$$\Delta\phi_n(x) = -\lambda_n\phi(x)$$

We can do the trick of multiplying by the m-th function $\phi_m(x)$ and integrating to get the coefficient C_m provided that the functions are *orthogonal* in the sense that

$$\int_a^b \phi_n(x)\phi_m(x)\,dx = 0$$

if $n \neq m$. We also need to have the complete list of functions, i.e., the only function which is orthogonal to all the $\phi_n(x)$ should be zero. In other words, we want a Hilbert space basis.

8.5. **Recurrence and transience.** The question is: Does Brownian motion make particle go off to ∞?

(1) Set up the probabilistic equation
(2) Convert to a differential equation by time reversal
(3) Solve the differential equation
(4) Reinterpret probabilitically

8.5.1. *set up.* We start at a point x which is off center between two circles (or spheres in dimensions ≥ 3)

$$\mathbf{x} \in B = \{\mathbf{x} \in \mathbb{R}^d \,|\, R_1 < ||x|| < R_2\}$$

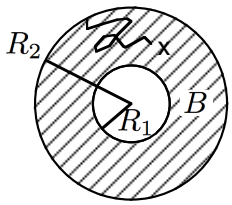

FIGURE 2. Will $\mathbf{x}$ reach the outer circle before reaching the inner circle?

Take the stopping time T to be the smallest time so that $\mathbf{X}_T \in \partial B$ given that $\mathbf{X}_0 = \mathbf{x} \in B$. We now want to know: What is the probability that $||\mathbf{X}_T|| = R_2$? The answer is

$$f(\mathbf{x}) = \mathbb{P}^{\mathbf{x}}(||\mathbf{X}_T|| = R_2) = \mathbb{E}^{\mathbf{x}}(g(\mathbf{X}_T))$$

where $g(\mathbf{y})$ is given by

$$g(\mathbf{y}) = \begin{cases} 1 & \text{if } ||\mathbf{y}|| = R_2 \\ 0 & \text{if } ||\mathbf{y}|| = R_1 \end{cases}$$

8.5.2. *differential equation.* By the time reversal argument explained last time, $f(\mathbf{x})$ is the solution of the differential equation

$$\Delta f = 0$$

on B with boundary condition

$$f(\mathbf{y}) = \begin{cases} 1 & \text{if } ||\mathbf{y}|| = R_2 \\ 0 & \text{if } ||\mathbf{y}|| = R_1 \end{cases}$$

Since everything is rotationally symmetric, we know that the solution will be a function of $||\mathbf{x}||$. It is also a function of $z = ||\mathbf{x}||^2 = \sum x_i^2$ which I much prefer since it has no nasty square roots.

$$f(\mathbf{x}) = \phi(z) = \phi(\sum x_i^2)$$

$$\frac{\partial f}{\partial x_i} = 2x_i\phi'(z)$$

$$\frac{\partial^2 f}{\partial x_i^2} = 2\phi'(z) + 4x_i^2\phi''(z)$$

Sum over all $i = 1, 2, \cdots, d$ to get

$$\Delta f(\mathbf{x}) = 2d\phi'(z) + 4z\phi''(z) = 0$$

8.5.3. *solution of diffeq.* Put $\psi(z) = \phi'(z)$. Then the equation is

$$2dim\psi(z) + 4z\psi'(z) = 0$$

where I replaced the dimension d by "dim" temporarily so that I can write this as:

$$4zd\psi/dz = -2dim\psi$$

$$\frac{d\psi}{\psi} = -\frac{dim}{2}\frac{dz}{z}$$

Integrate both sides to get

$$\ln\psi = -\frac{dim}{2}\ln z + C_0$$

or:

$$\psi = \phi' = K_0 z^{-dim/2}$$

where $K_0 = e^{C_0}$. Integrate to get $f = \phi$:

$$f(\mathbf{x}) = \phi(z) = K_0\frac{2z^{(2-d)/2}}{2-d} + C = K||\mathbf{x}||^{2-d} + C$$

if $d = dim \neq 2$ and $K = 2K_0/(2-d)$.

Now we put in the boundary conditions. First, $f(\mathbf{x}) = 0$ if $||\mathbf{x}|| = R_1$. This gives

$$C = -KR_1^{2-d}$$

The other boundary condition is $f(\mathbf{x}) = 1$ when $||\mathbf{x}|| = R_2$. This gives

$$1 = KR_2^{2-d} - KR_1^{2-d}$$

or

$$K = \frac{1}{R_2^{2-d} - R_1^{2-d}}$$

So, the solution (for $d \neq 2$) is

$$f(\mathbf{x}) = \frac{||\mathbf{x}||^{2-d} - R_1^{2-d}}{R_2^{2-d} - R_1^{2-d}}$$

If we put $d = 2$ we get $\frac{0}{0}$ and we can get the answer in the book by taking the limit as $d \to 2$ using l'Hostipal's rule. (That's called "dimensional regularization." It isn't rigorous but it works.) The answer is:

$$f(\mathbf{x}) = \frac{\ln ||\mathbf{x}|| - \ln R_1}{\ln R_2 - \ln R_1}$$

8.5.4. *interpret solution.* Remember that $f(\mathbf{x})$ is the probability that $||\mathbf{x}||$ will reach R_2 before it reaches R_1. So, we want to take the limit as $R_1 \to 0$ and $R_2 \to \infty$.

a) Take $R_1 \to 0$. When $d > 2$,

$$R_1^{2-d} = \frac{1}{R_1^{d-2}} \to \infty$$

So, $f(\mathbf{x}) \to 1$. Similarly, for $d = 2$,

$$\ln R_1 \to -\infty$$

So, $f(\mathbf{x}) \to 1$.

This means that, for $d \geq 2$, the probability is zero that the particle will ever return to the origin. When $d = 1$,

$$\lim_{R_1 \to 0} f(\mathbf{x}) = ||\mathbf{x}||/R_2 < 1$$

The particle has a chance to go to the origin and therefore it eventually will with probability one. Then it will keep coming back because it can't avoid probability one events.

b) Take $R_2 \to \infty$ When $d > 2$,

$$R_2^{2-d} = \frac{1}{R_1^{d-2}} \to 0$$

So,

$$f(\mathbf{x}) \to \frac{R_1^{2-d} - ||\mathbf{x}||^{2-d}}{R_1^{2-d}} = 1 - \left(\frac{R_1}{||\mathbf{x}||}\right)^{d-2} > 0$$

This means the particle has a chance to go to infinity. So, eventually it will with probability one. So, Brownian motion in dimensions > 2 is *transient*.

When $d = 2$,

$$\ln R_2 \to \infty$$

So,

$$f(\mathbf{x}) \to 0$$

The particle will never go to infinity. It will keep returning to the circle of radius R_1 about the origin no matter how small $R_1 > 0$ is. So, Brownian motion in $\mathbb{R}^2$ is *(neighborhood) recurrent*.

8.6. **Fractal dimension of the path.** The question is: What is the dimension of the path of Brownian motion?

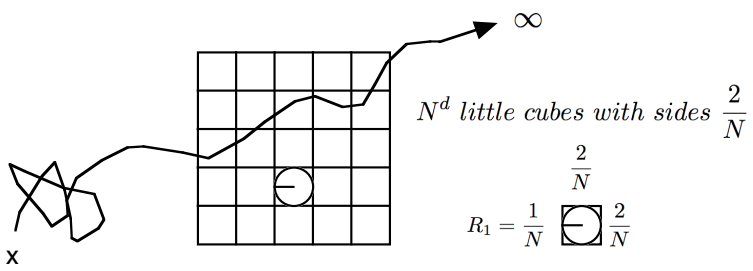

FIGURE 3. Count the number of little cubes needed to cover the path.

Take a $2 \times 2 \times \cdots \times 2$ cube in $\mathbb{R}^d$. Cut it into N^d little cubes with sides $2/N$ (so that it contains a disk of radius $1/N$). According to the definition, the box dimension of the path is given by counting the number of little squares needed to cover the path and looking at the exponent of N.

8.6.1. $d = 1$. In $\mathbb{R}^1$, the path is equal to all of the line $\mathbb{R}^1$. So, the dimension is 1.

8.6.2. $d = 2$. In $\mathbb{R}^2$ the path is dense. I.e., it gets arbitrarily close to every point. Therefore, we need all $N^d = N^2$ little cubes and the dimension is 2.

8.6.3. $d > 2$. For $d > 2$ we need to count. The expected number of little cubes that will be needed to cover the path is equal to the sum of the probability that the path will hit each little cube.

$$\mathbb{E}(\#\text{little cubes needed}) = \sum \mathbb{P}(\text{path hits one } (2/N)^d\text{-cube})$$

Since there are N^d little cubes this is approximately

$$N^d \mathbb{P}(\text{path hits one } (2/N)^d\text{-cube})$$

But we have the formula for the probability that a point will hit a sphere. So, I inserted the ratio between the volume of a cube and the volume of a ball and I got:

$$N^d \frac{2^d \Gamma\left(\frac{d+2}{2}\right)}{\pi^{d/2}} \mathbb{P}(\text{path hits ball of radius } \frac{1}{N})$$

We just calculated the probability of hitting the ball of radius $R_1 = 1/N$ before going off to infinity. (This was when we took the limit as $R_2 \to \infty$.) It was

$$\mathbb{P}(\text{hit ball of radius } \frac{1}{N}) = \left(\frac{R_1}{||x||}\right)^{d-2} = \frac{1}{||x||^{d-2}} \cdot N^{2-d}$$

So the number of cubes needed is a constant times $N^d N^{2-d} = N^2$. So the dimension of the path is 2.

8.7. **Scaling and the Cauchy distribution.** We skipped this because we talked about scaling at the beginning of the chapter and I don't think we need to know the Cauchy distribution.

8.8. **Drift.** *Brownian motion with (constant) drift* in $\mathbb{R}^d$ is given by

$$\mathbf{Y}_t = \mathbf{X}_t + \mu t$$

where $\mu \in \mathbb{R}^d$ is a vector.

Suppose we are given the information $\mathcal{F}_t$ up to time t. This is all contained in the single vector $\mathbf{Y}_t = \mathbf{x}$ in the sense that

$$\mathbb{E}^{\mathbf{x}}(-) = \mathbb{E}(- \mid \mathbf{Y}_t = \mathbf{x}) = \mathbb{E}(- \mid \mathcal{F}_t)$$

Then

$$\mathbf{Y}_{t+\delta t} = \mathbf{X}_{\delta t} + \mu \delta t + \mathbf{x}$$

where $\mathbf{X}_{\delta t}$ is a recentered standard Brownian motion.

Suppose that $f(\mathbf{x})$ is the particle density at $\mathbf{x}$ at time t. (Density is actually $f(t, \mathbf{x})$ where the time t is just not written.)

Here I converted to $d = 1$. Then

$$f(Y_{t+\delta t}) = f(x) + f'(x)(X_{\delta t} + \mu\delta t) + \frac{1}{2} f''(x)(X_{\delta t} + \mu\delta t)^2 + o((X_{\delta t} + \mu\delta t)^2)$$

I denoted the change in f by

$$\delta f = f(Y_{t+\delta t}) - f(x)$$

So, the expected value of this is

$$\mathbb{E}^x(\delta f) = f'(x)\mathbb{E}^x(X_{\delta t} + \mu\delta t) + \frac{1}{2} f''(x)\mathbb{E}^x((X_{\delta t} + \mu\delta t)^2) + o$$

The first expected value is

$$\mathbb{E}^x(X_{\delta t} + \mu\delta t) = \mu\delta t$$

Next, I used the formula $\mathbb{E}(Z^2) = \mathbb{E}(Z)^2 + Var(Z)$ to get

$$\begin{aligned}\mathbb{E}^x((X_{\delta t} + \mu\delta t)^2) &= \mathbb{E}^x(X_{\delta t} + \mu\delta t)^2 + Var(X_{\delta t} + \mu\delta t) \\ &= (\mu\delta t)^2 + Var(X_{\delta t}) + Var(\mu\delta t) \\ &= \mu^2\delta t^2 + \delta t\end{aligned}$$

and I pointed out that the term $\mu^2\delta t^2$ is negligible since it is a $o(\delta t)$. This also means that

$$o((X_{\delta t} + \mu\delta t)^2) = o(\mu^2\delta t^2 + \delta t) = o(\delta t)$$

and

$$\mathbb{E}^x(\delta f(x)) = f'(x)\mu\delta t + \frac{1}{2}f''(x)\delta t + o(\delta t)$$

Dividing by δt and taking limit as $\delta t \to 0$ we get

$$\dot{f}(x) = \mu f'(x) + \frac{1}{2}f''(x)$$

where the dot is time derivative and the primes are space derivatives. This was for $d = 1$. In higher dimensions we get

$$\frac{\partial f}{\partial t} = \sum_{i=1}^{d} \mu_i \frac{\partial f}{\partial x_i} + \frac{1}{2}\Delta f$$

MATH 56A: STOCHASTIC PROCESSES
CHAPTER 9

9. Stochastic integration

I will continue with the intuitive description of stochastic integrals that I started last week.

9.0. **the idea.** I already talked about the probabilitistic and analytic approach to Brownian motion. Stochastic integrals combine these methods. A key idea is Lévy's *quadratic variation* which is used in Kunita and Watanabe's [2] reformulation of stochastic integration.

9.0.1. *quadratic variation.* We want to define the stochastic integral

$$Z_t = \int_0^t Y_s \, dX_s$$

where X_s is Brownian motion in $\mathbb{R}^1$. However, there is a big problem because dX_s has *unbounded variation*[1]. In other words,

$$\int |dX_s| := \lim_{\delta t \to 0} \sum |X_{t_{i+1}} - X_{t_i}| = \infty.$$

Fortunately, we can still define the stochastic integral because the "quadratic variation" of X_t (denoted by $\langle X \rangle_t$) is bounded:

Theorem 9.1 (Lévy)**.**

$$\langle X \rangle_t = \int_0^t (dX_s)^2 := \lim_{\delta t \to 0} \sum (X_{t_{i+1}} - X_{t_i})^2 = t$$

with probability one.

Proof. (p. 207) It is easy to see that the quadratic variation is approximately equal to t since the summands have expected value:

$$\mathbb{E}((X_{t_{i+1}} - X_{t_i})^2) = t_{i+1} - t_i = \delta t$$

So the sum has expected value:

$$\mathbb{E}\left(\sum (X_{t_{i+1}} - X_{t_i})^2\right) = \sum \delta t = t$$

Date: December 4, 2006.

[1]The Riemann sum converges if and only the function has bounded variation.

The variance of each summand is[2]:

$$Var((X_{t_{i+1}} - X_{t_i})^2) = \mathbb{E}((X_{t_{i+1}} - X_{t_i})^4) - \mathbb{E}((X_{t_{i+1}} - X_{t_i})^2)^2 = 2\delta t^2$$

So, the sum have variance:

$$Var(\sum (X_{t_{i+1}} - X_{t_i})^2) = \sum 2\delta t^2 = 2t\delta t \to 0$$

This means that, in the limit, the sum has zero variance and is therefore not random. The value of this limit is almost sure equal to its expected value which is t. □

This theorem is usually written in the differential form

$$(dX_t)^2 = dt \tag{9.1}$$

For arbitrary increments δt of t this is

$$(\delta X_t)^2 := (X_{t+\delta t} - X_t)^2 = \delta t + o^{eff}(\delta t) \tag{9.2}$$

where I labeled the error term as an *effective little-oh*. Usual: $o(\delta t)/\delta t \to 0$ as $\delta t \to 0$. But effective little-oh means: If you take $N \approx 1/\delta t$ independent copies of $o^{eff}(\delta t)$ you get:

$$\sum_{1/\delta t \ copies} o^{eff}(\delta t) \to 0 \quad \text{as } \delta t \to 0 \tag{9.3}$$

These three equations (9.1), (9.2), (9.3) summarize the statement and proof of Lévy's theorem on quadratic variation of Brownian motion.

9.0.2. *Itō's formula.* Using quadratic variation we can "prove" Itō's formula.

Suppose that we have a particle density function $f(x)$ for $x \in \mathbb{R}$ and X_t is Brownian motion. The probabilistic argument said that we should look for the expected *present value* of f at the future position

[2]This is an easy calculation. The moment generating function for the standard normal distribution is

$$\mathbb{E}(e^{Xt}) = \int e^{xt - x^2/2} dx/\sqrt{2\pi} = e^{t^2/2} \int e^{-(x-t)^2/2} dx/\sqrt{2\pi} = e^{t^2/2}$$

The coefficient of t^{2n} in $\mathbb{E}(e^{Xt})$ is $\mathbb{E}(X^{2n})/(2n)!$ and the coefficient of t^{2n} in $e^{t^2/2}$ is $1/n!2^n$. Therefore, for $X \sim N(0,1)$,

$$\mathbb{E}(X^{2n}) = \frac{(2n)!}{n!2^n} = (2n-1)!! := 1 \cdot 3 \cdot 5 \cdot 7 \cdots (2n-1)$$

You need to multiply by σ^{2n} when $X \sim N(0, \sigma^2)$.

X_t. So, we assume that $f(x)$ is *not time dependent*. It only varies with position x. Do you remember the following formula?

$$f(X_{t+\delta t}) - f(X_t) = f'(X_t)(X_{t+\delta t} - X_t) + \frac{1}{2}f''(X_t)(X_{t+\delta t} - X_t)^2 + o(\delta t)$$

This can be abbreviated:

$$\delta f(X_t) = f'(X_t)\delta X_t + \frac{1}{2}f''(X_t)(\delta X_t)^2 + o(\delta t)$$

Use quadratic variation: $(\delta X_t)^2 = \delta t + o^{eff}(\delta t)$. Then:

$$\delta f(X_t) = f'(X_t)\delta X_t + \frac{1}{2}f''(X_t)\delta t + o^{eff}(\delta t)$$

Now take the sum from 0 to t. (We need to change t above to s so that s can be the variable going from 0 to t: $0 \le s \le t$.)

$$f(X_t) - f(X_0) = \sum f'(X_s)\delta X_s + \sum \frac{1}{2}f''(X_s)\delta s + \sum o^{eff}(\delta s)$$

Now, take the limit as $\delta s \to 0$. Then the last term goes to zero by (9.3) and we get *Itō's formula*:

$$f(X_t) - f(X_0) = \int_0^t f'(X_s)\, dX_s + \int_0^t \frac{1}{2}f''(X_s)\, ds \tag{9.4}$$

Here the stochastic integral is

$$\int_0^t f'(X_s)\, dX_s := \lim_{\delta s \to 0} \sum f'(X_s)\delta X_s$$

9.0.3. *discussion.* Why is this not a proof of Itō's formula? The main thing is that we haven't defined the stochastic integral:

$$Z_t = \int_0^t Y_s\, dX_s$$

We only showed that the traditional "limit of Riemann sum" definition makes sense and gives something which satisfies Itō's formula in the special case when $Y_t = f'(X_t)$ is the derivative of a twice differentiable function of standard Brownian motion X_t. In general we need the integral defined for *predictable* stochastic processes Y_s. This means Y_s must be $\mathcal{F}_s$-measurable and *left continuous*. Some people (e.g., our book) take Y_s to be right continuous. However, following my "bible" [4], it makes more intuitive sense to have information (X_t and $\mathcal{F}_t$) be right continuous and processes Y_t based on this information should be predictable.

9.1. **discrete stochastic integrals.** Stochastic integrals are constructed in three steps. First you have discrete time and finite state space (a finite Markov process). Then you have continuous time and finite state space (a continuous Markov chain). Then you take a limit.

The important properties of the construction are visible at each step:

- The construction is linear.
- The result is a martingale Z_t.
- $Z_t^2 - \langle Z \rangle_t$ is also a martingale where $\langle Z \rangle_t$ is the *quadratic variation* of Z_t.

Compare this with what you know about Brownian motion:

(1) X_t is a martingale.
(2) $X_t^2 - t$ is also a martingale.
(3) $\langle X \rangle_t = t$ by Lévy's theorem which we just proved.

9.1.1. *set up.* Take simple random walk on $\mathbb{Z}$. This gives a martingale X_n with $X_0 = 0$ and increments $X_{n+1} - X_n = \pm 1$ with equal probability. Suppose that Y_n is a *predictable process*, i.e., Y_n is $\mathcal{F}_{n-1}$-measurable. The discrete integral is

$$Z_n := \sum_{i=1}^{n} Y_i(X_i - X_{i-1}) = \sum_{i=1}^{n} Y_i \delta X_i$$

(This is supposed to resemble $\int Y dX$.)

The idea is that, at time n, you place a bet Y_{n+1} that X_n will increase. The money that you win or lose at that step is

$$Y_{n+1}(X_{n+1} - X_n)$$

Since you cannot see the future, Y_{n+1} is only $\mathcal{F}_n$-measurable.

9.1.2. *linearity.* This construction satisfies the following linearity condition:

$$\sum (aY_i + bV_i)\delta X_i = a \sum Y_i \delta X_i + b \sum V_i \delta X_i$$

In short, Z_n is a linear function of $\{Y_i\}$.

9.1.3. *martingale.*

Theorem 9.2. *Z_n is a martingale and $Z_0 = 0$.*

Proof. This is easy to verify:

$$\mathbb{E}(Z_n \,|\, \mathcal{F}_{n-1}) = Z_{n-1} + \mathbb{E}(Y_n(X_n - X_{n-1}) \,|\, \mathcal{F}_{n-1})$$

Since Y_n is $\mathcal{F}_{n-1}$-meaaurable, the last term vanishes:

$$\mathbb{E}(Y_n(X_n - X_{n-1}) \,|\, \mathcal{F}_{n-1}) = Y_n \mathbb{E}(X_n - X_{n-1} \,|\, \mathcal{F}_{n-1}) = 0$$

So,

$$\mathbb{E}(Z_n \,|\, \mathcal{F}_{n-1}) = Z_{n-1}$$

□

9.1.4. *quadratic variation.* The *quadratic variation* of Z_n is just the sum of squares of differences:

$$\langle Z \rangle_n := \sum_{i=1}^{n} (Z_i - Z_{i-1})^2 = \sum Y_i^2$$

since these differences are

$$Z_i - Z_{i-1} = Y_i(X_i - X_{i-1}) = \pm Y_i$$

Theorem 9.3. *Suppose that* $\mathbb{E}(Y_i^2) < \infty$ *for each* i*. Then* $(Z_n)^2 - \langle Z \rangle_n$ *is a martingale. In particular,*

$$Var(Z_n) = \mathbb{E}(Z_n^2) = \sum_{i=1}^{n} \mathbb{E}(Y_i^2)$$

Proof. The difference between Z_n^2 and the quadratic variation of Z_n is just the sum of the cross terms:

$$Z_n^2 - \langle Z \rangle_n = 2 \sum_{1 \le i < j \le n} Y_i Y_j (X_i - X_{i-1})(X_j - X_{j-1})$$

$$= Z_{n-1}^2 - \langle Z \rangle_{n-1} + 2 \sum_{i=1}^{n-1} \underbrace{Y_i Y_n (X_i - X_{i-1})}_{\mathcal{F}_{n-1}\text{-measurable}} \underbrace{(X_n - X_{n-1})}_{\mathbb{E}=0}$$

So,

$$\mathbb{E}(Z_n^2 - \langle Z \rangle_n \,|\, \mathcal{F}_{n-1}) = Z_{n-1}^2 - \langle Z \rangle_{n-1}$$

□

9.2. **Integration wrt Brownian motion.** We take W_t to be standard Brownian motion. This is also called the *Wiener process*, which might explain the use of the letter "W." We want to define the integral

$$Z_t = \int_0^t Y_s \, dW_s$$

where Y_s is a predictable process (left continuous $\mathcal{F}_t$-measurable) which we need to assume is *square summable* in the sense that

$$\int_0^t \mathbb{E}(Y_s^2) \, ds < \infty \tag{9.5}$$

for all t.

9.2.1. *simple processes.* The first step is to take a *step function* Y_t. This is also called a *simple predictable process*. The book calls it a "simple strategy" to emphasize the assumption that Y_t is given by a formula. "Simple" means that Y_t takes only a finite number of values: $0, Y_0, Y_1, \cdots, Y_n$

$$Y_t = \begin{cases} 0 & \text{if } t = 0 \\ Y_0 & \text{if } 0 < t \le t_1 \\ Y_1 & \text{if } t_1 < t \le t_2 \\ & \cdots \\ Y_n & \text{if } t_n < t \end{cases}$$

The stochastic integral is the function

$$Z_t = \int_0^t Y_s \, dW_s := \sum_{i=1}^{k} Y_{i-1}(W_{t_i} - W_{t_{i-1}}) + Y_k(W_t - W_{t_k}) \tag{9.6}$$

if $t_k < t \le t_{k+1}$.

Remark 9.4. *You can subdivide the intervals $(t_{i-1}, t_i]$ and the integral Z_t remains the same. For example, if you insert $t_{3/2}$ between t_1 and t_2 and put $Y_{3/2} = Y_1$ then*

$$Y_1(W_2 - W_1) = Y_1(W_{3/2} - W_1) + Y_{3/2}(W_2 - W_{3/2})$$

So the sum (9.6) remains the same after subdivision.

I want to go over the basic properties. Maybe I won't prove them.

(1) $Z_t = \int_0^t Y_s \, dW_s$ is linear in Y_t. I.e.,

$$\int_0^t (aX_s + bY_s) \, dW_s = a \int_0^t X_s \, dW_s + b \int_0^t Y_s \, dW_s$$

(2) Z_t is a martingale which is *square summable*, i.e., $\mathbb{E}(Z_t^2) < \infty$.

(3) $Z_t^2 - \langle Z \rangle_t$ is a martingale.

(4)

$$\mathbb{E}(Z_t^2) = \int_0^t \mathbb{E}(Y_s^2)\, ds$$

(So, Z_t is square summable if and only if Y_t is square summable.)

Here the *quadratic variation* $\langle Z \rangle_t$ is given by

$$\langle Z \rangle_t = \int_0^t Y_s^2\, ds$$

So, (3) $\Rightarrow$ (4): If $Z_t^2 - \langle Z \rangle_t$ is a martingale, then

$$\mathbb{E}(Z_t^2 - \langle Z \rangle_t) = Z_0^2 - \langle Z \rangle_0 = 0$$

So,

$$\mathbb{E}(Z_t^2) = \mathbb{E}(\langle Z \rangle_t) = \mathbb{E}\left(\int_0^t Y_s^2\, ds\right) = \int_0^t \mathbb{E}(Y_s^2)\, ds$$

Now, I am going to verify properties (2) and (3) (at least on paper). The key point is that *all cross terms have expectation zero.*

9.2.2. *vanishing expectation of cross terms.*

Theorem 9.5. *Z_t is a martingale for simple processes.*

Proof. The definition of a martingale is that $\mathbb{E}(Z_t \,|\, \mathcal{F}_s) = Z_s$ if $s < t$. By subdividing the interval, we can always assume that s is a jump time (Remark 9.4). By induction it suffices to show this when $s = t_k$ is the last jump time before t. In other words, we have to show that the last term in Equation (9.6) has expectation 0:

$$\mathbb{E}(Y_k(W_t - W_{t_k}) \,|\, \mathcal{F}_{t_k}) = Y_k \mathbb{E}(W_t - W_{t_k} \,|\, \mathcal{F}_{t_k}) = 0$$

The point is that this is a product where the first factor Y_k is determined when the second factor is still random with zero expectation. □

And now, here is a wonderful theorem that will save us a lot of time:

Theorem 9.6 (Meyer)**.** *Suppose that Z_t is a square summable martingale wrt $\mathcal{F}_t$ with $Z_0 = 0$. Then $Z_t^2 - \langle Z \rangle_t$ is also a martingale.*

In other words, (2) $\Rightarrow$ (3)!

Proof. The idea is summarized in the following motto (from [1], 1.5.8). *"When squaring sums of martingale increments and taking the expectation, one can neglect the cross-product terms."*

This theorem is supposed to proved property (3) in all cases simultaneously. So, Z_t could be anything. However, we can always subdivide

the interval $[0,t]$ into parts of length δt and get Z_t as a sum of increments:

$$Z_t = Z_t - Z_0 = \sum Z_{t_i} - Z_{t_{i-1}} = \sum \delta_i Z_t$$

The increments $\delta_i Z_t = Z_{t_i} - Z_{t_{i-1}}$ have expectation zero since Z_t is a martingale. When you square Z_t you get:

$$Z_t^2 = \sum (\delta_i Z_t)^2 + 2\sum_{i<j} \delta_i Z_t \delta_j Z_t$$

The sum of squares converges to the quadratic variation by definition:

$$\langle Z \rangle_t := \lim_{\delta t \to 0} \sum (\delta_i Z_t)^2$$

and the cross terms $\delta_i Z_t \delta_j Z_t$ have expectation zero because the first term is determined when the second term is random with expectation zero.

$$\mathbb{E}(\delta_i Z_t \delta_j Z_t \,|\, \mathcal{F}_{t_{j-1}}) = \delta_i Z_t \mathbb{E}(\delta_j Z_t \,|\, \mathcal{F}_{t_i}) = 0$$

and by the rule of iterated expectation,

$$\mathbb{E}(\delta_i Z_t \delta_j Z_t \,|\, \mathcal{F}_s) = \mathbb{E}(\mathbb{E}(\delta_i Z_t \delta_j Z_t \,|\, \mathcal{F}_{t_{j-1}}) \,|\, \mathcal{F}_s) = 0$$

for any $s \le t_{j-1}$. □

9.2.3. *general stochastic integral (wrt W_t).* Now suppose that Y_t is any (square summable) predictable process. Then we convert to a simple process by letting $Y_s^{(n)}$ be the average value of Y_t over the inverval $(\frac{k-1}{n}, \frac{k}{n}]$ if s lies in the next interval $(\frac{k}{n}, \frac{k+1}{n}]$. This is to insure that, at time $t = k/n$ when we choose $Y_s^{(n)}$, we only use information from the past and not from the future, i.e., it is predictable.

Since $Y_s^{(n)}$ is a simple predictable process, we can define

$$Z_t^{(n)} := \int_0^t Y_s^{(n)}\, dW_s$$

Without saying it, the book is using the following well-known theorem in real analysis applied to the measure space $(\Omega, \mathcal{F}_t, \mathbb{P})$.

Theorem 9.7. *The space of square summable real valued functions on any measure space is complete in the L^2 metric.*

The L^2 metric is just

$$||Z_t||_2 := \mathbb{E}(Z_t^2)$$

The martingales $Z_t^{(n)}$ form a Cauchy sequence in the L^2 norm. I.e.,

$$\mathbb{E}((Z_t^{(n)} - Z_t^{(m)})^2) = \mathbb{E}(\langle Z^{(n)} - Z^{(m)} \rangle_t) = \int_0^t (Y_s^{(n)} - Y_s^{(m)})^2\, ds \to 0$$

as $n, m \to \infty$. The book then uses the theorem about completeness of L^2 to concluded that the martingales $Z_t^{(n)}$ converge to some square summable process Z_t

$$Z_t = \int_0^t Y_s \, dW_s := \lim Z_t^{(n)}$$

Since the limit of martingales is a martingale, Z_t is a martingale. By Theorem 9.6 that is all we have to show (linearity being obvious).

9.3. **Itō's formula.** I will repeat the formula and do the examples in the book but I won't go over the proof since we already did it.

Theorem 9.8 (Itō's first formula)**.** *Suppose that W_t is standard Brownian motion and $f(x)$ is C^2, i.e., twice continuously differentiable. Then*

$$f(W_t) - f(W_0) = \int_0^t f'(W_s) \, dW_s + \frac{1}{2} \int_0^t f''(W_s) \, ds$$

I just want to point out that the naïve definition of the stochastic integral that we used in the earlier proof of this formula is equivalent to the rigorous definition that I just explained because $Y_t = f'(W_t)$ is a continuous function of t. Continuity implies that the average value over an interval (used in the rigorous definition) converges to the actual value at one end (used in our naïve definition).

9.3.1. *example 1.* Let $f(t) = t^2$. Then $f'(t) = 2t$ and $f''(t) = 2$. So,

$$f(W_t) - f(W_0) = W_t^2 = \int_0^t 2W_s \, dW_s + \frac{1}{2} \int_0^t 2 \, ds$$

$$\int_0^t W_s \, dW_s = \frac{1}{2} W_t^2 - \frac{1}{2} t$$

9.3.2. *geometric Brownian motion.* Take $f(t) = e^t = f'(t) = f''(t)$. Then Itō's formula is:

$$e^{W_t} - 1 = \int_0^t e^{W_s} \, dW_s + \frac{1}{2} \int_0^t e^{W_s} \, ds$$

If we write $X_t := e^{W_t}$ this becomes:

$$X_t - 1 = \int_0^t X_s \, dW_s + \frac{1}{2} \int_0^t X_s \, ds$$

or

$$dX_t = X_t \, dW_t + \frac{1}{2} X_t \, dt$$

9.4. **Extensions of Itō's formula.** The key ideas are covariation and the product rule.

9.4.1. *covariation.* This is also called the *covariance process.*

Definition 9.9. *The* covariation *of A_t and B_t is defined to be*

$$\langle A, B\rangle_t := \lim_{\delta t\to 0} \sum \delta_i A\, \delta_i B$$

$$= \lim_{\delta t\to 0} \sum (A_{t_i} - A_{t_{i-1}})(B_{t_i} - B_{t_{i-1}})$$

Properties

$$\langle A, A\rangle_t = \langle A\rangle_t \quad \text{(quadratic variation)}$$

$$d\langle A, B\rangle_t = dA_t dB_t \quad \text{(by definition)}$$

$$\langle A + B\rangle + t = \langle A\rangle_t + \langle B\rangle_t + 2\langle A, B\rangle_t$$

Quick proof:

$$\sum (\delta_i A + \delta_i B)^2 = \sum (\delta_i A)^2 + \sum (\delta_i B)^2 + 2\sum \delta_i A \delta_i B$$

9.4.2. *product rule.* The following formula holds without error. (See picture.)

$$\delta AB = A\delta B + B\delta A + \delta A\delta B$$

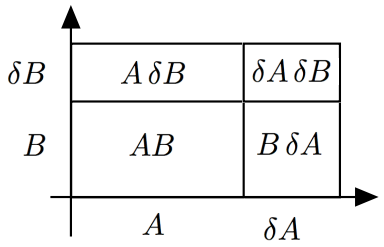

FIGURE 1. The term $\delta A\delta B$ becomes the covariation.

The infinitesmal version is

$$dAB = AdB + BdA + d\langle A, B\rangle_t$$

Example 9.10. *This is example 1 on p.209 which I also did at the end of the last section although I didn't have time to do it in class. Let* $A_t = B_t = W_t$. *Then*

$$d\langle W, W\rangle_t = d\langle W\rangle_t = dt$$

So, the product rule gives:

$$dW_t^2 = 2W_t\, dW_t + dt$$

9.4.3. *quadratic variation of* Z_t. is equal to the quadratic variation of the stochastic part of Z_t.

Lemma 9.11 (Lemma 1)**.** *If* f *is continuous with bounded variation (e.g. if* f *is differentiable) then*

$$\langle f\rangle_t = 0$$

(The quadratic variation of f *is zero.)*

Proof.

$$\langle f\rangle_t = \lim_{\substack{\delta t \to 0 \\ \delta f \to 0}} \sum (\delta f)^2 = \lim_{\delta f \to 0} \underbrace{|\delta f|}_{\to 0} \underbrace{\sum |\delta f|}_{\text{variation is bdd}} = 0 \cdot bdd = 0$$

□

Lemma 9.12 (Lemma 2)**.** $\langle f\rangle = 0 \Rightarrow \langle f, X\rangle = 0\ \forall X$.

Proof. If $\langle f, X\rangle_t > 0$ then

$$\langle X - af\rangle_t = \langle X\rangle_t - 2a \underbrace{\langle f, X\rangle_t}_{\text{fixed} > 0} + \underbrace{a^2 \langle f\rangle_t}_{0}$$

If we make a really big then we can make $\langle X - af\rangle_t < 0$. But this is impossible because quadratic variations are sums of squares! □

Here is the theorem we need:

Theorem 9.13. *Suppose that*

$$Z_t = \int_0^t X_s\, ds + \int_0^t Y_s\, dW_s$$

where X_s *integrable (i.e.,* $\int_0^t |X_s|\, ds < \infty$*). The second integral is the "stochastic part" of* Z_t. *Written infinitesmally:*

$$dZ_t = X_t\, dt + Y_t\, dW_t$$

The theorem is:

$$d\langle Z\rangle_t = Y_t^2 dt$$

Proof. Let $f = \int_0^t X_s\,ds$ and $g = \int_0^t Y_s\,dW_s$. Then $Z_t = f + g$. So, using the properties of covariation,

$$\langle Z\rangle_t = \langle f\rangle + \langle g\rangle + 2\langle f, g\rangle$$

But, $\langle f\rangle = 0$ by Lemma 1 (9.11) since f has bounded variation. And $\langle f, g\rangle = 0$ by Lemma 2 (9.12). So, $\langle Z\rangle_t = \langle g\rangle$ and

$$d\langle Z\rangle_t = (Y_t\,dW_t)^2 = Y_t^2\,dt$$

□

9.4.4. *Itō's second formula.* Suppose that $f(x)$ is C^2, i.e., twice continuously differentiable. Then, the Taylor series of f gives is

$$f(x+\delta x) - f(x) = f'(x)\delta x + \frac{1}{2}f''(x)(\delta x)^2 + o((\delta x)^2) \tag{9.7}$$

Now, substitute $x = Z_t$ where Z_t is as in the theorem above.

$$f(Z_{t+\delta t}) - f(Z_t) = f'(Z_t)\,\delta Z_t + \frac{1}{2}f''(Z_t)(\delta Z_t)^2 + o$$

The infinitesmal version is:

$$df(Z_t) = f'(Z_t)\,dZ_t + \frac{1}{2}f''(Z_t)\,d\langle Z\rangle_t$$

Substitute $dZ_t = X_t\,dt + Y_t\,dW_t$ and $d\langle Z\rangle_t = Y_t^2 dt$ and we get:

Theorem 9.14 (Itō II)**.**

$$df(Z_t) = f'(Z_t)X_t\,dt + f'(Z_t)Y_t\,dW_t + \frac{1}{2}f''(Z_t)Y_t^2\,dt$$

9.4.5. *Itō's third formula.* For this we need the vector version of the Taylor series (9.7) and we need to apply it to $x = (t, Z_t)$. Then f' becomes the *gradient* $\nabla f = (f_1, f_2)$ and

$$f'(x)\,dx = (f_1, f_2)\begin{pmatrix} dt \\ dZ_t\end{pmatrix} = \dot{f}(t, Z_t)\,dt + f'(t, Z_t)\,(X_t\,dt + Y_t\,dW_t)$$

and f'' becomes the *Hessian* $D^2 f = \begin{pmatrix} f_{11} & f_{12} \\ f_{21} & f_{22}\end{pmatrix}$. So,

$$\frac{1}{2}f''(x)(dx)^2 = (dt, dZ_t)\begin{pmatrix} f_{11} & f_{12} \\ f_{21} & f_{22}\end{pmatrix}\begin{pmatrix} dt \\ dZ_t\end{pmatrix}$$

$$= \frac{1}{2}f_{11}\underbrace{d\langle t\rangle}_{=0} + f_{12}\underbrace{d\langle t, Z_t\rangle}_{=0} + \frac{1}{2}f_{22}d\langle Z\rangle_t = \frac{1}{2}f''(t, Z_t)Y_t^2 dt$$

Putting these together we get:

Theorem 9.15 (Itō III)**.**

$$df(t, Z_t) = \dot{f}(t, Z_t)\, dt + f'(t, Z_t)\,(X_t\, dt + Y_t\, dW_t) + \frac{1}{2} f''(t, Z_t) Y_t^2 dt$$

At this point we jumped ahead to the last section 9.8 on the Black-Scholes formula.

9.5. **Continuous martingales.** We skipped this section. But this looks like a good place for me to put the proof of Lévy's theorem which implies that all continuous martingales are reparametrizations of Brownian motion.

Theorem 9.16 (Lévy)**.** *A continuous L^2 martingale M_t starting at 0 is standard Brownian motion if and only if $M_t^2 - t$ is a martingale.*

What follows is from Kunita and Watanabe [2] which is considered to be the most elegant. I also included proofs of the necessary lemmas.

9.5.1. *first step.*

Lemma 9.17. *Assuming the conditions of Theorem 9.16, $\langle M \rangle_t = t$.*

The proof of this lemma uses Meyer's theorem proved in class that

Theorem 9.18 (Meyer)**.** *If M_t is a continuous L^2 martingale then $\langle M \rangle_t$ is the unique continuous increasing process starting at 0 so that*

$$M_t^2 - \langle M \rangle_t$$

is a martingale.

Except that I didn't prove the uniqueness and I didn't define "increasing process."

9.5.2. *uniqueness of increasing process.*

Definition 9.19. *X_t is called an* increasing process *if*

$$t > s \Rightarrow X_t \geq X_s \quad a.s.$$

(1) Clearly, $X_t = t$ is a continuous increasing process.
(2) $\langle M \rangle_t$ is an increasing process starting at 0 since

$$\langle M \rangle_t - \langle M \rangle_s = \lim_{\delta t \to 0} \sum (M_{t_i} - M_{t_{i-1}})^2 \geq 0$$

And $\langle M \rangle_t$ is continuous if M_t is continuous and square summable (by definition of square summable).

Proof of uniqueness part of Meyer's Theorem. Suppose that A_t, B_t are continuous increasing processes starting at 0 and

$$M_t^2 - A_t, \quad M_t^2 - B_t$$

are martingales. Then the difference

$$A_t - B_t$$

is also a continuous martingale starting at 0 with bounded variation (for t bounded). By the following lemma this implies that $A_t = B_t$. So, the continuous increasing process that we need to subtract from M_t^2 to make it into a martingale is unique. □

Lemma 9.20. *Suppose M_t is a continuous martingale starting at 0 and M_t has bounded variation for bounded t. Then $M_t = 0$ for all t.*

Proof. By Lemma 1 (9.11), this implies that the quadratic variation of M_t is identically zero: $\langle M \rangle_t = 0$. Therefore, M_t^2 is also a martingale by the first part of Meyer's theorem that we already proved in (9.6). But $M_t^2 \geq 0$. So, $\mathbb{E}(M_t) = 0$ only if $M_t = 0$ almost surely. □

9.5.3. *Kunita-Watanabe.* One of the main results of [2] was to generalize Itō's formula to the case of L^2 martingales. Or perhaps it would be more fair to say that they formulated the theory of stochastic integrals in such a way that it easily extends to this case.

Theorem 9.21 (Kunita-Watanabe)**.** *If M_t is a continuous L^2 martingale and f is C^2 then*

$$f(M_t) - f(M_0) = \int_0^t f'(M_s)\, dM_s + \frac{1}{2}\int_0^t f''(M_s)\, d\langle M \rangle_s$$

Now we can prove Lévy's theorem. Suppose that $M_t^2 - t$ is a martingale. We know that $M_t^2 - \langle M \rangle_t$ is also a martingale. By Meyer's uniqueness theorem we can conclude that $\langle M \rangle_t = t$.

Now, let $f(x) = e^{ixz}, f'(x) = ize^{ixz}, f''(x) = -z^2 e^{ixz}$ where z is a formal variable. Then $f(M_t) = e^{iM_t z}$ and

$$\mathbb{E}(e^{iM_t z}) = 1 + i\mathbb{E}(M_t)z - \frac{1}{2}\mathbb{E}(M_t^2)z^2 - \frac{i}{3!}\mathbb{E}(M_t^3)z^3 + \frac{1}{4!}\mathbb{E}(M_t^4)z^4 + \cdots$$

gives all of the moments of M_t. So, it suffices to show that these moments are what they should be if M_t were Brownian motion. I.e., it suffices to show that

$$\mathbb{E}(e^{iM_t z}) = e^{-tz^2/2}$$

But the Kunita-Watanabe variation of Itō's formula gives:

$$\mathbb{E}(f(M_t) - f(M_0)) = \frac{1}{2}\int_0^t \mathbb{E}(f''(M_s))\,ds$$

since $d\langle M\rangle_s = ds$ and since anything predictable (like $f(M_s)$) times dM_s has expectation 0. Since $f(M_0) = e^0 = 1$ and $f''(M_s) = -z^2 e^{iM_t z}$ we have

$$\mathbb{E}(e^{iM_t z}) - 1 = \frac{1}{2}\int_0^t -z^2\mathbb{E}(e^{iM_s z})\,ds$$

Let $h(t) = \mathbb{E}(e^{iM_t z})$. Then

$$h(t) - 1 = \frac{-z^2}{2}\int_0^t h(s)\,ds$$

Differentiate both sides:

$$h'(t) = \frac{-z^2}{2}h(t)$$

This is just exponential growth: $h(t) = h(0)e^{-tz^2/2}$. But $h(0) = 1$ since $M_0 = 0$. So,

$$h(t) = \mathbb{E}(e^{iM_t z}) = e^{-tz^2/2}$$

as claimed.

9.6. **Girsanov transformation.** We skipped this section.

9.7. **Feynman-Kac.** The formula of Feynman and Kac gives another way to solve the Black-Scholes equation. First we need to understand how bonds grow if their rates are variable.

9.7.1. *variable bond rate.* Suppose the bond rate changes with time: $r(t)$.Then the value of your bonds will grow by

$$dY_t = r(t)Y_t\,dt \tag{9.8}$$

$$Y_t = Y_0\,exp\left(\int_0^t r(s)\,ds\right) \tag{9.9}$$

Why is (9.9) the solution of (9.8)?

$$Y_{t+dt} = Y_0\,exp\left(\int_0^t r(s)\,ds + \int_t^{t+dt} r(s)\,ds\right)$$

$$= Y_t\,exp(r(t)dt) = Y_t(1 + r(t)dt + \underbrace{r(t)^2dt^2/2 + \cdots}_{=0})$$

$$dY_t = Y_t r(t)dt$$

If we solve (9.9) for Y_0 we get:

$$Y_0 = Y_t\,exp\left(\int_0^t -r(s)\,ds\right)$$

9.7.2. *the stochastic process.* Now suppose we have a stochastic process Z_t satisfying the stochastic differential equation:

$$dZ_t = a(Z_t)dt + b(Z_t)dW_t$$

Lemma 9.22. *Z_t is a martingale if and only if $a(Z_t)$ is identically zero a.s.*

Proof. Z_t is a martingale iff $\mathbb{E}(dZ_t \,|\, \mathcal{F}_t) = a(Z_t)dt = 0$ □

Let J_t be given by

$$J_t := exp\left(\int_0^t -r(s, Z_s)\,ds\right)$$

This is how much one dollar at time t was worth at time 0 if the bond rate depends on time and on Z_t.

9.7.3. *the payoff.* If the payoff function is $f(Z_T)$ then the value at time 0 of this payoff is

$$g(Z_T)J_T$$

How much of this value is determined by time t $(0 \le t \le T)$?

$$J_T = exp\left(\int_0^T -r(s, Z_s)\, ds\right)$$

$$= exp\left(\int_0^t -r(s, Z_s)\, ds + \int_t^T -r(s, Z_s)\, ds\right)$$

$$= J_t\, exp\left(\int_t^T -r(s, Z_s)\, ds\right)$$

9.7.4. *the martingale.*

$$M_t := \mathbb{E}(g(Z_T)J_T \,|\, \mathcal{F}_t)$$

$$= J_t \underbrace{\mathbb{E}(g(Z_T)exp\left(\int_t^T -r(s, Z_s)\, ds\right) \,|\, \mathcal{F}_t)}_{V(t,Z_t)}$$

$$= J_t V(t, Z_t)$$

with $M_T = g(Z_T)$. This is a martingale by the law of iterated expectation:

$$\mathbb{E}(M_t \,|\, \mathcal{F}_s) = \mathbb{E}(\mathbb{E}(g(Z_T)J_T \,|\, \mathcal{F}_t) \,|\, \mathcal{F}_s) = \mathbb{E}(g(Z_T)J_T \,|\, \mathcal{F}_s) = M_s$$

Take the differential:

$$dM_t = dJ\, V + J\, dV$$

$$= J_t(-r(t, Z_t)V\, dt + J_t\left(\dot{V}\, dt + V'\, dZ_t + \frac{1}{2}V''\, d\langle M\rangle_t\right)$$

$$= J_t(-r(t, Z_t)V\, dt + J_t\left(\dot{V}\, dt + V'a\, dt + V'b\, dW_t + \frac{1}{2}V''b^2\, dt\right)$$

Since M_t is a martingale, the coefficient of dt must be zero (Lemma 9.22) Therefore, if $x = Z_t$then

$$-r(t, x)V(t, x) + \dot{V}(t, x) + a(x)V'(t, x) + \frac{b^2(x)}{2}V''(t, x) = 0 \tag{9.10}$$

$$V(T, x) = g(x) \tag{9.11}$$

Theorem 9.23. *The solution of the stochastic differential equation (9.10) with boundary condition (9.11) is given by*

$$V(t,x) = \mathbb{E}^x(g(Z_T)exp\left(\int_t^T -r(s,Z_s)\,ds\right))$$

where

$$dZ_t = a(Z_t)\,dt + b(Z_t)\,dW_t$$

9.7.5. *application to Black-Scholes.* If you apply this to the Black-Scholes equation (9.12) you get

$$r(t,x) = r$$
$$a(x) = rx$$
$$b(x) = \sigma x$$

So,

$$V(t,x) = \mathbb{E}^x((Z_T - K)_+ e^{-r(T-t)})$$
$$dZ_t = rZ_t\,dt + \sigma Z_t\,dW_t$$

These equations say that *the fair price of the option at time t is equal to the expected value of the option at time T adjusted for inflation assuming the stock has drift equal to the bond rate.*

9.8. **Black-Scholes.**

9.8.1. *the set-up.*

$$\begin{array}{ll} S_t & = \text{ value of one share of stock at time } t \\ dS_t & = \mu S_t dt + \sigma S_t\, dW_t \quad (= \text{return}) \\ \mu & = \text{drift (constant in this model)} \\ \sigma & = \text{volatility (also constant)} \\ W_t & = \text{standard Brownian motion} \end{array}$$

We are looking at a *European call option*. This is an option to buy one share of stock at price K at time T.

$$\begin{array}{ll} K & = \text{exercise price} = \text{strike price} \\ T & = \text{expiry date} \\ V(t,x) & := \text{fair price of the option at time } t \text{ given that } S_t = x \end{array}$$

We want to calculate $V(t, S_t)$. We know how much it will be worth at expiry:

$$V(T, x) = (x - K)_+ := \max(x - K, 0)$$

But how much is the option worth today?

9.8.2. *replicating portfolio.* The theory is that there is a portfolio O_t whose value at time T will be exactly $(S_T - K)_+$. Then the value of the option should be the present value of the portfolio. Black and Scholes assumed that there are no *arbitrage* opportunities. This implies that O_t is unique. In [1] it says that there are always sure ways to lose money. So, they don't assume that the value of O_t is unique. Instead it is proved that the fair price of the option is equal to the value of the *cheapest replicating portfolio*. Who is right is a matter of debate.

Fortunately, in this case, the replicating portfolio (also called a *hedging strategy*) is unique.

Our portfolio is just a combination of stocks and bonds but we only have one stock and one bond to choose from in this model. So,

$$O_t = X_t S_t + Y_t$$

$$\begin{array}{ll} X_t & = \text{number of shares of stock} \\ Y_t & = \text{money in bonds} \\ r & = \text{bond rate (assumed constant)} \end{array}$$

The problem is to find X_t and Y_t so that $O_t = V(t, S_t)$ for all $t \le T$.

Note: O_t needs to be *self-financing*. This means that we have to continuously reinvest all profits. So, O_t grows in two steps:

(a) Stocks and bonds increase in value. ($\Rightarrow$ more \$\$)
(b) You immediately reinvest the profit. (zero net gain in this step)

Step (a):

$$dO_t = X_t dS_t + rY_t\,dt$$

Step (b): Change X_t and Y_t by dX_t, dY_t so that there is no net change over what you got in step (a). Using the product rule this means:

$$dO_t = X_t dS_t + rY_t\,dt = S dX_t + X_t dS_t + d\langle X, S\rangle_t + dY_t$$

We need to have $V(t, S_t) = O_t$. Using Itō's third formula we get:

$$dV(t,S_t) = \dot{V}(t,S_t)\,dt + V'(t,S_t)\,dS_t + \frac{1}{2}V''(t,S_t)\sigma^2 S_t^2\,dt$$

$$= \dot{V}(t,S_t)\,dt + V'(t,S_t)\mu S_t dt + \underbrace{V'(t,S_t)\sigma S_t dW_t}_{\text{stochastic part}} + \frac{1}{2}V''(t,S_t)\sigma^2 S_t^2\,dt$$

If this is equal to

$$dO_t = X_t dS_t + rY_t\,dt = X_t\mu S_t dt + \underbrace{X_t\sigma S_t dW_t}_{\text{stochastic part}} + rY_t\,dt$$

then the stochastic parts must be equal. So,

$$X_t = V'(t, S_t)$$

Some people say it this way: If you are holding the option $V(t, S_t)$ and you sell off $X_t = V'(t, S_t)$ number of shares of stock then you have hedged away all of your risk (since the stochastic parts will cancel). Therefore, the financial instrument that you are holding: $V(t, S_t) - X_t S_t$ must be increasing in value at the bond rate which is exactly the case.

Now go back to the equation $dV(t, S_t) = dO_t$ and cross out the terms which are equal ($X_t dS_t$ and $V'(t, S_t)dS_t$) and divide by dt. Then we have:

$$rY_t = \dot{V}(t,S_t) + \frac{1}{2}V''(t,S_t)\sigma^2 S_t^2$$

But we also know that

$$Y_t = O_t - X_t S_t = V(t,S_t) - V'(t,S_t)S_t$$

So,

$$\dot{V}(t,S_t) + \frac{1}{2}V''(t,S_t)\sigma^2 S_t^2 - rV(t,S_t) + rV'(t,S_t)S_t = 0$$

which we can rewrite as:

$$\dot{V}(t,x) + \frac{1}{2}\sigma^2 x^2 V''(t,x) + rxV'(t,x) - rV(t,x) = 0 \tag{9.12}$$

This is the *Black-Scholes equation*. This can be solved using Feynman-Kac. But it can also be solved directly using some tricks.

9.8.3. *simplification of Black-Scholes equation.* The first step is to notice that the drift μ does not matter. It is not part of Equation (9.12). Therefore, we may assume that $\mu = 0$ and

$$dS_t = \sigma S_t \, dW_t \tag{9.13}$$

The next step is: We may assume that $r = 0$. This is the same as measuring value in terms of value at time T adjusted for inflation. When $r = 0$ the equation becomes:

$$\dot{V}_0(t,x) + \frac{1}{2}\sigma^2 x^2 V_0''(t,x) = 0 \tag{9.14}$$

When I say "we may assume $r = 0$" I mean that, if you can solve the $r = 0$ equation then you can also solve the original equation. Suppose $V_0(t,x)$ is the solution of this equation with boundary condition $V_0(T,x) = (x-K)_+$. Then the solution of the original equation (9.12) is

$$V(t,x) = e^{-r(T-t)} V_0(t, e^{r(T-t)}x) \tag{9.15}$$

The reason is that V_0 is in terms of time T dollars and x dollars today (at time t) is worth $e^{r(T-t)}x$ dollars at time T and the output V_0 is in terms of time T dollars. So, V_0 of those dollars is worth $e^{-r(T-t)}V_0$ in today's dollars.

If you don't believe it, you can just differentiate the expression in (9.15):

$$\begin{aligned} \dot{V} &= rV + e^{-r(T-t)}\dot{V}_0 - rxV_0' \\ V &= e^{-r(T-t)}V_0 \\ V' &= V_0' \\ V'' &= e^{r(T-t)}V_0'' \end{aligned}$$

Then

$$\begin{aligned} &\dot{V} + \frac{1}{2}\sigma^2 x^2 V'' + rxV' - rV \\ &= (rV + e^{-r(T-t)}\dot{V}_0 - rxV_0') + \frac{1}{2}\sigma^2 x^2 e^{r(T-t)}V_0'' + rxV' - rV \\ &= e^{-r(T-t)}\dot{V}_0 + \frac{1}{2}\sigma^2 x^2 e^{r(T-t)}V_0'' = 0 \end{aligned}$$

9.8.4. *solution of Black-Scholes.* Remember that $\mu = 0 = r$. Thus

$$dS_t = \sigma S_t \, dW_t \tag{9.16}$$

$$dV_0(t, S_t) = dO_t = X_t dS_t + rY_t = X_t \sigma S_t \, dW_t$$

So, $S_t, O_t, V_0(t, S_t)$ are all martingales. Therefore,

$$V_0(t, S_t) = \mathbb{E}(V_0(T, S_T) \mid \mathcal{F}_t)$$

But T is the payoff date. So, we get

$$V_0(T, x) = g(x) = (x - K)_+ = \max(x - K, 0)$$

where $g(x)$ represents the payoff function. So, the value of the option is just the expected payoff which will depend on the value of the stock at time T. So, we need to calculate S_T.

Equation (9.16) can be solved for S_t as follows.

$$\begin{aligned} d \ln S_t &= \frac{dS_t}{S_t} + \frac{1}{2}\left(\frac{-1}{S_t^2}\right)\sigma^2 S_t^2 dt \\ &= \sigma \, dW_t - \frac{\sigma^2}{2} \, dt \end{aligned}$$

So,

$$\ln S_t = \sigma W_t - \frac{\sigma^2 t}{2} + C$$

Plugging in $t = 0$ gives $C = \ln S_0$. So,

$$S_t = S_0 exp\left(\sigma W_t - \frac{\sigma^2 t}{2}\right)$$

Now attach a little timer on the stock certificate and set that timer to zero at time $T - t$. When the timer reaches t the expiry date has arrived:

$$\begin{aligned} V_0(T - t, x) &= \mathbb{E}(g(S_t) \mid S_0 = x) \\ &= \mathbb{E}\left(g(x \, e^{\sigma W_t - \sigma^2 t/2})\right) \end{aligned}$$

Letting $y = x \, e^{-\sigma^2 t/2}$ we get:

$$V_0(T - t, y \, e^{\sigma^2 t/2}) = \mathbb{E}\left((y \, e^{\sigma W_t} - K)_+\right)$$

But $\sigma W_t \sim N(0, \sigma^2 t)$. So we can use the following lemma with $b = \sigma\sqrt{t}, a = y = x \, e^{-\sigma^2 t/2} = xe^{-b^2/2}$.

Lemma 9.24. *Suppose that $X = ae^{bZ}$ where $Z \sim N(0, 1)$. Then the expected value of $g(X) = (X - K)_+$ is*

$$\mathbb{E}((X - K)_+) = ae^{b^2/2}\Phi\left(\frac{\ln(a/K) + b^2}{b}\right) - K\Phi\left(\frac{\ln(a/K)}{b}\right)$$

where Φ is the distribution function for $N(0, 1)$.

Theorem 9.25. *The value of the European call option at time $T-t$ is*

$$V_0(T-t,x) = x\Phi\left(\frac{\ln(x/K)+\sigma^2 t/2}{\sigma\sqrt{t}}\right) - K\Phi\left(\frac{\ln(x/K)-\sigma^2 t/2}{\sigma\sqrt{t}}\right)$$

if $S_t = x$ assuming the bond rate is zero and

$$V(T-t,x) = e^{-rt}V_0(t,e^{rt}x)$$

$$= x\Phi\left(\frac{\ln(x/K)+rt+\sigma^2 t/2}{\sigma\sqrt{t}}\right) - e^{-rt}K\Phi\left(\frac{\ln(x/K)+rt-\sigma^2 t/2}{\sigma\sqrt{t}}\right)$$

if the bond rate is a constant r.

Proof of Lemma 9.24. The expected value is given by an integral which is easy to compute:

$$\mathbb{E}((ae^{bZ}-K)_+) = \int_{-\infty}^{\infty}(ae^{bz}-K)_+\phi(z)\,dz$$

Since $ae^{bz} = K$ when $z = \frac{1}{b}\ln(K/a)$ we can write this as:

$$\int_{\frac{1}{b}\ln(K/a)}^{\infty} ae^{bz-z^2/2} - Ke^{-z^2/2}\,\frac{dz}{\sqrt{2\pi}}$$

For the first part you can change variables by $z = y+b, dz = dy$ to get

$$\int_{\frac{1}{b}\ln(K/a)}^{\infty} ae^{bz-z^2/2}\,\frac{dz}{\sqrt{2\pi}} = \int_{\frac{1}{b}\ln(K/a)-b}^{\infty} ae^{b^2/2-y^2/2}\frac{dy}{\sqrt{2\pi}}$$

Changing y to $-y$ this gives

$$ae^{b^2/2}\int_{-\infty}^{\frac{1}{b}\ln(a/K)+b} e^{-y^2/2}\frac{dy}{\sqrt{2\pi}} = ae^{b^2/2}\Phi\left(\frac{1}{b}\ln(a/K)+b\right)$$

The second term is easy:

$$\int_{\frac{1}{b}\ln(K/a)}^{\infty} -Ke^{-z^2/2}\,\frac{dz}{\sqrt{2\pi}} = -K\left(1-\Phi\left(\frac{1}{b}\ln(K/a)\right)\right)$$

$$= -K\Phi\left(\frac{1}{b}\ln(a/K)\right)$$

□

I'll stop here. (But in the lecture I have to go back to section 9.7 on the Feynman-Kac formula.)

"The one thing probabilists can do which analysts can't is *stop*..."
-Sid Port, analyst [3]

References

1. Ioannis Karatzas and Steven E. Shreve, *Brownian motion and stochastic calculus*, second ed., Graduate Texts in Mathematics, vol. 113, Springer-Verlag, New York, 1991.
2. H. Kunita and S. Watanabe, *On square integrable martingales*, Nagoya Math. J. **30** (1967), 209–245.
3. L. C. G. Rogers and D. Williams, *Diffusions, Markov processes, and martingales. Vol. 1*, second ed., Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics, John Wiley & Sons Ltd., Chichester, 1994, Foundations.
4. L. C. G. Rogers and David Williams, *Diffusions, Markov processes, and martingales. Vol. 2*, Cambridge Mathematical Library, Cambridge University Press, Cambridge, 2000, Itô calculus, Reprint of the second (1994) edition.