Lectures on Stochastic Processes

William G. Faris

November 8, 2001

Contents

Chapter 1

Random walk

1.1 Symmetric simple random walk

Let $X_0 = x$ and

$$X_{n+1} = X_n + \xi_{n+1}. \tag{1.1}$$

The ξ_i are independent, identically distributed random variables such that $P[\xi_i = \pm 1] = 1/2$. The probabilities for this random walk also depend on x, and we shall denote them by P_x. We can think of this as a fair gambling game, where at each stage one either wins or loses a fixed amount.

Let T_y be the first time $n \geq 1$ when $X_n = y$. Let $\rho_{xy} = P_x[T_y < \infty]$ be the probability that the walk starting at x ever gets to y at some future time.

First we show that $\rho_{12} = 1$. This says that in the fair game one is almost sure to eventually get ahead by one unit. This follows from the following three equations.

The first equation says that in the first step the walk either goes from 1 to 2 directly, or it goes from 1 to 0 and then must go from 0 to 2. Thus

$$\rho_{12} = \frac{1}{2} + \frac{1}{2}\rho_{02}. \tag{1.2}$$

The second equation says that to go from 0 to 2, the walk has to go from 0 to 1 and then from 1 to 2. Furthermore, these two events are independent. Thus

$$\rho_{02} = \rho_{01}\rho_{12}. \tag{1.3}$$

The third equation says that $\rho_{01} = \rho_{12}$. This is obvious.

Thus $\rho = \rho_{12}$ satisfies

$$\rho = \frac{1}{2}\rho^2 + \frac{1}{2}. \tag{1.4}$$

This is a quadratic equation that can also be written as $\rho^2 - 2\rho + 1 = 0$, or $(\rho - 1)^2 = 0$. Its only solution is $\rho = 1$.

How long does it take, on the average, to get ahead by this amount? Let $m_{xy} = E_x[T_y]$, the expected time that the random walk takes to get to y,

starting at x. We have just seen that if $x = 1$, then $T_2 < \infty$ with probability one. Let us do the same kind of computation for $m_{12} = E_1[T_2]$.

The first equation says that in the first step the walk either goes from 1 to 2 directly, or it goes from 1 to 0 and then must go from 0 to 2. Thus

$$m_{12} = \frac{1}{2}(1 + m_{02}) + \frac{1}{2}1. \tag{1.5}$$

The second equation says that to go from 0 to 2, the walk has to go from 0 to 1 and then from 1 to 2. Thus

$$m_{02} = m_{01} + m_{12}. \tag{1.6}$$

The third equation says that $m_{01} = m_{12}$. This is obvious.

Thus $m = m_{12}$ satisfies

$$m = \frac{1}{2}(1 + 2m) + \frac{1}{2}. \tag{1.7}$$

This is a linear equation that can also be written as $m = 1+m$. Its only solution is $m = \infty$. Thus we have seen that $m_{01} = \infty$.

At first this seems strange, but it is quite natural. In the fair game there is a small probability of a bad losing streak. It takes a long time to recover from the losing streak and eventually get ahead. Infinitely long, on the average.

The calculation above uses a rather subtle property of random walk. Namely, it was assumed that after the walk accomplishes the task of going from 0 to 1, then it has an equally difficult task of going from 1 to 2. Now this is delicate, because the walk going from 1 to 2 starts out at a random time, not at a fixed time. Namely, if we start the walk at 0 and set $T = T_1$ as the first time the walk gets to 1, then from that time on the walk is $X_T, X_{T+1}, X_{T+2}, \ldots$ with $X_T = 1$ by definition. The point is that this has the same distribution as the walk $X_0, X_1, X_2, \ldots$ with $X_0 = 1$. This fact is called the *strong Markov property*.

The strong Markov property has the following statement. Start the walk out at x and let $T = T_y$. Let B be a set of random walk paths. We must prove that

$$P_x[(X_T, X_{T+1}, X_{T+2}, \ldots) \in B \mid T < \infty] = P_y[(X_0, X_1, X_2, \ldots) \in B]. \tag{1.8}$$

This says that the walk does not care when and how it got to y for the first time; from that point on it moves just as if it were started there at time zero.

The proof is the following. We write

$$P_x[(X_T, X_{T+1}, X_{T+2}, \ldots) \in B, T < \infty] = \sum_{n=1}^{\infty} P_x[(X_T, X_{T+1}, X_{T+2}, \ldots) \in B, T = n]. \tag{1.9}$$

However

$$P_x[(X_T, X_{T+1}, X_{T+2}, \ldots) \in B, T = n] = P_x[(X_n, X_{n+1}, X_{n+2} \ldots) \in B, T = n]. \tag{1.10}$$

Now the event $T = n$ depends only on the walk $X_1, X_2, \ldots, X_n$, which depends only on the $\xi_1, \xi_2, \ldots, \xi_n$. On the other hand, the walk $X_n, X_{n+1}, X_{n+2}, \ldots$ with $X_n = y$ depends only on the $\xi_{n+1}, \xi_{n+1}, \ldots$. Thus the events are independent, and we have

$$P_x[(X_T, X_{T+1}, X_{T+2}, \ldots) \in B, T = n] = P_x[(X_n, X_{n+1}, X_{n+2} \ldots) \in B]P[T = n], \tag{1.11}$$

where $X_n = y$. Finally, this is

$$P_x[(X_T, X_{T+1}, X_{T+2}, \ldots) \in B, T = n] = P_y[(X_0, X_1, X_2 \ldots) \in B]P[T = n]. \tag{1.12}$$

The proof is finished by summing over n.

The picture of symmetric simple random walk that emerges is the following. Starting from any point, the probability of eventually getting to any other point is one. However the expected time to accomplish this is infinite. Thus an imbalance in one direction is always compensated, but this random process is incredibly inefficient and can take a huge amount of time to do it.

1.2 Simple random walk

Let $X_0 = x$ and

$$X_{n+1} = X_n + \xi_{n+1}. \tag{1.13}$$

The ξ_i are independent, identically distributed random variables such that $P[\xi_i = 1] = p$, $P[\xi_i = -1] = q$, and $P[\xi = 0] = r$. Here $p + q + r = 1$, so the walk can change position by only one step at a time. The probabilities for this random walk depend on x, and we shall denote them by P_x. We can think of this as a gambling game, where at each stage one either wins or loses a fixed amount.

Let T_y be the first time $n \geq 1$ when $X_n = y$. Let $\rho_{xy} = P_x[T_y < \infty]$ be the probability that the walk starting at x ever gets to y at some future time.

First we calculate ρ_{12}. The calculation again uses three equations.

The first equation says that in the first step the walk either goes from 1 to 2 directly, or it stays at 1 and then must go to 2, or it goes from 1 to 0 and then must go from 0 to 2. Thus

$$\rho_{12} = q\rho_{02} + r\rho_{12} + p. \tag{1.14}$$

The second equation says that to go from 0 to 2, the walk has to go from 0 to 1 and then from 1 to 2. Furthermore, these two events are independent. Thus

$$\rho_{02} = \rho_{01}\rho_{12}. \tag{1.15}$$

The third equation says that $\rho_{01} = \rho_{12}$. This is obvious.

Thus $\rho = \rho_{12}$ satisfies

$$\rho = q\rho^2 + r\rho + p. \tag{1.16}$$

This is a quadratic equation that can also be written as $q\rho^2 + (r-1)\rho + p = 0$, or $(\rho - 1)(q\rho - p) = 0$. If $q > 0$ its solutions are $\rho = 1$ and $\rho = p/q$.

If $q \leq p$, then it is clear that the probability $\rho_{01} = 1$, except in the case $p = q = 0$. The game is even or favorable, so one is eventually ahead.

If $p < q$, then it would seem plausible that the probability is $\rho_{01} = p/q$. The game is unfavorable, and there is some chance that there is an initial losing streak from which one nevers recovers. There are a number of ways of seeing that this is the correct root. If the root were one, then the process starting at any number $x \leq 0$ would be sure to eventually reach 1. However according to the strong law of large numbers the walk $X_n/n \to p - q < 0$ as $n \to \infty$. Thus the walk eventually becomes negative and stays negative. This is a contradiction.

How long does it take, on the average, to get ahead by this amount? Let $m_{xy} = E_x[T_y]$, the expected time that the random walk takes to get to y, starting at x. We have just seen that if $x = 1$ and $p \geq q$ (not both zero), then $T_2 < \infty$ with probability one. Let us do the same kind of computation for $m_{12} = E_1[T_2]$.

The first equation says that in the first step the walk either goes from 1 to 2 directly, or it remains at 1 and then goes to 2, or it goes from 1 to 0 and then must go from 0 to 2. Thus

$$m_{12} = q(1 + m_{02}) + r(1 + m_{12}) + p1. \tag{1.17}$$

The second equation says that to go from 0 to 2, the walk has to go from 0 to 1 and then from 1 to 2. Thus

$$m_{02} = m_{01} + m_{12}. \tag{1.18}$$

Here we are implicitly using the strong Markov property.

The third equation says that $m_{01} = m_{12}$. This is obvious.

Thus $m = m_{12}$ satisfies

$$m = q(1 + 2m) + r(1 + m) + p. \tag{1.19}$$

This is a linear equation that can also be written as $m = 1 + (1 + q - p)m$. Its solutions are $m = \infty$ and $m = 1/(p - q)$.

If $p \leq q$, then it is clear that the expectation $m_{12} = \infty$. The game is even or unfavorable, so it can take a very long while to get ahead, if ever.

If $q < p$, then it would seem plausible that the expected number of steps to get ahead by one is $m_{12} = 1/(p - q)$. The game is favorable, so one should have win relatively soon, on the average. There are a number of ways of seeing that this is the correct solution. One way would be to use the identity

$$E_1[T_2] = \sum_{k=0}^{\infty} k P_1[T_2 = k] = \sum_{k=1}^{\infty} P_1[T_2 \geq k]. \tag{1.20}$$

Then one can check that if $q < p$, then the probabilities $P_1[T_2 \geq k]$ go rapidly to zero, so the series converges. This rules out an infinite expectation.

1.3 Problems

1. For the symmetric simple random walk starting at zero, find the mean and standard deviation of X_n. Find an approximate expression for $P[X_n = y]$. (Hint: If n is even, then y is even, and if n is odd, then y is odd. So for fixed n the spacing between adjacent values of y is $\Delta y = 2$. Use the central limit theorem, and replace the integral with a Riemann sum with $\Delta y = 2$.)

2. For the simple random walk starting at zero, find the mean and standard deviation of X_n.

3. For the simple random walk starting at 1, compute ρ_{11}, the probability of a return to the starting point.

4. For the simple random walk starting at 1, compute m_{11}, the expected time to return to the starting point.

5. A gambler plays roulette with the same stake on each play. This is simple random walk with $p = 9/19$ and $q = 10/19$. The player has deep pockets, but must win one unit. What is the probability that he does this in the first three plays? What is the probability that he ever does this?

Chapter 2

Gambler's ruin and martingales

2.1 Gambler's ruin: fair game

Let $X_0 = x$ and

$$X_{n+1} = X_n + \xi_{n+1}. \tag{2.1}$$

The ξ_i are independent, identically distributed random variables such that $P[\xi_i = 1] = p$, $P[\xi_i = -1] = q$, and $P[\xi_i = 0] = r$. Here $p + q + r = 1$, so the walk can change position by only one step at a time. The probabilities for this random walk depend on x, and we shall denote them by P_x.

We can think of this as a gambling game, where at each stage one either wins or loses a fixed amount. However now we want to consider the more realistic case when one wants to win a certain amount, but there is a cap on the possible loss. Thus take $a < x < b$. The initial capital is x. We want to play until the time T_b when the earnings achieve the desired level b or until the time T_a when we are broke. Let T be the minimum of these two times. Thus T is the time the game stops.

There are a number of ways to solve the problem. One of the most elegant is the martingale method. A martingale is a fair game.

We first want to see when the game is fair. In this case the conditional expectation of the winnings at the next stage are the present fortune. Thus

$$E[X_{n+1} \mid X_n = z] = z. \tag{2.2}$$

However this just says that $q(z-1) + rz + p(z+1) = z$, or $p = q$.

It follows from this equation that

$$E[X_{n+1}] = \sum_z E[X_{n+1} \mid X_n = z]P[X_n = z] = \sum_z zP[X_n = z] = E[X_n]. \tag{2.3}$$

Therefore, starting at x we have

$$E_x[X_n] = x. \tag{2.4}$$

The expected fortune stays constant.

Let $T \wedge n$ be the minimum of T and n. Then $X_{T\wedge n}$ is the game stopped at T. It is easy to show that this too is a fair game. Either one has not yet won or lost, and the calculation of the conditional expectation is as before, or one has won or lost, and future gains and losses are zero. Therefore

$$E_x[X_{T\wedge n}] = x. \tag{2.5}$$

Now the possible values of the stopped game are bounded: $a \le X_{T\wedge n} \le b$ for all n. Furthermore, since $T < \infty$ with probability one, the limit as $n \to \infty$ of $X_{T\wedge n}$ is X_T. It follows from the dominated convergence theorem (described below) that

$$E_x[X_T] = x. \tag{2.6}$$

Thus the game remains fair even in the limit of infinite time.

From this we see that

$$aP_x[X_T = a] + bP_x[X_T = b] = x. \tag{2.7}$$

It is easy then to calculate, for instance, that the probability of winning is

$$P_x[X_T = b] = \frac{x-a}{b-a}. \tag{2.8}$$

2.2 Gambler's ruin: unfair game

If the original game is not fair, then we want to modify it to make it fair. The new game will be that game whose accumulated winnings at stage n is $f(X_n)$. The conditional expectation of the winnings in the new game at the next stage are the present fortune of this game. Thus

$$E[f(X_{n+1}) \mid X_n = z] = f(z). \tag{2.9}$$

This says that $qf(z-1) + rf(z) + pf(z+1) = f(z)$. One solution of this is to take f to be a constant function. However this will not give a very interesting game. Another solution is obtained by trying an exponential $f(z) = a^z$. this gives $q/a + r + pa = 1$. If we take $a = q/p$ this gives a solution. Thus in the following we take

$$f(X_n) = \left(\frac{q}{p}\right)^{X_n} \tag{2.10}$$

as the fair game. If $p < q$, so the original game is unfair, then in the new martingale game rewards being ahead greatly and penalized being behind only a little.

It follows from this equation that

$$E[f(X_{n+1})] = \sum_z E[f(X_{n+1}) \mid X_n = z]P[X_n = z] = \sum_z f(z)P[X_n = z] = E[f(X_n)]. \tag{2.11}$$

Therefore, starting at x we have

$$E_x[f(X_n)] = f(x). \tag{2.12}$$

The expected fortune in the new martingale game stays constant.

Notice that to play this game, one has to arrange that the gain or loss at stage $n+1$ is

$$f(X_{n+1}) - f(X_n) = \left(\left(\frac{q}{p} \right)^{\xi_{n+1}} - 1 \right) f(X_n). \tag{2.13}$$

If, for instance, $p < q$, then the possible gain of $(q-p)/p\, f(X_n)$ is greater than the possible loss of $(q-p)/q\, f(X_n)$; the multiplicative factor of q/p makes this a fair game.

Let $T \wedge n$ be the minimum of T and n. Then $X_{T \wedge n}$ is the game stopped at T. It is easy to show that this too is a fair game. Therefore the same calculation shows that

$$E_x[f(X_{T \wedge n})] = f(x). \tag{2.14}$$

Now if $p < q$ and $f(x) = (q/p)^x$, we have that $a \leq x \leq b$ implies $f(a) \leq f(x) \leq f(b)$. It follows that $f(a) \leq f(X_{T \wedge n}) \leq f(b)$ is bounded for all n with bounds that do not depend on n. This justifies the passage to the limit as $n \to \infty$. This gives the result

$$E_x[f(X_T)] = f(x). \tag{2.15}$$

The game is fair in the limit.

From this we see that

$$\left(\frac{q}{p} \right)^a P_x[X_T = a] + \left(\frac{q}{p} \right)^b P_x[X_T = b] = \left(\frac{q}{p} \right)^x. \tag{2.16}$$

It is easy then to calculate, for instance, that the probability of winning is

$$P_x[X_T = b] = \frac{(q/p)^x - (q/p)^a}{(q/p)^b - (q/p)^a}. \tag{2.17}$$

If we take $p < q$, so that we have a losing game, then we can recover our previous result for the probability of eventually getting from x to b in the random walk by taking $a = -\infty$. Then we get $(q/p)^{x-b} = (p/q)^{b-x}$.

2.3 The dominated convergence theorem

The *dominated convergence theorem* gives a useful condition on when one can take a limit inside an expectation.

Theorem 2.1 *Let X_n be a sequence of random variables. Suppose that the limit as $n \to \infty$ of X_n is X, for each outcome in the sample space. Let $Y \geq 0$ be a random variable with $E[Y] < \infty$ such that $|X_n| \leq Y$ for each outcome in the sample space and for each n. Then the limit as $n \to \infty$ of $E[X_n]$ is $E[X]$.*

It is perhaps almost too trivial to remark, but it is a fact that a constant M is a random variable with finite expectation. This gives the following special case of the dominated convergence theorem. This special case is sometimes called the *bounded convergence theorem*.

Corollary 2.1 *Let X_n be a sequence of random variables. Suppose that the limit as $n \to \infty$ of X_n is X, for each outcome in the sample space. Let $M \geq 0$ be a constant such that $|X_n| \leq M$ for each outcome in the sample space and for each n. Then the limit as $n \to \infty$ of $E[X_n]$ is $E[X]$.*

2.4 Martingales

The general definition of a martingale is the following. We have a sequence of random variables $\xi_1, \xi_2, \xi_3, \ldots$. The random variable Z_n is a function of $\xi_1, \ldots, \xi_n$, so that

$$Z_n = h_n(\xi_1, \ldots, \xi_n). \tag{2.18}$$

The martingale condition is that

$$E[Z_{n+1} \mid \xi_1 = a_1, \ldots, \xi_n = a_n] = h_n(a_1, \ldots, a_n). \tag{2.19}$$

Thus a martingale is a fair game. The expected value at the next stage, given the present and past, is the present fortune.

It follows by a straightforward calculation that

$$E[Z_{n+1}] = E[Z_n]. \tag{2.20}$$

In particular $E[Z_n] = E[Z_0]$.

The fundamental theorem about martingales says that if one applies a gambling scheme to a martingale, then the new process is again a martingale. That is, no gambling scheme can convert a fair game to a game where one has an unfair advantage.

Theorem 2.2 *Let $\xi_1, \xi_2, \xi_3, \ldots$ be a sequence of random variables. Let $X_0, X_1, X_2, \ldots$ be a martingale defined in terms of these random variables, so that X_n is a function of $\xi_1, \ldots, \xi_n$. Let W_n be a gambling scheme, that is, a function of $\xi_1, \ldots, \xi_n$. Let $Z_0, Z_1, Z_2, \ldots$ be a new process such that $Z_{n+1} - Z_n = W_n(X_{n+1} - X_n)$. Thus the gain in the new process is given by the gain in the original process modified by the gambling scheme. Then $Z_0, Z_1, Z_2, \ldots$ is also a martingale.*

Proof: The condition that X_n is a martingale is that the expected gain $E[X_{n+1} - X_n \mid \xi_1 = a_1, \ldots, \xi_n = a_n] = 0$. Let

$$W_n = g_n(\xi_1, \ldots, \xi_n) \tag{2.21}$$

be the gambling scheme. Then

$$E[Z_{n+1} - Z_n \mid \xi_1 = a_1, \ldots, \xi_n = a_n] = E[W_n(X_{n+1} - X_n) \mid \xi_1 = a_1, \ldots, \xi_n = a_n]. \tag{2.22}$$

On the other hand, this is equal to

$$g_n(a_1, \ldots, a_n)E[X_{n+1} - X_n \mid \xi_1 = a_1, \ldots, \xi_n = a_n] = 0. \tag{2.23}$$

Example: Say that $X_n = \xi_1 + \xi_2 + \cdots \xi_n$ is symmetric simple random walk. Let the gambling scheme $W_n = 2^n$. That is, at each stage one doubles the amount of the bet. The resulting martingale is $Z_n = \xi_1 + 2\xi_2 + 4\xi_3 + \cdots + 2^{n-1}\xi_n$. This game gets wilder and wilder, but it is always fair. Can one use this double the bet game to make money? See below.

One important gambling scheme is to quit gambling once some goal is achieved. Consider a martingale $X_0, X_1, X_2, \ldots$. Let T be a time with the property that the event $T \leq n$ is defined by $\xi_1, \ldots, \xi_n$. Such a time is called a *stopping time*. Let the gambling scheme W_n be 1 if $n < T$ and 0 if $T \leq n$. Then the stopped martingale is given by taking $Z_0 = X_0$ and $Z_{n+1} - Z_n = W_n(X_{n+1} - X_n)$. Thus if $T \leq n$, $Z_{n+1} - Z_n = 0$, while if $n < T$ then $Z_{n+1} - Z_n = X_{n+1} - X_n$. As a consequence, if $T \leq n$, then $Z_n = X_T$, while if $n < T$, then $Z_n = X_n$. This may be summarized by saying that $Z_n = X_{T \wedge n}$, where $T \wedge n$ is the minimum of T and n. In words: the process no longer changes after time T.

Corollary 2.2 *Let T be a stopping time. Then if X_n is a martingale, then so is the stopped martingale $Z_n = X_{n \wedge T}$.*

It might be, for instance, that T is the first time that the martingale X_n belongs to some set. Such a time is a stopping time. Then the process $X_{T \wedge n}$ is also a martingale.

Example: A gambler wants to use the double the bet martingale $Z_n = \xi_1 + 2\xi_2 + 4\xi_3 + \cdots + 2^{n-1}\xi_n$ to get rich. The strategy is to stop when ahead. The process $Z_{T \wedge n}$ is also a martingale. This process will eventually win one unit. However, unfortunately for the gambler, at any particular non-random time n the stopped process is a fair game. $Z_{T \wedge n}$ is either 1 with probability $1 - 1/2^n$ or $1 - 2^n$ with probability $1/2^n$. It looks like an easy win, most of the time. But a loss is a disaster.

If Z_n is a martingale, then the processes $Z_{T \wedge n}$ where one stops at the stopping time T is also a martingale. However what if $T < \infty$ with probability one, there is no limit on the time of play, and one plays until the stopping time? Does the game remain fair in the limit of infinite time?

Theorem 2.3 *If $T < \infty$ and if there is a random variable $Y \geq 0$ with $E[Y] < \infty$ such that for all n we have $|Z_{T \wedge n}| \leq Y$, then $E[Z_{T \wedge n}] \to E[Z_T]$ as $n \to \infty$.*

This theorem, of course, is just a corollary of the dominated convergence theorem. In the most important special case the dominating function Y is just a constant. This says that for a bounded martingale we can always pass to the limit. In general the result can be false, as we shall see in the following examples.

Example. Let X_n be symmetric simple random walk starting at zero. Then X_n is a martingale. Let T be the first n with $X_n = b > 0$. The process

$X_{T\wedge n}$ that stops when b is reached is a martingale. The stopped process is a martingale. However the infinite time limit is not fair! In fact $X_T = b$ by definition. A fair game is converted into a favorable game. However this is only possible because the game is unbounded below.

Example. Let X_n be simple random walk starting at zero, not symmetric. Then $(q/p)^{X_n}$ is a martingale. Let $b > 0$ and T be the first $n \geq 1$ with $X_n = b$. Then $(q/p)^{X_{T\wedge n}}$ is a martingale. If $p < q$, an unfavorable game, then this martingale is bounded. Thus we can pass to the limit. This gives $(q/p)^x = (q/p)^b P[T_b < \infty]$. On the other hand, if $q < p$, then the martingale is badly unbounded. It is not legitimate to pass to the limit. And in fact $(q/p)^x \neq (q/p)^b P[T_b < \infty] = (q/p)^b$.

Example: Consider again the double the bet gambling game where one quits when ahead by one unit. Here $Z_{T\wedge n} = 1$ with probability $1 - 1/2^m$ and $Z_{T\wedge n} = 1 - 2^n$ with probability $1/2^n$. Eventually the gambler will win, so the $Z_T = 1$. This limiting game is no longer a martingale, it is favorable to the gambler. However one can win only by having unlimited credit. This is unrealistic in this kind of game, since the losses along the way can be so huge.

2.5 Problems

1. HPS, Chapter 1, Exercise 25.

2. For the simple random walk with $p + q + r = 1$ and $p \neq q$ show that $X_n - (p - q)n$ is a martingale. This is the game in which an angel compensates the player for the average losses. Then let $a < x < b$ and let T be the first time the walk starting at x gets to a or b. Show that the stopped martingale $X_{T\wedge n} - (p - q)T \wedge n$ is a martingale. Finally, use $E_x[T] < \infty$ to show that $E_x[X_T] = (p - q)E_x[T] + x$. Compute $E_x[T]$ explicitly.

3. For the symmetric simple random walk with $p + q + r = 1$ and $p = q$ show that $X_n^2 - (1 - r)n$ is a martingale. The average growth of X_n^2 is compensated to make a fair game. Then let $a < x < b$ and let T be the first time the symmetric walk starting at x gets to a or b. Show that the stopped martingale $X_{T\wedge n}^2 - (1-r)T \wedge n$ is a martingale. Finally, use $E_x[T] < \infty$ to show that for the symmetric walk $E_x[X_T^2] = (1 - r)E_x[T] + x^2$. Compute $E_x[T]$ explicitly.

Chapter 3

The single server queue

3.1 A discrete time queue

Here the process is the following. Clients enter the queue. From time to time a client is served, if any are present. Let X_n be the number of clients in the queue just before service. Let W_n be the number of clients in the queue just after service. Then

$$W_n = (X_n - 1) \vee 0 \tag{3.1}$$

and

$$X_{n+1} = W_n + \xi_{n+1}, \tag{3.2}$$

where $\xi_{n+1} \geq 0$ are the number of new clients that enter. We shall assume that these are independent random variables whose values are natural numbers. Thus $P[\xi_i = k] = f(k)$.

The queue may be described either in terms of X_n or in terms of W_n. In the first description

$$X_{n+1} = (X_n - 1) \vee 0 + \xi_{n+1}. \tag{3.3}$$

In the second description

$$W_{n+1} = (W_n + \xi_{n+1}) \vee 0. \tag{3.4}$$

We shall use the first description. This process behaves like a random walk away from zero. Thus if $X_n \geq 1$, then $X_{n+1} = X_n + \xi_{n+1} - 1$. Thus it can step up by any amount, but it can step down by at most one. On the other hand, when $X_n = 0$, then $X_{n+1} = X_n + \xi_{n+1}$. An empty queue can only step up in size.

3.2 Emptying the queue

From now on we make the assumption that $P[\xi_i \geq 2] = 1 - f(0) - f(1) > 0$. If this is violated then the queue cannot possibly grow in size. We also make the

assumption that $P[\xi_i = 0] = f(0) > 0$. If this does not hold, then the queue cannot decrease in size.

We want to see if the queue is sure to eventually empty. For this we need to compute $\rho_{x0} = P_x[T_0 < \infty]$ for $x \geq 1$. First we note that

$$\rho_{10} = f(0) + \sum_{y=1}^{\infty} f(y)\rho_{y0}. \tag{3.5}$$

This just says that at the first step something happens. The second observation is that for $x \geq 1$ we have

$$\rho_{x0} = \rho_{10}^x. \tag{3.6}$$

This follows from the strong Markov property of random walk.

Let $\rho = \rho_{10}$. We have shown that

$$\rho = \sum_{y=0}^{\infty} f(y)\rho^y. \tag{3.7}$$

We can write this also as a fixed point equation

$$\rho = \Phi(\rho), \tag{3.8}$$

where $\Phi(t)$ is the generating function of the number of new entrants to the queue ξ_i.

This generating function has remarkable properties in the interval from 0 to 1. First, note that $\Phi(0) = f(0)$ and $\Phi(1) = 1$. Furthermore, $\Phi'(t) \geq 0$, so Φ is increasing from $f(0)$ to 1. Furthermore, $\Phi''(t) \geq 0$, so $\Phi(t)$ is a convex function. In fact, since there is some i with $i \geq 2$ and $f(i) > 0$, it follows that $\Phi''(t) > 0$ for $0 < t \leq 1$, so $\Phi(t)$ is strictly convex. Finally, $\Phi'(1) = \mu$, the expected number of new entrants to the queue. All these properties give a fairly complete picture of $\Phi(t)$.

There are two cases. If $\mu \leq 1$, then $\Phi(t) > t$ for $0 \leq t < 1$. So the only root of $\Phi(t) = t$ is $\rho = 1$. This says that the queue with one customer eventually becomes empty, with probability one. This is indeed what one would expect for a random walk with mean step size $\mu - 1 \leq 0$.

If, on the other hand, $\mu > 1$, then $\Phi(t) = t$ has a second root $\rho < 1$. This must be the probability of the queue with one customer eventually becoming empty. Then there is some probability $1 - \rho > 0$ that the queue never becomes empty. This is indeed what one gets by using the strong law of large numbers with the random walk with mean step size $\mu - 1 > 0$. In fact, the queue will eventually grow at a faster than linear rate.

Starting the queue from zero is exactly the same as starting it from 1. We have therefore proved the following theorem.

Theorem 3.1 *Consider the queue model with the above assumptions. Start the queue with x customers. If the mean number of entrants to the queue between service times is $\mu \leq 1$, then the queue eventually empties, with probability one.*

If this mean number is $\mu > 1$, then the probability that the queue ever empties in the future is $\rho^x < 1$ for $x \geq 1$ and is ρ for $x = 0$, where ρ is the unique root of $\Phi(t) = t$ that is less than one.

When the mean number of entrants to the queue between service times is $\mu \leq 1$, then the queue eventually empties. Let us see how long it takes on the average to empty. For this we need to compute $m_{x0} = E_x[T_0]$ for $x \geq 1$. First we note that

$$m_{10} = f(0)1 + \sum_{y=1}^{\infty} f(y)(1 + m_{y0}). \tag{3.9}$$

This just says that either one succeeds at the first step, or one has wasted one unit of time. The second observation is that for $x \geq 1$ we have

$$m_{x0} = x m_{10}. \tag{3.10}$$

This follows from the strong Markov property of random walk.

Let $m = m_{01}$. We have shown that

$$m = 1 + \mu m. \tag{3.11}$$

If $\mu \geq 1$, then this has only the solution $m = \infty$. However if $\mu < 1$, then there is also the solution $m = 1/(1 - \mu)$. This is the correct formula for the expected time. To see this, one needs to show that the expected time to for the random walk to get from 1 to 0 is finite. This can be done in various ways, such as by a martingale argument.

Here is a sketch of the martingale argument. Consider the martingale $X_{T \wedge n} - (\mu - 1)T \wedge n$ starting at $x \geq 1$. This is just the random walk stopped at the time T when it first gets to zero, with a correction term that compensates for the fact that the walk is not symmetric. Since it is a martingale, its expectation is the initial value x. This says that $E[X_{T \wedge n}] - (\mu - 1)E[T \wedge n] = x$. But $X_{T \wedge n} \geq 0$, and so its expectation is also greater than or equal to zero. This says $-(\mu - 1)E[T \wedge n] \leq x$. Since $\mu < 1$, this is equivalent to $E[T \wedge n] \leq x/(1 - \mu)$. Let $n \to \infty$. By the monotone convergence theorem (see below), $m = E[T] \leq x/(1 - \mu)$. So it cannot be infinite.

Theorem 3.2 *Consider the queue model with the above assumptions. Start the queue with x customers. If the mean number of customers entering the queue between service times is $\mu < 1$, then the expected time until the queue is empty again is equal to $x/(1 - \mu)$ for $x \geq 1$ and is $1/(1 - \mu)$ for $x = 0$.*

3.3 The monotone convergence theorem

Theorem 3.3 *Let $X_n \geq 0$ be a sequence of positive random variables such that $X_n \leq X_{n+1}$ for each n. Assume that for each outcome $X_n \to X$ as $n \to \infty$. Then $E[X_n] \to E[X]$ as $n \to \infty$.*

The monotone convergence theorem would be nothing new if we knew that the limiting function X had a finite expectation. In that case, we would have $0 \leq X_n \leq X$, and so we could apply the dominated convergence theorem. The point is that the theorem shows that if the X_n are positive and monotone increasing and $E[X_n]$ converges to a finite limit, then $E[X]$ is finite and has this limit as its value.

3.4 The stationary distribution

Consider the queue with $\mu < 1$. The theorem says that the mean time for a return to 0 is $1/(1-\mu)$. It would then seem reasonable that the proportion of time that the queue is empty would be $1-\mu$. In fact, for every state of the queue there should be a stable proportion of time that the queue spends in this state.

These numbers that predict the proportion of time in the state are probabilities $\pi(x)$. They are called the *stationary distribution* of the queue. The queue is defined by the numbers

$$f(y) = P[\xi_i = y]. \tag{3.12}$$

The problem is to go from the probabilities $f(y)$ to the probabilities $\pi(x)$. This should be set up in a way that works whenever $\mu < 1$.

The general requirement for a stationary distribution is

$$P[X_n = y] = P[X_{n+1} = y]. \tag{3.13}$$

If we define $\pi(y) = P[X_n = y]$, we obtain for the queue that

$$\pi(y) = \pi(0)f(y) + \sum_{x=1}^{y+1} \pi(x)f(y-x+1). \tag{3.14}$$

This equation may be solved recursively for $\pi(y+1)$ in terms of $\pi(0), \ldots, \pi(y)$. But it is not so easy to interpret the solutions. Therefore we transform it to another form from which it is apparent, for instance, that the solutions are positive.

Recall that expected number of entrants to the queue is

$$\mu = E[\xi_i] = \sum_{y=0}^{\infty} yf(y). \tag{3.15}$$

Let

$$g(z) = P[\xi_i \geq z] = \sum_{y=z}^{\infty} f(y). \tag{3.16}$$

Recall the alternative formula for the mean:

$$\mu = \sum_{z \geq 1} g(z). \tag{3.17}$$

Take $z \geq 1$. Sum the equation for the stationary distribution from 0 to $z-1$. This gives

$$\sum_{y=0}^{z-1} \pi(y) = \pi(0) \sum_{y=0}^{z-1} f(y) + \sum_{y=0}^{z-1} \sum_{x=1}^{y+1} \pi(x) f(y-x+1). \tag{3.18}$$

Interchange the order of summation. This gives

$$\sum_{y=0}^{z-1} \pi(y) = \pi(0)(1-g(z)) + \sum_{x=1}^{z} \sum_{y=x-1}^{z-1} \pi(x) f(y-x+1). \tag{3.19}$$

This can be rewritten as

$$\sum_{y=0}^{z-1} \pi(y) = \pi(0)(1-g(z)) + \sum_{x=1}^{z} \pi(x)(1-g(z-x+1)). \tag{3.20}$$

This can be solved for $\pi(z)$.

The final equation for the stationary distribution is that for each $z \geq 1$ we have

$$\pi(z)[1-g(1)] = \pi(0)g(z) + \sum_{x=1}^{z-1} \pi(x) g(z-x+1). \tag{3.21}$$

In more abstract language, this is a balance equation of the form

$$P[X_{n+1} < z, X_n \geq z] = P[X_{n+1} \geq z, X_n < z]. \tag{3.22}$$

This equation says that the rate of transition from having z or more clients to fewer than z clients is equal to the rate of transition from having fewer than z clients to z or more clients. The left hand says that to make the transition down you have to have exactly z clients and no new customers. The right hand size says that either you have 0 clients and z or more new customers, or you have $x \geq 1$ clients, serve 1, leaving $x-1$ and then add $z-(x-1)$ or more new customers.

These equations are easy to solve recursively. Suppose we know $\pi(0)$. We compute $\pi(1)f(0) = \pi(0)g(1)$, then $\pi(2)f(0) = \pi(0)g(2) + \pi(1)g(2)$, then $\pi(3)g(0) = \pi(0)g(3) + \pi(1)g(3) + \pi(2)g(4)$, and so on.

If the stationary probabilities are to exist, then the $\pi(0)$ must be chosen so that the sum of the $\pi(z)$ for all $z = 0, 1, 2, 3, \ldots$ is 1. To see when this works, sum the equation for $z = 1, 2, 3, \ldots$. Then interchange the order of summation. This gives the product of two independent sums on the right hand side. The result is

$$[1-\pi(0)][1-g(1)] = \pi(0)\mu + [1-\pi(0]](\mu - g(1)]. \tag{3.23}$$

It equates the rate of moving down by one from anywhere in the queue except zero to the rate for moving up to anywhere in the queue except zero. This simplifies to $[1-\pi(0)][1-g(1)] = \mu - [1-\pi(0)]g(1)$. This can balance only if $\mu < 1$. In fact, we see that

$$\pi(0) = 1 - \mu. \tag{3.24}$$

We have proved the following theorem.

Theorem 3.4 *Consider the queue model with $\mu < 1$. Then there is a stationary distribution satisfying $\pi(x) = P[X_n = x] = P[X_{n+1} = x]$. It satisfies the balance equation*

$$P[X_{n+1} < z, X_n \geq z] = P[X_{n+1} \geq z, X_n < z]. \tag{3.25}$$

Consider the probability $f(x)$ that exactly x clients enter between services and the cumulative probability $g(z) = \sum_{x \geq z} f(x)$. The balance equation takes the form

$$\pi(z) f(0) = \pi(0) g(z) + \sum_{x=1}^{z-1} \pi(x) g(z - x + 1). \tag{3.26}$$

for $z \geq 1$. Furthermore, $\pi(0) = 1 - \mu$.

3.5 Problems

1. Think of a queue process taking place in continuous time. Say that customers are entering the queue randomly at rate λ_1 per second. Thus number of customers who enter the queue in a time period of length t seconds is Poisson with mean $\lambda_1 t$. Say that the time between services is exponential with rate λ_2 per second. Show that the number of customers arriving in a time period between services is geometric with parameter $p = \lambda_2/(\lambda_1 + \lambda_2)$. Show that the expected number of customers arriving is $\mu = q/p$. Suppose that $\lambda_2 > \lambda_1$, so the service rate is faster than the incoming customer rate. Show that it follows that $p > 1/2$, so $\mu < 1$.

2. Now go to the discrete time model. Assume that the number of customers arriving in a time period between services is geometric with parameter p. Let $q = 1 - p$ and $\mu = q/p$. Suppose $p \leq 1/2$, so that $\mu \geq 1$. Suppose that there are x customers present initially. The manager has promised to quit his job the next time that the queue empties. What is the probability that this ever happens?

3. Continue with the discrete time model. Assume that the number of customers arriving in a time period between services is geometric with parameter p. Let $q = 1-p$ and $\mu = q/p$. Suppose $p > 1/2$, so that $\mu < 1$. Show that the stationary distribution of the queue is geometric with parameter $1-\mu$. Hint: Use $g(z) = qg(z-1)$ to show $\pi(z)p = q\pi(z-1)p + \pi(z-1)q^2 = q\pi(z-1)$.

4. Consider a stochastic process that satisfies detailed balance (time reversibility):

$$P[X_{n+1} = y, X_n = x] = P[X_{n+1} = x, X_n = y]. \tag{3.27}$$

   Show that it has a stationary measure:

$$P[X_{n+1} = y] = P[X_n = y]. \tag{3.28}$$

Chapter 4

The branching process

4.1 The branching model

The process is the following. There are X_n individuals at time n. The ith individual has a random number $\xi_i^{(n+1)}$ of children. Here $P[\xi_i^{(n+1)} = k] = f(k)$. Thus

$$X_{n+1} = \xi_1^{(n+1)} + \xi_2^{(n+1)} + \cdots + \xi_{X_n}^{(n+1)}. \tag{4.1}$$

Thus the next generation consists of all the children of the individuals of the previous generation.

One could also write this in a more abstract notation as

$$X_{n+1} = \sum_{i=1}^{\infty} \xi_i^{(n+1)} 1_{i \le X_n}. \tag{4.2}$$

4.2 Explosion

The first thing one can do to analyze the model is to look at expectations and variances. This turns out to be interesting, but ultimately misleading, in that it misses some of the essential stochastic features.

The first computation is that of the mean. Let $\mu = E[\xi_i^{(n+1)}]$ be the mean number of children of an individual. Since the conditional expectation of the number of children of z individuals is μz, we have

$$E[X_{n+1} \mid X_n = z] = z\mu. \tag{4.3}$$

It follows that

$$E[X_{n+1}] = \sum_z E[X_{n+1} \mid X_n = z] P[X_n = z] = \sum_z z\mu P[X_n = z] = \mu E[X_n]. \tag{4.4}$$

From this we see that

$$E_x[X_n] = x\mu^n. \tag{4.5}$$

If $\mu > 1$, then the mean size of the population grows exponentially.

The second computation is that of the variance. Let σ^2 be the variance of the number of children of an individual. Then the conditional variance of the number of children of z individuals is $\sigma^2 z$. We compute

$$E[(X_{n+1} - \mu z)^2 \mid X_n = z] = \sigma^2 z. \tag{4.6}$$

Write

$$(X_{n+1}-\mu E[X_n])^2 = (X_{n+1}-\mu z)^2+2(X_{n+1}-\mu z)(\mu z-\mu E[X_n])+(\mu z-\mu E[X_n])^2. \tag{4.7}$$

Take the conditional expectation of both sides, and note that the cross term has conditional expectation zero. This gives

$$E[(X_{n+1} - \mu E[X_n])^2 \mid X_n = z] = \sigma^2 z + \mu^2 (z - E[X_n])^2. \tag{4.8}$$

However

$$\mathrm{Var}(X_{n+1}) = \sum_z E[(X_{n+1} - E[X_{n+1}])^2 \mid X_n = z] P[X_n = z]. \tag{4.9}$$

It follows that

$$\mathrm{Var}(X_{n+1}) = \sigma^2 E[X_n] + \mu^2 \,\mathrm{Var}(X_n). \tag{4.10}$$

This equation says that the variance in the next generation is the mean size of the current generation times the variance of the number of children plus the variance of the current generation times the square of the mean number of children.

One can solve this explicitly. The solution is

$$\mathrm{Var}(X_n) = x\sigma^2 \mu^{n-1}(1 + \mu + \mu^2 + \cdots + \mu^{n-1}). \tag{4.11}$$

If $\mu \neq 1$, this is a geometric series with sum

$$\mathrm{Var}(X_n) = x\sigma^2 \mu^{n-1} \frac{\mu^n - 1}{\mu - 1}. \tag{4.12}$$

Theorem 4.1 *For the branching process the expected size of the nth generation is*

$$E_x[X_n] = x\mu^n \tag{4.13}$$

and the variance of the size of the nth generation is

$$\mathrm{Var}(X_n) = x\sigma^2 \mu^{n-1} \frac{\mu^n - 1}{\mu - 1}. \tag{4.14}$$

if $\mu \neq 1$ *and*

$$\mathrm{Var}(X_n) = x\sigma^2 n \tag{4.15}$$

if $\mu = 1$. *Furthermore, the random variable*

$$Z_n = \frac{X_n}{\mu^n} \tag{4.16}$$

is a martingale, and if $\mu > 1$ its variance

$$\mathrm{Var}(Z_n) = x\frac{\sigma^2}{\mu(\mu-1)}\left(1-\frac{1}{\mu^n}\right) \tag{4.17}$$

is bounded independent of n.

The fact that for $\mu > 1$ the ratio $Z_n = X_n/\mu^n$ is a martingale with bounded variance is important. It says that in the case of mean exponential growth the variability of the actual size of the population relative to its expected size is under control. There is a remarkable martingale theorem that says that under these circumstance there a random variable Z_∞ so that $Z_n \to Z_\infty$ as $n \to \infty$. Thus the asymptotic growth of the population is actually exponential, with a random coefficient $Z_\infty \geq 0$. As we shall now discuss in detail, with $\mu > 1$ it will happen that $Z_\infty > 0$ with strictly positive probability, but also $Z_\infty = 0$ with strictly positive probability.

Why does the average size calculation miss the point? Because even if $\mu > 1$, so that the expected population size is undergoing exponential growth, there is a possibility that the actual population size is zero. That is, there is a chance that fairly soon the population becomes extinct. There is no way to recover from this.

4.3 Extinction

From now on we make the assumption that $P[\xi_i = 1] = f(1) < 1$. If this is violated then each individual has exactly one child, and the number of individuals remains constant. It turns out that under this hypothesis the branching process is sure either to become extinct or to grow to infinity.

We want to see when the process will become extinct. For this we need to compute $\rho_{x0} = P_x[T_0 < \infty]$ for $x \geq 1$. First we note that

$$\rho_{10} = f(0) + \sum_{y=1}^{\infty} f(y)\rho_{y0}. \tag{4.18}$$

This just says that if there is one individual, then for the line to become extinct, each child must have a line that becomes extinct. The second observation is that for $x \geq 1$ we have

$$\rho_{x0} = \rho_{10}^x. \tag{4.19}$$

This is because the extinction events for the line of each child are independent.

Let $\rho = \rho_{01}$. We have shown that

$$\rho = \sum_{y=0}^{\infty} f(y)\rho^y. \tag{4.20}$$

We can write this also as a fixed point equation

$$\rho = \Phi(\rho), \tag{4.21}$$

where $\Phi(t)$ is the generating function of the number of new entrants to the queue ξ_i.

This generating function has remarkable properties in the interval from 0 to 1. First, note that $\Phi(0) = f(0)$ and $\Phi(1) = 1$. Furthermore, $\Phi'(t) \geq 0$, so Φ is increasing from $f(0)$ to 1. Furthermore, $\Phi''(t) \geq 0$, so $\Phi(t)$ is a convex function. Finally, $\Phi'(1) = \mu$, the expected number of children of an individual.

There are two cases. If $\mu \leq 1$, then $\Phi(t) > t$ for $0 \leq t < 1$. So the only root of $\Phi(t) = t$ is $\rho = 1$. This says that the descendents of the starting individual eventually die out, with probability one. This is indeed what one would expect from such pitiful reproductive capability.

If, on the other hand, $\mu > 1$, then $\Phi(t) = t$ has a second root $\rho < 1$. This root is unique. In fact, $\mu > 1$ implies that there is some $i \geq 2$ with $f(i) > 0$. It follows that $\Phi''(t) > 0$ for $0 < t \leq 1$, so $\Phi(t)$ is strictly convex.

This second root must be the probability that the line starting with one individual eventually dies out. To see this, consider ρ^{X_n}. This is a martingale, so $E_1[\rho^{X_n}] = \rho$. As $n \to \infty$, then either $X_n \to \infty$ or X_n is eventually zero. So ρ^{X_n} converges to the indicator function 1_A of the event A of eventual extinction. By the bounded convergence theorem $E_1[1_A] = \rho$. This says that $P_1[A] = \rho$.

Theorem 4.2 *Consider the branching model with the above assumptions. Start with x individuals. If the mean number of children of an individual is $\mu \leq 1$, then the line eventually dies out, with probability one. If this mean number of children is $\mu > 1$, then the probability that the line dies out is $\rho^x < 1$ for each $x \geq 1$. Here ρ is the unique root of $\Phi(t) = t$ that is less than one.*

4.4 Problems

1. HPS, Chapter 1, Problem 32
2. HPS, Chapter 1, Problem 33
3. HPS, Chapter 1, Problem 34

Chapter 5

Markov chains

5.1 The Markov property

Let $\mathcal{S}$ be a countable set. We consider random variables X_n with values in $\mathcal{S}$. The official definition of Markov chain is through the following Markov property:

$$P[X_{n+1} = b \mid X_0 = a_0, X_1 = a_1, \ldots, X_n = a_n] = P[X_{n+1} = b \mid X_n = a_n]. \tag{5.1}$$

In other words, the next step in the future depends on the past and present only through the present. We shall also assume that the process is time homogeneous, so that the transition probabilities are always the same. This says that

$$P[X_{n+1} = b \mid X_n = a] = P[X_1 = b \mid X_0 = a]. \tag{5.2}$$

Let

$$\pi(x) = P[X_0 = x] \tag{5.3}$$

be the initial probabilities. Let

$$P[X_{n+1} = y \mid X_n = x] = P(x, y) \tag{5.4}$$

be the transition probabilities. If these data are the same, then this is regarded as the same chain. In fact, the chain is often thought of as associated just with the transition probabilities.

Let us look at the Markov property in a special case, to see how it works. Say that we want to compute

$$P[X_{n+2} = b_2, X_{n+1} = b_1 \mid X_n = a] = P[X_{n+2} = b_2 \mid X_{n+1} = b_1, X_n = a]P[X_{n+1} = b_1 \mid X_n = a]. \tag{5.5}$$

We would be completely stuck unless we could use the Markov property

$$P[X_{n+2} = b_2 \mid X_{n+1} = b_1, X_n = a] = P[X_{n+2} = b_2 \mid X_{n+1} = b_1]. \tag{5.6}$$

Then we get

$$P[X_{n+2} = b_2, X_{n+1} = b_1 \mid X_n = a] = P[X_{n+2} = b_2 \mid X_{n+1} = b_1]P[X_{n+1} = b_1 \mid X_n = a]. \tag{5.7}$$

It follows by summing that

$$P[X_{n+2} = b_2 \mid X_n = a] = \sum_{b_1} P[X_{n+2} = b_2 \mid X_{n+1} = b_1]P[X_{n+1} = b_1 \mid X_n = a]. \tag{5.8}$$

There is a considerably more general statement of the Markov property. If G is an arbitrary event defined in terms of the values of $X_0, X_1, \ldots, X_n$, then

$$P[X_{n+1} = b_1, X_{n+2} = b_2, X_{n+3} = b_3, \ldots \mid G] = P_{X_n}[X_1 = b_1, X_2 = b_2, X_3 = b_3, \ldots]. \tag{5.9}$$

It is just as if the process started all over at X_n, none of the history in G is relevant, other than the value of X_n.

It is not too hard to show that

$$P[X_{n+k} = y \mid X_n = x] = P^k(x, y), \tag{5.10}$$

where P^k is the matrix power of the transition probability matrices. Thus there is a strong connection between Markov chains and linear algebra. We denote the process starting with probability density π by P_π, so

$$P_\pi[X_k = y] = \sum_x \pi(x)P^k(x, y). \tag{5.11}$$

In particular, if the process starts at x, then we write

$$P_x[X_k = y] = P^k(x, y). \tag{5.12}$$

Theorem 5.1 *Let $\mathcal{S}$ be a countable set. Let F be a function that sends a pair (x, r) with $x \in \mathcal{S}$ and r a number to a value $y = F(x, r)$ in $\mathcal{S}$. Let $\xi_1, \xi_2, \xi_3, \ldots$ be a sequence of independent identically distributed random variables. Let X_0 be a random variable with values in $\mathcal{S}$. Then the sequence of random variables X_n with values in $\mathcal{S}$ defined by*

$$X_{n+1} = F(X_n, \xi_{n+1}) \tag{5.13}$$

is a Markov chain.

5.2 The strong Markov property

The strong Markov property is an extension of the Markov property to a random time. This time must be stopping time. This means that the event that $T = n$ must be definable in terms of the values $X_0, X_1, X_2, \ldots, X_n$. In other words, to know when you stop, it is sufficient to know the past.

A typical example is when T is the first time in the future that the process arrives at some set. To know that $T = n$ you only need to know that $X_1, \ldots, X_{n-1}$ are not in the set and that X_n is in the set.

The strong Markov property says that if G is any event defined in terms of the value of T and of $X_0, X_1, \ldots, X_T$ that implies $T < \infty$, then

$$P[X_{T+1} = b_1, X_{T+2} = b_2, X_{T+3} = b_3, \ldots \mid G] = P_{X_T}[X_1 = b_1, X_2 = b_2, X_3 = b_3, \ldots] \tag{5.14}$$

It is just as if the process started all over at X_T. None of the history in G is relevant.

Here is a special case:

$$P[X_{T+1} = b_1, X_{T+2} = b_2, X_{T+3} = b_3, \ldots \mid T < \infty] = P_{X_T}[X_1 = b_1, X_2 = b_2, X_3 = b_3, \ldots] \tag{5.15}$$

Here is a typical application. Suppose the chain starting at x has the property that it can get to z only by first passing through y. Then

$$P_x[T_z < \infty] = P_x[T_z < \infty, T_y < \infty] = P_x[T_z < \infty \mid T_y < \infty] P[T_y < \infty]. \tag{5.16}$$

By the strong Markov property this is

$$P_x[T_z < \infty] = P_y[T_z < \infty] P_x[T_y < \infty]. \tag{5.17}$$

Introduce the notation $\rho_{xz} = P_x[T_z < \infty]$. Then this result says that if the chain starting at x can get to z only by first passing through y, then $\rho_{xz} = \rho_{xy}\rho_{yz}$.

5.3 Transient and recurrent states

First we fix some notation. Consider a Markov chain starting at x. Let T_y be the least $n \geq 1$ such that $X_n = y$. If there is no such n, then $T_y = \infty$. The *hitting probability* of y starting from x is

$$\rho_{xy} = P_x[T_y < \infty]. \tag{5.18}$$

Consider a Markov chain starting at x. Let $N(y)$ be the number of $n \geq 1$ such that $X_n = y$. We allow the possibility that $N(y) = \infty$.

Theorem 5.2 *For each $m \geq 1$*

$$P_x[N(y) \geq m] = \rho_{xy}\rho_{yy}^{m-1}. \tag{5.19}$$

Corollary 5.1 *If $\rho_{yy} < 1$, then*

$$P_x[N(y) = \infty] = 0. \tag{5.20}$$

Corollary 5.2 *If $\rho_{yy} = 1$, then*

$$P_x[N(y) = \infty] = \rho_{xy}. \tag{5.21}$$

A state of a Markov chain is *transient* if $\rho_{yy} < 1$ and *recurrent* if $\rho_{yy} = 1$. From the previous corollaries, if the chain is started at transient state, it returns finitely many times. If the chain is started at a recurrent state, it returns infinitely many times.

5.4 Recurrent classes

We say that x leads to y if $\rho_{xy} > 0$.

Theorem 5.3

$$E_x[N(y)] = \sum_{n=1}^{\infty} P^n(x, y). \tag{5.22}$$

Corollary 5.3 *The state x leads to the state y if and only if there exists some $k \geq 1$ with $P^k(x, y) > 0$.*

Proof: If $P^k(x, y) > 0$ for some $k \geq 1$, then $P_x[T_y \leq k] > 0$ and so $\rho_{xy} > 0$.

On the other hand, if $P^k(x, y) = 0$ for all $k \geq 1$, then $E_x[N(y)] = 0$, so $N(y) = 0$ almost surely. This can only happen if $\rho_{xy} = 0$.

Theorem 5.4 *If x is recurrent and x leads to y, then y is recurrent and y leads to x. Furthermore $\rho_{xy} = \rho_{yx} = 1$.*

Proof: Suppose x is recurrent and x leads to y. Then $\rho_{xx} = 1$ and $\rho_{xy} > 0$. Take the least n such that $P_x[T_y = n] > 0$. If the process gets from x to y in time n and does not return to x, then the process has left x and never returned. Thus $P_x[T_y = n](1 - \rho_{yx}) \leq 1 - \rho_{xx}$. This proves that $\rho_{yx} = 1$. In particular, y leads to x.

Start the process at x. Let A_r be the event that the process visits y between the rth and $r + 1$st visits to x (including the starting visit). Let A be the event $\exists r\, A_r$, which is the same as the event $T_y < \infty$. Then $P_x[A_r] = \theta$ is constant, by the strong Markov property. Furthermore, $\theta > 0$. The reason is that if $P_x[A_r] = 0$ for each r, then $P_x[A] = 0$, and so $\rho_{xy} = 0$. It follows that $P_x[A_r^c] = 1 - \theta < 1$. Since the events A_r are independent, $P_x[A^c] = P_x[\forall r\, A_r^c] = \prod_r P_x[A_r^c] = 0$. We conclude that $\rho_{xy} = P_x[A] = 1$.

We have proved that if x is recurrent and x leads to y, then $\rho_{yx} = 1$ and $\rho_{xy} = 1$. In particular y leads to x. Since $\rho_{yx} = 1$ and $\rho_{xy} = 1$, it follows that $\rho_{yy} = 1$. So y is recurrent.

Given a recurrent state, the set of all states that it leads to is called its *recurrent class*. We have the following theorem.

Theorem 5.5 *The states of a Markov chain consist of transient states and states belonging to a disjoint collection of recurrent classes.*

Let C be a recurrent class. Let x be a state. Let us write $\rho_C(x)$ for the probability that the process, starting from x, ever gets into the recurrent class C. Of course once it gets there, it remains there.

Theorem 5.6 *For each recurrent class C the hitting probability function ρ_C is a function that satisfies the equation*

$$P\rho_C = \rho_C \tag{5.23}$$

with the boundary condition that $\rho_C(y) = 1$ for each y in C and $\rho_C(x) = 0$ for each x belonging to another recurrence class $C' \neq C$.

This theorem shows how to find hitting probabilities of a recurrent class. The idea is to solve the equation $Pf = f$ with the boundary condition that f is one on the class and 0 on all other classes. This is a system of linear equations. The intuitive significance of the equation is exhibited by writing it as

$$f(x) = \sum_y P(x, y) f(y). \tag{5.24}$$

This shows that one is looking at the hitting probability as a function of the initial point and conditioning on the first step. Since this is using the initial point as the variable, this is sometimes called the backward equation or backward method.

In matrix language, one is finding a column vector that is an eigenvector of P with eigenvalue 1. The dimension of this space of column vectors is at least equal to the number of recurrent classes.

Remark: Say that one wants to find the hitting probability of some particular state. Then change to a new chain in which this state is an absorbing state. Then in this new chain one has a recurrent class with just one point. All the theory above applies.

5.5 Martingales

If we have a Markov chain with transition probability matrix P, then a function f with $Pf = f$ is called a harmonic function.

Theorem 5.7 *Let X_n be a Markov chain. Let f be a harmonic function on the state space. Then $Z_n = f(X_n)$ is a function of the Markov chain that is a martingale.*

Suppose that f is a bounded function satisfying $Pf = f$. Then we know from the martingale property that $f(x) = E_x[f(X_n)]$. There is a general theorem that a bounded martingale $Z_n = f(X_n)$ converges. It is also true that a martingale is constant on each recurrent class. Call its constant value on the recurrent class C by the name $f(C)$. Assume that with probability one the X_n eventually enters a recurrent class. Then it is not hard to see that

$$f(x) = \sum_C \rho_C(x) f(C). \tag{5.25}$$

This shows that in this case the bounded martingales that are functions of the process are just linear combinations of the hitting probability functions. Thus the martingale method gives nothing new.

On the other hand, from the point of interpretation the martingale method is quite different. It is a kind of forward method, in that one starts at a fixed x and follows the process forward in time, seeing where it eventually arrives.

There are examples where the martingale method is more general. Consider, for instance, simple random walk with probability p of going up and q of going

down. Assume that $p < q$, so this is a losing game. Then the solution of $Pf = f$ is $f(x) = (q/p)^x$. Say that we are interested in winning an amount b. So we start at $x < b$ and stop the random walk when we get to b. This gives a Markov chain X_n with b as an absorbing state. The martingale $(q/p)^{X_n}$ is bounded, since it can never be larger than $(q/p)^b$ or smaller than zero. So it converges as $n \to \infty$. The equation we obtain is $(q/p)^x = \rho_{xb}(q/p)^b + (1 - \rho_{xb})(q/p)^{-\infty}$. The second term is zero, so this is just $(q/p)^x = \rho_{xb}(q/p)^b$, a known equation. However notice that the term that is zero is associated not with another recurrent class, but with the process running off to $-\infty$.

5.6 The first passage time

Consider a Markov chain. Let A be a set, and let x be a point not in A. Let T_A be the first time the chain is in the set. Then $m_x = E_x[T_A]$ satisfies the equation $m_x = \sum_y P(x, y)(1 + m_y)$, provided that we make the convention that $m_y = 0$ for $y \in A$. This can be stated without any reference to the values of m on A by writing it in the form

$$m_x = \sum_{y \notin A} P(x, y)m_y + 1, \tag{5.26}$$

for $x \neq A$.

Let Q be the matrix P with the rows and columns corresponding to A taken out. Then

$$P_x[T_A > n] = \sum_y Q^n(x, y). \tag{5.27}$$

In matrix language this is $P_x[T_A > n] = Q^n 1(x)$, where Q^n is the matrix power and 1 is the column vector with all entries equal to 1. From this we see that the rate of decrease is bounded by a constant times the nth power of the largest eigenvalue of Q. Now we can calculate the expectation by summing over all $n = 0, 1, 2, 3, \ldots$. This gives

$$m_x = E_x[T_A] = \sum_y \sum_{n=0}^{\infty} Q^n(x, y) = \sum_y (1 - Q)^{-1}(x, y). \tag{5.28}$$

In matrix language this is $m_x = (1 - Q)^{-1}1(x)$. In other words, the column vector m satisfies $m = Qm + 1$. This is the same result as before.

5.7 Problems

1. HPS, Chapter 1, Problem 18
2. HPS, Chapter 1, Problem 19
3. HPS, Chapter 1, Problem 20
4. HPS, Chapter 1, Problem 22

Chapter 6

Markov chains: stationary distributions

6.1 Stationary distributions

Recall that a Markov process is determined by a transition probability matrix $P(x, y)$. It satisfies $0 \leq P(x, y) \leq 1$ and

$$\sum_y P(x, y) = 1 \tag{6.1}$$

for each x. Thus each row sums to one.

A *stationary distribution* is a probability mass function $\pi(x)$ such that

$$\sum_x \pi(x)P(x, y) = \pi(y) \tag{6.2}$$

for each y. Furthermore,

$$\sum_x \pi(x) = 1. \tag{6.3}$$

This equation has a more probabilistic interpretation. Let $\pi(y) = P[X_n = y$. It says that

$$\sum_x P[X_{n+1} = y \mid X_n = x]P[X_n = x] = P[X_n = y]. \tag{6.4}$$

In other words,

$$P[X_{n+1} = y] = P[X_n = y]. \tag{6.5}$$

The probabilities are stationary, or time invariant.

In matrix language the equation for a stationary distribution is $\pi P = \pi$. In other words, π is a row vector that is a left eigenvector of P with eigenvalue 1.

6.2 Detailed balance

The probability mass function $\pi(x)$ is said to satisfy *detailed balance* if

$$\pi(x)P(x,y) = \pi(y)P(y,x). \tag{6.6}$$

This equation may also be expressed in probabilistic language. It says that

$$P[X_{n+1} = y \mid X_n = x]P[X_n = x] = P[X_{n+1} = x \mid X_n = y]P[X_n = y]. \tag{6.7}$$

This can be expressed even more simply as

$$P[X_{n+1} = y, X_n = x] = P[X_{n+1} = x, X_n = y]. \tag{6.8}$$

From this it is clear that detailed balance is the same thing as time reversibility. There are exactly as many transitions from x to y as there are from y to x.

Theorem 6.1 *If π is a probability mass function that satisfies detailed balance, then it is a stationary distribution.*

Theorem 6.2 *If π satisfies detailed balance and is strictly positive, then*

$$P(x,y)P(y,z)P(z,x) = P(x,z)P(z,y)P(y,x). \tag{6.9}$$

Theorem 6.3 *If for some z the transition probability $P(x,z) > 0$ for all x, and if*

$$P(x,y)P(y,z)P(z,x) = P(x,z)P(z,y)P(y,x), \tag{6.10}$$

and if $c > 0$, then

$$\pi(x) = c\frac{P(z,x)}{P(x,z)} \tag{6.11}$$

satisfies

$$\pi(x)P(x,y) = \pi(y)P(y,x). \tag{6.12}$$

If this last theorem applies, and if the resulting π satisfies $\sum_x \pi(x) < \infty$, then the c may be chosen so that π is a stationary distribution.

6.3 Positive recurrence and stationary distributions

A recurrent state x of a Markov chain is *positive recurrent* if $m_x = E_x[T_x] < \infty$. It is *null recurrent* if $m_x = E_x[T_x] = \infty$.

Theorem 6.4 *If x is positive recurrent, and if x leads to y, then y is positive recurrent.*

According to this theorem, if one state in a recurrent class is positive recurrent, then all states in this class are positive recurrent. So we may speak of the class as being positive recurrent.

Theorem 6.5 *If a recurrent class C is positive recurrent, then it has a stationary distribution π_C given by*

$$\pi_C(x) = \frac{1}{m_x} \tag{6.13}$$

for x in C.

Theorem 6.6 *Consider a Markov chain with a collection of positive recurrent classes. For each recurrent class C there is a stationary distribution π. Let α_C be coefficients with $\alpha_C \geq 0$ and with $\sum_C \alpha_C = 1$. Then*

$$\pi(x) = \sum_C \alpha_C \pi_C(x) \tag{6.14}$$

is a stationary distribution. All stationary distributions are of this form.

We begin to get the following picture of a Markov chain. There are a collection of recurrent classes. For each recurrent class there is a bounded function f that solves $Pf = f$ and is one on the given class, zero on the others. For each positive recurrent class there is a stationary distribution π that solves $\pi P = \pi$ and is concentrated on the class.

In the case when there are only finitely many states, every recurrent class is positive recurrent.

6.4 The average time in a state

Theorem 6.7 *Consider a Markov chain starting at x. Let y be a positive recurrent state belonging to a positive recurrent class C. Let $N_n(y)$ be the number of visits to y in times 1 up to n. Then*

$$\lim_{n\to\infty} \frac{N_n(y)}{n} = 1_{T_y<\infty}\pi_C(y). \tag{6.15}$$

Corollary 6.1 *Consider a Markov chain starting at x. Let y be a positive recurrent state belonging to a positive recurrent class C. Then*

$$\lim_{n\to\infty} \frac{\sum_{k=1}^n P^k(x,y)}{n} = \rho_C(x)\pi_C(y). \tag{6.16}$$

In matrix language this result says that if all the recurrent classes C are positive recurrent, then

$$\lim_{n\to\infty} \frac{\sum_{k=1}^n P^k}{n} = \sum_C \rho_C \pi_C, \tag{6.17}$$

where the ρ_C are column vectors and the π_C are row vectors.

6.5 Examples

- Random walk. The random walk on the integers is transient except in the symmetric case, when it is null recurrent.
- Gambler's ruin. The gambler's ruin is a random walk on the integers where the game stops when a fixed amount is won or lost. All states are transient except for the winning and losing states. These are each a positive recurrent class consisting of one point.
- The single server queue. Let μ be the expected number of clients between services. If $\mu > 1$ it is transient; the queue length goes to infinity. If $\mu = 1$ it is null recurrent; the queue empties, but less and less frequently. If $\mu < 1$ it is positive recurrent. In this last case there is a stationary distribution.
- The branching process. There is one positive recurrent class consisting of zero individuals. All other states are transient. Let μ be the expected number of children of an individual. If $\mu > 1$ there is a possibility of a population explosion. If $\mu \leq 1$ there can only be extinction.

6.6 Convergence to a stationary distribution

The analysis of a Markov chain consists of two parts. If we start with a transient state, then it is interesting to see whether it reaches a recurrent class, and if so, which one. If it reaches a recurrent class, then it will stay there. So then the only question is how it spends its time.

If the class is a positive recurrent class, then the proportion of time in each state is given by the stationary distribution. We may as well consider the Markov chain to consist of just this one class. We always have that for each initial state x

$$\lim_{n\to\infty} \frac{\sum_{k=1}^{n} P^k(x,y)}{n} = \pi(y). \tag{6.18}$$

However in some cases there is a better result. This result needs a hypothesis that says that the process is sufficiently irregular in time.

Say, for instance, that for each state there is some non-zero probability $P(x,x) > 0$ of remaining at that state. Then if there is some non-zero probability $P^{n_0}(x,y) > 0$ of getting to another state y in n_0 steps, then for each $n \geq n_0$ there is a non-zero probability $P^n(x,y) > 0$ of getting to y in n steps. This is just because one could have been delayed at x for the first $n - n_0$ time steps and then gotten to y in the remaining n_0 time steps.

Theorem 6.8 *Suppose the Markov chain consists of a single positive recurrent class. Assume that for each pair of states x, y there exists a number n_0 such that for each $n \geq n_0$ we have*

$$P^n(x,y) > 0. \tag{6.19}$$

Then for each initial state x and final state z

$$\lim_{n\to\infty} P^n(x,z) = \pi(z). \tag{6.20}$$

Proof: Let X_n be the Markov chain starting at x. Let Y_n be another copy of the chain starting at a random point with probability distribution π. Run the two chains independently. Then after n steps the probability distribution of X_n is $P[X_n = z] = P^n(x,z)$ and the probability distribution of Y_n is $P[Y_n = z] = \pi(z)$.

This chain has a stationary distribution $\pi(x)\pi(y)$, in which the two components are independent and each individually have the stationary distribution. Therefore each state is positive recurrent. In particular, each state is recurrent.

It follows from the hypotheses of the theorem that each state (x,y) of this chain leads to each other state (z,w). Let a be any state. In particular, each state (x,y) leads to the state (a,a) on the diagonal. That is, $P_{(x,y)}[T_{(a,a)} < \infty] > 0$.

Now the crucial point is that the fact that (x,y) leads to (a,a) with probability greater than zero implies that (x,y) leads to (a,a) with probability one. That is, $P_{(x,y)}[T_{(a,a)} < \infty] = 1$.

Let T be the first $n \geq 1$ such that $X_n = Y_n$. Then since $T \leq T_{(a,a)}$, it follows that $T < \infty$ with probability one.

By the strong Markov property, $P[X_n = z \mid T \leq n] = P[Y_n = z \mid T \leq n]$. It follows that $P[X_n = z, T \leq n] = P[Y_n = z, T \leq n]$. Thus we have

$$P[X_n = z] - P[Y_n = z] = P[X_n = z, n < T] - P[Y_n = z, n < T]. \tag{6.21}$$

Now we can use the general inequality $|P[A,C] - P[B,C]| \leq P[C]$. This gives

$$|P^n(x,z) - \pi(z)| = |P[X_n = z] - P[Y_n = z]| \leq P[n < T]. \tag{6.22}$$

This is the key equation.

Since $T < \infty$ with probability one, it follows that $P[n < T]$ tends to zero as $n \to \infty$. Therefore each term in the right hand size approaches zero.

This theorem says in matrix language that under suitable hypotheses

$$\lim_{n\to\infty} P^n = 1\pi, \tag{6.23}$$

where the column vector 1 multiplies the row vector π to give a matrix.

The hypothesis of the theorem is intended to rule out some kind of periodicity that would be preserved over arbitrarily long times. A typical periodicity is where there is a division of the states into even states and odd states, and the process at each time moves from an even state to an odd one, or from an odd state to an even one.

The theorem by itself does not give a very good idea of how fast $P[n < T]$ approaches zero. All it does is reduce the problem of approach to an equilibrium to a first passage problem. This first passage time is the time for the doubled process to enter the set of diagonal elements (z,z). Here is a simple result along this line.

Corollary 6.2 *Consider a Markov chain consisting of one positive recurrent class with $N < \infty$ states. Assume that there is a number k such that $P^k(x, y) \geq \delta > 0$ for every pair of states x, y. Then*

$$|P^{nk}(x, z) - \pi(z)| \leq (1 - N\delta^2)^n. \tag{6.24}$$

Proof: Again let T be the hitting time for the diagonal. We estimate $P[n < T]$. Notice that from every initial state (x, y) and every final state (z, z) on the diagonal we have $P^k(x, z)P^k(y, z) \geq \delta^2$. Therefore, in every k steps we have a chance at least $\sum_z P^k(x, z)P^k(y, z) \geq N\delta^2$ of hitting the diagonal.

Let

$$Q = P - 1\pi. \tag{6.25}$$

Then

$$Q^n = P^n - 1\pi. \tag{6.26}$$

So the rate of convergence to the stationary distribution is determined by the eigenvalue λ of Q of largest absolute value. This is the eigenvalue of P of second largest absolute value. The difference $1 - |\lambda|$ is called the spectral gap. One of the most fundamental and difficult problems of Markov process theory is to estimate the spectral gap from below.

In the case of periodicity there can be eigenvalues of absolute value one other than the obvious eigenvalue 1. This is a nuisance with which one may not care to deal with. However it is easy to destroy periodicity by changing the process to wait a random amount of time before jumping.

6.7 Problems

1. Consider a Markov chain consisting of one positive recurrent class. Assume that there is a number k and a state z such that $P^k(x, z) \geq \delta > 0$ for every state x. Show that

$$|P^{nk}(x, z) - \pi(z)| \leq (1 - \delta^2)^n. \tag{6.27}$$

2. HPS, Chapter 2, Problem 6

3. HPS, Chapter 2, Problem 19

4. HPS, Chapter 2, Problem 20

Chapter 7

The Poisson process

7.1 The Bernoulli process

It is tempting to introduce the Poisson process as a limiting case of a Bernoulli process. The Bernoulli process is just repeatedly flipping a coin that has a probability p of coming up heads.

Let time be divided into intervals of length $\Delta t > 0$. Thus we consider times $\Delta t, 2\Delta t, \ldots, n\Delta t, \ldots$. At the end of each interval there is an event that is a success with probability $p = \lambda \Delta t$ and a failure with probability $q = 1 - \lambda \Delta t$. Let $N(t)$ be the number of successes up to time $t = n\Delta t$. Let T_r be the time of the rth success.

Then $N(t)$ has the binomial probabilities

$$P[N(t) = k] = \binom{n}{k} p^k (1-p)^{n-k}. \tag{7.1}$$

We can use $p = \lambda t/n$ to write this in the form

$$P[N(t) = k] = \binom{n}{k} \frac{1}{n^k} (\lambda t)^k \left(1 - \frac{\lambda t}{n}\right)^{n-k}. \tag{7.2}$$

On the other hand, T_r has the negative binomial probabilities

$$P[T_r = t] = P[N(t - \Delta t) = r - 1]p = \binom{n-1}{r-1} p^{r-1} (1-p)^{n-r} p \tag{7.3}$$

We can also write this as

$$P[T_r = t] = \binom{n-1}{r-1} \frac{1}{n^{r-1}} (\lambda t)^{r-1} \left(1 - \frac{\lambda t}{n}\right)^{n-r} \lambda \Delta t. \tag{7.4}$$

In particular, T_1 has the (shifted) geometric probabilities

$$P[T_1 = t] = P[N(t - \Delta t) = 0]p = (1-p)^{n-1} p = \left(1 - \frac{\lambda t}{n}\right)^{n-1} \lambda \Delta t. \tag{7.5}$$

It is worth recalling that for this geometric distribution

$$P[T_1 > t] = P[N(t) = 0] = (1-p)^n = \left(1 - \frac{\lambda t}{n}\right)^n. \tag{7.6}$$

7.2 The Poisson process

The Poisson process $N(t)$ is obtained by letting $n \to \infty$ and $\Delta t \to 0$ with $\lambda = p/\Delta t$ and $t = n\Delta t$ fixed. The calculations use the fact that

$$\binom{n}{k} \frac{1}{n^k} \to \frac{1}{k!} \tag{7.7}$$

as $n \to \infty$ with k fixed. They also use the fact that

$$\left(1 - \frac{\lambda t}{n}\right)^n \to e^{-\lambda t} \tag{7.8}$$

as $n \to \infty$ with λt fixed.

Let $N(t)$ be the number of successes up to time t. Let T_r be the time of the rth success. Then $N(t)$ has the Poisson probabilities

$$P[N(t) = k] = \frac{1}{k!}(\lambda t)^k e^{-\lambda t}. \tag{7.9}$$

On the other hand, T_r has the gamma probability density

$$P[t < T_r < t + dt] = \frac{1}{(r-1)!}(\lambda t)^{r-1} e^{-\lambda t} \lambda \, dt. \tag{7.10}$$

This is easy to remember, because

$$P[t < T_r < t + dt] = P[N(t) = r - 1]\lambda \, dt. \tag{7.11}$$

In particular, T_1 has the exponential probability density

$$P[t < T_1 < t + dt] = e^{-\lambda t} \lambda \, dt. \tag{7.12}$$

It is worth recalling that for this exponential distribution

$$P[T_1 > t] = P[N(t) = 0] = e^{-\lambda t}. \tag{7.13}$$

The Poisson process has the Markov property

$$P[N(t+s) = k + m \mid N(t) = k] = P[N(s) = m]. \tag{7.14}$$

If we take $m = 0$ and $k = 0$ we get

$$P[T_1 > t + s \mid T_1 > t] = P[T_1 > s]. \tag{7.15}$$

This is the famous Markov property of exponential waiting times.

The two key properties of the Poisson process are the following:

- Poisson increments. For each $s < t$ the random variable $N(t) - N(s)$ is Poisson with mean $\lambda(t - s)$.
- Independent increments. The increments $N(t) - N(s)$ corresponding to disjoint intervals are independent.

7.3 The Poisson paradox

Now there is an interesting paradox. The waiting times between jumps in the Poisson distribution have the exponential distribution with parameter λ. Let a be a fixed number. Let T_a be the first T_r with $T_r \geq a$. So if $r > 1$ we have $T_{r-1} < a \leq T_r = T_a$. Thus $T_a - a$ is only part of the waiting time $T_r - T_{r-1}$ associated with this r. Nevertheless, $T_a - a$ has the exponential distribution.

To see this, compute

$$P[T_a > a + s] = \sum_{r=1}^{\infty} P[N(a) = r - 1, N(a + s) = r - 1]. \tag{7.16}$$

Introducing conditional probability this becomes

$$P[T_a > a+s] = \sum_{r=1}^{\infty} P[N(a+s) = r-1 \mid N(a) = r-1]P[N(a) = r-1]. \tag{7.17}$$

However $P[N(a + s) = r - 1 \mid N(a) = r - 1] = e^{-\lambda s}$, by the Markov property. Thus

$$P[T_a > a + s] = e^{-\lambda s}. \tag{7.18}$$

How can $T_a - a$ have the exponential distribution, in spite of the fact that $T_a - a$ is only part of the waiting time up to the jump time T_a? The resolution of the paradox is that the waiting interval that happens to bracket a will not have the exponential distribution; in fact it will tend to be longer than the average. The remarkable thing is that this does not depend on the particular time a.

7.4 Combining Poisson processes

Theorem 7.1 *Say that $N_1(t)$ is a Poisson process with rate λ_1, and $N_2(t)$ is a Poisson process with rate λ_2. Say that these processes are independent. Let $N(t) = N_1(t) + N_2(t)$. Then $N(t)$ is a Poisson process with rate $\lambda = \lambda_1 + \lambda_2$.*

This theorem can be proved in several ways. The first is to use independence to compute the distribution

$$P[N(t) = n] = \sum_{k=0}^{n} P[N_1(t) = k, N_2(t) = n-k] = \sum_{k=0}^{n} P[N_1(t) = k][N_2(t) = n-k]. \tag{7.19}$$

This says that

$$P[N(t) = n] = \sum_{k=0}^{n} \frac{(\lambda_1 t)^k}{k!} e^{-\lambda_1 t} \frac{(\lambda_2 t)^{n-k}}{(n-k)!} e^{-\lambda_2 t}. \tag{7.20}$$

In other words,

$$P[N(t) = n] = \frac{1}{n!} \sum_{k=0}^{n} \binom{n}{k} (\lambda_1 t)^k (\lambda_2 t)^{n-k} e^{-\lambda_1 t - \lambda_2 t}. \tag{7.21}$$

It then follows from the binomial theorem that

$$P[N(t) = n] = \frac{1}{n!}(\lambda_1 t + \lambda_2 t)^n e^{-(\lambda_1 t + \lambda_2 t)}. \tag{7.22}$$

The second proof is much simpler. Consider the waiting times between jumps. The probability that $W_1 > t$ is $e^{\lambda_1 t}$. The probability that $W_2 > t$ is $e^{-\lambda_2 t}$. Let W be the minimum of the random variables. For the minimum to be larger than t, both individual times have to be larger than t. If they are independent, then this is the product of the probabilities. This gives the result $P[W > t] = e^{-(\lambda_1 t + \lambda_2 t)}$.

The third proof is simplest of all, though it could be criticized on the basis of rigor. Suppose that the probability of a jump in the first process in the time interval from t to $t + dt$ is $\lambda_1\, dt$. Similarly, suppose that the probability of a jump in the second process in same time interval is $\lambda_2\, dt$. If the jumps are independent, then the probability of two jumps in the time interval is $\lambda_1\, dt\ \lambda_2\, dt$, which is negligible. Therefore the probability that one or the other jumps is the sum $\lambda_1\, dt + \lambda_2\, dt$.

7.5 Problems

1. HPS, Chapter 3, Problem 3
2. HPS, Chapter 3, Problem 4
3. HPS, Chapter 3, Problem 5

Chapter 8

Markov jump processes

8.1 Jump rates

The idea of a Markov jump process is that there is a space of states. For every pair of states $x \neq y$ there is a jump rate $q_{xy} \geq 0$. This is the probability per second of a jump from x to y. Thus $q_{xy}\,dt$ is the conditional probability of a jump from x to y in time from t to $t + dt$, given that the particle is at x at time t.

Example: The Poisson process is an example. The state space consists of the natural numbers. The transition from x to $x + 1$ takes place at rate λ. All other transitions have rate zero.

Define

$$q_x = \sum_{y \neq x} q_{xy}. \tag{8.1}$$

This is the probability per second of a jump from x to anywhere else. Thus $q_x\,dt$ is the conditional probability of a jump from x in time from t to $t + dt$, given that the particle is at x at time t.

The time the particle remains at x is exponential with rate q_x. The expected time the particle remains at x is thus $1/q_x$, measured in seconds.

Define

$$q_{xx} = -q_x. \tag{8.2}$$

With this convention, there is a matrix q_{xy} defined for all pairs of states x, y. This matrix is called the generator of the jump process.

This matrix has the properties that all entries with $x \neq y$ have $q_{xy} \geq 0$, while $q_{xx} \leq 0$. Furthermore,

$$\sum_y q_{xy} = 0 \tag{8.3}$$

for each x. The row sums are zero.

8.2 Hitting probabilities

Consider the process starting at x. Let T_y be the first time the particle gets to y after it has jumped from x. If x is not an absorbing state, then we set

$$\rho_{xy} = P_x[T_y < \infty]. \tag{8.4}$$

On the other hand, if x is an absorbing state with $q_x = 0$, we set $\rho_{xy} = 0$ for $x \neq y$ and $\rho_{xx} = 1$.

We say that x leads to y if $\rho_{xy} > 0$.

A state y is called recurrent if $\rho_{yy} = 1$ and transient if $\rho_{yy} < 1$.

Theorem 8.1 *Let y be a recurrent state. Let ρ_{xy} be the probability of ever getting from x to y. Then*

$$\sum_z q_{xz}\rho_{zy} = 0. \tag{8.5}$$

The boundary conditions are that $\rho_{yy} = 1$ and $\rho_{xy} = 0$ if x does not lead to y.

In matrix language the equation says that $q\rho = 0$. Here the generator q is a square matrix, and ρ is a column vector.

Proof: The probability of a transition from x in time dt that eventually leads to y is given by solving

$$\rho_{xy} = \sum_{z \neq x} q_{xz}\, dt \rho_{zy} + (1 - q_x\, dt)\rho_{xy}. \tag{8.6}$$

This can also be written as an equality of the transition rates from x to y in the form

$$q_x \rho_{xy} = \sum_{z \neq x} q_{xz}\rho_{zy}. \tag{8.7}$$

This translates to the matrix equation in the theorem.

Note: If y is not a recurrent state, then one can compute the probability of ever getting from x to y by considering a modified chain in which y has been converted to an absorbing state.

8.3 Stationary distributions

Theorem 8.2 *The equation for a stationary distribution π is*

$$\sum_x \pi(x) q_{xy} = 0. \tag{8.8}$$

In matrix language the equation says that $\pi q = 0$. Here the generator q is a square matrix, and π is a row vector.

Proof: The rate of arriving at y is equal to the rate of leaving y. This says that

$$\sum_{x \neq y} \pi(x) q_{xy} = \pi(y) q_y. \tag{8.9}$$

This translates to the matrix equation in the theorem.

We define $m_x = E_x[T_x]$ and define a recurrent state to be positive recurrent if this is finite or if the state is absorbing.

Theorem 8.3 *The stationary probability associated with a positive recurrent state (other than an absorbing state) is given by*

$$\pi(x) = \frac{1/q_x}{m_x}. \tag{8.10}$$

This theorem says that the probability of being at x is the ratio of the average time in seconds spent at x before a jump to the average total time in seconds that it takes to jump from x, wander around, and then return.

Theorem 8.4 *If a distribution π satisfies detailed balance $\pi(x)q_{xy} = \pi(y)q_{yx}$ for all pairs of states, then it satisfies the equation for a stationary distribution.*

This equation says that the process is time reversible: The rate of transitions from x to y is the same as the rate from y to x.

8.4 The branching process

The branching process involves a situation where there are x individuals at a given time. Each individual has a rate λ of splitting in two and a rate μ of dying. So the birth rate is $x\lambda$ and the death rate is $x\mu$. Clearly 0 is an absorbing state. If $\lambda \leq \mu$ the process is sure to die out. On the other hand, if $\lambda > \mu$ there is a chance of a population explosion. If there are x individuals, the probability that the line of an individual dies out is $\rho_{x0} = \rho_{10}^x$, since the lines are independent. The probability of extinction is given by solving

$$(\lambda + \mu)\rho_{10} = \lambda\rho_{2\,0} + \mu \tag{8.11}$$

and using $\rho_{20} = \rho_{10}^2$. This can be written as a quadratic equation

$$\lambda\rho_{10}^2 - (\lambda + \mu)\rho_{10} + \mu = 0. \tag{8.12}$$

The solutions are $\rho_{10} = (\mu/\lambda)$ and $\rho_{10} = 1$. The probability of extinction is the non-trivial root μ/λ.

8.5 The N server queue

Here $N \geq 1$ is the number of servers. Customers come in at rate $q_{x\,x+1} = \lambda$. If there are $x \leq N$ customers in the queue, each gets served, at the same average rate. So the service rate is $q_{x\,x-1} = x\mu$. If there are $x > N$ customers, only the first N are being served. So the service rate is then $q_{x\,x-1} = N\mu$.

If $\lambda > N\mu$, then the queue is badly overloaded, and there is some probability that the queue never empties. To see this, take $x > N$. We solve the equation

for ρ_{xN}. This will tell us that we are not almost sure to get to N, and this in turn will tell us that the chain is transient. This equation for $N+1$ is

$$(\lambda + N\mu)\rho_{N+1\,N} = \lambda\rho_{N+2\,0} + N\mu. \tag{8.13}$$

On the other hand, when $x > N$ reducing the queue by two steps involves reducing it by one step and then by one more step. So $\rho_{N+2\,N} = \rho_{N+1\,N}^2$. So we have a quadratic equation for $\rho = \rho_{N+1\,N}$ of the form $\lambda\rho^2 - (\lambda + N\mu)\rho + N\mu = 0$. The solutions are 1 and $\rho = N\mu/\lambda$. The second root is the relevant one, and it is less than one.

On the other hand, if $\lambda < N\mu$, the service can keep up and there should be a stationary distribution. The detailed balance equation is

$$\pi(x+1)(x+1)\mu = \pi(x)\lambda \tag{8.14}$$

for $x + 1 \le N$ and

$$\pi(x+1)N\mu = \pi(x)\lambda \tag{8.15}$$

for $x + 1 > N$. From the second equation we see that we get a convergent series precisely when $\lambda/(N\mu) < 1$.

8.6 The Ehrenfest chain

The Ehrenfest chain consists of a process on the $d+1$ points $0, 1, 2, 3, \ldots, d$. The idea is that there are d particles, some in box 1 and some in box 0. The state of the chain is the number of particles in box 1.

The transition rate from x to $x+1$ is $\lambda(d-x)$. This is because if there are x particles in box 1 and $d-x$ particles in box 0, each of the $d-x$ particles has the same chance of making the jump. Each particle jumps at rate λ. So the total jump rate is $(d-x)\lambda$.

The transition rate from $x+1$ to x is $\mu(x+1)$. This is because if there are $x+1$ particles in box 1 and $d-x-1$ particles in box 0, each of the $x+1$ particles has the same chance of making the jump. Each particle jumps at rate μ. So the total jump rate is $(x+1)\lambda$. The Ehrenfest chain actually corresponds to the case when $\mu = \lambda$, but it is no trouble to handle the extra generality.

The detailed balance equation then says that

$$\pi(x+1)(x+1)\mu = \pi(x)(d-x)\lambda. \tag{8.16}$$

This equation can be solved for all the $\pi(x)$ in terms of $\pi(0)$. The result is that $\pi(x) = \binom{d}{x}(\lambda/\mu)^x\pi(0)$. The conclusion is that the stationary distribution

$$\pi(x) = \binom{d}{x}\left(\frac{\lambda}{\lambda+\mu}\right)^x\left(\frac{\mu}{\lambda+\mu}\right)^{d-x} \tag{8.17}$$

is binomial.

There is also a microscopic view of the Ehrenfest chain. First we consider a rather special two state Markov jump process. Its states are 0 and 1. The

transition rate in each direction is $\lambda > 0$. Clearly the stationary measure for such a process satisfies the detailed balance condition $\pi(0)\lambda = \pi(1)\mu$. This implies that the stationary distribution is assigns probability $\pi(1) = \lambda/(\lambda + \mu)$ to 1 and $\pi(0) = 1 - \pi(1) = \mu/(\lambda + \mu)$ to 0.

The microscopic Ehrenfest chain consists of d two state Markov jump processes. We think of each chain as describing what one particle does. This chain has a total of 2^d states. Its stationary distribution is clearly the independent distribution on the 2^d states, assigning probability $\lambda/(\lambda + \mu)$ to an individual particle being in state 1 rather than 0.

The previous macroscopic Ehrenfest chain is obtained by counting the number of the individual particles that are in state one. In the stationary distribution this is the number of successes in d trials, where the probability of success on each trial is $\lambda/(\lambda + \mu)$. So it must be binomial.

8.7 Transition probabilities

The transition probabilities

$$P_{xy}(t) = P_x[X(t) = y] \tag{8.18}$$

may in principle be computed from the jump rates.

One method is to use the *backward equation*

$$\frac{d}{dt}P_{xy}(t) = \sum_z q_{xz}P_{zy}(t). \tag{8.19}$$

This equation says that to go from x to y in time t one can jump in the first interval dt to z and then proceed to y, or one can not jump at all in this first interval. In matrix language this equation is

$$\frac{d}{dt}P(t) = qP(t). \tag{8.20}$$

Another method is to use the *forward equation*

$$\frac{d}{dt}P_{xy}(t) = \sum_z P_{xz}(t)q_{zy}. \tag{8.21}$$

This equation says that to go from x to y in time t one can jump to z just before t and then jump in the last interval dt to y, or one can already be at y and remain there. In matrix language this equation is

$$\frac{d}{dt}P(t) = P(t)q. \tag{8.22}$$

Example: Let us solve the forward equation for the Poisson process. In this case all the transitions are from x to $x + 1$ with the same rate λ. The forward equation says that for $x \leq y$

$$\frac{d}{dt}P_{xy}(t) = P_{x\,y-1}(t)\lambda - P_{xy}(t)\lambda. \tag{8.23}$$

When $y = x$ the first term is not present. The initial condition is $P_{xx}(0) = 1$ and $P_{xy}(0) = 0$ for all $y \neq x$.

To solve this equation, start with the case $y = x$. The solution is obviously $P_{xx}(t) = e^{-\lambda t}$. Then it is easy to show by induction as we consider $y = x + 1, x + 2, x + 3, \ldots$ that for $x \leq y$

$$P_{xy}(t) = \frac{(\lambda t)^{y-x}}{(y-x)!} \exp(-\lambda t). \tag{8.24}$$

8.8 The embedded Markov chain

The picture that emerges of a jump process is that a particle stays at x for some random time with exponential distribution with parameter q_x. If $q_x > 0$, then eventually it makes a transition to some other state y. To which state? The conditional probability that it makes a jump to state y, given that it has decided to jump, is q_{xy}/q_x. So it makes the jump according to this law. It then continues the process of waiting and jumping.

This leads to a formal definition of a Markov chain associated with the jump process. The conditional probability of a jump to y, given that there is a jump from x in time interval dt, is

$$Q_{xy} = \frac{q_{xy}}{q_x} \tag{8.25}$$

for $y \neq x$. If $q_x > 0$ we define $Q_{xx} = 0$. If $q_x = 0$, so that we have an absorbing state, then we define $Q_{xx} = 1$ and $Q_{xy} = 0$ for $y \neq x$. The matrix Q is the matrix of transition probabilities of a Markov chain. This is called the *embedded Markov chain* of the jump process.

The hitting probabilities ρ_{xy} may be calculated either with the jump process or with the embedded Markov chain. These give the same results. The reason is they are the probabilities of ever getting to a final state. Thus they do not depend at all on the timing.

The equation for the stationary distribution, on the other hand, must be expressed in terms of the jump rates. The probability $\pi(x)$ of being at a point depends not only on where the particle wanders, but on how long on the average it lingers at the state.

8.9 Problems

1. HPS, Chapter 3, Problem 10
2. HPS, Chapter 3, Problem 12
3. HPS, Chapter 3, Problem 13
4. HPS, Chapter 3, Problem 21

Chapter 9

The Wiener process

9.1 The symmetric random walk

It is tempting to introduce the Wiener process as a limit of a symmetric random walk process. This is just repeatedly flipping a fair coin and keeping track of the imbalance between heads and tails.

Let time be divided into intervals of length $\Delta t > 0$. Let space be divided into intervals of length

$$\Delta x = \sigma\sqrt{\Delta t}. \tag{9.1}$$

Here $\sigma > 0$ is a positive constant that measures the amount of diffusion. Often $\sigma^2/2$ is called the *diffusion constant*. For each interval there is a random variable ξ_i that is ± 1 with probabilities one half for either sign. These are thus identically distributed random variables with mean zero and variance one.

We take the random walk process to be

$$W_t = \xi_1 \Delta x + \xi_2 \Delta x + \cdots \xi_n \Delta x \tag{9.2}$$

where $t = n\Delta t$. Then W_t has mean zero and variance $n(\Delta x)^2 = \sigma^2 t$. Then $X(t)$ has the binomial probabilities associated to $(n + k)/2$ positive steps and $(n - k)/2$ negative steps:

$$P[X(t) = x] = \binom{n}{(n+k)/2} \left(\frac{1}{2}\right)^n \tag{9.3}$$

where $x = k\Delta x$.

9.2 The Wiener process

The Wiener process is obtained by letting $n \to \infty$ and $\Delta t \to 0$ with $n\Delta t = t$ fixed. Thus $W(t)$ has mean zero and variance $\sigma^2 t$ and is Gaussian.

Now this argument does not prove that there is such a mathematical object as the Wiener process. But Wiener proved that it exists, and while the proof

is somewhat technical, there are many versions now available that are not so terribly difficult.

Without going into the construction of the process, we can write properties that characterize it.

The two key properties of the Wiener process are:

- Gaussian increments. For each $s < t$ the random variable $W(t) - W(s)$ is Gaussian with mean zero and variance $\sigma^2(t - s)$.
- Independent increments. The increments $W(t) - W(s)$ corresponding to disjoint intervals are independent.

It is often convenient to specify that the Wiener process has a particular value at $t = 0$, for instance $W(0) = 0$. Sometimes we may specify another value. It is also convenient to think of the Wiener process as defined for all real t. This can be arranged by imagining another Wiener process going backward in time and joining the two at time zero.

9.3 Continuity and differentiability

Wiener proved that the process can be chosen so that the function that sends t to $W(t)$ is continuous with probability one. This is another property that should be assumed.

The continuity becomes less obvious when one realizes that the Wiener process is not differentiable. This is made plausible by the following remark.

Theorem 9.1 *Let $W(t)$ be the Wiener process. Let $h > 0$. Then the variance of the difference quotient is*

$$\operatorname{Var}\left(\frac{W(t+h) - W(t)}{h}\right) = \frac{\sigma^2}{h}. \tag{9.4}$$

Hence the variance of the difference quotient approaches infinity as $h \to 0$.

9.4 Stochastic integrals

Even though the Wiener process is not differentiable, we would like to make sense of integrals of the form

$$\int_{-\infty}^{\infty} f(t)W'(t)\,dt. \tag{9.5}$$

The derivative of the Wiener process is called *white noise*. In order to get a well-defined random variable out of white noise, one has to integrate it.

Theorem 9.2 *Let $f(t)$ be a function that is square-integrable:*

$$\int_{-\infty}^{\infty} |f(t)|^2 \, dt < \infty. \tag{9.6}$$

Then there is a corresponding Gaussian random variable

$$\int_{-\infty}^{\infty} f(t) \, dW(t) \tag{9.7}$$

with mean zero and variance

$$\sigma^2 \int_{-\infty}^{\infty} |f(t)|^2 \, dt. \tag{9.8}$$

Theorem 9.3 *Let $f(t)$ and $g(t)$ be square-integrable functions. Then the covariance of the associated stochastic integrals is*

$$\mathrm{Cov}\left(\int_{-\infty}^{\infty} f(t) \, dW(t), \int_{-\infty}^{\infty} g(s) \, dW(s)\right) = \sigma^2 \int_{-\infty}^{\infty} f(t) g(t) \, dt. \tag{9.9}$$

If you like delta functions, you can think of this result as saying that for white noise the covariance of $W'(t)$ and $W'(s)$ is $\sigma^2 \delta(t-s)$. However it is not necessary to talk this way.

9.5 Equilibrium statistical mechanics

Equilibrium statistical mechanics is based on a very simple principle: lower energy states are more likely than higher energy states. This principle is used to define a probability density. Then the energy is a random variable.

Here are the details. The energy of a system is a function $H(x)$ of the state x of the system. The probability density is supposed to be a decreasing function of the energy. Furthermore, it should have the property that if the energy $H(x) = H_1(x_1) + H_2(x_2)$ is the sum of two terms that depend on different coordinates, then the corresponding probabilities should be independent.

A formula that makes this work is to take the probability density to be

$$\frac{1}{Z(\beta)} e^{-\beta H(x)}, \tag{9.10}$$

That is, the probability is a decreasing exponential function of the energy. The constant

$$Z(\beta) = \int e^{-\beta H(x)} \, dx \tag{9.11}$$

is chosen to make this a probability density.

This rule satisfies the factorization property needed for independence:

$$\frac{1}{Z\beta)} e^{-\beta(H_1+H_2)} = \frac{1}{Z_1(\beta)} e^{-\beta H_1} \frac{1}{Z_2(\beta)} e^{-\beta H_2}. \tag{9.12}$$

Now the only remaining problem is to identify the coefficient β. This is a long story, but the result is that

$$\beta = \frac{1}{kT}, \tag{9.13}$$

where kT is the absolute temperature measured in energy units. Perhaps one could think of this as the definition of absolute temperature.

Once we have the probability density, we can think of H as a random variable. The expectation of H is

$$E(\beta) = \frac{1}{Z(\beta)} \int H(x) e^{-\beta H(x)} \, dx. \tag{9.14}$$

Lemma 9.1 *The normalization constant $Z(\beta)$ satisfies*

$$\frac{d}{d\beta} Z(\beta) = -E(\beta) Z(\beta). \tag{9.15}$$

Theorem 9.4 *The rate of change of the energy with respect to β is*

$$\frac{d}{d\beta} E(\beta) = -\frac{1}{Z(\beta)} \int H(x)^2 e^{-\beta H(x)} \, dx + E(\beta)^2 = -\operatorname{Var}(H). \tag{9.16}$$

This shows that the expected energy $E(\beta)$ is a decreasing function of β. Thus the expected energy is an increasing function of the temperature.

9.6 The Einstein model of Brownian motion

The Einstein model of Brownian motion describes the motion of a particle. The particle is bombarded randomly by molecules. In the Einstein model the x component of the particle as a function of time is a Wiener process with a certain diffusion constant $\sigma^2/2$. The fact that the motion is so irregular is due to the fact that the molecular bombardment is extremely rapid and random.

This model is very strange from the point of view of physics. In physics the fundamental physical law is that mass times acceleration equals force. In the Einstein model the diffusing particle not only does not have an acceleration, it does not even have a velocity. Therefore it is remarkable that Einstein was able to do physics with this model. His great achievement was to derive a formula for the diffusion constant.

The formula involves the absolute temperature T, measured in degrees Kelvin. There is a constant k, called Boltzmann's constant, that converts temperature units into energy units. The formula actually involves the energy quantity kT, which is proportional to temperature. Recall that energy units are force times distance. The formula also involves a frictional coefficient that measures the drag on the particle due to the surrounding fluid. If the particle is dragged through the fluid, it experiences a frictional force proportional to the speed

with which it is dragged. This proportionality constant γ has the units of force over velocity.

The Einstein formula is

$$\frac{1}{2}\sigma_E^2 = \frac{kT}{\gamma}. \tag{9.17}$$

The diffusion constant has the units of distance over velocity, that is, distance squared over time. This reflects the fact that the distance over which diffusion takes place is proportional, on the average, to the square root of the time.

Because of this famous theory, the Wiener process is sometimes called Brownian motion. However this is misleading, because one could imagine various models of the physical process of Brownian motion. But there is only one mathematical Wiener process.

Here is how Einstein found his formula. He used an amazing trick. He considered an external force in which the Brownian motion is taking place. This force f is taken to be constant. It does not matter how large or small it is, as long as it is not zero! This force balances the frictional force to produce an average terminal velocity a. Again this can be arbitrarily small. Thus Einstein's model for the displacement $X(t)$ of a diffusing particle is really

$$X(t) - X(0) = at + W(t) - W(0), \tag{9.18}$$

where for the purposes of the derivation a is non-zero, but possibly very small. Strictly speaking the particle has no velocity, since the difference quotient of the displacement $(X(t+\Delta t) - X(t))/\Delta t$ does not have a limit as $\Delta t \to 0$. However the expectation of the difference quotient does, and limit of the expectation is a. So a is the velocity in some average sense.

The external force f on the particle makes it achieve the terminal average velocity a. The relation between the force and the terminal average velocity is

$$f - \gamma a = 0. \tag{9.19}$$

This just says that the external force and the average frictional force sum to zero when the terminal average velocity is reached. That is, the assumption is that the average acceleration is zero.

Now assume that there is a stationary density for the particles satisfying detailed balance. The particles diffuse, but they also drift systematically in the direction of the force at an average velocity of a. We imagine that the motion takes place in a bounded interval of space. At the endpoints the particles just reflect off, so this is like keeping the particles in a box. It may help to think of the interval as being a vertical column, and the force f being the force of gravity in a downward direction. Then the density will be higher at the bottom than at the top. The particles drift downward, but since there are more of them at the bottom, the net diffusion is upward.

The detailed balance equation, as we shall see later, is

$$a\rho(x) - \frac{1}{2}\sigma^2 \frac{\partial \rho(x)}{\partial x} = 0. \tag{9.20}$$

This says that the systematic motion of particles in the direction of the velocity a plus the diffusion in the direction from higher density to lower density gives a balance of zero transport of particles.

Now the fundamental principle of equilibrium statistical mechanics is that the probability is determined by the formula $(1/Z)\exp(-H/(kT))$, where H is the energy and kT is the absolute temperature, measured in energy units. In some sense this is the fundamental definition of temperature. For our problem the energy $H = -fx$, given by force times distance. Thus

$$\rho(x) = \frac{1}{Z}\exp(\frac{fx}{kT}). \tag{9.21}$$

Inserting this in the detailed balance equation gives

$$\frac{1}{2}\sigma^2\frac{f}{kT} = a. \tag{9.22}$$

The Einstein relation comes from eliminating a/f from these equations. However Einstein's derivation is really too clever. We shall see that there is another method that more straightforward and intuitive. This involves a slightly more detailed model of Brownian motion, due to Ornstein and Uhlenbeck.

9.7 Problems

1. The transition probability density $p^t(x, y)$ for the Wiener process satisfies

$$E_x[f(W(t))] = \int_{-\infty}^{\infty} p^t(x, y)f(y)\,dy. \tag{9.23}$$

   Write an explicit formula for $p^t(x, y)$.

2. Prove the theorems about stochastic integrals for the case when the functions $f(t)$ and $g(s)$ are piecewise constant.

3. Suppose that the energy $H(x) = \frac{1}{2}cx^2$ is quadratic. Show that in equilibrium the expected energy $E(\beta)$ is proportional to the absolute temperature T. (Recall that $1/\beta = kT$.) What is the proportionality constant? How does it depend on c.

4. HPS, Chapter 4, Problem 18

5. HPS, Chapter 4, Problem 19

6. HPS, Chapter 5, Problem 15

Chapter 10

The Ornstein-Uhlenbeck process

10.1 The velocity process

We shall describe the Ornstein-Uhlenbeck velocity process as a process describing the velocity of a diffusing particle. (This process is also called the Langevin process.) The same mathematics works for any diffusion process with linear drift.

The law of motion is mass times acceleration equals force. After multiplying by dt we get a stochastic differential equation

$$m\,dV(t) = -\gamma V(t)\,dt + dW(t). \tag{10.1}$$

Here $m > 0$ is the mass and $\gamma > 0$ is the friction coefficient. This is Newton's second law of motion. Both sides have the dimensions of momentum. The $dW(t)$ represents a random force. The corresponding diffusion constant σ_{OU}^2 has the dimensions of momentum squared over time.

Let

$$\alpha = \frac{\gamma}{m} \tag{10.2}$$

be the relaxation rate. Then the equation becomes

$$dV(t) = -\alpha V(t)\,dt + \frac{1}{m}dW(t). \tag{10.3}$$

For a linear force law like this it is easy to solve the equation explicitly by a stochastic integral. The solution is

$$V(t) = e^{-\alpha t}V(0) + \frac{1}{m}\int_0^t e^{-\alpha(t-s)}\,dW(s). \tag{10.4}$$

Theorem 10.1 *The Ornstein-Uhlenbeck velocity process starting at v is a random variable $V(t)$ defined for $t \geq 0$ that is Gaussian with mean*

$$E[V(t)] = e^{-\alpha t} v \tag{10.5}$$

and variance

$$\mathrm{Var}(V(t)) = \frac{\sigma_{OU}^2}{2\alpha m^2}(1 - e^{-2\alpha t}). \tag{10.6}$$

Proof: The mean comes from taking the mean in the first term in the solution. The mean of the stochastic integral in the second term is zero. On the other hand, the variance of the first term is zero. The variance of the stochastic integral in the second term is computed by using the properties of stochastic integrals.

Theorem 10.2 *The Ornstein-Uhlenbeck process is correlated over a time roughly equal to $2/\alpha$. Its covariance for $t_1 \geq 0$ and $t_2 \geq 0$ is*

$$\mathrm{Cov}(V(t_1), V(t_2)) = \frac{\sigma_{OU}^2}{2\alpha m^2}(e^{-\alpha|t_2 - t_1|} - e^{-\alpha(t_1 + t_2)}) \tag{10.7}$$

Proof: The covariance comes from the stochastic integral in the second term of the solution.

If we wait for a long time, we see that the mean is zero and the covariance of the process is

$$\mathrm{Cov}(V(t_1), V(t_2)) = \frac{\sigma_{OU}^2}{2\alpha m^2} e^{-\alpha|t_2 - t_1|}. \tag{10.8}$$

This identifies the stationary measure as the Gaussian measure with mean zero and variance $\sigma_{OU}^2/(2\alpha m^2)$. This relation is due to a balance between the fluctuation σ_{OU}^2/m^2 and the dissipation α.

From these results we can identify σ_{OU}^2. Consider the process in equilibrium, that is, in the stationary measure. The expectation of the kinetic energy is given by equilibrium statistical mechanics. Since the kinetic energy is quadratic, it is just

$$E[\frac{1}{2} m V(t)^2] = \frac{1}{2} kT. \tag{10.9}$$

This comes from treating the velocity as a Gaussian variable with density proportional to $\exp(-\frac{1}{2}mv^2/(kT))$. That is, the velocity is Gaussian with mean zero and variance kT/m. Since the variance of the velocity is also $(1/2)\sigma_{OU}^2/(\alpha m^2)$, this allows us to identify

$$\frac{1}{2}\sigma_{OU}^2 = \alpha m k T = \gamma k T. \tag{10.10}$$

10.2 The Ornstein-Uhlenbeck position process

The result of the above discussion is a model for the velocity of a diffusing particle. In order to compare the results with the Einstein model, one must find the position. This is given by

$$X(t) = \int_0^t V(s)\, ds. \tag{10.11}$$

However this is easy to figure out. Integrate the original stochastic differential equation directly. This gives

$$m(V(t) - V(0)) = -\gamma(X(t) - X(0)) + W(t) - W(0). \tag{10.12}$$

Solving for the position we get

$$X(t) - X(0) = \frac{1}{\gamma}(W(t) - W(0)) - \frac{m}{\gamma}(V(t) - V(0)). \tag{10.13}$$

This is not the same as the process described by Einstein, but it is rather close. In fact, if we can neglect the second term, we get that

$$X(t) - X(0) \approx \frac{1}{\gamma}(W(t) - W(0)). \tag{10.14}$$

Thus in this approximation the position process is a Wiener process with variance $(\sigma_{OU}^2/\gamma^2)t = 2(kT/\gamma)t$. This is the same as the prediction $\sigma_E^2 t$ of the Einstein theory. However in this case the Einstein formula was obtained by applying statistical mechanics to the velocity instead of to the position.

How good is this approximation? The process $V(t) - V(0)$ has a stationary distribution. This stationary distribution has mean zero and variance $\sigma_{OU}^2/(2\gamma m)$. So the process in the second term has a stationary distribution with variance $\sigma_{OU}^2\, m/(2\gamma^3) = (1/2)(m/\gamma)\sigma_E^2$. So this shows that as soon as t is larger than $(1/2)(m/\gamma)$, this is a good approximation for the displacement.

10.3 Stationary Gaussian Markov processes

There is a sense in which the stationary Ornstein-Uhlenbeck velocity process is the nicest of all possible stochastic processes. It is a stationary stochastic process with continuous time parameter and with continuous real values. Thus the covariance function for the process at two times s and t is of the form $r(t-s)$, where $r(t) = r(-t)$ is a symmetric function. It is also a Gaussian process with mean zero. Thus the covariance function completely describes the process. Finally, it is a Markov process. This forces the covariance to have the decaying exponential form $r(t) = r(0)e^{-\alpha|t|}$ for some $\alpha > 0$. Thus only one parameter is needed to describe the process: the decay rate α.

10.4 Problems

1. The transition probability density $p^t(v, w)$ for the Ornstein-Uhlenbeck velocity $V(t)$ process satisfies

$$E_v[f(V(t))] = \int_{-\infty}^{\infty} p^t(v, w) f(w)\, dw. \tag{10.15}$$

   Write an explicit formula for $p^t(v, w)$.

2. Show directly that

$$V(t) = \frac{1}{m} \int_{-\infty}^{t} e^{-\alpha(t-s)}\, dW(s) \tag{10.16}$$

   has the covariance of the stationary Ornstein-Uhlenbeck velocity process.

3. Find the covariance function for the Ornstein-Uhlenbeck position process $X(t) - X(0)$. (Hint: Write the process as a sum of two stochastic integrals. Then the covariance is a sum of four terms.)

4. Is the Ornstein-Uhlenbeck velocity process a Markov process? Discuss.

5. Is the Ornstein-Uhlenbeck position process a Markov process? Discuss.

Chapter 11

Diffusion and drift

11.1 Stochastic differential equation

We can look at stochastic differential equations of a more general form

$$dX(t) = a(X(t))\,dt + dW(t). \tag{11.1}$$

Here $W(t)$ is the Wiener process with variance $\sigma^2 t$. The function $a(x)$ is called the *drift*. If we think of x as displacement, then $a(x)$ is a kind of average velocity. The meaning of this stochastic differential equation is given by converting it to an integral equation

$$X(t) - X(t_0) = \int_{t_0}^{t} a(X(s))\,ds + W(t) - W(t_0). \tag{11.2}$$

Both sides of this equation are intended to be well-defined random variables.

If the function $a(x)$ is not linear, then it is difficult to find an explicit solution. However under appropriate hypotheses one can prove that a solution exists by iterating the integral equation and showing that these iterates converge to a solution.

11.2 Diffusion equations

If we start this process with $X(0) = x$, then we can try to compute

$$E_x[f(X(t))] = \int_{-\infty}^{\infty} p^t(x,y) f(y)\,dy. \tag{11.3}$$

Here p^t is the probability density for getting from x to near y in time t.

Theorem 11.1 *Let*

$$L = \frac{1}{2}\sigma^2 \frac{\partial^2}{\partial x^2} + a(x)\frac{\partial}{\partial x}. \tag{11.4}$$

Then

$$u(x,t) = E_x[f(X(t))] = \int_{-\infty}^{\infty} p^t(x,y) f(y)\, dy \tag{11.5}$$

satisfies the backward equation

$$\frac{\partial u}{\partial t} = Lu \tag{11.6}$$

with initial condition $u(x,0) = f(x)$.

Theorem 11.2 *Let*

$$L^\dagger = \frac{1}{2}\sigma^2 \frac{\partial^2}{\partial y^2} - \frac{\partial}{\partial y}(a(y)\,\cdot). \tag{11.7}$$

Thus $L^\dagger f = (1/2)\sigma^2 f'' - (a(y)f)'$. *Let*

$$\rho(y,t) = \int_{-\infty}^{\infty} \rho_0(x) p^t(x,y)\, dx \tag{11.8}$$

be the probability density as a function of y, *when the process is started with density* $\rho_0(x)$. *Then* $\rho(y,t)$ *satisfies the* forward equation *(or* Fokker-Planck equation*)*

$$\frac{\partial \rho}{\partial t} = L^\dagger \rho \tag{11.9}$$

with initial condition $\rho(y,0) = \rho_0(y)$.

Note that the forward equation may be written

$$\frac{\partial}{\partial t}\rho + \frac{\partial}{\partial y} J(y) = 0, \tag{11.10}$$

where

$$J(y) = -\frac{1}{2}\sigma^2 \frac{\partial}{\partial y}\rho(y) + a(y)\rho(y) = 0 \tag{11.11}$$

is the *probability current.*

11.3 Stationary distributions

Corollary 11.1 *Let*

$$L^\dagger = \frac{1}{2}\sigma^2 \frac{\partial^2}{\partial y^2} - \frac{\partial}{\partial y}(a(y)\,\cdot). \tag{11.12}$$

The equation for a stationary probability density $\rho(y)$ *is*

$$L^\dagger \rho = 0. \tag{11.13}$$

Corollary 11.2 *Say that a probability density satisfies* detailed balance*:*

$$J(y) = -\frac{1}{2}\sigma^2 \frac{\partial}{\partial y}\rho(y) + a(y)\rho(y) = 0. \tag{11.14}$$

Then it is a stationary probability density.

The nice thing is that it is possible to solve the detailed balance equation in terms of an integral. Let

$$I(y) = \int_{y_0}^{y} a(z)\, dz. \tag{11.15}$$

The solution is then

$$\rho(y) = C \exp(\frac{2}{\sigma^2} I(y)). \tag{11.16}$$

Example: Suppose that the drift is constant: $a(y) = a$ and $I(y) = ay$. Then the stationary distribution is exponential

$$\rho(y) = C \exp(\frac{2}{\sigma^2} y). \tag{11.17}$$

The total integral cannot be one, unless there are some boundary conditions that place a limit to the exponential growth. We shall see that limiting the process to a bounded interval with reflecting boundary conditions at the end points will give a stationary process.

Example: Suppose that the drift is linear: $a(y) = -\alpha y$ and $I(y) = -\alpha y^2/2$. Then

$$\rho(y) = C \exp(-\frac{\alpha}{\sigma^2} y^2). \tag{11.18}$$

If $\alpha > 0$ this makes sense as a Gaussian. This is in some sense the nicest of all stochastic processes: it is Markov, stationary, and Gaussian. We have already encountered it as the Ornstein-Uhlenbeck velocity process.

11.4 Boundary conditions

One can also think of the motion as taking place in an interval. Then one has to specify what happens when the diffusing particle gets to an end point.

The first case is that when the particle reaches the boundary point x^* it is sent to a storage place. This is the case of an absorbing boundary. The boundary conditions for this case are $u(x^*, t) = 0$ for the backward equation and $\rho(x^*, t) = 0$ for the forward equation. Usually when there is an absorbing boundary the process is transient, except for the absorbing state. Notice that the detailed balance equation, when combined with the boundary condition of vanishing at a point, gives only the zero solution. This is not eligible to be a probability density.

The second case is when the particle is reflected when it reaches the boundary point x^*. This is the case of a reflecting boundary. The boundary conditions for this case are $\partial u/\partial x(x^*, t) = 0$ for the backward equation and $J(x^*, t) = a(x^*, t) - (\sigma^2/2)\partial\rho/\partial y(x^*, t) = 0$ for the forward equation. If both boundaries are reflecting, then the process is usually recurrent.

If one has a one dimensional diffusion with reflecting boundary conditions, then solving for a stationary probability density is easy. Detailed balance says that the current is zero, and the reflecting boundary condition says that the current continues to be zero at the boundary. An example would be the Wiener

process with two reflecting boundaries. Then the statationary probability density is uniform.

11.5 Martingales and hitting probabilities

A function $f(X(t))$ of the diffusion process is a martingale if the equation $Lf = 0$ is satisfied. The general solution to this equation is given by

$$f(x) = C_1 \int_{x_0}^{x} \exp(-\frac{2}{\sigma^2} I(y))\, dy + C_2. \tag{11.19}$$

In particular, hitting probabilities define martingales. Thus if $x_0 < x < x_1$, and $f(x)$ is the probability of hitting x_1 before x_0, then $f(x)$ satisfies this equation with $f(x_0) = 0$ and $f(x_1) = 1$. Thus

$$f(x) = C_1 \int_{x_0}^{x} \exp(-\frac{2}{\sigma^2} I(y))\, dy. \tag{11.20}$$

Furthermore, we require that

$$f(x_1) = C_1 \int_{x_0}^{x_1} \exp(-\frac{2}{\sigma^2} I(y))\, dy = 1. \tag{11.21}$$

This fixes the constant C_1.

This technique also works for investigating recurrence and transience. Thus, for instance, take the case when $x_1 = +\infty$. If the integral of $\exp(-(2/\sigma^2)I(y))$ converges near $+\infty$, then this shows that there is a strictly positive probability of running off to $+\infty$. Thus the process is transient. If the integral does not converge, then the probability of going to $+\infty$. before reaching a is zero. The boundary condition at $+\infty$ is cannot be satisfied. There is a similar criterion that stays that if the integral of $\exp(-(2/\sigma^2)I(y))$ converges near $-\infty$, then there is a strictly positive probability of running off to $-\infty$. So the only way to have recurrence is to have both integrals diverge, at $+\infty$ and at $-\infty$.

It is amusing to compare this with the condition for positive recurrence, that is, for having a stationary measure. This condition says that the integral of $\exp((2/\sigma^2)I(y))$ converges at $+\infty$ and at $-\infty$. However the sign in front of the drift coefficient in the exponential is opposite. This condition is a stronger condition than the condition for recurrence. The Wiener process is an example that is recurrent but not positive recurrent.

Example: Consider diffusion on the half line $x \geq 0$ with constant drift $a \neq 0$. The solution to the martingale equation is $f(x) = -C_1 e^{-ax} + C_2$. Say that we are interested in the probability that the particle escapes to $+\infty$ before getting to 0. The solution that vanishes at 0 is $f(x) = C_1(1 - e^{-ax})$. If $a > 0$, then the solution is $f(x) = (1 - e^{-ax})$. On the other hand, if $a < 0$ the solution is $f(x) = 0$.

11.6 Problems

1. Recurrence of the Wiener process. Show that starting at any point x, the probability of reaching a point b is one.

2. Transience of the Wiener process with drift. Show that if the drift is negative, then starting at any point x, the probability of reaching a point b with $x < b$ is less than one. Calculate this probability.

3. Gambler's ruin. Consider diffusion on an interval $a < x < b$ with absorbing boundary conditions. Take the case of constant negative drift. Find the probability of reaching b as a function of x.

4. A queue. Consider diffusion on the interval $0 < x < \infty$ with reflecting boundary condition at zero. Take the case of a constant negative drift. Find the stationary probability density.

5. The double well. Consider diffusion on the line with drift $a(x) = bx - cx^3$, where $b > 0$ and $c > 0$. Sketch the drift $a(x)$. This is positive recurrent, so it has a stationary probablity density $\rho(x)$. Solve for $\rho(x)$ and sketch it.

6. Punctuated equilibrium. Consider the diffusion in the preceding problem. Take b large with b/c fixed. Sketch a typical sample path. Discuss how the process combines features of the Ornstein-Uhlenbeck process with features of the two-state Markov process.

Chapter 12

Stationary processes

12.1 Mean and covariance

Finally we leave the realm of Markov processes. We go to the realm of probability where we study processes through their covariance functions.

Consider a stochastic process $X(t)$ defined for real t. (We could also consider a stochastic process X_n defined for integer n; much of the theory is parallel.)

The *mean* of the process is

$$\mu(t) = E[X(t)]. \tag{12.1}$$

The *covariance* of the process is

$$r(s,t) = \mathrm{Cov}(X(s), X(t)). \tag{12.2}$$

The *variance* of the process is of course $r(t,t)$.

Not every function can be a variance function. It is obvious that $r(s,t) = r(t,s)$, so this symmetry property is essential. Furthermore, it is not hard to prove from the Schwarz inequality that $|r(s,t)| \le \sqrt{r(s,s)}\sqrt{r(t,t)}$. However there is another deeper property.

Theorem 12.1 *For each integrable function f the covariance $r(s,t)$ has the positivity property*

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(s)r(s,t)f(t)\,ds\,dt \ge 0. \tag{12.3}$$

Proof: Consider the linear combination

$$\int f(t)X(t)\,dt. \tag{12.4}$$

Its variance is greater than equal to zero.

12.2 Gaussian processes

In general, knowing the mean and covariance of a process is not enough to determine the process. However for a Gaussian process this always works.

A *Gaussian* process is a process that has the property that for each $t_1, \ldots, t_n$ the random variables $X(t_1), \ldots, X(t_n)$ have a joint Gaussian (normal) distribution.

The mean vector of $X(t_1), \ldots, X(t_n)$ is the vector $\mu(t_1), \ldots, \mu(t_n)$ with components $\mu(t_j)$. The covariance matrix of $X(t_1), \ldots, X(t_n)$ is the matrix with entries $r(t_i, t_j)$. Since the mean vector and covariance matrix of jointly Gaussian random variables determine the density function, a Gaussian process is determined by its mean and covariance functions. (Technically, this is true only for properties of the process that are determined by what happens at a finite number of time instants.)

Thus in the following there are two points of view. The first is to think of a general process with finite mean and covariance functions. Then these functions are giving seriously incomplete information about the process. However there are circumstances where one has a non-Gaussian process, but one still gets useful information from knowing the covariance. The second is to think of a Gaussian process. This is much more special, but one is sure that the mean and covariance is giving a complete description.

12.3 Mean and covariance of stationary processes

A process is stationary if for every function $f(x_1, \ldots, x_n)$ for which the expectations are defined we have

$$E[f(X(t_1), X(t_2), \ldots, X(t_n))] = E[f(X(t_1 - s), X(t_2 - s), \ldots, X(t_n - s))]. \tag{12.5}$$

In particular the mean satisfies $\mu(t) = \mu(0)$ and so

$$E[X(t)] = \mu, \tag{12.6}$$

independent of t. We shall usually consider a situation when this constant is zero.

The covariance of the process satisfies $r(t_1, t_2) = r(t_1 - t_2, 0)$. Thus we may write $r(t)$ for $r(t, 0)$ and obtain

$$\mathrm{Cov}(X(t_1), X(t_2)) = r(t_1 - t_2). \tag{12.7}$$

Clearly the function $r(t) = r(-t)$ is symmetric. The variance of the process is of course the constant $r(0)$. We have the inequality $|r(t)| \leq r(0)$. Furthermore, the covariance $r(s - t)$ has the positivity property

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(s) r(s - t) f(t)\, ds\, dt \geq 0. \tag{12.8}$$

12.4 Convolution

The *convolution* of two functions f and g is the function $f * g$ defined by

$$(f * g)(t) = \int_{-\infty}^{\infty} f(t-s)g(s)\,ds. \tag{12.9}$$

This is a commutative product, since a change of variable gives

$$(f * g)(t) = \int_{-\infty}^{\infty} f(u)g(t-u)\,du. \tag{12.10}$$

Ordinarily we want to assume that one of the functions is integrable and the other at least bounded.

The interpretation of $f * g$ is that it is a weighted integral of translates of f with weight function g. Of course it can also be thought of as a weighted integral of translates of g with weight function f.

Sometimes we want one of the functions to satisfy $f(t) = 0$ for $t < 0$. In this case we say that the convolution is *causal*. Then the convolution takes the form

$$(f * g)(t) = \int_{-\infty}^{t} f(t-s)g(s)\,ds \tag{12.11}$$

or

$$(f * g)(t) = \int_{0}^{\infty} f(u)g(t-u)\,du. \tag{12.12}$$

We see that in this case the convolution $(f*g)(t)$ only depends on $g(s)$ for $s \le t$, or on $g(t-u)$ for $u \ge 0$. In other words, the convolution $f * g$ at time t is a weighted integral of the values of g in the past of t.

12.5 Impulse response functions

Let $Y(t)$ be a stationary process with mean zero. Let $h(t)$ be a real integrable function, which we call the *impulse response function*. Then we define a new process by the convolution of the impulse response function with the process. The result

$$X(t) = \int_{-\infty}^{\infty} h(t-u)Y(u)\,du = \int_{-\infty}^{\infty} h(v)Y(t-v)\,dv \tag{12.13}$$

is another stationary process. If the first process has covariance function r_Y, the new process has covariance function r_X. We can express r_X in terms of r_Y by the formula

$$r_X(t-s) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} h(t-u)h(s-v)r_Y(u-v)\,du\,dv. \tag{12.14}$$

The impulse response function is causal if $h(t) = 0$ for all $t < 0$. Then we may write the new process as

$$X(t) = \int_{-\infty}^{t} h(t-u)Y(u)\,du = \int_{0}^{\infty} h(v)Y(t-v)\,dv \tag{12.15}$$

The covariances are then related by

$$r_X(t-s) = \int_{-\infty}^{t}\int_{-\infty}^{s} h(t-u)h(s-v)r_Y(u-v)\,du\,dv. \tag{12.16}$$

One important but rather singular case is when $Y(t) = dW(t)/dt$ is white noise. We take $h(t)$ to be a real square integrable function. Then the convolution

$$X(t) = \int_{-\infty}^{\infty} h(t-u)\,dW(u).$$

is another stationary process. Since the white noise process has covariance $\sigma^2\delta(u-v)$, the covariance of the resulting process is

$$r_X(t-s) = \sigma^2 \int_{-\infty}^{\infty} h(t-u)h(s-u)\,du.$$

Again the impulse response function is causal if $h(t) = 0$ for all $t < 0$. Then we may write the new process as

$$X(t) = \int_{-\infty}^{t} h(t-u)\,dW(u). \tag{12.17}$$

The covariance is given by

$$r_X(t-s) = \sigma^2 \int_{-\infty}^{\min(s,t)} h(t-u)h(s-u)\,du. \tag{12.18}$$

In particular, for $t \geq 0$ this is

$$r_X(t) = \sigma^2 \int_{-\infty}^{0} h(t-u)h(-u)\,du. \tag{12.19}$$

Example: We have the standard example of the Ornstein-Uhlenbeck process. This is given by

$$X(t) = \int_{-\infty}^{t} e^{-\alpha(t-s)}\,dW(s). \tag{12.20}$$

Thus the impulse response function is $h(t) = e^{-\alpha t}$ for $t \geq 0$ and zero elsewhere. The corresponding covariance function is

$$r_X(t) = \sigma^2 \int_{-\infty}^{0} e^{-\alpha(t-s)}e^{\alpha s}\,ds = \frac{\sigma^2}{2\alpha}e^{-\alpha t} \tag{12.21}$$

for $t \geq 0$.

We can ask the converse question: Given a covariance function $r(t)$, can it be represented by some causal impulse response function $h(t)$ in the form

$$r(t) = \sigma^2 \int_{-\infty}^{0} h(t-u)h(-u)\, du \tag{12.22}$$

for $t \geq 0$. This would say that this covariance function could be explained as that of a causal process formed from white noise. The answer turns out to be somewhat surprising. It will appear in the next chapter. However here is one easy remark. We shall see that it can be accomplished by an impulse response function with $h(0) \neq 0$ only for covariances with a slope discontinuity at zero. The reason is that we can compute the right hand derivative to be

$$r'(0+) = -\frac{1}{2}\sigma^2 h(0)^2. \tag{12.23}$$

Example: Take $c > 0$ and consider the covariance function

$$r(t) = \frac{c^2}{t^2 + c^2} \tag{12.24}$$

This cannot be explained by a causal impulse response function with $h(0) \neq 0$. It is too smooth. We shall see in the next chapter that it cannot be explained by any causal impulse response function.

12.6 Problems

1. HPS, Chapter 4, Problem 9
2. HPS, Chapter 4, Problems 20(b)(c)
3. HPS, Chapter 4, Problem 20(d)
4. HPS, Chapter 5, Problem 6
5. HPS, Chapter 5, Problem 7.
6. Consider the identity $r'(0+) = -(1/2)\sigma^2 h(0)^2$ for the case of the Ornstein-Uhlenbeck process with $r(t) = r(0)e^{-\alpha|t|}$ and $h(t) = e^{-\alpha t}$. Show that this gives the fluctuation-dissipation relation.

Chapter 13

Spectral analysis

13.1 Fourier transforms

Let f be a complex function of time t such that the quantity

$$M^2 = \int_{-\infty}^{\infty} |f(t)|^2 \, dt < \infty. \tag{13.1}$$

Define the *Fourier transform* $\hat{f}$ as another complex function of angular frequency λ given by

$$\hat{f}(\lambda) = \int_{-\infty}^{\infty} e^{-i\lambda t} f(t) \, dt. \tag{13.2}$$

Then

$$M^2 = \int_{-\infty}^{\infty} |\hat{f}(\lambda)|^2 \, \frac{d\lambda}{2\pi} < \infty \tag{13.3}$$

with the same constant M^2.

The big theorem about Fourier transforms is the *inversion formula*. It says that every square integrable function of t can be represented as as an inverse Fourier transform, that is, as an integral involving functions $e^{i\lambda t}$ for varying angular frequencies λ.

Theorem 13.1 *Let $f(t)$ be a square-integrable function, and let $\hat{f}(\lambda)$ be its Fourier transform. Then*

$$f(t) = \int_{-\infty}^{\infty} e^{i\lambda t} \hat{f}(\lambda) \, \frac{d\lambda}{2\pi}. \tag{13.4}$$

Recall that $e^{i\lambda t} = \cos(\lambda t) + i \sin(\lambda t)$, so this is an expansion of an arbitrary function of t in terms of trigonometric functions $\cos(\lambda t)$ and $\sin(\lambda t)$ with varying angular frequencies λ.

If $f(t)$ is real, then $\hat{f}(-\lambda) = \overline{\hat{f}(\lambda)}$, so if we write $\hat{f}(\lambda) = a(\lambda) - ib(\lambda)$ and $\hat{f}(-\lambda) = a(\lambda) + ib(\lambda)$, we get the inversion formula in *real form* as

$$f(t) = \int_0^\infty [\cos(\lambda t)a(\lambda) + \sin(\lambda t)b(\lambda)] \, \frac{d\lambda}{\pi}. \tag{13.5}$$

If in addition $f(t)$ is even, so that $f(-t) = f(t)$, then $\hat{f}(\lambda)$ is real and even, and so $\hat{f}(\lambda) = a(\lambda)$ and $b(\lambda) = 0$. Thus we have the *real even form*

$$f(t) = \int_{-\infty}^\infty \cos(\lambda t)\hat{f}(\lambda) \, \frac{d\lambda}{2\pi} = \int_0^\infty \cos(\lambda t)\hat{f}(\lambda) \, \frac{d\lambda}{\pi}. \tag{13.6}$$

The real expressions do much to demystify the Fourier transform, but in practice everybody uses the complex notation.

13.2 Convolution and Fourier transforms

The Fourier transform is particularly useful for simplifying convolutions. Recall that the convolution of f and g is the function $f * g$ defined by

$$(f * g)(t) = \int_{-\infty}^\infty f(t - s)g(s) \, ds. \tag{13.7}$$

The fundamental result is that the Fourier transform of the convolution is given by

$$\widehat{(f * g)}(\lambda) = \hat{f}(\lambda)\hat{g}(\lambda). \tag{13.8}$$

Thus the Fourier transform of a convolution is the product of the Fourier transforms.

Another useful fact is that the Fourier transform of the function $\overline{f(-t)}$ is the complex conjugate $\overline{\hat{f}(\lambda)}$ of the Fourier transform of $f(t)$.

13.3 Smoothness and decay

Here are two more big theorems about Fourier transforms.

Theorem 13.2 *Let $f(t)$ be a square integrable function, that is, a function such that*

$$\int_{-\infty}^\infty |f(t)|^2 \, dt < \infty. \tag{13.9}$$

Then its Fourier transform $\hat{f}(\lambda)$ is also square integrable. Furthermore,

$$\int_{-\infty}^\infty |f(t)|^2 \, dt = \int_{-\infty}^\infty |\hat{f}(\lambda)|^2 \, \frac{d\lambda}{2\pi}. \tag{13.10}$$

There is a corresponding result for the inverse Fourier transform.

Theorem 13.3 *Let $f(t)$ be an integrable function, that is, a function such that*

$$\int_{-\infty}^{\infty} |f(t)|\, dt < \infty. \tag{13.11}$$

Then its Fourier transform $\hat{f}(\lambda)$ is bounded and continuous. There is a corresponding result for the inverse Fourier transform.

Proof: It is clear that

$$|\hat{f}(\lambda)| = |\int_{-\infty}^{\infty} e^{-i\lambda t} f(t)\, dt| \leq \int_{-\infty}^{\infty} |f(t)|\, dt \tag{13.12}$$

for each λ. Since the bound on the right hand side does not depend on λ, the function $\hat{f}$ is bounded.

The fact that $\hat{f}$ is continuous is proved in more advanced treatments. (It is an immediate consequence of the dominated convergence theorem.)

The second theorem has some useful corollaries.

Corollary 13.1 *Let f be a function such $f(t)$ and $t^m f(t)$ are integrable. Then each derivative of the Fourier transform $\hat{f}(\lambda)$ up to order m is bounded and continuous. There is a corresponding result for the inverse Fourier transform.*

Proof: The Fourier transform of $(-it)^m f(t)$ is the mth derivative of $\hat{f}(\lambda)$.

Corollary 13.2 *Let f be a function such $f(t)$ and its mth derivative $f^{(m)}(t)$ are integrable. Then the Fourier transform $\hat{f}(\lambda)$ is bounded and also satisfies a bound $|\hat{f}(\lambda)| \leq C/|t|^m$. It thus goes to zero at infinity at least at this rate. There is a corresponding result for the inverse Fourier transform.*

Proof: The Fourier transform of the mth derivative of $f(t)$ is $(i\lambda)^m \hat{f}(\lambda)$.

What do these corollaries mean intuitively? Let us look at the case of the inverse Fourier transform. If $\hat{f}(\lambda)$ and $\lambda^m \hat{f}(\lambda)$ are integrable, then the inverse Fourier transform $f(t)$ has m continuous derivatives. This says that if there are few high frequencies λ in $\hat{f}(\lambda)$, then the only kind of function $f(t)$ that you can synthesize out of sines and cosines is a rather smooth function. In particular, to synthesize a discontinuous function $f(t)$ you need a lot of high frequencies, so $\hat{f}(\lambda)$ cannot even be integrable.

On the other hand, if $\hat{f}(\lambda)$ and its derivatives through order m are integrable, then the inverse Fourier transform $f(t)$ approaches zero at infinity at least as fast as a constant times $1/|t|^m$. This says that the smoothness of the Fourier transform $\hat{f}(\lambda)$ produces a lot of cancelation in $f(t)$ at long distances. However a sharp frequency cutoff in $\hat{f}(\lambda)$ leaves persistent oscillations in $f(t)$ that do not approach zero rapidly.

13.4 Some transforms

Most Fourier transforms are difficult to compute. However there are a few cases where it is quite easy. Here are some of them. Notice that whenever you have computed a Fourier transform, you have also computed an inverse Fourier transform. The only change is a sign and a factor of 2π. In these examples each of the functions whose Fourier transform is being computed is a probability density.

Take $\alpha > 0$. The Fourier transform of the function that is $f(t) = \alpha e^{-\alpha t}$ for $t \geq 0$ and zero elsewhere is

$$\int_0^\infty e^{-i\lambda t} \alpha e^{-\alpha t}\, dt = \frac{\alpha}{\alpha + i\lambda}. \tag{13.13}$$

Notice that the function $f(t)$ is discontinuous, and its Fourier transform $\hat{f}(\lambda)$ is not integrable.

It follows that the Fourier transform of the function $(1/2)(f(t) + f(-t)) = (1/2)\alpha e^{-\alpha|t|}$ is $(1/2)(\hat{f}(\lambda) + \overline{\hat{f}(\lambda)})$, which gives

$$\int_{-\infty}^\infty e^{-i\lambda t} \alpha e^{-\alpha|t|}\, dt = \frac{\alpha^2}{\alpha^2 + \lambda^2}. \tag{13.14}$$

In this case the Fourier transform $\hat{f}(\lambda)$ is integrable, and the function $f(t)$ is continuous. On the other hand, the derivative of $f(t)$ is not continuous, and correspondingly $\lambda\hat{f}(\lambda)$ is not integrable.

Take $a > 0$. The Fourier transform of the function that is $f(t) = 1/a$ $0 \leq t \leq a$ and zero elsewhere is

$$\int_0^a e^{-i\lambda t} \frac{1}{a}\, dt = \frac{1}{ia\lambda}(1 - e^{-ia\lambda}). \tag{13.15}$$

The function $f(t)$ is discontinuous, and its Fourier transform $\hat{f}(\lambda)$ is not integrable.

It follows that the Fourier transform of the function $(1/2)(f(t) + f(-t)) = 1/(2a)$ for $-a \leq t \leq t$, zero elsewhere, is $(1/2)(\hat{f}(\lambda) + \overline{\hat{f}(\lambda)})$, which gives

$$\int_{-a}^a e^{-i\lambda t} \frac{1}{2a}\, dt = \frac{\sin(a\lambda)}{a\lambda}. \tag{13.16}$$

The function is not continuous, and its Fourier transform is not integrable.

As an example of the convolution theorem, take $f(t) = 1/(2a)$ for $|t| \leq a$ and zero otherwise. Then the convolution $(f * f)(t)$ has Fourier transform $\hat{f}(\lambda)^2$. This says that

$$\int_{-2a}^{2a} e^{-i\lambda t} \frac{1}{2a}(1 - \frac{1}{2a}|t|)\, dt = \frac{\sin^2(a\lambda)}{(a\lambda)^2}. \tag{13.17}$$

This time the Fourier transform $\hat{f}(\lambda)$ is integrable, and the function $f(t)$ is continuous. However, the derivative of $f(t)$ is not continuous, and correspondingly $\lambda\hat{f}(\lambda)$ is not integrable.

13.5 Spectral densities

Let $X(t)$ be a real stationary process, indexed by the time parameter t. Assume that the mean is zero, so E[X(t)]=0. Then the *covariance function* is given by

$$r(t-s) = E[X(t)X(s)]. \tag{13.18}$$

If $r(t)$ is the covariance function of a real stationary process, then $r(t)$ is real and even. Define the *spectral density* function to be $\hat{r}(\lambda)$. This is real and even. Then the covariance function may be expressed as an integral over frequencies by

$$r(t) = \int_{-\infty}^{\infty} \cos(\lambda t)\hat{r}(\lambda)\,\frac{d\lambda}{2\pi} = \int_{0}^{\infty} \cos(\lambda t)\hat{r}(\lambda)\,\frac{d\lambda}{\pi}. \tag{13.19}$$

Note that in the present treatment the spectral density is always the density with respect to $d\lambda/(2\pi)$. Whenever one does an integral over the frequency parameter one must use this weight, with the $1/(2\pi)$ factor.

It is a remarkable fact (demonstrated below) that for a covariance function the spectral density is positive: $\hat{r}(\lambda) \geq 0$. Thus one can think of the variance as the integral over variances associated with each frequency:

$$r(0) = \int_{-\infty}^{\infty} \hat{r}(\lambda)\,\frac{d\lambda}{2\pi} = \int_{0}^{\infty} \hat{r}(\lambda)\,\frac{d\lambda}{\pi}. \tag{13.20}$$

Since $X(t)$ is stationary random, there is no reason to believe that it would have a Fourier transform that is a function. However one can define the *truncated Fourier transform*

$$\hat{X}_T(\lambda) = \int_{-T}^{T} e^{-i\lambda t} X(t)\,dt \tag{13.21}$$

and this is a well-defined complex-valued random variable. Its variance is $E[|X_T(\lambda)|^2]$. There is no reason to expect this variance to have a limit as $T \to \infty$. However we shall consider the limit of the *normalized variance* $E[|X_T(\lambda)|^2]/(2T)$.

Theorem 13.4 *Consider a mean zero stationary process $X(t)$ with integrable correlation function $r(t)$. Then the corresponding spectral density $\hat{r}(\lambda) \geq 0$ for all λ. This positivity has an interpretation in terms of the process. Namely, let $\hat{X}_T(\lambda)$ be the truncated Fourier transform of the process. Then*

$$\frac{1}{2T} E[|X_T(\lambda)|^2] \to \hat{r}(\lambda) \tag{13.22}$$

as $T \to \infty$.

This theorem proves that the spectral density function is positive. Furthermore, it gives an interpretation of the spectral density function as a limit of the normalized variance of the truncated Fourier transform $\hat{X}_T(\lambda)$ as $T \to \infty$.

Proof: We can calculate the normalized variance

$$\frac{1}{2T}E[|X_T(\lambda)|^2] = \frac{1}{2T}\int_{-T}^{T}\int_{-T}^{T} e^{-i\lambda(t-s)}r(t-s)\,dt\,ds. \tag{13.23}$$

Change to variables $u = t - s$ and $v = t + s$. We obtain

$$\frac{1}{2T}E[|X_T(\lambda)|^2] = \frac{1}{4T}\int_{-2T}^{2T}\int_{-2T+|u|}^{2T-|u|} e^{-i\lambda u}r(u)\,dv\,du. \tag{13.24}$$

This gives the explicit formula

$$\frac{1}{2T}E[|X_T(\lambda)|^2] = \int_{-2T}^{2T}(1 - \frac{|u|}{2T})e^{-i\lambda u}r(u)\,du. \tag{13.25}$$

For each u the limit of the integrand as $T \to \infty$ is just $e^{-i\lambda u}r(u)$. Hence the limit of the integrals is

$$\lim_{T\to\infty}\frac{1}{2T}E[|X_T(\lambda)|^2] = \int_{-\infty}^{\infty} e^{-i\lambda u}r(u)\,du = \hat{r}(\lambda). \tag{13.26}$$

13.6 Frequency response functions

Let $Y(t)$ be a stationary process. Let $h(t)$ be a real function, which we call the *impulse response function*. Then the convolution

$$X(t) = \int_{-\infty}^{\infty} h(t-u)Y(u)\,du \tag{13.27}$$

is another stationary process. If the first process has covariance function r_Y, the new process has covariance function r_X. We can express r_X in terms of r_Y by the formula

$$r_X(t-s) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} h(t-u)h(s-v)r_Y(u-v)\,du\,dv. \tag{13.28}$$

This is a messy integral. However consider the Fourier transform $\hat{h}(\lambda)$ which we call the *frequency response function*. Then it is easy to calculate that

$$\hat{r}_X(\lambda) = |\hat{h}(\lambda)|^2\hat{r}_Y(\lambda). \tag{13.29}$$

Thus the covariance of the new process has a very simple description in the frequency domain.

Let $h(t)$ be an impulse response function that is causal, that is, $h(t) = 0$ for $t < 0$. Then the relation between the processes is

$$X(t) = \int_{-\infty}^{t} h(t-u)Y(u)\,du. \tag{13.30}$$

We can now express r_X in terms of r_Y by the formula

$$r_X(t-s) = \int_{-\infty}^{t} \int_{-\infty}^{s} h(t-u)h(s-v)r_Y(u-v)\,du\,dv. \tag{13.31}$$

What is causality in terms of the frequency response function? In that case

$$\hat{h}(\lambda) = \int_0^{\infty} e^{-i\lambda t}h(t)\,dt. \tag{13.32}$$

The special feature now is that one is allowed to take λ complex, with negative imaginary part, and the integral still converges, in fact very rapidly. Write $\lambda = \lambda_1 + i\lambda_2$ with $\lambda_2 < 0$. Then for $t \geq 0$

$$|e^{-i\lambda t}| = e^{\lambda_2 t} \leq 1 \tag{13.33}$$

is not only bounded by one, it also decreases exponentially as $t \to \infty$. So the integral is very rapidly convergent.

Now look at the case where the input is white noise. Let h be a square integrable function. Then the general (not necessarily causal) convolution

$$X(t) = \int_{-\infty}^{\infty} h(t-u)\,dW(u) \tag{13.34}$$

is a stationary process. This has covariance function r_X. We can express r_X by

$$r_X(t) = \sigma^2 \int_{-\infty}^{\infty} h(t-u)h(-y)\,du. \tag{13.35}$$

Again this is simpler when expressed in terms of the frequency response function. We see that

$$\hat{r}_X(\lambda) = |\hat{h}(\lambda)|^2\sigma^2. \tag{13.36}$$

The white noise has spectral density σ^2, a constant. Since all frequencies contribute equally, this is a justification for the terminology white noise.

Let $h(t)$ be an impulse response function that is causal, that is, $h(t) = 0$ for $t < 0$. Then the relation between the processes is

$$X(t) = \int_{-\infty}^{t} h(t-u)\,dW(u). \tag{13.37}$$

We can now express $r_X(t)$ for $t \geq 0$ by

$$r_X(t) = \sigma^2 \int_{-\infty}^{0} h(t-u)h(-u)\,du. \tag{13.38}$$

The special feature of the frequency response function $\hat{h}(\lambda)$ in the causal case is again that it is defined for all λ in the lower half plane.

Example: We have the standard example of the Ornstein-Uhlenbeck process. This is given by

$$dX(t) = \int_{-\infty}^{t} e^{-\alpha(t-s)}\, dW(s). \tag{13.39}$$

Thus the impulse response function is $h(t) = e^{-\alpha t}$ for $t \geq 0$ and zero elsewhere. The corresponding covariance function is

$$r_X(t) = \sigma^2 \int_{-\infty}^{0} e^{-\alpha(t-s)} e^{\alpha s}\, ds = \frac{\sigma^2}{2\alpha} e^{-\alpha t} \tag{13.40}$$

for $t \geq 0$. The frequency response function is

$$\hat{h}(\lambda) = \frac{1}{i\lambda + \alpha}, \tag{13.41}$$

which is defined for all λ in the lower half plane. The corresponding spectral density is

$$\hat{r}_X(\lambda) = \sigma^2 |\hat{h}(\lambda)|^2 = \sigma^2 \frac{1}{\lambda^2 + \alpha^2} = r(0) \frac{2\alpha}{\lambda^2 + \alpha^2}. \tag{13.42}$$

In the physics literature this is called *Lorentzian*. Recall that integrals involving λ involve the weight $d\lambda/(2\pi)$. So the Lorentzian in the literature is often

$$\frac{\hat{r}_X(\lambda)}{2\pi} = r(0) \frac{1}{\pi} \frac{\alpha}{\lambda^2 + \alpha^2}. \tag{13.43}$$

13.7 Causal response functions

It is interesting to ask when a correlation function $r(t)$ comes from convolving a causal impulse response function with white noise. This is equivalent to the representation

$$r(t) = \sigma^2 \int_{-\infty}^{0} h(t-s) h(-s)\, ds. \tag{13.44}$$

for $t \geq 0$. Here $h(t)$ is a square-integrable function such that $h(t) = 0$ for all $t < 0$.

There is an interesting theorem that describes when this happens. To describe this theorem, we need a preliminary definition Take $\epsilon > 0$ and define the approximate delta function δ_ϵ by

$$\delta_\epsilon(\mu) = \frac{1}{\pi} \frac{\epsilon}{\mu^2 + \epsilon^2}. \tag{13.45}$$

This function has integral one and is very peaked near zero when ϵ is small.

Theorem 13.5 *Let $r(t)$ be a covariance function such that the corresponding spectral density $\hat{r}(\lambda)$ has a logarithm $\log \hat{r}(\lambda)$ such that for each real λ and for each real $\epsilon > 0$ the convolution*

$$(\delta_\epsilon * \log \hat{r})(\lambda) = \int_{-\infty}^{\infty} \delta_\epsilon(\mu - \lambda) \log \hat{r}(\mu)\, d\mu > -\infty. \tag{13.46}$$

Then $r(t)$ may be represented by a square integrable causal impulse response function.

The hypothesis of the theorem ensures that $\hat{r}(\lambda)$ is never too close to zero. Thus it cannot be zero on an interval, and it cannot approach zero too rapidly at infinity. Otherwise $\log \hat{r}(\lambda)$ is $-\infty$ on an interval, or approaches $-\infty$ rapidly at infinity. This would in turn make the convolution integral diverge to $-\infty$.

It is only possible to give an indication of the proof of the theorem. The condition that $r(t)$ may be represented in this way may be stated in terms of the corresponding frequency impulse function. Recall the $\hat{r}(\lambda)$ is even and positive. Furthermore $\hat{h}(\lambda)$ satisfies $\hat{h}(-\lambda) = \overline{\hat{h}(\lambda)}$. The representation takes the form

$$\hat{r}(\lambda) = \sigma^2 |\hat{h}(\lambda)|^2. \tag{13.47}$$

The causality condition is that $h(\lambda)$ may be extended to all λ with negative imaginary part. Another way to write this representation is

$$\hat{r}(\lambda) = \sigma\overline{\hat{h}(\lambda)}\, \sigma\hat{h}(\lambda) \tag{13.48}$$

Thus to see whether the correlation $r(t)$ comes from a causal convolution with white noise, the task is to see whether the spectral density $\hat{r}(\lambda)$ is a product of a function defined in the lower half plane with its complex conjugate.

Sums are easier than products. So the idea is to express

$$\hat{r}(\lambda) = \exp(\log \hat{r}(\lambda)) \tag{13.49}$$

and try to write $\log \hat{r}(\lambda)$ as the sum of a function with its complex conjugate. However under the hypothesis of the theorem, we can write this as a limit using

$$(\delta_\epsilon * \log \hat{r})(\lambda) \to \log \hat{r}(\lambda) \tag{13.50}$$

as $\epsilon \to 0$.

Now the trick is to use the decomposition

$$\delta_\epsilon(\mu) = \frac{1}{2\pi i}\frac{1}{\mu - \lambda - i\epsilon} - \frac{1}{2\pi i}\frac{1}{\mu - \lambda + i\epsilon} \tag{13.51}$$

of the approximate delta function as the sum of two complex conjugates. Then we define

$$\sigma\hat{h}(\lambda - i\epsilon) = \exp(-\frac{1}{2\pi i}\int_{-\infty}^{\infty} \frac{1}{\mu - \lambda + i\epsilon} \log \hat{r}(\mu)\, d\mu). \tag{13.52}$$

Then

$$\sigma\overline{\hat{h}(\lambda - i\epsilon)} = \exp(\frac{1}{2\pi i}\int_{-\infty}^{\infty} \frac{1}{\mu - \lambda - i\epsilon} \log \hat{r}(\mu)\, d\mu). \tag{13.53}$$

It follows that

$$\sigma^2|\hat{h}(\lambda)|^2 = \sigma^2\overline{\hat{h}(\lambda)}\hat{h}(\lambda) = \exp(\int_{-\infty}^{\infty} \delta_\epsilon(\mu - \lambda) \log \hat{r}(\mu)\, d\mu). \tag{13.54}$$

Take $\epsilon \to 0$. In the limit we get

$$\sigma^2 |\hat{h}(\lambda)|^2 = \exp(\log \hat{r}(\lambda)) = r(\lambda). \tag{13.55}$$

Example: Take the Ornstein-Uhlenbeck example with $\alpha > 0$. Here the impulse response function is causal and given by

$$h(t) = e^{-\lambda t} \tag{13.56}$$

for $t \geq 0$. Therefore the frequency response function

$$\hat{h}(\lambda) = \frac{-i}{\lambda - i\alpha} \tag{13.57}$$

extends to the lower half plane. The spectral density is

$$\hat{r}(\lambda) = \sigma^2 \frac{1}{\lambda^2 + \alpha^2}. \tag{13.58}$$

The corresponding correlation function is

$$r(\lambda) = \frac{\sigma^2}{2\alpha} e^{-\alpha|t|}. \tag{13.59}$$

Example: Take the example

$$r(t) = \frac{c^2}{c^2 + t^2}. \tag{13.60}$$

The corresponding spectral density is

$$\hat{r}(\lambda) = \pi c e^{-c|\lambda|}. \tag{13.61}$$

The integral does not converge, and so there is no causal representation. The problem is that there are not enough high frequencies in the spectral density. The time correlations are too smooth to have a causal explanation.

See the book Gaussian Processes, Function Theory, and the Inverse Spectral Problem, by Dym and McKean. It has much more on this subject.

13.8 White noise

The reason for the term white noise is that the spectral density of the white noise process is a constant σ^2. This is the density with respect to $d\lambda/(2\pi)$, so the contribution from the variance from any frequency interval of fixed length is the same, the integral of $\sigma^2\, d\lambda/(2\pi)$ over the interval. All frequencies are equally represented, hence the term white. Of course there is no such process, at least not if we require it to be realized in the form of a function of t.

We can approximate white noise $dW(t)/dt$ by difference quotients of the Wiener process. Take $a > 0$ and define

$$X(t) = \frac{W(t) - W(t-a)}{a}. \tag{13.62}$$

This is obtained as a Wiener stochastic integral by using a causal impulse response function $h(t)$ that is $1/a$ on the interval from zero to a. The corresponding frequency response function is

$$\hat{h}(\lambda) = (1 - e^{-ia\lambda})/(ia\lambda). \tag{13.63}$$

The corresponding spectral density of the $X(t)$ process is

$$\hat{r}(\lambda) = \sigma^2 |\hat{h}(\lambda)|^2 = \sigma^2 4 \sin^2(a\lambda/2)/(a^2\lambda^2). \tag{13.64}$$

Thus the correlation function of the $X(t)$ process is

$$r(t) = \sigma^2 (1/a)(1 - |t|/a) \tag{13.65}$$

for $|t| \leq a$, and zero elsewhere.

In applications people do not often worry about whether their process is really white noise for all frequencies, just over a sufficiently large range of frequencies. So, for instance, if we take a to be small, then as long as λ is much less than π/a, the spectral density for the $X(t)$ process will be quite uniform over this range. So this might be a model for an actual white noise process in an experimental situation.

13.9 $1/f$ noise

The frequency λ that we have been using up to now is actually the angular frequency $\lambda = 2\pi f$, where f is the frequency. The frequency f tells how many complete oscillations per second; the angular frequency λ describes how many radians per second.

The subject of $1/f$ noise (or flicker noise) is more mysterious. This refers to experimental situations where the spectral density at low frequencies λ is approximately proportional to $1/\lambda$. The paradox is that $1/\lambda$ is not integrable near $\lambda = 0$, and so the total variance would be infinite.

One explanation of white noise is that it is noise due to the additive effect of independent processes on various time scales. Think of α as determining a time rate (so that $1/\alpha$ is a time duration). The Ornstein-Uhlenbeck process has a fixed rate, so

$$r(t) = r(0)e^{-\alpha|t|} \tag{13.66}$$

and

$$\hat{r}(\lambda) = r(0)\frac{2\alpha}{\alpha^2 + \lambda^2} \tag{13.67}$$

are determined by this parameter. A process with a wide variety of rates might be determined by

$$\hat{r}(\lambda) = \int_0^\infty p(\alpha)\frac{2\alpha}{\alpha^2 + \lambda^2}\, d\alpha. \tag{13.68}$$

Take the case when $p(\alpha) = c/\alpha$ for $0 < a \leq \alpha < b < \infty$ and is zero elsewhere. The a is a lower cutoff on the rate of decay, and b is an upper cutoff. Notice

that this gives a lot of weight to slow processes with a decay rate near a. This gives

$$\hat{r}(\lambda) = \int_a^b \frac{c}{\alpha} \frac{2\alpha}{\alpha^2 + \lambda^2} \, d\alpha. \tag{13.69}$$

This simplifies to

$$\hat{r}(\lambda) = \frac{2\pi c}{\lambda} \int_a^b \frac{1}{\pi} \frac{\lambda}{\alpha^2 + \lambda^2} \, d\alpha = \frac{1}{\lambda} 2\pi c \int_a^b \delta_\lambda(\alpha) \, d\alpha. \tag{13.70}$$

If $0 < a$ is small compared to λ, and if b is large compared to λ, then for this range of λ the integral is approximately equal to $1/2$. Thus in this range

$$\hat{r}(\lambda) \approx \frac{\pi c}{\lambda}. \tag{13.71}$$

So this provides a possible covariance that could simulate $1/f$ noise over a large range. Of course this is just an empirical fit, not an explanation. The expression above shows that there can be a covariance that gives $1/f$ noise over some range; it does not give the underlying mechanism.

If we want to see what this looks like in terms of time correlations, we can write

$$r(t) = \int_a^b \frac{c}{\alpha} e^{-\alpha|t|} \, d\alpha. \tag{13.72}$$

This can also be written

$$r(t) = \int_{a|t|}^{b|t|} \frac{c}{s} e^{-s} \, ds = c[\log(b|t|)e^{-b|t|} - \log(a|t|)e^{-a|t|} + \int_{a|t|}^{b|t|} \log(s) e^{-s} \, ds]. \tag{13.73}$$

If $a|t|$ is much smaller then one and $b|t|$ is much larger than one, then this becomes

$$r(t) \approx c[-\log(a|t|) - \gamma]. \tag{13.74}$$

Here γ is the integral

$$\gamma = -\int_0^\infty \log(s) e^{-s} \, ds. \tag{13.75}$$

The expression for the covariance shows that for this range of time the correlation decreases with time at a logarithmic rate. This is a very slow rate of decrease.

Much more information about $1/f$ noise may be found in Kogan's book on Electronic Noise and Fluctuations in Solids, and elsewhere.

13.10 Problems

1. HPS, Chapter 6, Problem 21
2. HPS, Chapter 6, Problem 22