Course Notes for Stochastic Processes

by Russell Lyons
Based on the book by Sheldon Ross

These are notes that I used myself to lecture from. A modified version was handed to the students, which is reflected in various changes of fonts and marginal hacks in this version. These things were not in their version. In particular, certain things were omitted and they were given space to write things that either were in my notes or on which I expanded in class.

The first part of the course contains some material that is not taught when one semester is devoted to the whole course.

Prerequisites: Undergraduate probability, up through joint density of continuous random variables. You should be comfortable with undergraduate real analysis/advanced calculus, meaning proofs and "epsilonics", in order to understand some of the derivations, although you will almost never have to do epsilonics yourself. You will be asked to do calculations as well as derivations in this course. It will be crucial to understand probabilistic concepts; they make calculations much easier and strengthen your intuition. An introduction to measure theory is not needed and will not be assumed, but would add to your understanding if you happen to have had it or are taking it concurrently.

The **textbook** is by S. Ross, *Stochastic Processes*, 2nd ed., 1996. We will cover Chapters 1–4 and 8 fairly thoroughly, and Chapters 5–7 in part. Other books that will be used as sources of examples are *Introduction to Probability Models*, 7th ed., by Ross (to be abbreviated as "PM") and *Modeling and Analysis of Stochastic Systems* by V.G. Kulkarni (to be abbreviated as "MASS"). You do not need get them. The material of the course is extremely useful in practice, and also a lot of fun. We will give examples that are designed to illustrate both of these (not always at the same time).

Grades will be based on weekly homework, class participation, two exams, and a final exam (Tuesday, May 6, 12:40-2:40 p.m.).

These notes follow the book fairly closely. In particular, all numbering (such as of sections and theorems) follows that in the book. However, the notes often provide proofs that are shorter or more conceptual than the ones in the book. The book tends to prefer proofs that rely on calculation, despite the excellent intuition and concepts that are introduced. On the other hand, these notes are sometimes sketchy, with more details to be given in class. (There are blank spaces often left for you to fill in details as we go.) Sometimes, entire chapters are done differently in these notes than in the book.

Occasionally, we need to assign numbers to equations that do not appear in the book. These will be preceded by "N" (for "notes").

5" Definition of stochastic process. Examples and graphs. ...

Example MASS 1.3 (Single-Server Queue). Here we begin with 2 stochastic processes as input and study several others derived from them. Suppose that the nth customer arrives at time A_n and, once service begins, takes time S_n to be served. There is a single server who serves the customers in the order of their arrival, each one until finished. We want to study $Q(t)$, the number of customers in the system at time t; the time of departure D_n of the nth customer; and the waiting time of the nth customer, $W_n := D_n - A_n$.

Draw graph of arrival and departure times on the horizontal axis with length of queue on the vertical axis. Put down A_n and D_n first, both being increas-
3" ing processes. ...

Chapter 1

Preliminaries

§1.1. Probability.

The axioms of probability are that there is a "sample space" Ω (or S) containing all possible outcomes and a function P that assigns to subsets of Ω (called "events") a number in $[0, 1]$ such that

(i) $P(\Omega) = 1$ and

(ii) if $E_1, E_2, \ldots$ are (pairwise) disjoint events, then

$$P\Big(\bigcup_{n=1}^{\infty} E_n\Big) = \sum_{n=1}^{\infty} P(E_n).$$

[Actually, sometimes only certain subsets of Ω can be given a probability, but that will not concern us. Part of the development of measure theory elucidates this issue.] A probability space is such a pair, (Ω, P).

The axioms imply the particularly useful consequences:

1" (i) $P(\varnothing) = 0$. ...

1" (ii) If $E \subseteq F$, then $P(E) \le P(F)$. ...

1" (iii) $P(E^{\mathrm{c}}) = 1 - P(E)$. ...

(iv) (subadditivity) For any events E_n, we have

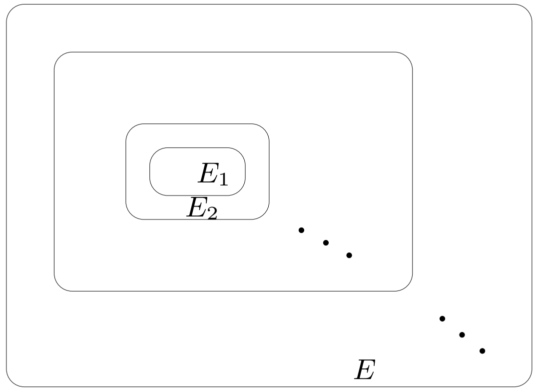
$$P\Big(\bigcup_{n=1}^{\infty} E_n\Big) \le \sum_{n=1}^{\infty} P(E_n).$$

1" ...

1" (v) If $P(E_n) = 0$, then $P(\bigcup_n E_n) = 0$. If $P(E_n) = 1$, then $P(\bigcap_n E_n) = 1$. ...

Proposition 1.1.1. *If $E_n \uparrow E$ or $E_n \downarrow E$, then $P(E_n) \to P(E)$.*

Proof. In the first case that $E_n \uparrow E$, write E as a disjoint union $\bigcup_n (E_{n+1} \setminus E_n)$. (See the
2" figure.) ... When $E_n \downarrow E$, use $E_n^{\mathrm{c}} \uparrow E^{\mathrm{c}}$. ... ∎



If you flip a sequence of coins and the nth coin has chance $1/n^2$ of landing H, will you get an infinite number of heads? What if the chance is $1/n$? To answer these questions, we prove the Borel–Cantelli lemmas.

3" `Explain limsup and liminf of sequence of sets. ...`

Proposition 1.1.2 (First Borel–Cantelli Lemma). *If $\sum_n P(E_n) < \infty$, then $P(E_n$ i.o.$) = 0$.*

Proof. We have

$$P(E_n \text{ i.o.}) = P\Big(\bigcap_{n\ge 1}\bigcup_{k\ge n} E_k\Big) = \lim_n P\Big(\bigcup_{k\ge n} E_k\Big) \le \liminf_n \sum_{k\ge n} P(E_k) = 0. \qquad \blacksquare$$

2" ...

Proposition 1.1.3 (Second Borel–Cantelli Lemma). *If $\sum_n P(E_n) = \infty$ and $\{E_n\}$ are (mutually) independent, then $P(E_n$ i.o.$) = 1$.*

Proof. We have

$$\begin{aligned} P(E_n \text{ i.o.}) &= \lim_n P\Big(\bigcup_{k\ge n} E_k\Big) = \lim_n \Big[1 - P\Big(\bigcap_{k\ge n} E_k^{\mathrm{c}}\Big)\Big] \\ &= \lim_n \Big[1 - \prod_{k\ge n}\big(1 - P(E_k)\big)\Big] \ge \limsup_n \Big[1 - \prod_{k\ge n} e^{-P(E_k)}\Big] \quad \text{since } 1 - x \le e^{-x} \\ &= \limsup_n \Big[1 - e^{-\sum_{k\ge n} P(E_k)}\Big] = 1. \qquad \blacksquare \end{aligned}$$

`Counterexamples without independence.`

3" `... Draw the tangent line to illustrate the inequality. Prove using def-`
2" `inition of` log. `...`

§1.2. Random Variables.

A ***(real-valued) random variable*** is a function $X : \Omega$ (or S) $\to \mathbb{R}$. Its ***(cumulative) distribution function*** (c.d.f.) $F = F_X$ is $F(x) := P[X \le x]$. The c.d.f. determines the
2" distribution of X. ... Often the tail probability function $\overline{F}(x) := 1 - F(x) = P[X > x]$ is useful. We use the notation $X \sim F$, especially when F has a name, like $\mathrm{Bin}(n, p)$ or $\mathrm{Unif}[0, 1]$. When two random variables X and Y have the same c.d.f., we write $X \overset{\mathscr{D}}{=} Y$; here, X and Y need not be defined on the same probability space. In case we have a collection of identically distributed random variables X_i, we often write X for a random variable with the same distribution as all of the X_i.

If the range of X is countable, we call X ***discrete***; if its values are isolated, then F_X is a step function. If no value has positive probability, X is ***continuous***; this is the same as saying that F_X is a continuous function. The random variable X could be neither discrete nor continuous. For example, if we flip a coin and get H, then set $X := 0$; but if
1" we get T, then choose $X \sim \mathrm{Unif}[0, 1]$. What is F_X? ... We could define X in different ways that give the same distribution (though not the same random variable): for example, $X := YZ$, where $Y \sim \mathrm{Bern}(1/2)$, $Z \sim \mathrm{Unif}[0, 1]$, and Y, Z independent; or $X := W^+$, where $W \sim \mathrm{Unif}[-1, 1]$.

0" If $\exists f : \mathbb{R} \to \mathbb{R}$ such that $\forall x \;\; F_X(x) = \int_{-\infty}^{x} f(s)\,ds$, ... then X is ***absolutely continuous*** [called "continuous" in the book] and f is its ***probability density function***. In this case, $f(x) = F_X'(x)$ and $P[X \in B] = \int_B f(x)\,dx$. Almost always, we will use the
0" case that B is an interval. Defn for more general B. ...

For two random variables X and Y on the same probability space, their ***joint dis-
2" tribution function*** is $F_{X,Y}(x, y) := P[X \le x, Y \le y]$. ... If

$$F_{X,Y}(x, y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f_{X,Y}(s, t)\,dt\,ds,$$

then $f_{X,Y}$ is called the ***joint density*** of X and Y; we also say that X and Y are ***jointly absolutely continuous***.

Note that $F_X(x) = F_{X,Y}(x, \infty)$ (no densities are being assumed here). Explain
3" limits not along sequences. ... We have that X and Y are independent $\iff$ $\forall x, y \;\; F_{X,Y}(x, y) = F_X(x) F_Y(y) \iff \forall A, B \;\; P[X \in A, Y \in B] = P[X \in A] P[Y \in B]$.

We are generally interested in random variables and their distributions, not the underlying probability spaces on which they are defined.

§1.3. Expected Value.

The ***expectation*** of X is defined by $E(X) := \int_{-\infty}^{\infty} x\,dF_X(x)$. However, we will only say what this means in two cases: If X is absolutely continuous, then $E(X) = \int_{-\infty}^{\infty} x f(x)\,dx$. If X is discrete, then $E(X) = \sum_x xP[X = x]$. Give idea of Stieltjes
2" integral with $\int h(x)\,dF_X(x)$ to explain notation. ... If we want to find the expectation of $h(X)$, we don't need to find the distribution of $h(X)$; instead, we can use the distribution of X directly, as it can be shown that

$$E\big(h(X)\big) = \int_{-\infty}^{\infty} h(x)\,dF_X(x).$$

This change-of-variable formula is very handy! In particular,

$$P[X \in A] = \int_A dF_X(x).$$

It can also be shown that $E\big(\sum_{i=1}^n X_i\big) = \sum_{i=1}^n E(X_i)$. This linearity is actually a key property of expectation; the proper definition of expectation using measure theory makes the proof of linearity easy. Since $E[Z] \ge 0$ when $Z \ge 0$, it follows that $E[Y] \ge E[X]$ when $Y \ge X$. Another particular case of the previous formula for $E\big(h(X)\big)$ uses $h(x) := \int_{-\infty}^{\infty} g(x, y)\,dy$, which gives

$$\begin{aligned} E\Big[\int_{-\infty}^{\infty} g(X, y)\,dy\Big] &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x, y)\,dy\,dF_X(x) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x, y)\,dF_X(x)\,dy \\ &= \int_{-\infty}^{\infty} E[g(X, y)]\,dy. \end{aligned}$$

In other words, we can interchange expectation and integral. Define the ***variance*** of X as

$$\operatorname{Var}(X) := E\Big[\big(X - E(X)\big)^2\Big] = E(X^2) - E(X)^2$$

and the ***covariance*** of X and Y as

$$\operatorname{Cov}(X, Y) := E\Big[\big(X - E(X)\big)\big(Y - E(Y)\big)\Big] = E(XY) - E(X)E(Y).$$

0" ... Recall that $\operatorname{Cov}(X, Y) = 0$ if (but not only if) X and Y are independent. We have

$$\operatorname{Var}\Big(\sum_{i=1}^n X_i\Big) = \sum_{i=1}^n \operatorname{Var}(X_i) + \sum_{i \ne j} \operatorname{Cov}(X_i, X_j).$$

0" ... In particular, if X_i are independent, then $\operatorname{Var}\big(\sum_1^n X_i\big) = \sum_{i=1}^n \operatorname{Var}(X_i)$.

If X and Y have a joint density, $f_{X,Y}$, then it can be shown that

$$E[h(X, Y)] = \iint h(x, y) f_{X,Y}(x, y)\,dx\,dy.$$

§1.4. Moment Generating, Characteristic Functions, and Laplace Transforms.

We will occasionally need the ***moment generating function***

$$E[e^{tX}] = \sum_{n \ge 0} \frac{E[X^n]}{n!} t^n.$$

(Although we will not pay close attention to when this equality holds, it does in all situations we will encounter. For example, if the left-hand side is finite for some $t_+ \ge 0$ and some $t_- \le 0$, then equality holds for all $t \in [t_-, t_+]$.)

The reason for the name is that

$$E[X^n] = \left(\frac{d}{dt}\right)^n E[e^{tX}]\Big|_{t=0}.$$

1" ... We can differentiate inside the expectation.

EXAMPLE: $\mathrm{Exp}(\lambda)$. Recall that $X \sim \mathrm{Exp}(\lambda)$ (λ is called the ***parameter*** or ***rate***) if it has probability density function

$$f_X(x) = \begin{cases} \lambda e^{-\lambda x} & \text{if } x \ge 0, \\ 0 & \text{if } x < 0. \end{cases}$$

3" ...

3" Thus, for $X \sim \mathrm{Exp}(\lambda)$, we have $E[X] = 1/\lambda$ and $\mathrm{Var}(X) = 1/\lambda^2$. ...

In the case of the exponential distribution, one could also derive the moments from the fact that $\int_0^\infty x^n e^{-x}\, dx = n!$ (which follows by induction and integration by parts). Conversely, one can derive the value of this integral by differentiating the moment generating function for $\mathrm{Exp}(1)$.

§1.5. Conditional Expectation.

Use an example for all the following, such as X the number of the first die and Y the sum of two dice. We can wait to do examples until after (1.5.1).

Suppose that X and Y are discrete. Then

$$P[X = x \mid Y = y] = \frac{P[X = x, Y = y]}{P[Y = y]}.$$

We can regard this as a function of x or of y. As a function of x, it gives the distribution function of a random variable since $\sum_x P[X = x \mid Y = y] = 1$. It is called the ***conditional distribution of*** X ***given*** $Y = y$. It has, thus, an expectation,

$$\sum_x x P[X = x \mid Y = y] =: E[X \mid Y = y].$$

2" . . .

If we regard $P[X = x \mid Y = y]$ as a function of y, then we may pre-compose it with
2" Y to get a random variable denoted $P[X = x \mid Y]$. . . . This random variable has an expectation:

$$E\big[P[X = x \mid Y]\big] = P[X = x]. \tag{N1}$$

1" . . . This is a special case of the so-called law of total probability. We can also pre-compose
2" the function $y \mapsto E[X \mid Y = y]$ with Y to get a random variable denoted $E[X \mid Y]$. . . . We have

$$E[X] = E\big[E[X \mid Y]\big]. \tag{1.5.1}$$

This is called the ***tower property***.

2" Proof: Write it out. ...

1" We may regard (N1) as a special case of (1.5.1): . . .

These ideas extend to all random variables. For the case that X and Y are jointly absolutely continuous, the density of X given $Y = y$ is $x \mapsto f_{X,Y}(x, y)/f_Y(y)$ (by Exercise 2, this is a probability density function). Think of this as follows. Note that

$$f_X(x)\,dx = P\big[X \in (x, x + dx)\big]$$

so that

$$\begin{aligned}\frac{f_{X,Y}(x, y)}{f_Y(y)} &= \frac{P\big[X \in (x, x + dx), Y \in (y, y + dy)\big]/(dx\,dy)}{P\big[Y \in (y, y + dy)\big]/dy} \\ &= P\big[X \in (x, x + dx) \mid Y \in (y, y + dy)\big]/dx \\ &= P\big[X \in (x, x + dx) \mid Y = y\big]/dx.\end{aligned}$$

This probability density has an expectation, which we write as $E[X \mid Y = y]$. When we pre-compose this function of y with Y, we get the random variable $E[X \mid Y]$. Equation
↓ (1.5.1) holds too:

$$
\begin{aligned}
E\big[\underbrace{E[X \mid Y]}_{\text{function of } Y}\big] &= \int_{-\infty}^{\infty} E[X \mid Y = y]\, dF_Y(y) = \int_{-\infty}^{\infty} E[X \mid Y = y] f_Y(y)\, dy \\
&= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x \frac{f_{X,Y}(x,y)}{f_Y(y)} dx \ f_Y(y)\, dy \\
&= \int_{-\infty}^{\infty} x \int_{-\infty}^{\infty} f_{X,Y}(x,y)\, dy\, dx = \int_{-\infty}^{\infty} x f_X(x)\, dx \quad \text{[by Exercise 2]} \\
&= E(X).
\end{aligned}
$$

2" ↑

We also want to define $E[X \mid Y]$ when only one of X or Y is absolutely continuous. First, if X is any random variable and A is an event of positive probability, then the function

$$x \mapsto P[X \le x \mid A]$$

is the c.d.f. of a random variable, because it is non-decreasing, tends to 0 as $x \to -\infty$ and to 1 as $x \to \infty$, and is continuous from the right; its expectation is denoted $E[X \mid A]$. This allows us to define $P[X \le x \mid Y = y]$ and $E[X \mid Y = y]$ when Y is discrete. (It can be shown using measure theory that if X above is absolutely continuous, then so is the random variable with the above c.d.f.) Second, if A is an event and Y is an absolutely continuous random variable, we define

$$P[A \mid Y = y] := \lim_{\epsilon \to 0} \frac{P\big(A \cap \{Y \in (y-\epsilon, y+\epsilon)\}\big)}{P\big[Y \in (y-\epsilon, y+\epsilon)\big]}$$

when the limit exists. When X is discrete and Y is absolutely continuous, this allows us to define

$$E[X \mid Y = y] := \sum_x x P[X = x \mid Y = y].$$

Once we have defined $E[X \mid Y = y]$ in either of these cases, we define $E[X \mid Y]$ as before by pre-composing with Y.

In all cases, it can be shown that (1.5.1) still holds.

When X and Y are both discrete, then one can verify the following conditional change-of-variable formula (Exercise 6):

$$E\big[h(X,Y)\big] = E\big[E[h(X,Y) \mid Y]\big] = \int_{-\infty}^{\infty} E[h(X,y) \mid Y = y]\, dF_Y(y). \tag{N2}$$

In fact, this holds in complete generality.

In particular,

$$P(A) = E[\mathbf{1}_A] = E\big[E[\mathbf{1}_A \mid Y]\big] = E\big[P(A \mid Y)\big].$$

1" ... Also, it follows from (1.5.1) and (N2) that

$$E\big[h(Y)E[X \mid Y]\big] = E\big[E[h(Y)X \mid Y]\big] = E[h(Y)X].$$

1" ... We can condition on several random variables, too:

$$E\big[E[X \mid Y_1, Y_2, \ldots]\big] = E[X].$$

All these uses of (1.5.1) are extremely useful!

Two random variables X and Y are independent iff the conditional distribution of X given $Y = y$ is equal to the unconditional distribution of X (for all $y \in A$ for some set A where $P[Y \in A] = 1$).

We now give some applications of conditioning.

Example PM 3.15 (Analyzing the Quick-Sort Algorithm). Given a list of distinct numbers $x_1, \ldots, x_n$, the goal is to place them in increasing order, that is, to ***sort*** them, as quickly as possible. (A list is the same as a sequence.) The quick-sort algorithm works as illustrated in an example: Suppose that the original list is 10, 5, 8, 2, 1, 4, 7. Choose one at random, say, 4. Compare 4 to the others: $\{2,1\},\ 4,\ \{10,5,8,7\}$. Now apply the same procedure to the set < 4 and the set > 4:

$\to 1,2,4,\ \{10,5,8,7\} \to$ choose at random from 2nd set, say 7:

$\to 1,2,4,5,7,\ \{10,8\} \to 1,2,4,5,7,8,10.$

The number of comparisons here was $6 + 1 + 3 + 1 = 11$. This is a random "divide and conquer" algorithm. How well does it do? The slowest would be if we always pick the smallest or the largest one; then every pair must be compared and it takes $\sim n^2/2$ comparisons. The fastest possible would be if every time, the median were chosen; then the number of comparisons would be

$$\sim n + \frac{n}{2} \times 2 + \frac{n}{4} \times 4 + \cdots \quad (\sim \log_2 n \text{ terms}) \sim n \log_2 n.$$

It turns out that this is quite close to M_n, the expected number of comparisons in quick-sort!

↓ To calculate M_n, condition on the rank of the initial value selected:

$$M_n = \sum_{j=1}^{n} E[\text{number of comparisons} \mid \text{initial value is } j\text{th smallest}] P[\text{initial value is } j\text{th smallest}]$$
$$= \sum_{j=1}^{n} (n-1+M_{j-1}+M_{n-j}) \cdot \frac{1}{n} = n-1+\frac{2}{n}\sum_{k=1}^{n-1} M_k.$$

Thus

$$nM_n = n(n-1) + 2\sum_{k=1}^{n-1} M_k.$$

To solve this recursion, substitute $n-1$ for n and subtract:

$$nM_n - (n-1)M_{n-1} = 2(n-1) + 2M_{n-1},$$

whence

$$nM_n = (n+1)M_{n-1} + 2(n-1),$$

which is

$$\frac{M_n}{n+1} = \frac{M_{n-1}}{n} + \frac{2(n-1)}{n(n+1)}.$$

Iterating gives

$$\frac{M_n}{n+1} = 2\sum_{n \ge k \ge 1} \frac{k-1}{k(k+1)} = 2\sum_{k=1}^{n}\left[\frac{2}{k+1} - \frac{1}{k}\right]$$
$$\sim 2(2\log n - \log n) = 2\log n,$$

whence

$$M_n \sim 2n \log n.$$

7" ↑ Note that $2 > (\log 2)^{-1}$.

`Remark:` $M_n = 2(n+1)\bigl(-2 + \log n + \gamma + 3/n + O(1/n^2)\bigr)$.

Example 1.5(e) (The Ballot Theorem). In an election, A receives n votes and B receives m votes, $n > m$. If all orderings of the $n+m$ votes are equally likely, then

$$P[\text{A always (strictly) ahead of B}] = \frac{\mathtt{n-m}}{\mathtt{n+m}}.$$

↓

Proof. Let $P_{n,m}$ be the desired probability. Then

$$\begin{aligned}P_{n,m} &= P[\text{A always ahead} \mid \text{A gets last vote}]P[\text{A gets last vote}] \\ &\qquad + P[\text{A always ahead} \mid \text{B gets last vote}]P[\text{B gets last vote}] \\ &= P_{n-1,m} \cdot \frac{n}{n+m} + P_{n,m-1} \cdot \frac{m}{n+m}.\end{aligned}$$

Here, we make the convention that $P_{n-1,m} := 0$ if $n = m + 1$; note that this fits our formula nicely, so we needn't consider that case separately when we claim that our formula fits the equation. `Note why we conditioned on the last vote, rather than the first.` Now use induction on $n + m$. ∎

4" ↑

`This also holds assuming only that all cyclic orderings of the votes are equally likely. In other words, given a cyclic order, the number of linear orders is` $n + m$`, of which` $n - m$ `have A ahead always. This requires a different argument; see the proof by the cycle lemma of Dvoretzky and Motzkin.`

Example 1.5(a) (The Sum of a Random Number of Random Variables). Let X_i be i.i.d. ($i \ge 1$) and N be a random variable with values in $\mathbb{N} := \{0, 1, 2, \ldots\}$, independent of all X_i. Let $Y := \sum_{i=1}^{N} X_i$, which means $Y(\omega) := \sum_{i=1}^{N(\omega)} X_i(\omega)$.

Examples:

- **Queueing:** $N :=$ the number of customers arriving in a specific time period, $X_i :=$ the service time required by the ith customer. Then $\sum_1^N X_i =$ the total service time required by customers arriving in that time period.
- **Risk Theory:** $N :=$ the number of claims arriving at an insurance company in a given week, $X_i :=$ the amount of the ith claim. Then $\sum_1^N X_i =$ the total liability for that week.
- **Population Model:** $N :=$ the number of plants of a given species in a certain area, $X_i :=$ the number of seeds produced by the ith plant.

To compute moments of Y, we compute the moment generating function:

$$E\big[e^{tY}\big] = E\big[E[e^{tY} \mid N]\big].$$

↓ Now

$$\begin{aligned} E[e^{tY} \mid N = n] &= E\Big[e^{t\sum_1^N X_i} \Bigm| N = n\Big] = E\Big[e^{t\sum_1^n X_i} \Bigm| N = n\Big] \text{ by (N3) below} \\ &= E\Big[e^{t\sum_1^n X_i}\Big] \text{ by independence} \\ &= E[e^{tX}]^n \text{ by independence,} \end{aligned}$$

where $X \overset{\mathscr{D}}{=} X_i$. Therefore $E[e^{tY}] = E\big[E[e^{tX}]^N\big]$,

$$\frac{d}{dt}E[e^{tY}] = E[Ye^{tY}] = E\big[NE[e^{tX}]^{N-1}E[Xe^{tX}]\big],$$

$$\begin{aligned} \frac{d^2}{dt^2}E[e^{tY}] &= E[Y^2 e^{tY}] \\ &= E\big[N(N-1)E[e^{tX}]^{N-2}E[Xe^{tX}]^2\big] + E\big[NE[e^{tX}]^{N-1}E[X^2e^{tX}]\big], \end{aligned}$$

so

$$\begin{aligned} E[Y] &= E\big[NE[X]\big] = E[N]E[X], \\ E[Y^2] &= E\big[N(N-1)\big]E[X]^2 + E[N]E[X^2], \end{aligned}$$

and

$$\begin{aligned} \operatorname{Var}(Y) &= E[N]E[X^2] + \big\{E[N^2] - E[N] - E[N]^2\big\}E[X]^2 \\ &= E[N]\operatorname{Var}(X) + \operatorname{Var}(N)E[X]^2. \end{aligned}$$

7" ↑

▷ `Read pp. 1--9, 15--16, 18, and 20--24 in the book.`

Given an event A of positive probability and a random variable X with finite expectation, we have defined $E[X \mid A]$ as $E[Y]$, where Y has the distribution of X given A. There is another definition of conditional probability used in measure-theory-based courses, which we will occasionally find useful:

$$E[X \mid A] = E[X\mathbf{1}_A]/P(A). \tag{N3}$$

Note that this equation immediately gives linearity of conditional expectation. To see (N3), we may assume that $X \ge 0$ by decomposing $X = X^+ - X^-$: see Exercise 10. Then we can use the tail formula, Exercise 11 (p. 46, 1.1 in the book), as follows:

$$\begin{aligned}
E[Y] &= \int_0^\infty P[Y > y]\,dy = \int_0^\infty P[X > y \mid A]\,dy = \int_0^\infty P[A,\ X > y]/P(A)\,dy \\
&= \frac{1}{P(A)} \int_0^\infty E[\mathbf{1}_A \mathbf{1}_{[X>y]}]\,dy = \frac{1}{P(A)} E\Big[\int_0^\infty \mathbf{1}_A \mathbf{1}_{[X>y]}\,dy\Big] \\
&= \frac{1}{P(A)} E\Big[\mathbf{1}_A \int_0^\infty \mathbf{1}_{[X>y]}\,dy\Big] = \frac{1}{P(A)} E\Big[\mathbf{1}_A X\Big].
\end{aligned}$$

§1.6. The Exponential Distribution, Lack of Memory, and Hazard Rate Functions.

If $X \sim \mathrm{Exp}(\lambda)$, then $\overline{F}_X(x) = e^{-\lambda x}$. Such a random variable is ***memoryless:***

$$P(X > s + t \mid X > t) = P(X > s) \quad \text{for all } s, t \ge 0. \tag{N4}$$

1" … In other words, the conditional distribution of $X - t$ given that $X > t$ is the same as the unconditional distribution of X.

Example: A post office has 2 clerks. The customer service time of each clerk is $\mathrm{Exp}(\lambda)$. Neither clerk is busy. One customer arrives at a random time and, while that customer is still being served, another customer arrives at a random time and begins service with the other clerk. What is the chance that the first customer finishes first? (Note that in problems like this, we assume a lot of independence that is not stated explicitly. What
1" such assumptions can you identify?) …

Solution. The answer is 1/2 by the (strong) memoryless property and symmetry. Symmetry here is the following principle: If two random variables T_1 and T_2 are IID, then $P[T_1 < T_2] = P[T_2 < T_1]$. The proof is that $(T_1, T_2) \overset{\mathscr{D}}{=} (T_2, T_1)$. If, in addition, T_1 is continuous, then these probabilities must equal 1/2, because $P[T_1 = T_2] = E\big[P[T_1 = T_2 \mid T_2]\big]$
1" and for all t, we have $P[T_1 = T_2 \mid T_2 = t] = P[T_1 = t \mid T_2 = t] = P[T_1 = t] = 0$. ...

The ***strong memoryless property*** says that if $X \sim \operatorname{Exp}(\lambda)$ and $Y \geq 0$ is independent of X, then for all $s \geq 0$, we have $P[X > s + Y \mid X > Y] = P[X > s]$. The book does not mention that the strong memoryless property needs to be established. We prove this by writing the conditional probability as a quotient; calculate both numerator and denominator by conditioning on Y. E.g., $P[X > s + Y, X > Y] = P[X > s + Y] = E\big[P[X > s + Y \mid Y]\big]$ and $P[X > s + Y \mid Y = y] = P[X > s + y \mid Y = y] = e^{-\lambda(s+y)}$ by independence, whence $P[X > s + Y] = E[e^{-\lambda(s+Y)}]$. Likewise, $P[X > Y] = E[e^{-\lambda Y}]$. Thus, the quotient is $e^{-\lambda s}$.

For this problem, let A_1 and A_2 be the arrival times and S_1, S_2 be the service times. We want to show symmetry, i.e., for all $t > 0$, we want

$$P[A_1 + S_1 - A_2 > t \mid A_1 < A_2 < A_1 + S_1] = P[S_2 > t \mid A_1 < A_2 < A_1 + S_1] = e^{-\lambda t}$$

and that $A_1 + S_1 - A_2$ is independent of S_2 given $A_1 < A_2 < A_1 + S_1$, which follows from
1" the assumed mutual independence. ... We have

$$\begin{aligned} P[A_1 + S_1 - A_2 > t \mid A_1 < A_2 < A_1 + S_1] &= P[S_1 > A_2 - A_1 + t \mid 0 < A_2 - A_1 < S_1] \\ &= e^{-\lambda t} \end{aligned}$$

by the strong memoryless property, where we use the random variables S_1 and $A_2 - A_1$
2" with respect to the probability measure where $A_2 - A_1$ is conditioned to be positive. ...

Alternatively, we can formulate an even stronger memoryless property: if $X \sim \operatorname{Exp}(\lambda)$ and $Y, Z \geq 0$ with X, Y and Z being mutually independent, then

$$P[X > Z + Y \mid X > Y] = P[X > Z].$$

We prove this by re-using some of what we proved above, namely, $P[X > Y] = E\big[e^{-\lambda Y}\big]$, $P[X > Z] = E\big[e^{-\lambda Z}\big]$ and

$$P[X > Z + Y,\ X > Y] = P[X > Z + Y] = E\big[e^{-\lambda(Z+Y)}\big] = E\big[e^{-\lambda Z}\big]\, E\big[e^{-\lambda Y}\big].$$

1" ...

We can use this stronger memoryless property to give another solution:

$$\begin{aligned} P[A_1 + S_1 > A_2 + S_2 \mid A_1 < A_2 < A_1 + S_1] &= P[S_1 > S_2 + (A_2 - A_1) \mid 0 < A_2 - A_1 < S_1] \\ &= P[S_1 > S_2] = 1/2. \end{aligned}$$

1" ...

EXAMPLE 1.6(A). A post office has 2 clerks. The customer service time of each clerk is $\mathrm{Exp}(\lambda)$. You enter and are first in line, with both clerks already serving customers. What is the chance that both customers currently being served will be finished before you are?

↓

Solution. Answer: The intuition is 1/2 by the strong memoryless property and symmetry, measuring time from when the first customer leaves. To be rigorous, we will measure time from when you arrive: as in the previous example, the strong memoryless property allows us to assume that both previous customers began service when you entered. Let their service times be S_1 and S_2, while yours is X. The chance that you are *not* the last to finish is

$$\begin{aligned} P\big[X + \min\{S_1, S_2\} < \max\{S_1, S_2\}\big] &= P\big[\max\{S_1, S_2\} > X + \min\{S_1, S_2\}\big] \\ &= P[S_2 > X + S_1 \mid S_2 > S_1]\, P[S_2 > S_1] + \\ &\qquad P[S_1 > X + S_2 \mid S_1 > S_2]\, P[S_1 > S_2] \\ &= P[S_2 > X](1/2) + P[S_1 > X](1/2) = 1/2. \end{aligned}$$

3" ↑

EXAMPLE: Consider two teams of gladiators. At each round, each team sends one of its gladiators to battle. Each gladiator has a fixed strength, and when two fight, the probability of winning is proportional to the strength. The loser never plays again. Show that the order in which teams send gladiators to battle does not change the probability of having a surviving gladiator.

↓

Solution. Replace each gladiator by a light bulb with exponential lifetime whose mean equals the strength of that gladiator. This replacement satisfies the terms of the competition by the strong memoryless property. Now consider turning on a light bulb when it competes. If it burns out, it is thrown away. If it outlasts the other team's light bulb, then it can be turned off until used again, or it can continue to the next match. In any case, a given bulb has some lifetime whatever we decide to do with it (think of predestination, where the future is determined though unknown to us), so the team that wins is the one whose sum of bulb lifetimes is largest. To be less colorful, let $X_1, X_2, \ldots, X_n$ be the exponential random variables for one team. For each $\omega \in \Omega$, the team's total available competition time is $\sum_{i=1}^n X_i(\omega)$, regardless of the order of competition. If $Y_1, \ldots, Y_m$ are the other team's exponential random variables, then the first team wins with probability

3" ↑ $P\big[\sum_{i=1}^n X_i > \sum_{j=1}^m Y_j\big]$.

Note that (N4) is the same as $\overline{F}_X(s+t) = \overline{F}_X(s)\overline{F}_X(t)$. Thus, $\log \overline{F}_X$ satisfies the functional equation

$$g(x+y) = g(x) + g(y) \quad (x, y \geq 0).$$

Also, $\log \overline{F}_X$ is right-continuous (i.e., continuous *from* the right), written $\log \overline{F}_X \in C_{\mathrm{r}}(\mathbb{R}^+)$.

1" ... To show that the exponential random variables are the *only* memoryless random variables, we show that this equation has only the linear solutions $g(x) = cx$ provided $g \in C_{\mathrm{r}}(\mathbb{R}^+)$. (This result will be useful later, too.) Here are the steps:

2" (1) Let $c := g(1)$. Then $g(x) = cx$ for $x \in \mathbb{Q}^+$. ...

1" (2) Since $g \in C_{\mathrm{r}}(\mathbb{R}^+)$, we're done. ...

In fact, the same result holds if g is assumed only to be right-continuous at 0: then
1" actually $g \in C_{\mathrm{r}}(\mathbb{R}^+)$ since $g(x_0 + h) - g(x_0) = g(h)$. ...

For later use, we note that if g is bounded in some interval $[0, \delta]$ ($\delta > 0$), say, by M, then $|g(x)| = |g(nx)|/n \leq M/n$ for $0 \leq x \leq \delta/n$, whence g is right-continuous at 0.

`Exponential as a limit of geometric, which is the discrete memoryless ran-`
3" `dom variable:` $n^{-1}\mathrm{Geom}(\lambda/n) \Rightarrow \mathrm{Exp}(\lambda)$. ...

In general, if X has a probability density function, the ***failure*** or ***hazard rate*** function $\lambda_X(t)$ is $\lambda_X(t) := f_X(t)/\overline{F}_X(t)$. Thus $P\big[X \in (t, t+dt) \mid X > t\big] \approx \lambda_X(t)\, dt$, which explains the name.

§1.8. Some Limit Theorems.

WLLN. *If X_i are i.i.d. with mean $\mu \in (-\infty, \infty)$, then for all $\epsilon > 0$, we have*

$$\lim_{n\to\infty} P\Big(\frac{1}{n}\sum_{i=1}^n X_i \in (\mu - \epsilon, \mu + \epsilon)\Big) = 1.$$

SLLN. *If X_i are i.i.d. with mean $\mu \in [-\infty, \infty]$, then*

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

CLT. *If X_i are i.i.d. with finite mean μ and finite variance σ^2, then $\forall a \in \mathbb{R}$*

$$\lim_{n\to\infty} P\Big(\frac{1}{\sigma\sqrt{n}}\sum_{i=1}^n (X_i - \mu) \leq a\Big) = \int_{-\infty}^a \frac{1}{\sqrt{2\pi}} e^{-x^2/2}\, dx.$$

I.e.,

$$\frac{1}{\sigma\sqrt{n}}\sum_{i=1}^{n}(X_i-\mu)\Rightarrow \mathrm{N}(0,1).$$

(Let the random variables X_n have c.d.f. F_n and Y have c.d.f. F. We write that $X_n \Rightarrow Y$, $X_n \Rightarrow F$, or $F_n \Rightarrow F$ if $F_n(a) \to F(a)$ at every $a \in \mathbb{R}$ where $F(a)$ is continuous. This is called ***convergence in distribution***, ***convergence in law***, or ***weak convergence***. The last name is because this kind of convergence follows from a.s. convergence; that is, if $P[X_n \to Y] = 1$, then $X_n \Rightarrow Y$. The WLLN is about weak convergence to a constant random variable, while the SLLN is about a.s. convergence. If X_n and Y are integer valued, then $X_n \Rightarrow Y$ iff $P[X_n = k] \to P[Y = k]$ for all $k \in \mathbb{Z}$.)

We will use the following generalization only once:

CLT of Lindeberg. *Let X_i be independent, $F_i :=$ the c.d.f. of X_i. Suppose that $E(X_i) = 0$, $\mathrm{Var}(X_i) = \sigma_i^2 < \infty$,*

$$s_n^2 := \sum_{i=1}^{n}\sigma_i^2,$$

and

$$\forall t > 0 \qquad \lim_{n\to\infty}\frac{1}{s_n^2}\sum_{i=1}^{n}\int_{|x|\ge t s_n} x^2\,dF_i(x) = 0.$$

Then

$$\frac{1}{s_n}\sum_{i=1}^{n}X_i \Rightarrow \mathrm{N}(0,1).$$

2" Note that this *is* a generalization. ...

Poisson Convergence (The Law of Rare Events). *For every $\lambda > 0$, we have*

$$\mathrm{Bin}(n, \lambda/n) \Rightarrow \mathrm{Pois}(\lambda)$$

as $n \to \infty$. (See Exercise 14 (Exercise 1.3, p. 46), which extends this a bit.) More generally, suppose that $\forall n$ $X_{n,i}$ $(1 \le i \le n)$ are independent random variables with values in $\mathbb{N}$ such that

$$p_{n,i} := P[X_{n,i} = 1]$$

and

$$\varepsilon_{n,i} := P[X_{n,i} \ge 2]$$

satisfy

$$\sum_{i=1}^{n} p_{n,i} \to \lambda \in [0, \infty],$$

$$\max_{1 \le i \le n} p_{n,i} \to 0,$$

and

$$\sum_{i=1}^{n} \varepsilon_{n,i} \to 0.$$

Then

$$\sum_{i=1}^{n} X_{n,i} \Rightarrow \mathrm{Pois}(\lambda).$$

Note: There are two special interpretations: $\mathrm{Pois}(0)$ means the distribution of the random variable that is identically 0; and $\mathrm{Pois}(\infty)$ means the distribution of the random variable that is identically ∞. In the latter case, to say that random variables X_n converge weakly to ∞ means that for all $t < \infty$, we have $F_{X_n}(t) \to 0$ as $n \to \infty$.

The Monotone Convergence Theorem (MCT). *If $X_n \to X$ a.s. and $0 \le X_n \le X$ a.s., then $E[X_n] \to E[X]$.*

The Lebesgue Dominated Convergence Theorem (LDCT). *If $X_n \to X$ a.s., $|X_n| \le Y$, and $E[Y] < \infty$, then $E[X_n] \to E[X]$.*

The Bounded Convergence Theorem (BCT). *The LDCT for Y a constant.*

`Definition of independent increments and stationary increments for a stochastic process. We will be dealing with two stochastic processes that have independent and stationary increments: ones that jump (Poisson processes) and`
3" `ones that don't (Brownian motion). ...`

We'll finish with a fun fact:

EXAMPLE 1.9(A). There are n beads arranged on a circular necklace. Number them 1 through n. An ant starts at one of them, say, number 1, and takes a simple random walk
1" on the beads. ... For each $k \neq 1$, what is the chance that bead number k is visited only after all the other beads have been visited?

$\downarrow$

Solution. Surprisingly, it is the same for all k, whence it is $1/(n-1)$. To see this, consider the first time that either bead $k \pm 1$ is reached (counting mod n). At this time, what matters is whether the other bead $k \mp 1$ is reached before bead k. This does not depend
3" $\uparrow$ on the sign and clearly does not depend on k.

$\triangleright$ `Read pp. 35--36, 37--39, and 41--42 in the book.`

Chapter 2

The Poisson Process

Poisson processes are examples of point processes, which are models for random distributions of "particles" (called "points") in "space". E.g., this might include defects on a surface, raisins in cookies or cereal, misprints in books, stars in space, arrival times of phone calls at an exchange, etc. `Motivation for independent, stationary increments.` The theory of Poisson processes when space is 1-dimensional, or more precisely, $\mathbb{R}^+$, is especially important for its connections to many other stochastic processes. In this case, "space" is usually called "time" and the "points" are usually called "events".

Informally, we call $\langle N(t)\rangle_{t\geq 0}$ a counting process if $N(t)$ is the (finite) number of "events" occurring in $(0,t]$. Thus, $N(t)-N(s)=$ the number of events in $(s,t]$. Formally, the definition of a ***counting process*** is that

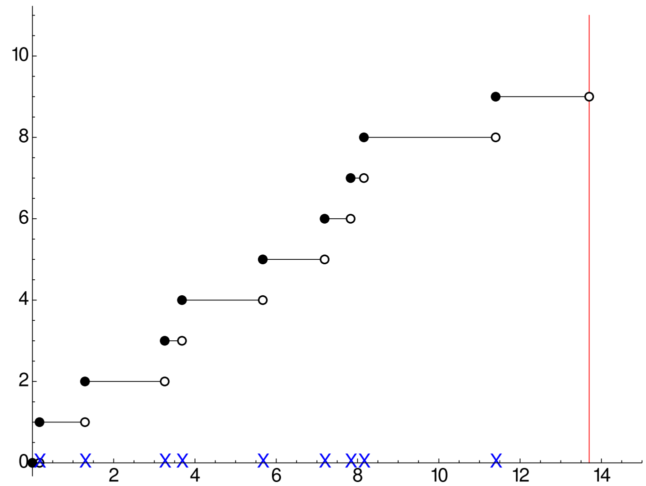
- $N(t)\in\mathbb{N}$,
- $s<t\Rightarrow N(s)\leq N(t)$, and
- $N(\cdot)$ is right-continuous (with probability 1).

A counting process is called ***simple*** if it never jumps by more than 1 (with probability 1). Thus, if $N(\cdot)$ is a simple counting process, then $N(t)-N(s)$ is equal to the number of jumps that occur in $(s,t]$; we usually refer to the location of a jump as an event.

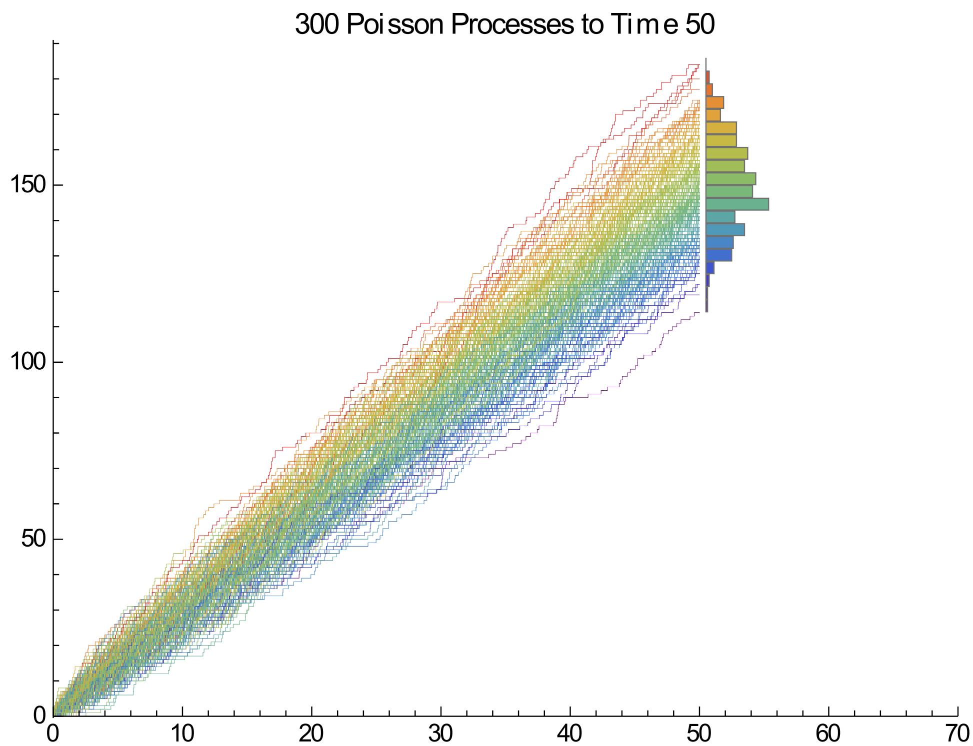
THEOREM. *Suppose that $N(\cdot)$ is a simple counting process with independent, stationary increments. Suppose that $P\big[N(0)=0\big]=1$ and $P\big[\forall t\ \ N(t)=0\big]=0$. Then $\exists\lambda\in(0,\infty)$ such that $\forall t\ \ N(t)\sim\mathrm{Pois}(\lambda t)$.*

DEFINITION. A process satisfying these hypotheses is called a ***Poisson process with rate*** λ.

`Demo and figure. Note normal limit: adding Poisson random variables or adding i.i.d. increments.`



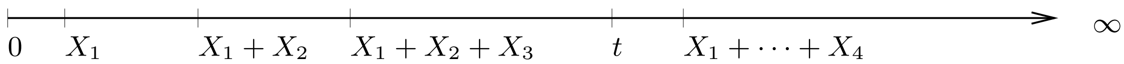
A sample of a Poisson process.



Proof. In order to apply the Poisson convergence theorem, fix t and let $X_{n,i}$ be the indicator that there is an event in $\big((i-1)t/n, it/n\big]$ for $1 \le i \le n$. Because $N(\cdot)$ is simple and
1" $N(0) = 0$, it follows that $Y_n := \sum_i X_{n,i} \uparrow N(t)$ a.s. as $n \to \infty$. ... Therefore $Y_n \Rightarrow N(t)$
1" as $n \to \infty$. By Exercise 14, it follows ... that $N(t)$ is a Poisson random variable. Let $m(t)$
1" denote its mean. Then $m(t) < \infty$ and $m(s+t) = m(s) + m(t)$. ... Since m is monotonic,
0.5" we get $m(t) = \lambda t$ for some $\lambda \in [0, \infty)$. ... It follows that $\lambda > 0$. ... ∎

We now show that such a process exists. Let X_n be i.i.d. $\mathrm{Exp}(\lambda)$. Set

$$N(t) := \sup\Big\{n\,;\ \sum_{k=1}^{n} X_k \le t\Big\}.$$

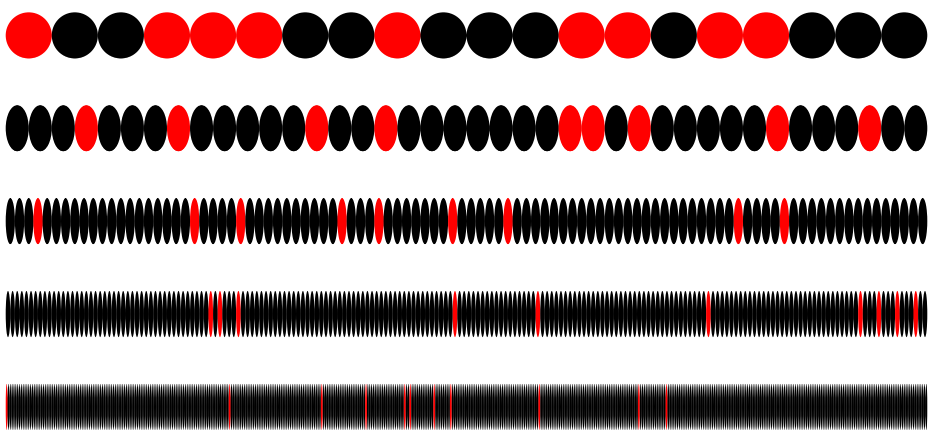


Since $\sum_1^n X_k/n \to 1/\lambda$ a.s., we have $N < \infty$ a.s. Clearly N jumps by 1 and, by the
1.5" memoryless property, has independent, stationary increments. . . . Finally,

$$P\big[N(t) = 0\big] = P[X_1 > t] = e^{-\lambda t},$$

whence λ is the rate of the Poisson process N. This proves existence of the process. The sequence $\langle X_n\,;\ n \ge 1\rangle$ is called the sequence of ***interarrival times***.

`Give the intuition in terms of coins and the limits of geometric/binomial random variables.`



A Poisson process is a scaling limit of Bernoulli trials.

Let $\langle N(t)\,;\ t \ge 0\rangle$ be a stochastic process with independent, stationary increments that is right-continuous and all of whose discontinuities are jump discontinuities. Assume $N(0) \equiv 0$.* The following properties can be shown:

• For all $s > 0$, if an event A is defined in terms of $N(t)$ for $t \le s$ and another event B is defined in terms of the increments $N(t) - N(s)$ for $t > s$, then A and B are independent.

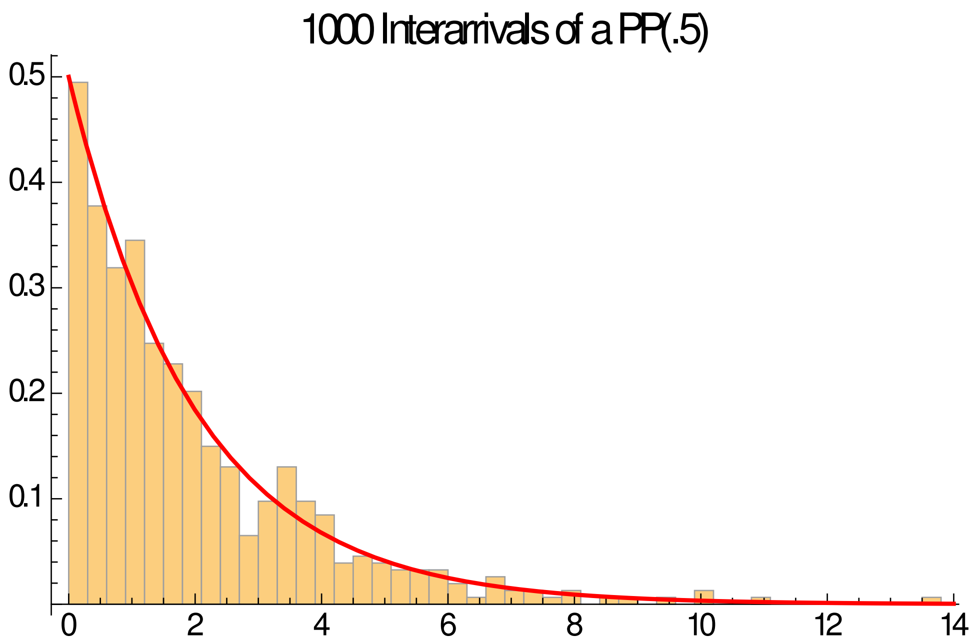
1" . . .

* Such stochastic processes that are also nonnegative are called ***subordinators***. The condition of no jump discontinuities is the same as the existence of left limits everywhere once we assume right-continuity. Another name for right-continuous functions with left limits is ***càdlàg***, which is an abbreviation for the French phrase, “continue à droite, limites à gauche”.

• The Markov property holds: for all $s > 0$, if an event B is defined in terms of $N(t)$
1" for $t > s$, then $P\big(B \bigm| N(t), t \le s\big) = P\big(B \bigm| N(s)\big)$. ...

• The strong Markov property holds: Suppose τ is a random variable with values in
$[0, \infty)$ such that for all s, the event that $\tau \le s$ depends only on $N(t)$ for $t \le s$. Then if
an event A is defined in terms of $N(t)$ for $t \le \tau$ and another event B is defined in terms
of $N(\tau + t) - N(\tau)$ for $t \ge 0$, then A and B are independent. Furthermore, the law of
1" $\langle N(\tau + t) - N(\tau)\, ;\ t \ge 0\rangle$ is the same as the law of $\langle N(t)\, ;\ t \ge 0\rangle$. ...

Now it is easily calculated (see p. 64) that the first arrival time of a Poisson
1" process with rate λ has an $\mathrm{Exp}(\lambda)$ distribution. ... By the strong Markov property, all
the interarrival times have an $\mathrm{Exp}(\lambda)$ distribution and, in fact, the Poisson process is of
2" the type constructed. ...



Thus, λ uniquely determines the law of the Poisson process. Here, "law" (or "distribution") is the analogue for processes of "c.d.f." for a real-valued random variable. There are, in fact, two interpretations of "law" for processes. The simpler one is the collection of joint distributions of $\langle N(t_i)\rangle$ for all finite $\{t_i\} \subset \mathbb{R}$. To be more explicit, these are called the ***finite-dimensional marginals*** of the process N. This is often sufficient; but note that N and N' may have the same finite-dimensional marginals, yet N may be right-continuous
1" ↕ and N' not be. Example: Poisson process made left-continuous.

Thus, sometimes one introduces a space of functions that each $N(\cdot)$ belongs to. For example, for counting processes, N, we have that a.s., $N \in C_{\rm r}\big([0,\infty)\big)$. Then the ***law*** of N is the collection of probabilities $P[N \in A]$, where $A \subseteq C_{\rm r}\big([0,\infty)\big)$.* In fact, in this case, one can show that the finite-dimensional marginals determine the law in this case, even if we use only rational times, because the values at rational times combined with right-continuity determine the function values at all positive times.

In any case, when we say that two processes have the same law, it means that all relevant probabilities are the same for the two processes: They are indistinguishable probabilistically. (Of course, this does not mean they are the same process, just as we do not say that two fair coins are the same coin.)

An equivalent definition of Poisson process is given in the book:

THEOREM. *Suppose that $N(\cdot)$ is a counting process with independent, stationary increments. If $\exists\lambda \in (0,\infty)$ such that $\forall t\ \ N(t) \sim \mathrm{Pois}(\lambda t)$, then $N(\cdot)$ is a Poisson process with rate λ.*

Proof. That $P\big[N(0) = 0\big] = 1$ is clear from $N(0) \sim \mathrm{Pois}(0)$. Also, $P\big[\forall t\ \ N(t) = 0\big] \le P\big[N(s) = 0\big] = e^{-\lambda s}$ for all $s \ge 0$, whence $P\big[\forall t\ \ N(t) = 0\big] = 0$. It remains to show that $N(\cdot)$ is simple. It suffices to show that for each $k \ge 1$, $N(\cdot)$ does not jump by more than
1" 1 in $[0,k]$. ... Fix k and let A be the event that there is a jump by more than 1 in $[0,k]$. Let $A_{n,i}$ be the event that $N(ik/n) - N\big((i-1)k/n\big) \ge 2$. Then $A \subseteq \bigcup_{i=1}^n A_{n,i}$ for each n. Because

$$N(ik/n) - N\big((i-1)k/n\big) \sim \mathrm{Pois}(\lambda k/n),$$

1" it follows that $P(A_{n,i}) = 1 - e^{-\lambda k/n} - (\lambda k/n)e^{-\lambda k/n} = o(\lambda k/n)$ as $n \to \infty$. ... Hence $P(A) \le \sum_{i=1}^n P(A_{n,i}) \le n \cdot o(\lambda k/n) = o(1)$, i.e., $P(A) = 0$. ∎

In fact, we do not need the assumption of independent increments in the preceding theorem, nor that the process be a counting process:

LEMMA. *If $X \sim \mathrm{Pois}(\lambda)$ and Y is independent of X with $X + Y \sim \mathrm{Pois}(\lambda + \mu)$, then $Y \sim \mathrm{Pois}(\mu)$.*

Proof. Recall that the probability generating function of a $\mathrm{Pois}(\lambda)$ random variable is $z \mapsto e^{\lambda(z-1)}$. By independence of X and Y, the p.g.f. of $X + Y$ divided by the p.g.f. of X is the p.g.f. of Y. Since this quotient equals $z \mapsto e^{\mu(z-1)}$, it follows that $Y \sim \mathrm{Pois}(\mu)$. ∎

---

* Technically, one needs to introduce a σ-field on $C_{\rm r}\big([0,\infty)\big)$ and then $A \mapsto P[N \in A]$ is a probability measure on this σ-field.

Corollary. *Suppose that $N(\cdot)$ is a right-continuous process with independent increments. If $\exists\lambda \in (0,\infty)$ such that $\forall t\ \ N(t) \sim \text{Pois}(\lambda t)$, then $N(\cdot)$ is a Poisson process with rate λ.*

Proof. Apply the lemma to $N(\cdot)$: Given $0 < s < t$, let $X := N(s)$ and $Y := N(t) - N(s)$ in the preceding paragraph. The assumptions that Y is independent of X, $X \sim \text{Pois}(\lambda s)$, and $X + Y = N(t) \sim \text{Pois}(\lambda t)$ imply that $Y \sim \text{Pois}\big(\lambda(t-s)\big)$. In other words, $N(\cdot)$ has stationary increments. Furthermore, $X(t) \in \mathbb{N}$ a.s. for every $t \in \mathbb{Q}$, whence by right-continuity, $X(t) \in \mathbb{N}$ for every $t \geq 0$ a.s. In addition, $Y \geq 0$ a.s., so (using the rationals again) $N(\cdot)$ is a counting process. ∎

▷ `Read §2.2, pp. 64--66 in the book.`

Let $N_1(\cdot), \ldots, N_k(\cdot)$ be real-valued stochastic processes defined on the same probability space and indexed by a set T of "times". The processes are called ***(mutually) independent*** if for any events A_i $(1 \leq i \leq k)$ such that A_i depends only on $N_i(\cdot)$, the events A_i are independent; this is equivalent to the following condition: for all $J \in \mathbb{N}$, all $t_{i,j} \in T$ and all $A_{i,j} \subseteq \mathbb{R}$ $(1 \leq i \leq k,\ 1 \leq j \leq J)$, we have

$$P\big[\forall i\ \ \forall j\ \ N_i(t_{i,j}) \in A_{i,j}\big] = \prod_i P\big[\forall j\ \ N_i(t_{i,j}) \in A_{i,j}\big].$$

Example PM 5.3.6 (Estimating Software Reliability). New software is tested for time t. After the whole run is complete, the bugs discovered are fixed. What error rate remains? Suppose the bugs cause errors like a Poisson process with rate λ_i $(i \geq 1)$. Suppose also that they are independent. If $\psi_i(t)$ is the indicator that bug i has *not* caused an error by time t, then we want to estimate

$$\Lambda(t) := \sum_i \lambda_i \psi_i(t).$$

(By Exercise 22, the remaining bugs cause errors together at the times of a Poisson process with rate $\Lambda(t)$.)

↓ Naturally, the bugs with small λ_i are those remaining. Let $M_j(t) :=$ the number of bugs that caused exactly j errors up to time t. If $X_i(t)$ is the indicator that bug i has caused exactly one error by time t, then

$$E[\Lambda(t)] = \sum \lambda_i e^{-\lambda_i t},$$
$$E[M_1(t)] = E\Big[\sum X_i\Big] = \sum \lambda_i t e^{-\lambda_i t},$$

whence

$$E\left[\Lambda(t) - \frac{M_1(t)}{t}\right] = 0.$$

Thus, $M_1(t)/t$ may be a good estimate of $\Lambda(t)$. In this way, we estimate the unknown (unobserved) by the known (observed). This magic is made possible by probability and is the foundation of statistics. `Like the game of choosing the higher of 2 numbers when we are allowed to see only one.` To see how good this estimate is, compute

$$\operatorname{Var}\Big(\Lambda(t) - \frac{M_1(t)}{t}\Big) = \frac{E[M_1(t) + 2M_2(t)]}{t^2}$$

(after lengthy calculations, shown below). Thus, we may estimate the error by

$$\sqrt{M_1(t) + 2M_2(t)}/t.$$

(Clearly, we should test until $M_1(t)$ and $M_2(t)$ are small compared to t^2.)

For example, suppose that at $t = 100$, we discover 20 bugs, of which 2 cause 1 error and 3 cause 2 errors. Then $\Lambda(100) \approx \frac{2}{100} \pm \frac{\sqrt{8}}{100}$.

Here are the calculations: Recall that the variance of a $\operatorname{Bin}(1, p)$-random variable is $p(1-p)$. Since $\psi_i(t) \sim \operatorname{Bin}(1, e^{-\lambda_i t})$ and $X_i(t) \sim \operatorname{Bin}(1, \lambda_i t e^{-\lambda_i t})$, we have

$$\begin{aligned}
\operatorname{Var}\Big(\Lambda(t) - \frac{M_1(t)}{t}\Big) &= \operatorname{Var}\Big(\sum_i \big(\lambda_i \psi_i(t) - X_i(t)/t\big)\Big) \\
&= \sum_i \operatorname{Var}\big(\lambda_i \psi_i(t) - X_i(t)/t\big) \\
&= \sum_i \Big[\lambda_i^2 \operatorname{Var}\big(\psi_i(t)\big) + \frac{1}{t^2} \operatorname{Var}\big(X_i(t)\big) - 2\frac{\lambda_i}{t} \operatorname{Cov}\big(\psi_i(t), X_i(t)\big)\Big] \\
&= \sum_i \Big[\lambda_i^2 e^{-\lambda_i t}(1 - e^{-\lambda_i t}) + \frac{1}{t^2} \lambda_i t e^{-\lambda_i t}(1 - \lambda_i t e^{-\lambda_i t}) + 2\frac{\lambda_i}{t} e^{-\lambda_i t} \lambda_i t e^{-\lambda_i t}\Big] \\
&\qquad \text{since } \psi_i(t) X_i(t) \equiv 0 \\
&= \sum_i \Big(\lambda_i^2 e^{-\lambda_i t} + \frac{1}{t} \lambda_i e^{-\lambda_i t}\Big) \\
&= \sum_i \lambda_i^2 e^{-\lambda_i t} + \frac{E\big(M_1(t)\big)}{t^2}.
\end{aligned}$$

On the other hand, if $Y_i(t)$ denotes the indicator that bug i has caused exactly 2 errors by time t, then

$$E\big(M_2(t)\big) = \sum_i E\big(Y_i(t)\big) = \sum_i \frac{(\lambda_i t)^2}{2!} e^{-\lambda_i t},$$

6" ↑ whence we obtain the desired formula.

6" . . .

Suppose that each event of a Poisson process with rate λ is classified independently as type i ($1 \le i \le K$) with probability p_i, where $\sum_{i=1}^{K} p_i = 1$. We also assume that the type classification is independent of the times of the events. Let $N_i(t)$ be the number of type-i events by time t.* Here's an amazing fact:

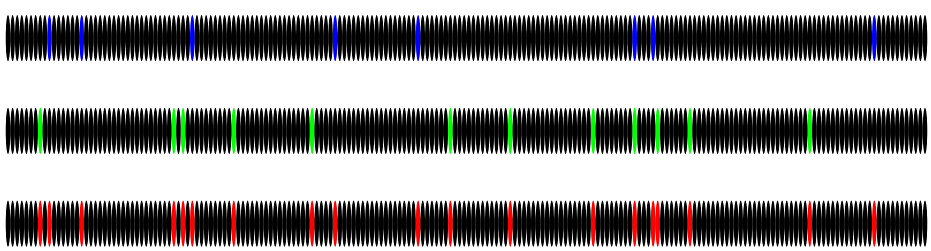
THEOREM. *For each i, $N_i(\cdot)$ is a Poisson process with rate λp_i and these processes are mutually independent.*

`Demo`

Proof. This proof will use the following idea: Suppose we want to show that a random variable X has the same distribution as a random variable Y. One way to do this would be to work with X and manipulate its distribution function to show that it is the same as that of Y. A second way would be to work with Y from the start and show it has the same distribution as X. These strategies are obvious (and not the only useful strategies), but it can be confusing when we do it the second way in the present context. We will use this same idea repeatedly after the proof of this theorem.

Let $\widetilde{N}_i(\cdot)$ be independent Poisson processes with rates λp_i. Set $\widetilde{N}(t) := \sum \widetilde{N}_i(t)$. By Exercise 22 (p. 89, 2.5, extended), $\widetilde{N}(\cdot)$ is a Poisson process with rate λ. Call the events of $\widetilde{N}(\cdot)$ type i if they come from $\widetilde{N}_i(\cdot)$. Also by Exercise 22, the first event of $\widetilde{N}(\cdot)$ is of type i with probability p_i, independently of the time of the first event. By the strong Markov property applied to all $\widetilde{N}_i(\cdot)$, the same holds for the second event of $\widetilde{N}(\cdot)$,
1" independently of the first, etc. ... Thus, $\langle \widetilde{N}_i(\cdot)\,;\, 1 \le i \le k\rangle$ comes from $\widetilde{N}(\cdot)$ by the same classification procedure that gives $\langle N_i(\cdot)\rangle$ from $N(\cdot)$. Therefore $\langle \widetilde{N}_i(\cdot)\rangle \overset{\mathscr{D}}{=} \langle N_i(\cdot)\rangle$, from which the theorem follows. ∎

`Give the intuition in terms of coins and limits.`



Combining two Bernoulli processes.

---

* More formally, let Z_n be i.i.d. random variables for $n \ge 1$, independent also of the Poisson process, with $P[Z_n = i] = p_i$. We call the nth event type Z_n. Then $N_i(t)$ is the number of events by time t whose type is i.

This proof is short, but subtle. (The proof of Exercise 22 is also short when done the best way.) A calculational proof of the above theorem, if one wants one, proceeds along the following lines: Look at the interarrival times of $N_i(\cdot)$. The combination of the memoryless property of the geometric distribution and of the exponential distribution shows that the interarrival times are i.i.d. with distribution equal to that of the sum of $\text{Geom}(p_i)$ independent $\text{Exp}(\lambda)$ random variables. What is this distribution? We claim that it is $\text{Exp}(\lambda p_i)$. [Note that this follows immediately from the theorem, but we are trying to give a direct proof.] One method to prove this is calculational; e.g., calculate the moment generating function and use the result that the m.g.f. determines the distribution uniquely. A simpler way is to verify the memoryless property and calculate the mean.
1" ... (And intuitively, if we think of an exponential random variable as a scaling limit of geometric random variables, it follows from the fact that a geometric sum of geometrics is geometric.) In any case, once we have this done, it follows that for each i, the process $N_i(\cdot)$ is a $\text{Pois}(\lambda p_i)$ process. To show that the processes are mutually independent requires verifying a statement that is already complicated to state, still more to prove. Actually, it is not too hard to prove, but it is messy. So we will just write out a very simple case: Given any numbers n_i and writing $n := \sum_{i=1}^K n_i$, we have

$$
\begin{aligned}
&P\big(N_1(t) = n_1, N_2(t) = n_2, \ldots, N_K(t) = n_K\big) \\
&\quad = P\big(N_1(t) = n_1, N_2(t) = n_2, \ldots, N_K(t) = n_K \;\big|\; N(t) = n\big) P\big(N(t) = n\big) \\
&\quad = \binom{n}{n_1\ n_2\ \cdots\ n_K} \prod_{i=1}^K p_i^{n_i} \cdot e^{-\lambda t} \frac{(\lambda t)^n}{n!} \\
&\quad = \frac{n!}{n_1!\, n_2! \cdots\ n_K!} \prod_{i=1}^K p_i^{n_i} \cdot e^{-\lambda t} \frac{(\lambda t)^n}{n!} \\
&\quad = \prod_{i=1}^K e^{-\lambda p_i t} \frac{(\lambda p_i t)^{n_i}}{n_i!} \\
&\quad = \prod_{i=1}^K P\big(N_i(t) = n_i\big).
\end{aligned}
$$

This is only the simplest case of independence, since here all the times were the same, t. But this should be enough to give an idea of how much pain is saved by the conceptual proof above.

Corollary. *The sum of independent Poisson random variables is a Poisson random variable whose mean is the sum of the means. If a* Pois(λ) *number of objects is classified independently as type i with probability p_i each, then the number of type-i objects is* Pois(λp_i) *and these numbers are independent.*

Proof. The first part follows from Exercise 22 by embedding the Poisson random variables
2" in Poisson processes as their values at time 1. ... Note that this proof is accomplished without any real calculation (if Exercise 22 was done the best way) and gives the result of Exercise 5(a) (p. 47, 1.8). The second part follows from the theorem also by embedding
1" and gives the result of Exercise 5(b). ... Again, this requires no further calculation. ∎

Example MASS 5.16. No device works perfectly. Suppose that a Geiger counter fails to register an arriving radioactive particle with probability 0.1, independently of everything else. Suppose also that radioactive particles arrive at the counter according to a Pois(1000/sec) process. If during a certain 1/100 sec, the counter registered 4 particles, what is the probability that actually more than 5 arrived?

↓

Solution. This is the probability that it missed at least 2. The missed particles form a Pois(100/sec) process, so the chance that there were at least 2 in that interval is

$$1 - (e^{-1}1^0/0! + e^{-1}1^1/1!) = 1 - 2/e = 0.264^+.$$

Alternatively, one can start with the fact that the number of arriving particles in that
2" ↑ interval is Pois(10) and then use the corollary.

If we classify the events of a Poisson process into only two types, "keep" and "remove", then we often refer to the kept events as a ***thinned*** process.

Another amazing fact about Poisson processes is that, for every $t > 0$, given that $N(t) = n$, the n events in $(0, t]$ are distributed the same as n i.i.d. Unif$[0, t]$ points that we'll call "dots" (Theorem 2.3.1).

2" If we think about Bernoulli trials, then this is actually quite intuitive. ... Moreover, we see that to estimate the success probability in n trials, we can do no better than use the total number of successes: Given that there are k successes, the distribution of those k is uniform on the n trials. In particular, it does not depend on p. Similarly, to estimate λ given $N(s)$ for $s \le t$, we can do no better than use $N(t)$. These are so-called sufficient statistics.

In fact, another way to construct a Poisson process is to choose i.i.d. $Y_0, Y_1, Y_2, \ldots \sim$ Pois(λ) and, given $\langle Y_i \rangle_{i=0}^\infty$, choose Y_i independent Unif$[i, i+1]$ random variables. The

resulting set of dots on $[0, \infty)$ gives the arrival times. (Here, the positive integers could be replaced by any sequence of times increasing to ∞, with a corresponding change in the Poisson parameters.) To see that this is a Poisson process with
0" rate λ, we need only check that increments are independent and stationary. ...

To check independence, it's enough to show that $\langle N(t_i) - N(t_{i-1}) \rangle_{i=1}^{r}$ are independent
1" for $0 = t_0 < t_1 < \cdots < t_r = 1$. ... But these numbers are obtained by taking Y_0 dots, each independently having probability $t_i - t_{i-1}$ of falling in $(t_{i-1}, t_i]$; so we may apply the above corollary.

0" Now we check that $N(s+t) - N(s) \overset{\mathscr{D}}{=} N(t)$. This is clear for $s + t \le 1$. ... To see that it holds for $s + t \le 2$, it suffices to show that the dots in $[0, 2]$ are uniform and
0" independent, given that their number is $Y_0 + Y_1 \sim \mathrm{Pois}(2\lambda)$. ... Stated another way, it suffices to show that if we choose $\mathrm{Pois}(2\lambda)$ dots in $[0, 2]$ independently and uniformly, then the numbers in $[0, 1]$ and $[1, 2]$ are independent $\mathrm{Pois}(\lambda)$ and given that a point falls in one
1" of the two halves, it is uniformly distributed in that half. ... But note that if we choose $\mathrm{Pois}(2\lambda)$ dots in $[0, 2]$ independently and uniformly, then each has chance $1/2$ of falling in $[0, 1]$, so by the corollary, the number that fall in $[0, 1]$ is $\mathrm{Pois}(\frac{1}{2} \cdot 2\lambda) = \mathrm{Pois}(\lambda)$ and they are clearly i.i.d. $\mathrm{Unif}[0, 1]$; similarly for those in $[1, 2]$; and these are independent of each other. Likewise, we see it holds for $s + t \le n$ for every n.

A calculational proof of almost the same statement is given in the book on p. 67. It is not long.

Example 2.3(a). Suppose that travelers arrive at a train station according to the times of a $\mathrm{Pois}(\lambda)$ process during the interval $[0, T]$. If the train leaves at time T, what is the expected total waiting time of the passengers (i.e., the sum of all the waiting times)?

↓

Solution. Condition on the number of passengers and then use their uniform distribution.

3" ↑

Example MASS 5.11. Suppose that customers arrive at an automatic teller machine (ATM) according to the times of a $\mathrm{Pois}(\lambda)$ process. The ATM records the start and finish times of each customer's service, but not when the customers arrive (if they join a queue). Suppose that the ATM is opened for business one day at 7:00am and that the log that day turns out to begin as follows:

Customer No.	Service Start Time	Service Completion Time
0	7:30	7:34
1	7:34	7:40
2	7:40	7:42
3	7:45	7:50

What is the expected arrival time of Customer 1 given the above information?

↓

Solution. Let A_n be the arrival time of customer n. We want

$$E\big[A_1 \bigm| A_0 = 7{:}\,30, A_1 \le 7{:}\,34, A_2 \le 7{:}\,40, A_3 = 7{:}\,45\big].$$

(The only other information we are given is the service times—i.e., we can reconstruct the table from the service times and the conditioned event—, but these are independent of the arrival times. We are making an implicit assumption of such independence; e.g., in principle, it could be that a customer who starts service just after arriving takes longer to be served, needing time to think about how much to withdraw.) This is the same as

$$E\big[A_1 \bigm| A_0 = 7{:}\,30, A_1 \le 7{:}\,34, A_2 \le 7{:}\,40, N(7{:}\,40) - N(7{:}\,30) = 2\big].$$

We begin by calculating the conditional distribution of A_1. We also will convert time to minutes after 7:30 and count arrivals only after that. We have for $x \le 4$,

$$\begin{aligned} P\big[A_1 \le x \bigm| A_1 \le 4, A_2 \le 10, N(10) = 2\big] &= P\big[A_1 \le x \bigm| A_1 \le 4, N(10) = 2\big] \\ &= \frac{P\big[A_1 \le x, A_1 \le 4 \bigm| N(10) = 2\big]}{P\big[A_1 \le 4 \bigm| N(10) = 2\big]} \\ &= \frac{P\big[A_1 \le x \bigm| N(10) = 2\big]}{P\big[A_1 \le 4 \bigm| N(10) = 2\big]}. \end{aligned}$$

Now we use the theorem to calculate that

$$P\big[A_1 \le x \bigm| N(10) = 2\big] = 1 - P\big[A_1 > x \bigm| N(10) = 2\big] = 1 - \left(\frac{10 - x}{10}\right)^2 = \frac{20x - x^2}{100}.$$

Therefore

$$P\big[A_1 \le x \bigm| A_1 \le 4, A_2 \le 10, N(10) = 2\big] = \frac{20x - x^2}{64}.$$

This allows us to calculate the expectation as

$$\int_0^4 x \frac{20 - 2x}{64} dx = 11/6.$$

Converting to the original time scale, this gives exactly 7:31:50am. (That this is independent of λ should have been anticipated once we formulated it as depending only on probabilities conditional on $N(10) = 2$. It is earlier than 7:32, which it would be if customer 2 did not arrive before 7:34, but there is a chance that he did. Likewise, it is later than 7:31:20 because there is a chance that customer 2 arrived after 7:34. Therefore,
8" ↑ $P[A_2 \le 7 : 34] = 1/4$.)

Example PM 5.10 (The Coupon Collector's Problem). There are m different types of coupons. Each time you collect one, it has probability p_j ($1 \le j \le m$) of being of type j, independently of the past. How many coupons do you expect to have to collect in order to have a complete set?

↓

Solution. We use a method known as "Poissonization": it consists in introducing Poisson random variables or processes where there are none apparent in the problem.

We may suppose that the coupons are collected at the times of a Poisson process $N(\cdot)$ with rate 1. Classifying by type of coupon decomposes this into m *independent* Poisson processes with rates p_j. Let $X(j)$ be the *first* waiting time of the jth process. Then $X := \max_j X(j)$ is the time when a complete collection is amassed. We want to know how many coupons, Y, have been collected at this time. Now $X = \sum_{i=1}^{Y} T_i$, where $\langle T_i \rangle$ are the interarrival times of $N(\cdot)$. Therefore, $E[X] = E[Y]E[T] = E[Y]$. (Note that although $Y = N(X)$, we cannot calculate $E[Y] = E\big[E[N(X) \mid X]\big]$ easily, since $N(X)$ is at least m and so does not have a Poisson distribution.*)

Thus, it remains to calculate $E[X]$. Now $\forall t > 0$

$$P[X \le t] = P[\forall j \;\; X(j) \le t] = \prod_{j=1}^{m} (1 - e^{-p_j t}).$$

Therefore

$$E[X] = \int_0^\infty P[X > t] \, dt = \int_0^\infty \Big[1 - \prod_{j=1}^{m} (1 - e^{-p_j t})\Big] \, dt.$$

(To get the maximum amount of fun out of this example, show that if $p_j \equiv \frac{1}{m}$, then $E[X] = m \sum_{i=1}^{m} 1/i \sim m \log m$ by changing variables to $x := 1 - e^{-t/m}$. Another way to do this particular case is to calculate $E[X]$ by the result of Exercise 26(b). Then show
7" ↑ that $E[X]$ is a strict minimum when $p_j \equiv \frac{1}{m}$.)

We now study counting processes that may not have stationary increments, called ***(nonhomogeneous) Poisson processes***:

---

* However, $\langle N(t) - t \rangle$ is a martingale and X is a stopping time with respect to an appropriate filtration, so the optional stopping theorem gives $E\big[N(X)\big] = E[X]$.

THEOREM. *Suppose that $N(\cdot)$ is a simple counting process with independent increments,*[†] *$N(0) \equiv 0$, and $\forall t$ $P[N(\cdot)$ jumps at $t] = 0$. Then $\forall t$ $\exists m(t) < \infty$ and*

$$\forall s < t \qquad N(t) - N(s) \sim \mathrm{Pois}\big(m(t) - m(s)\big).$$

Also, m is continuous.

If $m(t) = \int_0^t \lambda(s)\, ds$ for some function $\lambda(\cdot)$, then $\lambda(\cdot)$ is called the ***intensity function*** of $N(\cdot)$.

`Intuition: when there is an intensity function, we use coins of varying probabilities of heads.`

Proof. A similar proof as the stationary case works, except that we use the Poisson convergence theorem in greater generality. Fix t and let $X_{n,i}$ be the indicator that there is an event in the interval $\big((i-1)t/n, it/n\big]$ for $1 \le i \le n$. We first show that

$$\max_{1 \le i \le n} P[X_{n,i} = 1] \to 0.$$

If this were not true, then since every bounded sequence has a convergent subsequence, we could find $t_0 \in [0, t]$, integers i_k and n_k, and $p > 0$ such that $\forall k$ $P[X_{n_k,i_k} = 1] \ge p$ and
1" $i_k t/n_k \to t_0$. ... We can enlarge the intervals so as to be decreasing and contain t_0, while still having length tending to 0. But then $P\big[N(\cdot)$ jumps at $t_0\big] \ge p > 0$, a contradiction.
2" ...

The fact that $N(\cdot)$ is simple guarantees that $\sum_i X_{n,i} \uparrow N(t)$ a.s., whence the Poisson convergence theorem implies that $N(t)$ is a Poisson random variable. Let $m(t) \in [0, \infty)$ be its mean.

We finally show that m is continuous. Given t_0, let $s < t_0 < t$. Since $N(s) \le N(t_0) \le N(t)$, we know that $m(s) \le m(t_0) \le m(t)$. The number of events in $(s, t]$ is $\mathrm{Pois}\big(m(t) - m(s)\big)$: By the above argument starting at time s, it is Poisson; its mean is $E\big[N(t) - N(s)\big] = m(t) - m(s)$. If $m(t) - m(s) \not\to 0$ as $t - s \to 0$, then the probability of an event occurring in $(s, t] \not\to 0$ either, making the probability of an event at t_0 positive, a
1" contradiction. ... ∎

As before, an equivalent definition of nonhomogeneous Poisson process is the following:

† We will not actually use the full strength of the assumption of independent increments, only that the events of having no arrivals within disjoint intervals are independent.

Theorem. *Suppose that $N(\cdot)$ is a counting process with independent increments and $N(0) \equiv 0$. If there is a continuous function $m\colon [0,\infty) \to [0,\infty)$ such that*

$$\forall s, t \geq 0 \;\; N(s+t) - N(s) \sim \operatorname{Pois}\bigl(m(s+t) - m(s)\bigr),$$

then $N(\cdot)$ is a nonhomogeneous Poisson process.

Proof. Looking at small intervals surrounding a fixed time, we see that the probability that
1" $N(\cdot)$ jumps at that fixed time is 0. ... To prove that $N(\cdot)$ is simple, it suffices, as before,
1" to show that for each $k \geq 1$, $N(\cdot)$ does not jump by more than 1 in $[0, k]$. ... Fix k. The Maclaurin series with remainder tells us that for $x \geq 0$, we have $e^x = 1 + x + e^\xi x^2/2$ for some $\xi \in [0, x]$. Since $e^\xi \leq e^x$, we obtain upon dividing by e^x that $1 - e^{-x} - xe^{-x} \leq x^2/2$ for $x \geq 0$. Let $\epsilon > 0$. Because $m(\cdot)$ is continuous, m is uniformly continuous on $[0, k]$, so $\exists \delta > 0$ such that if $s, t \in [0, k]$ satisfy $|s - t| < \delta$, then $|m(s) - m(t)| < \epsilon$. Let A be the event that there is a jump by more than 1 in $[0, k]$. Let $A_{n,i}$ be the event that $N(ik/n) - N\bigl((i-1)k/n\bigr) \geq 2$. Then $A \subseteq \bigcup_{i=1}^n A_{n,i}$ for each n. Let $n > 1/\delta$. Because

$$N(ik/n) - N\bigl((i-1)k/n\bigr) \sim \operatorname{Pois}\bigl(m(ik/n) - m((i-1)k/n)\bigr),$$

it follows that

$$P(A_{n,i}) \leq \bigl(m(ik/n) - m((i-1)k/n)\bigr)^2/2 \leq \epsilon\bigl(m(ik/n) - m((i-1)k/n)\bigr)/2.$$

Hence $P(A) \leq \sum_{i=1}^n P(A_{n,i}) \leq \epsilon m(k)/2$. Since this holds for all $\epsilon > 0$, it follows that $P(A) = 0$. ∎

We can weaken this in the same way we did for the homogeneous case:

Theorem. *Suppose that $N(\cdot)$ is a right-continuous process with independent increments and $N(0) \equiv 0$. If there is a continuous function $m\colon [0,\infty) \to [0,\infty)$ such that*

$$\forall t \geq 0 \;\; N(t) \sim \operatorname{Pois}\bigl(m(t)\bigr),$$

then $N(\cdot)$ is a nonhomogeneous Poisson process. ∎

Example PM 5.20. Dogbert runs a hotdog stand. He observes that customers arrive at an increasing rate from opening time at 8:00am until 11:00am, then at a steady rate until 1:00pm, and then at a decreasing rate until closing at 5:00pm. He models the arrival times as a nonhomogeneous Poisson process with piecewise linear continuous intensity (with 3 pieces). He measures that on average, the number of customers before 11am is 37.5, the number at lunch (the steady period) is 40, and the number after 1pm is 64.

(a) Assume Dogbert's model. What is the expected number of customers arriving between 8:30am and 9:30am?

(b) What is the probability that no customers arrive between 8:30am and 9:30am?

↓

Solution. The lunch rate per hour is 20, the 8am rate is 5, and (not needed) the
7" ↑ 5pm rate is 12. This gives (a) 10 and (b) e^{-10}.

Discuss why nonhomogeneous Poisson processes might be reasonable models in general.

Here is a way to construct nonhomogeneous Poisson processes. Let $m\colon [0,\infty) \to [0,\infty)$ be a continuous, nondecreasing function with $m(0) = 0$. Suppose that $\widetilde{N}(\cdot)$ is a homogeneous Poisson process of rate 1. Define $N(t) := \widetilde{N}\big(m(t)\big)$. Then $N(\cdot)$ is a Poisson
2" process with mean function m. ...

Using measure theory, one can prove that the mean function $m(\cdot)$ uniquely determines
1" $N(\cdot)$ (in law). ...

It is easy to check that the sum of a finite number of independent (nonhomogeneous)
1" Poisson processes is a Poisson process. ... For simplicity, we state the following for Poisson processes with continuous intensity:

Theorem. *Let $N(\cdot)$ be a Poisson process with continuous intensity $\lambda(\cdot)$. Let $p_i(\cdot)$ ($i = 1, \ldots, k$) be continuous functions with values in $[0,1]$ and $\sum_{i=1}^k p_i(t) = 1$ for all $t \ge 0$. Classify each event as type i with probability $p_i(t)$ if it occurs at time t, independently of other events.* Then the type-i events form a Poisson process with intensity $\lambda_i(t) := \lambda(t)p_i(t)$ and these k Poisson processes are mutually independent.*

The theorem holds even if $p_i(\cdot)$ are not continuous, but we won't prove that.

Proof. Let $N_i(\cdot)$ be independent Poisson processes with intensity $\lambda_i(t)$. Then their sum has the same law as $N(\cdot)$. It suffices to show that the events are classified independently as given.

Consider a very small interval $(t-h, t+h]$. With probability $o(h)$, it has ≥ 2 events.
2" ... Thus, the probability that there exists an event in $(t-h, t+h]$ of type i given that
.5" there exists an event in $(t-h, t+h]$ is ...

$$\frac{\int_{t-h}^{t+h} \lambda_i(s)\,ds}{\int_{t-h}^{t+h} \lambda(s)\,ds} + o(1),$$

.5" ... whence an event at time t is type i with probability $p_i(t)$. ... This is independent of
1" other events, as seen by considering disjoint intervals. ... ∎

* More formally, if the events of $N(\cdot)$ are $S_1 < S_2 < \cdots$, then for all $n \ge 1$, the conditional probability that the first n events have types $i_1, \ldots, i_n$ given $N(\cdot)$ is $\prod_{k=1}^n p_i(S_k)$.

Example 2.3(b), 2.4(b) (The M/G/∞ Queue). There is a standard scheme for coding the type of queue considered. The last of the three symbols indicates the number of servers (here, ∞); they are always assumed to have i.i.d. service times. The first symbol indicates the type of arrival stream: “M” stands for “memoryless”, which means that the arrivals form a homogeneous Poisson process. The middle symbol indicates the type of service distribution; “M” would be exponential, while “G” is “general”, with c.d.f. equal to G.

Let $N_1(t)$ be the number of customers that have completed service by time t and $N_2(t)$ be the number still in service at time t. What is their joint distribution? Is $N_1(\cdot)$ a Poisson process?

↓

Solution. Now $N_1(t) + N_2(t)$ is the Poisson arrival process. Let its rate be λ. A customer that arrives at time $s \in (0, t]$ has completed service by time t with probability $G(t - s)$. Fix t_0. Then on $(0, t_0]$, we see the arrival Poisson process with an event at time s classified as “done” or “in service” with probability $G(t_0 - s)$ and $\overline{G}(t_0 - s)$. This gives that

$$N_1(t_0) \sim \mathrm{Pois}\Big(\int_0^{t_0} \lambda G(u)\,du\Big),$$

0" . . .

$$N_2(t_0) \sim \mathrm{Pois}\Big(\int_0^{t_0} \lambda \overline{G}(u)\,du\Big),$$

(here, we should assume that G is continuous to apply the theorem) and $N_1(t_0)$, $N_2(t_0)$ are independent (we changed variable to $u := t_0 - s$). (The same holds for $N_i(t)$ for $t \neq t_0$, but those are not meaningful random variables.) (Alternatively, we could have used the fact that unordered arrivals are independent, uniform on $(0, t_0]$ given their number. Then we could have used the earlier corollary, which would avoid the assumption that G is continuous.)

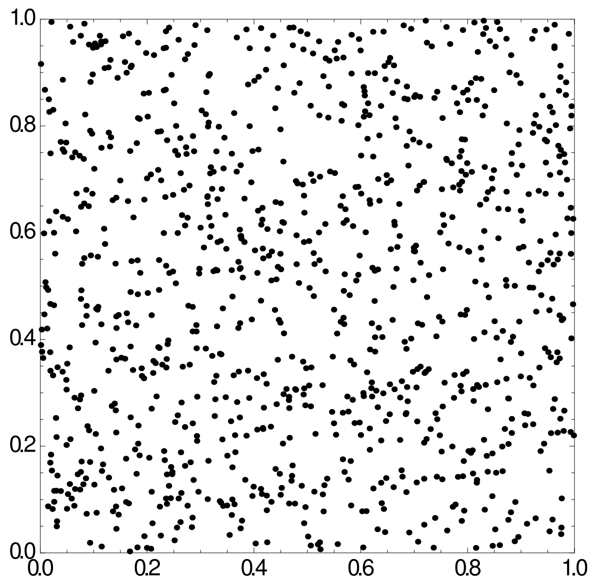
Of course, $N_2(\cdot)$ is not an increasing process, but $N_1(\cdot)$ is. In fact, it is a counting process that jumps by 1 and starts at 0. Also, it has no chance of jumping at any prespecified (deterministic) time, so to prove that $N_1(\cdot)$ is, in fact, a Poisson process, we need to show that it has independent increments. If we consider any finite collection of disjoint intervals and classify arrivals according to during which of these intervals they have completed service, or none (draw two timelines, one for arrivals and the other for completion of service), then the theorem gives what we want. (Here, we can classify all arrivals for all time, but after the last interval, they will all be classified as “none”.) We also have that the intensity function of $N_1(\cdot)$ is $\lambda G(\cdot)$ by what we have already calculated. (Alternatively, we could again have used the fact that unordered arrivals are independent, uniform on $(0, t_0]$ given their number for a t_0 large enough that $[0, t_0]$ includes all the intervals in question. Then we could have used the earlier corollary by the same classification as here, which would again avoid the assumption that G is continuous.)

Note that if $\int_0^\infty \overline{G}(u)\,du = \infty$, i.e., G has infinite expectation, then we may change time in $N_2(\cdot)$ to get a process $N(\cdot)$ such that $\forall t\ N(t) \sim \mathrm{Pois}(t)$, yet $N(\cdot)$ is not even a counting process, much less a Poisson process. From our earlier theorem, it follows that
5" $\uparrow$ $N_2(\cdot)$ does not have independent increments.

$\triangleright$ `Read (if you wish an alternative development) pp. 69--70 and §2.4 (pp. 78--82) in the book.`

We can think of a Poisson process as a random set of points in $[0, \infty)$. This leads us to consider random points in other settings, such as euclidean space. A ***point process*** is a random (finite or infinite) set of points; equivalently, it is a stochastic process N indexed by sets in euclidean space: $N(A)$ is the number of points in $A \subseteq \mathbb{R}^d$. Clearly, if $\langle A_i \rangle$ are disjoint, then $N(\bigcup A_i) = \sum N(A_i)$. We *assume* that if A is bounded, then $N(A) < \infty$ a.s. (i.e., we make this part of our hypotheses without stating this assumption explicitly, or in other words, these are the only point processes we will study).

`Demo`



A sample of a Poisson point process of intensity 1000.

Denote the size (length, area, volume, etc.) of A by $|A|$.

THEOREM. *Let $N(\cdot)$ be a point process such that when $\langle A_i \rangle_{i=1}^r$ are disjoint, $\langle N(A_i)\rangle_{i=1}^r$ are independent and such that $N(A)$ has a distribution depending only on $|A|$. Then $\exists \lambda \in [0,\infty)$ such that $\forall A \;\; N(A) \sim \mathrm{Pois}\big(\lambda|A|\big)$.*

This is called a ***Poisson point process with intensity*** λ.

More generally, we have:

THEOREM. *Let $N(\cdot)$ be a point process such that $\langle A_i \rangle_{i=1}^r$ disjoint $\Rightarrow \langle N(A_i)\rangle_{i=1}^r$ independent and $\forall x \;\; P\big[N(\{x\}) = 0\big] = 1$. Then there exists a function $\mu \ge 0$ on subsets such that*

$$\begin{gathered} \forall A \quad N(A) \sim \mathrm{Pois}\big(\mu(A)\big), \\ A \textit{ bounded } \Rightarrow \mu(A) < \infty, \\ \langle A_i \rangle \textit{ disjoint } \Rightarrow \mu\Big(\bigcup A_i\Big) = \sum \mu(A_i), \\ \textit{and} \quad \forall x \quad \mu(\{x\}) = 0. \end{gathered}$$

Conversely, for all such μ, there is such a point process.

$\downarrow$

Proof. $\Rightarrow$: Similar to before, but subdivide euclidean space by regions.

$\Leftarrow$: Start with a subdivision of euclidean space by regions. In region A, take a $\mathrm{Pois}\big(\mu(A)\big)$ number of dots distributed independently according to $\mu/\mu(A)$, i.e., $P[\text{dot} \in B] = \mu(B)/\mu(A)$ for $B \subseteq A$. This really needs measure theory for full justification. The proof that this works is as before. ∎

4" $\uparrow$

§2.5. Compound Poisson Random Variables and Processes.

If $X_i \sim F$ are i.i.d. and N is a Pois(λ) random variable independent of all X_i, then the sum

$$W := \sum_{i=1}^{N} X_i$$

is called a ***compound Poisson random variable*** with parameters λ and F. Similarly, if $N(\cdot)$ is a Poisson process independent of all X_i (which are still i.i.d.), then

$$W(t) := \sum_{i=1}^{N(t)} X_i$$

is called a ***compound Poisson process***. Thus, each $W(t)$ is a compound Poisson random variable. E.g., $N(\cdot)$ might describe the times of insurance claims and X_i the amounts of the claims. As another example, the special case where $X_i \sim$ Bern(p) gives the thinned Poisson processes considered before. Note that the compound Poisson random variable with parameters λ and Bern(p) is a Pois(λp) random variable.

Example 2.5(a). Suppose that X_s $(s \ge 0)$ are independent random variables but not necessarily identically distributed. Let $\langle S_i \rangle$ be the event times of a Poisson process $N(\cdot)$, independent of all X_s. Interestingly, it turns out that

$$W(t) := \sum_{i=1}^{N(t)} X_{S_i},$$

although not necessarily a compound Poisson process, is, for each t, a compound Poisson random variable!

↓

Solution. For each t, condition on $N(t)$ and use the fact that the event times are independent and uniform on $[0, t]$. If $N(\cdot)$ is Pois(λ) and $X_s \sim F_s$, then $W(t)$ has parameters λt and F, where

$$F(x) := \frac{1}{t} \int_0^t F_s(x)\, ds.$$

Even if we allow X_i in the definition of compound Poisson process to have different
3" ↑ distributions, we may still not get a compound Poisson process.

Chapter 3

Renewal Theory

We now generalize Poisson processes to counting processes where the interarrival times are i.i.d. with an arbitrary nonnegative distribution. Let $\langle X_i \rangle_{i=1}^{\infty}$ be i.i.d. ≥ 0, $P[X = 0] < 1$, $\mu := E[X] \in (0, \infty]$, $S_n := \sum_{i=1}^{n} X_i$. Since $S_n/n \to \mu$ a.s. by the SLLN, we may define (a.s.)

$$N(t) := \max\{n\,;\; S_n \le t\}$$

for $t \ge 0$. This counting process is called a ***renewal process.*** We even allow X_i to take the value $+\infty$ with positive probability, but this will be useful only in the chapter on Markov chains.

The time S_n is called the ***n*th renewal** more often than it is called an arrival. The reason for the name is that if we count time from S_n onwards, then process starts afresh, independent of the past, in the sense that if $f(x_1, x_2, x_3, \ldots)$ is a function, then $f(X_{n+1}, X_{n+2}, X_{n+3}, \ldots)$ has the same distribution regardless of n, and is also independent of $X_1, X_2, \ldots, X_n$.

If X_1 has (or may have) a different distribution than all the X_n for $n \ge 2$, then the process is called a ***delayed renewal process***.

Examples:

• Replace light bulbs when they burn out, assuming that only a single bulb is lit at each instant.

• Cars passing a fixed location in one direction on a two-lane road. (Since some distance between cars is necessary, a Poisson process would not be as accurate.)

• If customer arrival times in a queueing process form a renewal process, then the times of the starts of successive busy periods ("busy" means someone is being served) gen-
2" erate a second (delayed) renewal process. ... In case the arrival times are exponential, then also the times of the starts of successive free periods (no customers) determine a renewal
3" process. But not in general: ... Suppose that the arrivals occur with interarrival times 1 or 5 and that the service time is always 2, with only 1 server. Suppose that the arrival times start as 1, 2, 3, 8, 13. Let $S_1, S_2, S_3, \ldots$ be

the beginnings of the free periods. Then $S_1 = 7$ and $S_2 = 10$. In particular, $P[S_2 - S_1 = 3] > 0$. However, $P[S_3 - S_2 = 3 \mid S_2 - S_1 = 3] = 0$: indeed, $S_2 - S_1 = 3$ implies that the last customer arrival before S_2 was at $S_2 - 2$, whence the next customer arrival will be at $S_2 - 2 + 5 = S_2 + 3$, so that $P[S_3 - S_2 \geq 5 \mid S_2 - S_1 = 3] = 1$. Thus, $S_3 - S_2$ and $S_2 - S_1$ are dependent. (Note that S_1 would be the delay.)

Deciding whether certain times are renewals can be tricky. Although we need to verify only that the times between (proposed) renewals are i.i.d., it is usually easier to verify the stronger property that what happens after every renewal has the same distribution and that it is independent of what happens before that renewal.

We have the following important and intuitive property:

PROPOSITION 3.3.1. *$\lim_{t\to\infty} N(t)/t = 1/\mu$ a.s.*

In words, the rate of renewals is $1/\mu$.

Proof. If $P[X = \infty] > 0$, then $\mu = \infty$ and $\lim_{t\to\infty} N(t) < \infty$ a.s., whence the result is
1.5" clear. Otherwise, we compare to the previous and the next arrival times: ... Important picture here $S_{N(t)} \leq t < S_{N(t)+1}$, so

$$\frac{S_{N(t)}}{N(t)} \leq \frac{t}{N(t)} < \frac{S_{N(t)+1}}{N(t)+1} \cdot \frac{N(t)+1}{N(t)}.$$

The left-hand side tends to μ, as does the first term on the right-hand side, while the last term tends to 1. ∎

Note that if $N(\cdot)$ is a delayed renewal process, then the same result holds (provided
2" $P[X_1 < \infty] = 1$). ...

We start with a straightforward example.

EXAMPLE PM 7.5. A battery has a lifetime that is Unif$[30, 60]$ in hours. If a battery is replaced as soon as it fails, what is the long-run rate at which batteries are replaced?

↓

3" ↑ *Solution.* We have $\mu = 45$ hours, so the rate is one battery every 45 hours.

Next is a somewhat more complicated example.

EXAMPLE ON P. 154, PROBLEM 3.9(A),(B). Customers arrive at a single-server bank at the times of a Poisson process with rate λ. If the server is free, an arriving customer enters the bank; otherwise, the customer goes elsewhere rather than waits (the customer is "lost"). Let the service time be random with c.d.f. G. (This is called an M/G/1/1 queue, where the 4th number indicates the capacity of the system.)

(a) At what rate do customers enter the bank?

(b) What proportion of arrivals actually enter the bank?

↓

Solution. (a) Consider starts of busy periods as renewals. By the memoryless property, the mean time between entering customers is $\mu = \mu_G + 1/\lambda$, whence the rate is $1/\mu = \lambda/(1 + \lambda\mu_G)$.

(b) Let N_A be the arrival process and N_E the entering process. Then $\lim N_E(t)/N_A(t) =$
5" ↑ $\lim(N_E(t)/t)/(N_A(t)/t) = 1/(1 + \lambda\mu_G)$.

In the preceding examples, we used Proposition 3.3.1 in the way one would expect. However, it can also be used to deduce probabilities where there is no renewal process; rather, one introduces a renewal process and uses Proposition 3.3.1 in the reverse direction. This will become clearer in an example. It is a useful method more generally.

EXAMPLE PM 7.8. The following game is played by n players. A spinner has n outcomes. Outcome i has probability p_i, where $\sum_{i=1}^n p_i = 1$. Also given are $k_i \in \mathbb{Z}^+$. The spinner is spun until some i appears k_i times in a row; player i is then declared the winner. Determine each player's chance of winning and the expected number of spins in a game.

Solution. Suppose that this game is played repeatedly. By the SLLN, the probability that i wins is the long-run proportion of games that i wins. Consider wins per game divided by spins per game to get that i wins with probability

$$r_i \Big/ \sum_{j=1}^{n} r_j,$$

2" where $r_j :=$ rate per spin that j wins. ... For each j, wins by player j constitute renewals. Thus, by Proposition 3.3.1, $r_j = 1/$(expected number of spins until j wins), so by Exercise 15 (p. 50, 1.18), $r_j = (1 - p_j)/(p_j^{-k_j} - 1)$. Therefore,

$$P[i \text{ wins a game}] = \frac{(1 - p_i)/(p_i^{-k_i} - 1)}{\sum_{j=1}^n (1 - p_j)/(p_j^{-k_j} - 1)}.$$

Also, the endings of games constitute a renewal process, so by Proposition 3.3.1, the expected number of spins per game is 1/(rate per spin at which games end)

$$= 1\Big/\sum_{j=1}^{n} r_j = 1\Big/\sum_{j=1}^{n}(1-p_j)/(p_j^{-k_j}-1).$$

E.g., if $n = 2$ and $k_2 = 1$, then the game is fair iff $p_1 = 2^{-1/k_1}$.

E.g., if we draw cards with replacement from a standard deck, then the expected number of cards until we draw 4 consecutive cards of the same suit is 85. `This is problem 3.24.`

Renewal processes "begin anew" after each renewal, just as Poisson processes do after each event. (In addition, Poisson processes begin anew at each instant by the memoryless property.) We justified that in one sense earlier. Here is another formal statement and proof for the first renewal; induction gives the same for later renewals.

THEOREM. *If $N(\cdot)$ is a renewal process with first arrival time X_1, then $\langle N(X_1+t)-1\,;\ t \ge 0\rangle$ has the same distribution as $N(\cdot)$, even conditional on X_1.*

Proof. They are both counting processes, so we have to show that they have the same finite-dimensional marginals. Let $0 < t_1 < t_2 < \cdots < t_n < \infty$ and $k_1, k_2, \ldots, k_n$ be nonnegative integers. We have to show that

$$P\big[\forall i\ N(X_1+t_i)-1 = k_i \bigm| X_1\big] = P\big[\forall i\ N(t_i) = k_i\big].$$

3" Write out the left-hand side in terms of S_{k_i+1} and S_{k_i+2}. ... ∎

The function $m(t) := E\big[N(t)\big]$ is called the ***renewal function*** of the process. The previous theorem implies that

$$\begin{aligned} m(t) &= E\big[N(t)\big] = E\big[E[N(t) \mid X_1]\big] \\ &= E\Big[\big(1+m(t-X_1)\big)\mathbf{1}_{[X_1\le t]}\Big] = \int_{[0,t]} \big(1+m(t-x)\big)\,dF_X(x) \\ &= F_X(t) + \int_{[0,t]} m(t-x)\,dF_X(x). \end{aligned}$$

2" ... This is called the ***renewal equation***, but it is usually too hard to solve (for $m(\cdot)$ in terms of $F_X(\cdot)$).

Nevertheless, we can prove that Proposition 3.3.1 holds also in expectation:

Theorem 3.3.4 (The Elementary Renewal Theorem). $\lim_{t\to\infty} m(t)/t = 1/\mu$.

In order to prove this, we will take expectation of

$$S_{N(t)+1} = \sum_{i=1}^{N(t)+1} X_i$$

and get

$$E\big[S_{N(t)+1}\big] = \mu\big(m(t)+1\big). \tag{3.3.3}$$

↓ Note that the analogous equation is not true for $E[S_{N(t)}]$, even if X is ex-
1" ↑ ponential.

However, (3.3.3) doesn't follow from our preceding work (Example 1.5(a)) since $N(t)+1$ is *not* independent of $\langle X_i\rangle$. However, for each n, the event $[N(t)+1=n]$ *is* independent of $\langle X_i\,;\ i \ge n+1\rangle$. Thus, (3.3.3) follows from the following theorem that extends Example 1.5(a). [Note that (3.3.3) is trivial if $\mu = \infty$ since $N(t)+1 \ge 1$.]

We are paying close attention in the statement and proof to hypotheses involving finiteness of expectations because later we will use this to deduce that some expectations are infinite. The notation $\wedge$ is used for minimum; later, we will use $\vee$ for maximum.

Theorem 3.3.2 (Wald's Equation). *Let X_n be random variables all with the same mean $\mu \in (-\infty, \infty]$. Suppose that N is an $\mathbb{N}$-valued random variable such that*

$$\forall n \ge 0\ \ \forall i \ge 1\ \ [N=n] \text{ is independent of } X_{n+i}.$$

If either

(a) *all $X_n \ge 0$ or*

(b) $E[N] < \infty$ *and* $\sup_n E|X_n| < \infty$,

then

$$E\Big[\sum_{n=1}^{N} X_n\Big] = \mu \cdot E[N].$$

Proof. Fix $n \in \mathbb{N}$. Let $I_n := \mathbf{1}_{[N\ge n]} = \mathbf{1} - \sum_{i=0}^{n-1} \mathbf{1}_{[N=i]}$. Suppose temporarily that X_n is bounded. Since X_n and $\mathbf{1}_{[N=i]}$ are independent for $i \le n-1$, we have

$$E[X_n I_n] = E\Big[X_n\big(\mathbf{1} - \sum_{i=0}^{n-1} \mathbf{1}_{[N=i]}\big)\Big] = EX_n - \sum_{i=0}^{n-1} EX_n E\mathbf{1}_{[N=i]} = E[X_n]E[I_n].$$

It follows from this result that even if X_n is not bounded, then, since $\big(|X_n| \wedge M\big) \operatorname{sgn} X_n$ is bounded and independent of $\mathbf{1}_{[N=i]}$ for $i \le n-1$, we have

$$E\big[\big(|X_n| \wedge M\big) \operatorname{sgn} X_n I_n\big] = E\big[\big(|X_n| \wedge M\big) \operatorname{sgn} X_n\big] E[I_n].$$

Taking $M \to \infty$, we get that $E[X_n I_n] = E[X_n] \cdot E[I_n]$: in case (a), we use the MCT, whereas in case (b), we use the LDCT. Likewise, $E\big[|X_n| I_n\big] = E\big[|X_n|\big] E[I_n]$.

Therefore, in case (a), we have by the MCT that

$$\begin{aligned} E\Big[\sum_{n=1}^{N} X_n\Big] &= E\Big[\sum_{n=1}^{\infty} X_n I_n\Big] = \sum_{n=1}^{\infty} E[X_n I_n] = \sum_{n=1}^{\infty} E[X_n] E[I_n] = \mu \sum_{n=1}^{\infty} E[I_n] \\ &= \mu \sum_{n=1}^{\infty} P[N \ge n] = \mu E[N]. \end{aligned}$$

In case (b), calculate first $E\Big[\sum |X_n I_n|\Big] \le E[N] \cdot \sup E|X_n| < \infty$, so we may apply the
1" LDCT to justify the previous calculation. ... ∎

↓ *Proof of Theorem 3.3.4.* We are going to prove this by proving two inequalities,
1" ↑ a lower bound on the liminf and an upper bound on the limsup.

We first show $\liminf m(t)/t \ge 1/\mu$. Since $S_{N(t)+1} > t$, we have $\mu(m(t)+1) =$
.5" $E[S_{N(t)+1}] > t$, whence we get the inequality. ...

For the other direction, that is, $\limsup m(t)/t \le 1/\mu$, the difficulty is that $X_{N(t)+1}$ may be very large (we want to use an upper bound on $S_{N(t)+1}$). Thus, we use the method of truncation: Fix $M \in (0, \infty)$ and define

$$\overline{X}_n := X_n \wedge M.$$

This gives a new renewal process $\overline{N}(t) \ge N(t)$ with mean $\overline{m}(t) \ge m(t)$. Write $\mu_M := E[X \wedge M]$. Note that $\lim_{M \to \infty} \mu_M = \mu$ by the MCT. Now (3.3.3) gives

$$\mu_M(\overline{m}(t)+1) = E\Big[\overline{S}_{\overline{N}(t)+1}\Big] \le t + M,$$

so $\limsup \overline{m}(t)/t \le 1/\mu_M$. Therefore $\limsup m(t)/t \le 1/\mu_M$. Since M is arbitrary, we get $\limsup m(t)/t \le 1/\mu$. ∎

The same holds for delayed renewal processes, provided $X_1 < \infty$ a.s.: Conditional on X_1, the expected number of renewals of the delayed process $N_D(\cdot)$ by time t is 0 if $X_1 > t$ and is $1+m(t-X_1)$ otherwise, where m is the renewal function for the non-delayed renewal process determined by $X_2, X_3, \ldots$. Therefore

$$m_D(t) := E[N_D(t)] = E[(1 + m(t - X_1))\mathbf{1}_{[X_1 \le t]}].$$

Since $m(t)/t \to 1/\mu$, there is some constant c such that for all large t, we have $(1+m(t))/t \le$
2" c. Therefore we can apply the BCT to conclude that $m_D(t)/t \to 1/\mu$. ...

▷ `Read pp. 98--108 in the book.`

According to both Proposition 3.3.1 and the elementary renewal theorem, $N(t)$ is approximately t/μ. In fact, we can say more: it is approximately normally distributed. This is not an instance of the CLT, since $N(t)$ is not a sum. However, $N(\cdot)$ is related to sums, so we will be able to deduce it from the usual CLT.

Theorem 3.3.5. *Let $N(\cdot)$ be a renewal process whose interarrival times have finite mean μ and finite standard deviation σ. Then as $t \to \infty$,*

$$\frac{N(t) - t/\mu}{(\sigma/\mu)\sqrt{t/\mu}} \Rightarrow \mathrm{N}(0, 1).$$

Note that t/μ is not (usually) the mean of $N(t)$, but we saw in the elementary renewal theorem that the mean is asymptotic to t/μ. We can also see intuitively why the standard deviation of $N(t)$ is asymptotic to the denominator above: To first order, $N(t)$ is about t/μ, so

$$N(t) \approx t/\mu \approx S_{N(t)}/\mu \approx S_{\lfloor t/\mu \rfloor}/\mu.$$

Since $S_{\lfloor t/\mu \rfloor}$ has a standard deviation of $\sigma\sqrt{\lfloor t/\mu \rfloor}$, this tells us that the standard deviation of $N(t)$ is roughly $\sigma\sqrt{t/\mu}/\mu$.

Proof. Given any real y, write $r_t := t/\mu + y(\sigma/\mu)\sqrt{t/\mu}$. Also, write $r'_t := \lceil r_t \rceil$. We have

$$\left[\frac{N(t) - t/\mu}{(\sigma/\mu)\sqrt{t/\mu}} < y\right] = [N(t) < r_t] = [N(t) < r'_t] = [S_{r'_t} > t]$$

$$= \left[\frac{S_{r'_t} - r'_t\mu}{\sigma\sqrt{r'_t}} > \frac{t - r'_t\mu}{\sigma\sqrt{r'_t}}\right].$$

Now

$$\frac{t - r_t\mu}{\sigma\sqrt{r_t}} = \frac{-y\sqrt{t/\mu}}{\sqrt{r_t}} = \frac{-y}{\sqrt{(\mu/t)r_t}} = -y\left(1 + \frac{y\sigma}{\sqrt{t\mu}}\right)^{-1/2} \to -y$$

1" as $t \to \infty$. ... Since $|r'_t - r_t| < 1$, the same holds with r'_t in place of r_t: divide both
1" numerator and denominator by $\sqrt{t}$. ... Therefore the CLT tells us that the probability of
2" the event above tends to $1 - \Phi(-y) = \Phi(y)$, where Φ is the c.d.f. of $\mathrm{N}(0, 1)$. ... ∎

EXAMPLE MASS 8.13. Suppose that a part in a machine can be obtained from two different sources, A and B. Each time the part fails, it is replaced by a new one, but the sources are i.i.d., coming from A with probability 0.3 and from B with probability 0.7. (The source is also independent of everything else.) Lifetimes of parts are exponentially distributed; if the source is A, the mean is 8 days, while if B, the mean is only 5 days. However, parts from A take 1 day to install, while those from B take only 1/2 day to install. Installation times are not random. What is the approximate distribution of the number of part failures during a year?

↓

Solution. If X is the interfailure time, then $E[X] = 6.55$ days, $E[X^2] = 82.175$ days^2, whence $\mathrm{Var}(X) = 39.2725$ days^2. This gives an answer of $\mathrm{N}(55.725, 51.01)$. Note that
5" ↑ replacements of parts are not renewals, but failures are.

↓ The bus-waiting paradox: Suppose bus times are deterministic and alternate between 1 minute and 10 minutes. Thus, half of the buses take 10 minutes. But if you go out at a random time, you are more likely to get a bus that takes longer. Most (10/11) of the time it seems that buses take 10 minutes. The paradox lies partly in conflating two different measures of "time": real time or counting buses. When the interarrival times are random,
4" ↑ if we condition on the interarrival times, the same thing holds, of course.

Recall from Exercise 38 that $X_{N(t)+1}$ is stochastically larger than X (i.e., $\overline{F}_{X_{N(t)+1}} \geq \overline{F}_X$). This is intuitive from the viewpoint that longer intervals have a greater chance of capturing a given point, t. As $t \to \infty$, we should expect the length $X_{N(t)+1}$ to converge in
1" law to a size-biased version of X, ... where we say that $\widehat{X}$ is a ***size-biased version*** of X if

$$F_{\widehat{X}}(x) = \frac{1}{E[X]} \int_0^x s \, dF_X(s) : \qquad \text{think } dF_{\widehat{X}}(x) = \frac{1}{E[X]} x \, dF_X(x).$$

1" (Examples: If $X \sim \frac{1}{2}\delta_1 + \frac{1}{2}\delta_{10}$, ... then $\widehat{X} \sim \frac{1}{11}\delta_1 + \frac{10}{11}\delta_{10}$. If $X \sim \frac{1}{3}\delta_1 + \frac{2}{3}\delta_{10}$, then $\widehat{X} \sim \frac{1}{21}\delta_1 + \frac{20}{21}\delta_{10}$.) This is the same as

$$\text{for all bounded } h \qquad E\big[h(\widehat{X})\big] = \frac{1}{E[X]} E\big[X h(X)\big],$$

1" as shown by the change-of-variable formula. ... E.g.,

$$E[\widehat{X}] = E[X^2]/E[X].$$

Furthermore, if we let

$$A(t) := t - S_{N(t)}$$

be the ***age*** of the process at time t, then we expect that the law of $A(t)$ given $X_{N(t)+1} = x$ converges to $\operatorname{Unif}[0, x)$, and thus that

$$\big(A(t), X_{N(t)+1}\big) \Rightarrow \mathcal{L}\big((\operatorname{Unif}[0, \widehat{X}), \widehat{X})\big). \tag{N1}$$

Actually, if X is a ***lattice random variable***, i.e., $\exists d > 0$ such that $P[X \in d\mathbb{Z}] = 1$, then this cannot be true because of periodicity; instead, if d is the largest real number such that $P[X \in d\mathbb{Z}] = 1$, where d is called the ***period*** of X, then

$$\big(A(t), X_{N(t)+1}\big) \Rightarrow \mathcal{L}\big((\operatorname{Unif}_d[0, \widehat{X}), \widehat{X})\big) \tag{N2}$$

as $t \to \infty$ in $d\mathbb{Z}$, where $\operatorname{Unif}_d[0, dn)$ is the uniform distribution on $\{0, d, \ldots, (n-1)d\}$. These two limit results, (N1) for nonlattice random variables and (N2) for lattice random variables, are true when $E[X] < \infty$, but we won't prove them.

Example (Poisson Process). Suppose that $X \sim \operatorname{Exp}(\lambda)$. Then $S_{N(t)+1} - t \sim \operatorname{Exp}(\lambda)$. What is the distribution of $A(t)$?

↓

Solution. By embedding the Poisson process in a point process on all of $\mathbb{R}$, we see that it is exponential truncated at t, i.e., the minimum of an exponential random variable and t. Alternatively, use that $\big[A(t) > s\big]$ is the event that there are no arrivals in $[t - s, t]$. For
3" ↑ one more solution, see Exercise 44. Note what happens as $t \to \infty$.

Assume that X is nonlattice Then as we said above,

$$A(t) \Rightarrow \mathcal{L}\big(\operatorname{Unif}[0, \widehat{X})\big),$$

i.e., if g has the form $g = \mathbf{1}_{(-\infty, x]}$, then

$$\begin{aligned}
\lim_{t\to\infty} E\Big[g\big(A(t)\big)\Big] &= E\Big[g\big(\operatorname{Unif}[0, \widehat{X})\big)\Big] = E\Big[\,E\big[g\big(\operatorname{Unif}[0, \widehat{X})\big) \;\big|\; \widehat{X}\big]\Big] \\
&= E\Big[\frac{1}{\widehat{X}} \int_0^{\widehat{X}} g(s)\,ds\Big] = \frac{1}{\mu} E[X \cdot \frac{1}{X} \int_0^X g(s)\,ds\Big] \\
&= \frac{1}{\mu} E\Big[\int_0^\infty g(s)\mathbf{1}_{[X>s]}\,ds\Big] = \frac{1}{\mu}\int_0^\infty g(s)\overline{F}(s)\,ds.
\end{aligned} \tag{N3}$$

Actually, this holds for other g as well. We also need $\mu = E[X] < \infty$ in order to define $\widehat{X}$. However, this is not needed for the final result, as long as we interpret $1/\mu = 0$ when $\mu = \infty$. The full theorem is as follows:

Theorem 3.4.2 (Probabilistic Form of the Key Renewal Theorem). *If F is not lattice, $\mu = E[X] \le \infty$, and $g\overline{F}$ is directly Riemann integrable on $[0,\infty)$, then*

$$\lim_{t\to\infty} E\Big[g\big(A(t)\big)\Big] = \frac{1}{\mu}\int_0^\infty g(s)\overline{F}(s)\,ds.$$

Likewise, if F is lattice with period d, $\mu = E[X] \le \infty$, and $\sum_{n=0}^\infty g(nd)\overline{F}(nd)$ exists and is finite, then

$$\lim_{n\to\infty} E\Big[g\big(A(nd)\big)\Big] = \frac{d}{\mu}\sum_{n=0}^\infty g(nd)\overline{F}(nd).$$

Here, we say that a function L is ***directly Riemann integrable*** on $[0,\infty)$ if the upper and lower Riemann integrals of L over all of $[0,\infty)$ are equal and finite, when using equally spaced divisions of $[0,\infty)$ for integrating over $[0,\infty)$. It can be shown that
0" besides $L \in C_{\mathrm{c}}\big([0,\infty)\big)$, ... it suffices that L be a decreasing nonnegative function with $\lim_{x\to\infty}\int_0^x L(t)\,dt < \infty$.

The lattice case actually can be written in the very same form as the nonlattice case, as long as we restrict t to $d\mathbb{Z}$.

It is not hard to check that (N1), (N2), and the key renewal theorem hold for delayed renewal processes since they hold for renewal processes (with the usual caveat on X_1 being finite).

We still have to justify heuristically the lattice case. This follows similar reasoning as above (when $\mu < \infty$): Since

$$A(nd) \Rightarrow \mathcal{L}\big(\mathrm{Unif}_d[0,\widehat{X})\big),$$

we have

$$\begin{aligned}
\lim_{n\to\infty} E\Big[g\big(A(nd)\big)\Big] &= E\Big[g\big(\mathrm{Unif}_d[0,\widehat{X})\big)\Big] = E\Big[E\big[g\big(\mathrm{Unif}_d[0,\widehat{X})\big) \mid \widehat{X}\big]\Big] \\
&= E\Big[\frac{1}{\widehat{X}/d}\sum_{n=0}^{\widehat{X}/d-1} g(nd)\Big] = \frac{1}{\mu}E[X\cdot\frac{d}{X}\sum_{n=0}^{X/d-1} g(nd)\Big] \\
&= \frac{d}{\mu}E\Big[\sum_{n=0}^\infty g(nd)\mathbf{1}_{[X>nd]}\Big] = \frac{d}{\mu}\sum_{n=0}^\infty g(nd)\overline{F}(nd).
\end{aligned}$$

Note that the ***residual life*** or ***excess*** at t, defined to be $Y(t) := S_{N(t)+1} - t$, intuitively has the same limit law as $A(t)$ in the nonlattice case, since it is $\mathrm{Unif}(0,\widehat{X}]$, which is the same as $\mathrm{Unif}[0,\widehat{X})$. (In the lattice case, there is a difference since $A(t)$ cannot be equal to $\widehat{X}$, while $Y(t)$ can be.) This makes sense intuitively if we look backwards in time.

The statement you will see of the key renewal theorem in the book and in other books looks quite different. It is phrased purely analytically, with no apparent probabilistic content:

THEOREM 3.4.2 (ANALYTIC FORM OF THE KEY RENEWAL THEOREM). *If F is not lattice, $\mu = E[X] \le \infty$, and h is directly Riemann integrable on $[0, \infty)$, then*

$$\lim_{t\to\infty} \int_{[0,t]} h(t-x)\, dm(x) = \frac{1}{\mu} \int_0^\infty h(s)\, ds.$$

Likewise, if F is lattice with period d, $\mu = E[X] \le \infty$, and $\sum_{n=0}^\infty h(nd)$ exists and is finite, then

$$\lim_{n\to\infty} \sum_{k=0}^{n} h\big((n-k)d\big)\big[m(kd) - m\big((k-1)d\big)\big] = \frac{d}{\mu} \sum_{n=0}^{\infty} h(nd).$$

Here, we are using the notion of Stieltjes integral with respect to $m(\cdot)$, which is defined just as it was with respect to c.d.f.'s.

To discuss this form of the theorem, we will use the fact that $\int h(x)\, dm_1(x) + \int h(x)\, dm_2(x) = \int h(x)\, d(m_1 + m_2)(x)$.

When the theorem is applied in a probabilistic context, it is usually more useful to have it stated probabilistically. But here is the heuristic reason why the theorems are the same. Write

$$F_n := F_{S_n}.$$

We have, for every function g that is 0 on $(-\infty, 0)$,

$$\begin{aligned}
E\Big[g\big(A(t)\big)\Big] &= E\Big[g\big(A(t)\big) \sum_{n\ge 0} \mathbf{1}_{[N(t)=n]}\Big] \\
&= E\Big[\sum_{n\ge 0} g\big(A(t)\big) \mathbf{1}_{[S_n \le t, S_{n+1} > t]}\Big] \\
&= \sum_{n\ge 0} E\Big[g\big(t - S_n\big) \mathbf{1}_{[S_n \le t, S_{n+1} > t]}\Big] \\
&= \sum_{n\ge 0} E\Big[g\big(t - S_n\big) \mathbf{1}_{[S_{n+1} > t]}\Big] \\
&= \sum_{n\ge 0} E\Big[E\Big[g\big(t - S_n\big) \mathbf{1}_{[S_{n+1} > t]} \;\Big|\; S_n\Big]\Big] \\
&= \sum_{n\ge 0} E\Big[g\big(t - S_n\big)\, P\Big[S_{n+1} > t \;\Big|\; S_n\Big]\Big] \\
&= \sum_{n\ge 0} E\Big[g\big(t - S_n\big) \overline{F}(t - S_n)\Big] \\
&= \sum_{n\ge 0} \int_{[0,t]} g\big(t-s\big) \overline{F}(t-s)\, dF_n(s) \\
&= g(t)\overline{F}(t) + \int_{[0,t]} g\big(t-s\big) \overline{F}(t-s)\, dm(s)
\end{aligned}$$

since $F_0(s) = \mathbf{1}_{[0,\infty)}$ and

$$m(s) = E\big[N(s)\big] = \sum_{n\ge 1} P\big[N(s) \ge n\big] = \sum_{n\ge 1} P[S_n \le s] = \sum_{n\ge 1} F_n(s).$$

Now let $t \to \infty$ and use the probabilistic form of the key renewal theorem. To get the theorem
1" in the analytic form above, use $g := h/\overline{F}$. ...

Theorem 3.4.1 (Blackwell's Renewal Theorem). *For a nonlattice renewal process,*

$$E\big[\textit{number of renewals in } (t, t+a]\big] = m(t+a) - m(t) \to a/\mu$$

as $t \to \infty$. If X is lattice with period d, then

$$E[\textit{number of renewals at } nd] \to d/\mu$$

as $n \to \infty$. In the lattice case, if only one renewal can occur at a given time, this is equivalent to

$$P[\textit{renewal at } nd] \to d/\mu.$$

This follows from the analytic form of the key renewal theorem by using $h := \mathbf{1}_{[0,a)}$
2" in the nonlattice case and $h := \mathbf{1}_{\{0\}}$ in the lattice case. ...

Example 3.5(a) (Application of Delayed Renewal Processes to Patterns). Let X_n be i.i.d. and discrete. Given a ***pattern***, i.e., a sequence of possible values of X, say, $\langle x_1, x_2, \ldots, x_k \rangle$, let $N(t)$ be the number of times the pattern occurs by time $\lfloor t \rfloor$. E.g., if the pattern is $\langle 0, 1, 0, 1 \rangle$ and the sequence $\langle X_n \rangle$ is $\langle 1, 0, 1, 0, 1, 0, 1, 1, 1, 0, 1, 0, 1, \ldots \rangle$, then the pattern occurs at times $5, 7, 13, \ldots$ and $N(13) = 3$. Clearly $N(\cdot)$ is a delayed renewal process. What's the expected time μ between patterns?

Solution. By Blackwell's theorem,

$$\frac{1}{\mu} = \lim_{n\to\infty} P[\text{pattern at time } n] = \prod_{i=1}^{k} P[X = x_i].$$

1" ...

Remark. We could also use simply the elementary renewal theorem:

$$\frac{1}{\mu} = \lim_{n\to\infty} \frac{m(n)}{n} = \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} P[\text{pattern at time } i].$$

Then we use the following fact:

$$\text{If } \lim_{k\to\infty} a_k = a \in \mathbb{R}, \text{ then } \lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} a_k = a.$$

The limit of the averages is called the ***Cesàro limit*** of the sequence $\langle a_k \rangle$; it can exist even when the sequence itself does not converge.

To prove the above fact, let $\epsilon > 0$. Choose K so that $|a_k - a| < \epsilon$ for $k > K$. Let $M := \max\{|a_k - a|\,;\ k \le K\}$. Then for $n > KM/\epsilon$,

$$\left|\frac{1}{n}\sum_{k=1}^{n} a_k - a\right| \le \frac{1}{n}\sum_{k=1}^{n} |a_k - a| = \frac{1}{n}\sum_{k=1}^{K} |a_k - a| + \frac{1}{n}\sum_{k=K+1}^{n} |a_k - a| \le \frac{KM}{n} + \frac{(n-K)^+}{n}\epsilon < 2\epsilon.$$

This proves the result. The same result holds if $a = \pm\infty$.

If a coin has probability p of H, what is the expected number of tosses until the pattern HTHT occurs?

By the above, the expected time from HTHT to the next occurrence is $1/p^2q^2$, where $q := 1 - p$. Note that this is less than the expected time to see the first HTHT because it helps to have the last HT from HTHT that can be part of the next HTHT, whereas if we wait from HTHT to the next HTHT that does not overlap with the given HTHT, it is the same as starting from nothing. Now, to get to the first HTHT, one must first see HT. Also, to get from HT to HTHT has the same expected number of tosses as to get from HTHT to the next HTHT, i.e., $1/p^2q^2$, so (the expected time to the first HTHT) = (the expected time to HT) $+\,1/p^2q^2$. Since (the expected time to HT) = (the expected time
3" between HT's) $= 1/pq$, we get $1/pq + 1/p^2q^2$ as our answer. ...

Note how this method also provides another solution to Exercise 15 (p. 50, 1.18 in the
2" book). ...

Similar reasoning works with an underlying i.i.d. process having more than 2 possible outcomes. E.g., if $P(\text{outcome } j) = p_j$, then

$$E[\text{time to } 012301] = E[\text{time to } 01] + \frac{1}{p_0^2 p_1^2 p_2 p_3} = \frac{1}{p_0 p_1} + \frac{1}{p_0^2 p_1^2 p_2 p_3}.$$

Suppose that a system is on for time Z_1, then off for time Y_1, then on for time Z_2, then off for time Y_2, etc. We assume that the intervals of on times and of off times are closed on the left and open on the right. We suppose that (Z_n, Y_n) are i.i.d., but for each n, Z_n and Y_n may be dependent. Still, the partial sums of $\langle Z_n + Y_n\,;\ n \ge 1\rangle$ form a renewal process. The times $Z_1, Z_1 + Y_1, Z_1 + Y_1 + Z_2, Z_1 + Y_1 + Z_2 + Y_2, Z_1 + Y_1 + Z_2 + Y_2 + Z_3, \ldots$ form what is called an ***alternating renewal process***.

THEOREM 3.4.4. *For an alternating renewal process, if $Z+Y$ has finite mean and is nonlattice then*

$$\lim_{t\to\infty} P[\textit{system is on at time } t] = \frac{E[Z]}{E[Z]+E[Y]}.$$

Proof. Let $N(\cdot)$ be the renewal process corresponding to $\langle Z_n + Y_n \rangle$ and $A(\cdot)$ be the associated age process. We have $P[\text{on at } t] = E\Big[P\big[\text{on at } t \bigm| A(t), N(t)\big]\Big]$. Now for $a \ge 0$,

$$\begin{aligned}
P[\text{on at } t \mid A(t)=a, N(t)=n] &= P[\text{on at } t \mid A(t)=a, S_n \le t, S_{n+1} > t] \\
&= P[Z_{n+1} > a \mid S_n = t-a, S_n \le t, S_{n+1} > t] \\
&= P[Z_{n+1} > a \mid S_n = t-a, Z_{n+1}+Y_{n+1} > a] \\
&= P[Z_{n+1} > a \mid Z_{n+1}+Y_{n+1} > a] \\
&= \overline{F}_Z(a)/\overline{F}_{Z+Y}(a).
\end{aligned}$$

Therefore, $P[\text{on at } t] = E\Big[\overline{F}_Z\big(A(t)\big)/\overline{F}_{Z+Y}\big(A(t)\big)\Big]$. We can apply the key renewal theorem with $g(s) := \overline{F}_Z(s)/\overline{F}_{Z+Y}(s)$ (and $F := F_{Z+Y}$) since $\overline{F}_Z$ is a nonnegative decreasing function with finite integral $E[Z]$. This tells us that $P[\text{on at } t]$ converges to

$$\frac{1}{E[Z+Y]} \int_0^\infty \frac{\overline{F}_Z(s)}{\overline{F}_{Z+Y}(s)} \overline{F}_{Z+Y}(s)\, ds. \qquad \blacksquare$$

2" . . .

Of course,

$$\lim_{t\to\infty} P[\text{off at time } t] = 1 - \frac{E[Z]}{E[Z]+E[Y]} = \frac{E[Y]}{E[Z]+E[Y]}.$$

Note that $\lim_{t\to\infty} P[\text{on at } t]$ is equal to the long-run expected proportion of time that the system is on, since if $I(t)$ is the indicator that the system is on at time t, then this long-run expected proportion is

$$\lim_{t\to\infty} E\left[\frac{1}{t}\int_0^t I(s)\, ds\right] = \lim_{t\to\infty} \frac{1}{t}\int_0^t E\big[I(s)\big]\, ds = \lim_{t\to\infty} \frac{1}{t}\int_0^t P[\text{on at } s]\, ds.$$

As for sums, we have Cesàro limits for integrals with a very similar proof:

$$\text{If } \lim_{s\to\infty} f(s) = a, \text{ then } \lim_{t\to\infty} \frac{1}{t}\int_{s=0}^t f(s)\, ds = a.$$

In Exercise 56, you will be asked to show that also in the lattice case, the long-run proportion of time that the system is on equals $E[Z]/\big(E[Z]+E[Y]\big)$.

As usual, the same results hold for delayed alternating renewal processes provided that the delay is finite a.s.

Example MASS 8.29. Let an M/G/1/1 queue have arrival rate λ. If $Q(t)$ denotes the number of customers in the system at time t (which is either 0 or 1), find $\lim_{t\to\infty} P\big[Q(t) = 1\big]$.

↓

Solution. If we interpret $Q(t) = 1$ as on time and $Q(t) = 0$ as off time, then we see a (delayed) alternating renewal process. (It could be done the other way, too. Recall the example of a delayed renewal process at the beginning of the chapter.) The mean on time is μ_G and the mean off time is $1/\lambda$ (by the memoryless property), so the answer is
2" ↑ $\mu_G/(\mu_G + 1/\lambda) = \lambda\mu_G/(1 + \lambda\mu_G)$.

Example MASS 8.30. Let a G/M/1/1 queue have service rate μ. If $Q(t)$ denotes the number of customers in the system at time t, find $\lim_{t\to\infty} P\big[Q(t) = 1\big]$.

↓

Solution. If we interpret $Q(t) = 1$ as on time and $Q(t) = 0$ as off time, then we see a (delayed) alternating renewal process. (This will not work the other way, however.) The mean on time (length of busy period) is $1/\mu$, but the mean off time is more difficult to calculate. We will calculate the mean cycle time, which is the denominator anyway. Condition that a certain busy period has length s. The arrivals starting at the beginning of this service period form a renewal process with interarrival distribution $G(\cdot)$. The total cycle time is then the time that this renewal process first exceeds s. By (3.3.3), this has expectation $\mu_G(m_G(s) + 1)$. Therefore, the unconditioned expected cycle time is $\int_0^\infty \mu_G(m_G(s) + 1)\mu e^{-\mu s}\,ds$, which gives the answer $1/\big(\mu^2\mu_G \int_0^\infty (m_G(s) + 1)e^{-\mu s}\,ds\big)$.

4" ↑ We did not actually need the service distribution to be memoryless.

Note that these answers agree for an M/M/1/1 queue.

Example 3.4(a). A store stocks a certain commodity. It tries to keep the amount on hand in the interval $[s, S]$. It does this by ordering $S - x$ whenever the amount on hand dips to $x < s$, but not ordering otherwise. Thus, the store restocks to level S after such an order, which is assumed to be executed and received instantaneously. Customers arrive at the times of a renewal process with nonlattice interarrival distribution F. Each customer independently buys an amount with distribution G. (If a customer wants to buy more than is on hand, then the customer buys only what is on hand.) What is the limiting distribution of the inventory level (the amount on hand), which is considered always to be in $[s, S]$?

↓

Solution. Let $V(t)$ denote the inventory at time t. We will calculate $\lim_{t\to\infty} P\big[V(t) \ge x\big]$. Fix $x \in [s, S]$. Say that the system is on when $V(t) \ge x$ and off otherwise. Then we see a nonlattice alternating renewal process, with the beginning of a cycle at the time of each restocking order. In order to apply Theorem 3.4.4, let D_k be i.i.d. with the purchase distribution G. For every y, define

$$L_y := \min\Big\{n\,;\ \sum_{k=1}^{n} D_k > S - y\Big\}.$$

Then in each cycle, the number of customers until the inventory falls below x has distribution equal to that of L_x, while the number of customers in the total cycle has distribution equal to that of L_s. Also, let $X_i \sim F$ be i.i.d. independent of D_k. Then in each cycle, the time until the inventory falls below y has distribution equal to that of $\sum_{i=1}^{L_y} X_i$. Thus,

$$\lim_{t\to\infty} P\big[V(t) \ge x\big] = \frac{E\big[\sum_{i=1}^{L_x} X_i\big]}{E\big[\sum_{i=1}^{L_s} X_i\big]} = \frac{E[L_x]\mu_F}{E[L_s]\mu_F} = \frac{E[L_x]}{E[L_s]}.$$

Now if $N_G(\cdot)$ is the renewal process defined by $\langle D_k \rangle$, then $L_y = N_G(S-y) + 1$, whence $E[L_y] = m_G(S-y) + 1$ in the notation of the renewal function for $N_G(\cdot)$. In conclusion,

$$\lim_{t\to\infty} P\big[V(t) \ge x\big] = \frac{m_G(S-x)+1}{m_G(S-s)+1}.$$

(If $x = S$, this is still correct: the numerator is 1, and indeed $V(t) = S$ generally holds for intervals of t. If we wanted $x > S$, this is not correct, but then $L_x = 0$, so $E[L_x] = 0$. If $x = s$, this is correct because $V(t) \ge s$ always. If $x < s$, then $L_x = L_s$, so the formula again has to be modified.) Of course, this isn't so explicit, but in any particular case, one can calculate numerically $m_G(\cdot)$ by iterating the renewal equation,

$$m_G(t) = G(t) + \int_{[0,t]} m_G(t-x)\, dG(x).$$

`7" ↑`

▷ `Read pp. 109--119 in the book.`

Now we go the other way: we use Theorem 3.4.4 in order to calculate the means.

Example 3.5(b). Suppose that a machine has n components, each of which functions during the on times of an independent alternating renewal process. More precisely, component i functions for an $\text{Exp}(\lambda_i)$ time, then is down for an $\text{Exp}(\mu_i)$ time. (Note: I switched the notation from the book, which gave the means, not the parameters.) The machine as a whole functions as long as at least one component functions. What is the mean time between breakdowns? What is the mean length of a functioning period?

↓

Solution. Note that the breakdowns of the machine constitute a delayed renewal process. When the periods of *functioning* are considered as the *off* periods, we see a delayed alternating renewal process. This relies on the memoryless property, because the breakdowns do not coincide with a change of state of all the components at once. (We could also use the ends of the non-functioning periods as renewals, but this is harder to see. To see it, realize that we can choose which component becomes functional at those times always with the same distribution, but that what happens until the end of such a functioning period is complicated. In any case, the functioning periods then become the on periods.)

Component i has a limiting down probability $(1/\mu_i)/(1/\lambda_i + 1/\mu_i) = \lambda_i/(\lambda_i + \mu_i)$. Since the components operate independently, the limiting down probability of the machine is the product $\prod_{i=1}^n \lambda_i/(\lambda_i + \mu_i)$. By Theorem 3.4.4, this equals the mean down time divided by the mean cycle time. Now the mean cycle time is what we want to know. By the memoryless property, the down periods have distribution $\text{Exp}\big(\sum_{i=1}^n \mu_i\big)$, whence the mean down length is $1/\sum_i \mu_i$. Therefore, the mean cycle time is $\big(\sum_j \mu_j \prod_{i=1}^n \lambda_i/(\lambda_i + \mu_i)\big)^{-1}$. The mean length of an up period is the difference between the mean cycle time and the
4" ↑ mean down period.

Suppose that at the nth renewal (i.e., nth event of a renewal process), we receive a reward R_n. We allow R_n to depend on X_n, but assume that (X_n, R_n) are i.i.d. The total reward earned by time t is

$$R(t) := \sum_{n=1}^{N(t)} R_n.$$

The stochastic process $R(\cdot)$ is called a ***renewal-reward process***. For example, a renewal process is a renewal-reward process; and a compound Poisson process is a renewal-reward process.

THEOREM 3.6.1. *If $E\big[|R|\big] < \infty$ and $E[X] < \infty$, then as $t \to \infty$, we have*

$$\frac{R(t)}{t} \to \frac{E[R]}{E[X]} \quad a.s.$$

and

$$\frac{E\big[R(t)\big]}{t} \to \frac{E[R]}{E[X]}.$$

Proof. The first statement is rather easy to prove, but the 2nd will require some work. For the first, write the quotient as

$$\frac{R(t)}{t} = \frac{R(t)}{N(t)}\frac{N(t)}{t} \to E[R] \cdot \frac{1}{E[X]} \quad \text{a.s.}$$

by the SLLN, using the fact that $N(t) \to \infty$ as $t \to \infty$ and Proposition 3.3.1.

For the second statement, we use truncation, as in the proof of Theorem 3.3.4, the elementary renewal theorem. In order to do that, we need to decompose R_n as $R_n = R_n^+ - R_n^-$, where $R_n^\pm \ge 0$. Defining $R^\pm(t) := \sum_{n=1}^{N(t)} R_n^\pm$, we have $R(t) = R^+(t) - R^-(t)$,
2" so we see that it suffices to prove the result when $R_n \ge 0$. . . .

Assume now that $R_n \ge 0$. We have that $[N(t)+1 = n]$ is independent of $\langle R_i\,;\ i > n\rangle$.
1" . . . Therefore Wald's equation gives us

$$E\left[\sum_{n=1}^{N(t)+1} R_n\right] = (m(t)+1)E[R].$$

Since $m(t)/t \to 1/E[X]$, the result desired follows if $E[R_{N(t)+1}]/t \to 0$ as $t \to \infty$. Rather than show this directly, we take an easier approach. Namely, this certainly holds if R_n is bounded. Therefore, given $M < \infty$, we have

$$\liminf_{t\to\infty} E[R(t)]/t \ge \lim_{t\to\infty} E\left[\frac{1}{t}\sum_{n=1}^{N(t)}(R_n \wedge M)\right] = E[R \wedge M]/E[X].$$

Taking the limit as $M \to \infty$ and using the MCT, we obtain $\liminf_{t\to\infty} E[R(t)]/t \ge$
1" $E[R]/E[X]$. . . . On the other hand, we have

$$\limsup_{t\to\infty} E[R(t)]/t \le \lim_{t\to\infty} E\left[\frac{1}{t}\sum_{n=1}^{N(t)+1} R_n\right] = \lim_{t\to\infty}\frac{m(t)+1}{t}E[R] = E[R]/E[X].$$

1" Putting these inequalities together gives us the desired limit. . . . ∎

The reward need not be given exactly at the renewal times. It could accumulate during the renewal cycle. For example, as long as we define $R(t)$ to lie between $\sum_{n=1}^{N(t)} R_n$ and $\sum_{n=1}^{N(t)+1} R_n$, then the theorem applies to $R(t)$. This is because, as shown during the
1" proof, $\sum_{n=1}^{N(t)} R_n/t$ and $\sum_{n=1}^{N(t)+1} R_n/t$ have the same limit. ...

Example 3.6(a). Consider an alternating renewal process. Suppose that a reward accumulates at a unit rate during the on periods, but not during the off periods. Then Theorem 3.6.1 tells us that a.s., the long-run proportion of time that the system is on is
1" equal to the same limit as in Theorem 3.4.4. ... So this gives us another way to interpret all our results about alternating renewal processes. It also tells us that the on period during a cycle need not be an interval at the beginning of the cycle. Finally, it says that even in the lattice case, the long-run proportion that the system is on is equal to the same limit as in Theorem 3.4.4.

Example 3.6(c). An amusement park ride only starts when there are N passengers waiting. Passengers arrive at the times of a renewal process with mean interarrival time μ. The management must endure the grumbling of waiting passengers, which it quantifies as a cost of c dollars per unit time per waiting passenger. Also, it costs K dollars each time the ride is started. What is the average cost per unit time of this operation, and what N minimizes it?

↓

Solution. The mean cycle time is $N\mu$ and the mean cost during a cycle is, if S_n are the arrival times, $E[\sum_{n=1}^{N}(S_N - S_n)c] + K = \sum_{n=1}^{N}(N-n)\mu c + K = c\mu N(N-1)/2 + K$. Dividing, we get $c(N-1)/2 + K/(N\mu)$. This is minimized when N is one of the integers
4" ↑ nearest to $\sqrt{2K/(c\mu)}$.

Example PM 7.11. Let X_n be the lifetimes of items assumed i.i.d. with c.d.f. F. It may be that failure of an item is costly, and so replacement is done when an item has reached age T if it has not yet failed. Suppose that each replacement costs c_r and each failure costs an additional c_f. Show that the long-run cost per unit time is (a.s. and in mean)

$$\frac{c_r + c_f F(T)}{\int_0^T \overline{F}(x)\,dx}.$$

↓

Solution. Each replacement constitutes a renewal. The expected cycle length is thus

$$E[X \wedge T] = \int_0^\infty P[X \wedge T > x]\,dx = \int_0^T \overline{F}(x)\,dx,$$

while the expected cost during a cycle is $c_r + c_f F(T)$. Now apply Theorem 3.6.1.

For example, if $X \sim \mathrm{Unif}[0,a]$, then the long-run cost rate is $(2ac_r + 2c_f T)/(2aT - T^2)$ for $0 \le T \le a$, and this is minimized at

$$\frac{T}{a} = -\frac{c_r}{c_f} + \sqrt{\left(\frac{c_r}{c_f} + 1\right)^2 - 1}.$$

4" ↑

▷ Read pp. 125--126, 128--129, 133--137 in the book.

One way that delayed renewal processes arise is by beginning our observation of a renewal process at time t, rather than at time 0. Then the first renewal time is $Y(t)$ later, where $Y(t)$ is the residual life at time t. Now if t happens to be large, $\mu < \infty$, and the interarrival time $X \sim F$ is nonlattice then we know that $Y(t)$ is approximately a uniform pick on $(0, \widehat{X}]$. Use $g := \mathbf{1}_{[0,x]}$ in the key renewal theorem to get

$$\lim_{t\to\infty} P\big(A(t) \le x\big) = \lim_{t\to\infty} E\big[g\big(A(t)\big)\big] = \frac{1}{\mu}\int_0^\infty g(s)\overline{F}(s)\,ds = \frac{1}{\mu}\int_0^x \overline{F}(s)\,ds$$

if X is nonlattice Define

$$F_e := \text{ the c.d.f. of } \mathrm{Unif}[0, \widehat{X}).$$

If we write $U \sim \mathrm{Unif}[0, \widehat{X})$, then

$$F_e(x) = P[U \le x] = E[(x \wedge \widehat{X})/\widehat{X}] = E[x \wedge X]/\mu = \frac{1}{\mu}\int_0^x \overline{F}(s)\,ds$$

by Exercise 13 in the homework. Thus, we have proved that $A(t) \Rightarrow F_e$ from the key renewal theorem. To see more formally that $Y(t) \Rightarrow F_e$ as we argued intuitively before, use Theorem 3.4.4:* Fix x. Let the system be "on" at time t when $Y(t) > x$ and "off" when $Y(t) \le x$. The beginnings of the on periods are the ends of the off periods; we use this convention even if an on period has length 0 (i.e., is empty). In this way, the renewals in the alternating renewal process are the same as the renewals in the original renewal process we started with. Thus, $\lim_{t\to\infty} P\big[Y(t) \le x\big] = E[x \wedge X]/\mu = F_e(x)$.

Let's consider a delayed renewal process with initial time $\mathrm{Unif}(0, \widehat{X}]$; the other interarrival times have the same distribution as X. This is called the ***equilibrium renewal process*** associated to the original process because of the following result:

* This also follows from (N1), which one can prove similarly to Exercise 47 by considering the c.d.f. of $\alpha A(t) + \beta X_{N(t)+1}$ for real α, β and using the Cramér-Wold device.

THEOREM 3.5.2. *Let $N_e(\cdot)$ be a nonlattice equilibrium renewal process. Then $N_e(\cdot)$ has stationary increments and $\forall t \ge 0$ $m_e(t) = t/\mu$ and $Y_e(t) \sim F_e$.*

2" *Proof.* Since $Y(t) \Rightarrow F_e$ as $t \to \infty$, we have that for $s \ge 0$, ...

$$N(t+s) - N(t) \Rightarrow N_e(s) \quad \text{as } t \to \infty.$$

In fact, this is true jointly in all $s \ge 0$; for example, if $s_1, s_2 \ge 0$, then

$$\big(N(t+s_1) - N(t), N(t+s_1+s_2) - N(t)\big) \Rightarrow \big(N_e(s_1), N_e(s_1+s_2)\big) \quad \text{as } t \to \infty.$$

Thus, given $s_1, s_2 \ge 0$, we have

$$\begin{aligned}
\mathcal{L}\Big(N_e(s_1+s_2) - N_e(s_1)\Big) &= \lim_{t\to\infty} \mathcal{L}\Big(N(t+s_1+s_2) - N(t+s_1)\Big) \\
&= \lim_{t\to\infty} \mathcal{L}\Big(N(t+s_2) - N(t)\Big) \\
&= \mathcal{L}\big(N_e(s_2)\big),
\end{aligned}$$

i.e., $N_e(\cdot)$ has stationary increments.

1" In particular, $m_e(kt) = km_e(t)$ for all positive integers k and every fixed t. ... Therefore, by the elementary renewal theorem applied to the equilibrium renewal process, which is a delayed renewal process, we have $m_e(t)/t = m_e(kt)/(kt) \to 1/\mu$ as $k \to \infty$.

Finally, because the increments of $N_e(\cdot)$ are stationary, $\mathcal{L}\big(Y_e(s)\big) = \mathcal{L}\big(Y_e(0)\big) = F_e$.

1" ... ∎

1" `Is a Poisson process an equilibrium renewal process? ...`

▷ `Read pp. 125--126, 128--129, 131--132 in the book.`

Chapter 4

Markov Chains

We now go beyond having real-valued random variables. In this chapter, we consider stochastic processes indexed by $\mathbb{N}$ (or $\mathbb{Z}^+$) and which can take values in a finite or countable set called the ***state space***. For simplicity, the states will often be labelled $0, 1, 2, \ldots$, but there may be no numerical significance to the labels.

What replaces independence of increments is the ***Markovian property*** that the future and the past are independent given the present: given $n, r, i_0, i_1, \ldots, i_{n+r}$ with $P[X_n = i_n] > 0$, consider the past event $A := [X_0 = i_0, X_1 = i_1, \ldots, X_{n-1} = i_{n-1}]$ and the future event $B := [X_{n+1} = i_{n+1}, X_{n+2} = i_{n+2}, \ldots, X_{n+r} = i_{n+r}]$. Then

$$P(A, B \mid X_n = i_n) = P(A \mid X_n = i_n)P(B \mid X_n = i_n). \tag{N1}$$

These events A and B are very basic events. If we sum over all possibilities, then we see that the same equation holds for all pairs of events A and B where A depends only on X_j
.5" for $j < n$ and B depends only on X_k for $k > n$. ...

The Markovian property is equivalent to the following property: $\forall n\ \ \forall i_0, \ldots, i_{n+1}$ with $P[X_0 = i_0, \ldots, X_n = i_n] > 0$,

$$P[X_{n+1} = i_{n+1} \mid X_0 = i_0, \ldots, X_n = i_n] = P[X_{n+1} = i_{n+1} \mid X_n = i_n]. \tag{N2}$$

↓ To see this, suppose first that the Markov property (N1) holds; let A be as in the definition and let $B := [X_{n+1} = i_{n+1}]$. Then the left-hand side of (N2) is equal to

$$P(B \mid A,\ X_n = i_n) = \frac{P(A, B \mid X_n = i_n)}{P(A \mid X_n = i_n)} = P(B \mid X_n = i_n)$$

by (N1), which shows (N2). Conversely, suppose that (N2) holds. The calculation we just did shows that (N1) holds for A as in the definition and for B of the special form $B := [X_{n+1} = i_{n+1}]$. Summing over possibilities shows that it also holds for every A that depends only on times before the present, n. We can also reformulate (N1) as

$$P(B \mid A,\ X_n = i_n) = P(B \mid X_n = i_n). \tag{$*$}$$

Now suppose that B is general, as in the definition. Let A depend only on the past, X_j for $j < n$. Then we have

$$
\begin{aligned}
P(B \mid A,\ X_n = i_n) &= \prod_{j=1}^{r} P(X_{n+j} = i_{n+j} \mid A,\ X_n = i_n, X_{n+1} = i_{n+1}, \ldots, X_{n+j-1} = i_{n+j-1}) \\
&= \prod_{j=1}^{r} P(X_{n+j} = i_{n+j} \mid X_{n+j-1} = i_{n+j-1})
\end{aligned}
$$

5" ↑ by (∗). Since this does not depend on A, it follows that (N1) holds.

The right-hand side of (N2) is known as a ***transition probability***. The analogue of stationary increments is that this doesn't depend on n, only on i_n and i_{n+1}; that is,

$$P[X_{n+1} = j \mid X_n = i] =: p_{ij}$$

does not depend on n. In this case, the process is called a ***(homogeneous) Markov chain***. From the transition probabilities and the initial distribution $p_i := P[X_0 = i]$, we can calculate all probabilities:

$$P[X_0 = i_0, X_1 = i_1, \ldots, X_n = i_n] = p_{i_0} p_{i_0 i_1} p_{i_1 i_2} \ldots p_{i_{n-1} i_n}.$$

3" ...

It's really better to say that a Markov chain is a collection of probability measures P_i, representing the chain when it starts in state i, with the property that

$$
\begin{aligned}
&P_{i_0}[X_1 = i_1, \ldots, X_{n-1} = i_{n-1}, X_n = i_n, X_{n+1} = i_{n+1}, X_{n+2} = i_{n+2}, \ldots, X_{n+r} = i_{n+r}] \\
&\quad = P_{i_0}[X_1 = i_1, \ldots, X_{n-1} = i_{n-1}, X_n = i_n] P_{i_n}[X_1 = i_{n+1}, \ldots, X_r = i_{n+r}]
\end{aligned}
$$

for all n, r, and $i_0, \ldots, i_{n+r}$. This avoids problems of p_i possibly being 0 for some (even most) i. Again, by summing over all possibilities, we see that for all pairs of events A and B where A depends only on X_m for $m < n$ and B depends only on X_k for $k > n$, we have

$$P_i[A,\ X_n = j,\ B] = P_i[A,\ X_n = j] P_j(\tau_n B).$$

Here, we define τ_n of an event that depends only on X_k for $k > n$ by subtracting n from all the indices of the random variables; i.e.,

$$\tau_n[X_{n+1} = i_{n+1}, X_{n+2} = i_{n+2}, \ldots, X_{n+r} = i_{n+r}] := [X_1 = i_{n+1}, X_2 = i_{n+2}, \ldots, X_r = i_{n+r}]$$

and similarly for unions of such events.

Example (I.I.D. Trials). If X_n are i.i.d., then $\langle X_n \rangle$ is a Markov chain.

Example 4.1(c) (Sums of I.I.D. $\mathbb{Z}$-valued Random Variables). Here, the state
2" space is $\mathbb{Z}$. ...

Example 4.1(a) (The M/G/1 Queue). Let $X_n :=$ the number of customers in the system when the nth customer leaves the system, and $X_0 := 0$. The memoryless property
2" of the arrival stream shows that this is a Markov chain. ... Now

$$X_{n+1} = \begin{cases} X_n - 1 + Y_{n+1} & \text{if } X_n \geq 1, \\ Y_{n+1} & \text{if } X_n = 0, \end{cases}$$

where $Y_{n+1} :=$ the number of arrivals during the period of service of the $(n+1)$st customer. (Note that those customers who arrive during a free period are not counted by any of the Y_n. Only customers who have to wait in queue are counted by some Y_n.) Thus, Y_n are i.i.d. and for $j \in \mathbb{N}$, with λ denoting the rate of the arrivals, we have

$$\begin{aligned} P[Y = j] &= E\Big[P[Y = j \mid \underbrace{\text{service time}}_{Z \sim G}]\Big] \\ &= E\Big[e^{-\lambda Z}(\lambda Z)^j / j!\Big] = \int_0^\infty e^{-\lambda x} \frac{(\lambda x)^j}{j!} \, dG(x). \end{aligned}$$

Thus,

$$\begin{aligned} p_i &= \begin{cases} 1 & \text{if } i = 0, \\ 0 & \text{otherwise,} \end{cases} \\ p_{0j} &= \int_0^\infty e^{-\lambda x} \frac{(\lambda x)^j}{j!} \, dG(x) \quad (j \geq 0), \\ p_{i,i-1+j} &= p_{0j} \qquad (i \geq 1,\ j \geq 0), \\ p_{i,j} &= 0 \qquad (i \geq 2,\ j \leq i-2). \end{aligned}$$

▷ `Read pp. 163--165 in the book.`

Note: In Example 4.1(d), Ross says that the case of summands $X_i = \pm 1$ is "simple random walk". Usually, this is called "nearest-neighbor random walk" and the term "simple" is reserved for the case when $P[X_i = 1] = \frac{1}{2}$. We will *not* use Ross's terminology.

Example 4.4(a) (The Gambler's Ruin Problem). A gambler needs \$$N$ but has only \$$i$ $(1 \leq i \leq N - 1)$. He plays games that give him chance p of winning \$1 and $q := 1 - p$ of losing \$1 each time. When his fortune is either \$$N$ or 0, he stops. What is his chance of success?

↓

Solution. The gambler's fortune is a Markov chain on $\{0, 1, \ldots, N\}$. Let α_i be the probability of success. Then $\alpha_0 = 0$, $\alpha_N = 1$, and

$$\text{for } 1 \le i \le N-1 \qquad \alpha_i = p\alpha_{i+1} + q\alpha_{i-1},$$

which gives

$$\alpha_{i+1} - \alpha_i = \frac{q}{p}(\alpha_i - \alpha_{i-1}).$$

Therefore

$$\alpha_{i+1} - \alpha_i = \left(\frac{q}{p}\right)^i (\alpha_1 - \alpha_0) = \left(\frac{q}{p}\right)^i \alpha_1.$$

To determine α_1, add these up:

$$1 = \alpha_N = \sum_{i=0}^{N-1} (\alpha_{i+1} - \alpha_i) = \sum_{i=0}^{N-1} \left(\frac{q}{p}\right)^i \alpha_1,$$

so

$$\alpha_1 = 1 \Big/ \sum_{i=0}^{N-1} \left(\frac{q}{p}\right)^i.$$

By adding only the equations for $\alpha_1 - \alpha_0, \alpha_2 - \alpha_1, \ldots, \alpha_i - \alpha_{i-1}$, we get

$$\alpha_i = \sum_{j=0}^{i-1} \left(\frac{q}{p}\right)^j \Big/ \sum_{j=0}^{N-1} \left(\frac{q}{p}\right)^j.$$

6" ↑ For example, when $p = \frac{1}{2}$, we have $\alpha_i = i/N$.

For an application to statistics, see Exercise 4.30, p. 224 (with answer in the back).

§4.2. Chapman–Kolmogorov Equations and Classification of States.

Let $p_{ij}^{(n)}$ be the n-step transition probabilities, i.e.,

$$p_{ij}^{(n)} := P_i[X_n = j].$$

1" Note that ...

$$p_{ij}^{(n+m)} = \sum_k P_i[X_{n+m} = j, X_n = k] = \sum_k P_i[X_n = k] P_k[X_m = j] = \sum_k p_{ik}^{(n)} p_{kj}^{(m)}.$$

Notice that this is matrix multiplication: If $P^{(n)} := (p_{ij}^{(n)})$, then the above equation is $P^{(n+m)} = P^{(n)} P^{(m)}$, whence $P^{(n)} = P^n$, where $P := (p_{ij})$.

We say that state j is ***accessible*** from state i if $\exists n \ge 0\ p_{ij}^{(n)} > 0$. If i and j are accessible from each other, we say they ***communicate*** and write $i \leftrightarrow j$. It is not hard
3" to check that $\leftrightarrow$ is an equivalence relation. ... If there is only one equivalence class, the Markov chain is called ***irreducible***. The ***period*** of state i is the g.c.d. of $\{n \ge 0\,;\ p_{ii}^{(n)} > 0\}$, written $d(i)$. If $d(i) = 1$, then state i is called ***aperiodic***.

PROPOSITION 4.2.2. *If $i \leftrightarrow j$, then $d(i) = d(j)$.*

Proof. It suffices to show that $d(j) \mid d(i)$. [The symbol $k \mid n$ here stands for "divides" and means that $n/k \in \mathbb{Z}$.] Let $p_{ii}^{(s)} > 0$ and choose m, n such that $p_{ij}^{(m)} > 0$ and $p_{ji}^{(n)} > 0$. First, we have $p_{jj}^{(n+m)} \ge p_{ji}^{(n)} p_{ij}^{(m)} > 0$, so $d(j) \mid (n+m)$. Second, we have

$$p_{jj}^{(n+s+m)} \ge p_{ji}^{(n)} p_{ii}^{(s)} p_{ij}^{(m)} > 0,$$

so $d(j) \mid (n+s+m)$. Therefore, $d(j) \mid s$, so $d(j) \mid d(i)$. ∎

Let $f_{ij}^{(n)}$ be the probability that the first* transition into j is at time n when the chain starts in state i: $f_{ij}^{(0)} := 0$ and for $n \ge 1$,

$$f_{ij}^{(n)} := P_i\Big[X_n = j \text{ and } \forall k \in [1, n-1]\ \ X_k \ne j\Big].$$

Then $f_{ij} := \sum_{n=1}^{\infty} f_{ij}^{(n)}$ is the probability of *ever* making a transition into state j when the chain starts in i. We call j ***recurrent*** if $f_{jj} = 1$ and ***transient*** otherwise. The function

$$G(i,j) := \sum_{n=0}^{\infty} p_{ij}^{(n)},$$

1" the expected number of visits to j for the chain started at i, ... is called the ***Green function*** of the Markov chain.

---

* f stands for "first"

Proposition 4.2.3. *State j is transient iff $G(j,j) < \infty$. If state j is transient, then a.s. the number of visits to j starting from j is finite, while if state j is recurrent, then a.s. the number of visits to j starting from j is infinite.*

Proof. Each visit to j is followed (at some time) by another visit to j with probability
2" f_{jj}. Hence the number of visits is geometric with mean $(1-f_{jj})^{-1}$. … But we already know that the mean number of visits to j when the chain starts in j is $G(j,j)$. Therefore $(1-f_{jj})^{-1} = G(j,j)$, so that $f_{jj} < 1$ iff $G(j,j) < \infty$ iff the geometric distribution of the
1" number of visits to j has finite mean and hence is finite a.s. … ∎

Corollary. *If a Markov chain has only finitely many states, then some state is recurrent.*

2" …

Corollary 4.2.4. *If $i \leftrightarrow j$ and i is recurrent, then j is recurrent.*

Proof. Fix m and n such that $p_{ij}^{(m)} > 0$ and $p_{ji}^{(n)} > 0$. Then

$$\forall s \ge 0 \quad p_{jj}^{(n+s+m)} \ge p_{ji}^{(n)} p_{ii}^{(s)} p_{ij}^{(m)},$$

whence $\sum_s p_{jj}^{(n+s+m)} \ge p_{ji}^{(n)} p_{ij}^{(m)} G(i,i) = \infty$. ∎

Proposition. *If i is recurrent and j is accessible from i, then $f_{ij} = 1$ and $i \leftrightarrow j$.*

Proof. Let $X_0 = i$ and fix n such that $p_{ij}^{(n)} > 0$. Let $A_0 := [X_n = j]$ and let $T_1 := \min\{k \ge n\,;\ X_k = i\}$. Let $A_1 := [X_{T_1+n} = j]$ and $T_2 := \min\{k \ge T_1 + n\,;\ X_k = i\}$. In general, set $A_r := [X_{T_r+n} = j]$ and $T_{r+1} := \min\{k \ge T_r + n\,;\ X_k = i\}$. Then $\langle A_r \rangle$ are independent
2" and each have probability $p_{ij}^{(n)}$, so one of them occurs. … Thus, $f_{ij} = 1$; since $f_{ii} = 1$, it
1" follows that i is accessible from j. … ∎

Actually, we are using something stronger than the Markov property here and in the proof of Proposition 4.2.3, namely, a special case of what is called the ***strong Markov property***. It always holds for discrete-time Markov chains, and usually, but not always, for continuous-time ones. It says the following. Given a Markov chain $\langle X_n \rangle$, call a random variable N with values in $\mathbb{N} \cup \{\infty\}$ a ***stopping time*** if for all n, the event $[N = n]$ (or, if you prefer, its indicator $\mathbf{1}_{[N=n]}$) depends only on $X_0, X_1, \ldots, X_n$, written $[N = n] \in \sigma\big(X_0, X_1, \ldots, X_n\big)$; in other words, there are functions $\phi_n \colon \mathbb{N}^{n+1} \to \{0,1\}$ such that $N = n$ iff $\phi_n\big(X_0, X_1, \ldots, X_n\big) = 1$. Write $\psi(i,B) := P_i\Big[\big(X_0, X_1, \ldots\big) \in B\Big]$. The strong Markov property says that if N is a stopping time and B is an event, then

$$P\Big[\big(X_N, X_{N+1}, \ldots\big) \in B \;\Big|\; X_0, X_1, \ldots, X_N\Big] = \psi(X_N, B)$$

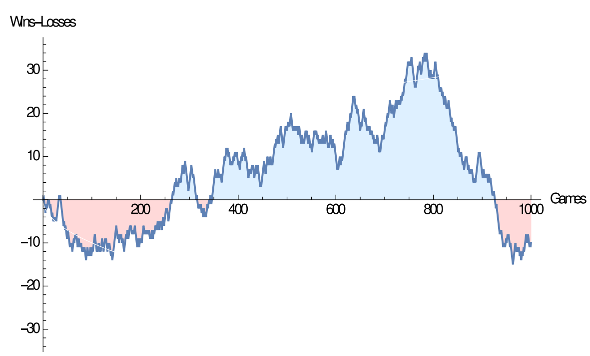
on the event that $N < \infty$. In the cases we are using, X_N is a fixed state, so this is easier to interpret. Note that the conditioning on $X_0, X_1, \ldots, X_N$ implicitly includes conditioning on N. The proof of the strong Markov property is not hard: Given $i_0, i_1, \ldots$ and n with $\phi_n(i_0, i_1, \ldots, i_n) = 1$, the event $[N = n]$ is implied by the event $[X_0 = i_0, X_1 = i_1, \ldots, X_n = i_n]$, whence

$$\begin{aligned}
&P\Big[(X_N, X_{N+1}, \ldots) \in B \;\Big|\; N = n, X_0 = i_0, X_1 = i_1, \ldots, X_n = i_n\Big] \\
&\quad = P\Big[(X_n, X_{n+1}, \ldots) \in B \;\Big|\; X_0 = i_0, X_1 = i_1, \ldots, X_n = i_n\Big] \\
&\quad = P\Big[(X_n, X_{n+1}, \ldots) \in B \;\Big|\; X_n = i_n\Big] \\
&\quad = \psi(i_n, B)
\end{aligned}$$

by the Markov property and homogeneity.

An irreducible Markov chain is called ***transient*** or ***recurrent*** according as its states are.

In particular, the Markov chain of Example 1.9(a) is recurrent and a.s. every bead is visited.



EXAMPLE 4.2(A). Consider the Markov chain on $\mathbb{Z}$ such that, for a given p and all i, $p_{i,i+1} = p$ and $p_{i,i-1} = 1 - p$. If $p \neq \frac{1}{2}$, then the SLLN shows that the chain is transient. We show that for $p = 1/2$, it is recurrent. Note that it has period 2. Now

$$p_{00}^{(2n)} = \binom{2n}{n} \frac{1}{2^{2n}}.$$

2" $\ldots$ Stirling's approximation $n! \sim \sqrt{2\pi n}(n/e)^n$ yields

$$p_{00}^{(2n)} \sim \frac{1}{\sqrt{\pi n}}, \tag{N3}$$

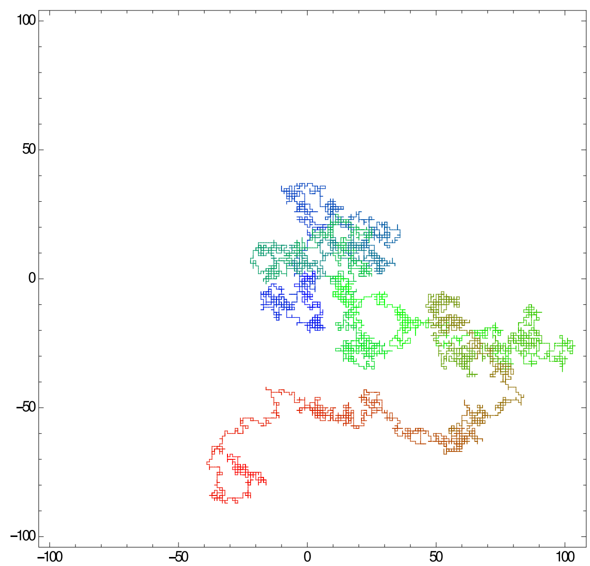
whence $G(0,0) = \infty$.

Second proof. Let $q := 1 - p$. Then $p_{00}^{(2n)} = \binom{2n}{n} p^n q^n$. Now

$$\binom{-\frac{1}{2}}{n} = (-1)^n \binom{2n}{n} \frac{1}{2^{2n}}.$$

2" ... Therefore $G(0,0) = \sum_{n \ge 0} \binom{-1/2}{n} (-4pq)^n = (1-4pq)^{-1/2} = |1-2p|^{-1}$. Thus, ...
2" $G(0,0) = \infty$ iff $p = 1/2$. Also, we see that $f_{00} = 2(p \wedge q)$. ...

Third proof when $p = 1/2$. Let $a := f_{10}$. By symmetry, we have $a = f_{-1,0}$, whence $f_{00} = a$.
.5" ... We also have $a = (1 + f_{20})/2$ ... and $f_{20} = f_{21} f_{10} = a^2$, ... whence $(a-1)^2 = 0$, so $a = 1$.



10,000 steps of 2D random walk.

One of the most famous theorems in probability extends this to higher dimensions:

PÓLYA'S THEOREM. *Simple random walk on the lattice $\mathbb{Z}^d$ is transient iff $d \ge 3$.*

This follows from:

PROPOSITION. *For simple random walk on $\mathbb{Z}^d$,*

$$p_{\mathbf{00}}^{(2n)} \sim 2\Big(\frac{d}{4\pi n}\Big)^{d/2}$$

as $n \to \infty$.

Idea of proof: Let $N_i(n)$ be the number of steps among the first n in direction i. By the WLLN, $N_i(2n) \sim 2n/d$ and $P\big[\forall i \;\; N_i(2n) \text{ is even}\big] \to 2^{-(d-1)}$. Given the values of $N_i(2n)$,
2" the d coordinates of X_{2n} are independent, so the result follows from (N3). ...

§4.3. Limit Theorems.

Let μ_{ii} denote the expected number of transitions needed to return to i starting from i. Let $N_j(n)$ be the number of visits to j by time n. These visits form a delayed renewal process if j is recurrent. Actually, even when j is transient, the visits form a delayed renewal process, albeit one with only a finite number of renewals a.s. Thus, with the convention that $a/\infty := 0$ for any finite a, our results on renewal theory give:

THEOREM 4.3.1. *If $i \leftrightarrow j$, then*

(i) *when the Markov chain starts from state i, we have* $\lim_{n\to\infty} \dfrac{N_j(n)}{n} = \dfrac{1}{\mu_{jj}}$ *a.s.;*

(ii) $\lim_{n\to\infty} \dfrac{1}{n} \sum_{k=1}^{n} p_{ij}^{(k)} = \dfrac{1}{\mu_{jj}}$*;*

(iii) *when j is aperiodic, we have* $\lim_{n\to\infty} p_{ij}^{(n)} = \dfrac{1}{\mu_{jj}}$*;*

(iv) $\lim_{n\to\infty} p_{jj}^{(nd)} = \dfrac{d}{\mu_{jj}}$, *where $d := d(j)$ is the period of j.*

↓ The results we are using are: (i) Proposition 3.3.1; (ii) The elementary renewal theorem; (iii) and (iv) Blackwell's renewal theorem. Here, if j (and so i) is transient, then
3" ↑ we consider separately the two conditions that j is visited or not.

We call a recurrent state i ***positive recurrent*** if $\mu_{ii} < \infty$ and ***null recurrent*** otherwise.

PROPOSITION. *Every finite-state Markov chain has a state that is positive recurrent.*

2" . . .

PROPOSITION 4.3.2. *If $i \leftrightarrow j$ and i is null recurrent, then so is j.*

Proof. Let k and ℓ be such that $p_{ij}^{(k)} > 0$ and $p_{ji}^{(\ell)} > 0$. Let $d = d(i) = d(j)$. Since i is null recurrent, Theorem 4.3.1(iv) tells us that

$$0 = \lim_{n\to\infty} p_{ii}^{(nd+k+\ell)} \ge \limsup_{n\to\infty} p_{ij}^{(k)} p_{jj}^{(nd)} p_{ji}^{(\ell)} = p_{ij}^{(k)} p_{ji}^{(\ell)} \cdot \frac{d}{\mu_{jj}},$$

whence $\mu_{jj} = \infty$. ∎

As was the case for renewal processes, stationary Markov chains arise from limiting probabilities used as initial distributions. Recall that $\langle X_n \rangle$ is stationary if $\forall k \ge 0$ $\langle X_n, X_{n+1}, \ldots, X_{n+k} \rangle$ has a joint distribution that is the same for each n. The homogeneous Markov property shows that this follows for all k if it holds for $k = 0$ and $n \in \{0, 1\}$:

$$\begin{aligned} \forall j \quad p_j = P[X_0 = j] = P[X_1 = j] &= \sum_{i=0}^{\infty} P[X_0 = i, X_1 = j] \\ &= \sum_{i=0}^{\infty} p_i p_{ij}. \end{aligned}$$

4" ... We call an initial distribution $\langle p_j \rangle$ ***stationary*** if this holds.

An irreducible Markov chain is called ***positive recurrent*** or ***null recurrent*** according as its states are.

THEOREM 4.3.3. *Consider an irreducible aperiodic Markov chain and write*

$$\pi_j := \lim_{n \to \infty} p_{ij}^{(n)} = \frac{1}{\mu_{jj}}.$$

The following are equivalent:

(i) *the chain is positive recurrent;*

(ii) $\exists$ *a stationary probability distribution;*

(iii) $\langle \pi_j \rangle$ *is the unique stationary probability distribution.*

In this case, if $x_i \ge 0$, $c := \sum_j x_j > 0$, *and* $\forall j \;\; x_j = \sum_i x_i p_{ij}$, *then* $c < \infty$ *and* $\forall i \;\; x_i = c\pi_i$.

LEMMA (FATOU'S LEMMA FOR SERIES). *If* $\alpha_n(j) \ge 0$, $\lim_{n \to \infty} \alpha_n(j) = \alpha(j)$, *and* $\lim\limits_{n \to \infty} \sum\limits_j \alpha_n(j) = \alpha$, *then* $\alpha \ge \sum\limits_j \alpha(j)$.

Proof. $\forall J \;\; \sum\limits_{j \le J} \alpha(j) = \lim\limits_n \sum\limits_{j \le J} \alpha_n(j) \le \lim\limits_n \sum\limits_{j=0}^{\infty} \alpha_n(j) = \alpha$, so $\sum\limits_{j=0}^{\infty} \alpha(j) \le \alpha$. ∎

LEMMA (LDCT FOR SERIES). *If* $|\alpha_n(j)| \le \beta(j)$, $\sum_{j=0}^{\infty} \beta(j) < \infty$, *and* $\lim\limits_{n \to \infty} \alpha_n(j) = \alpha(j)$, *then*

$$\lim_{n \to \infty} \sum_{j=0}^{\infty} \alpha_n(j) = \sum_{j=0}^{\infty} \alpha(j).$$

Proof. We have

$$\left| \sum_{j=0}^{\infty} \alpha_n(j) - \sum_{j=0}^{\infty} \alpha(j) \right| \le \sum_{j=0}^{\infty} \left| \alpha_n(j) - \alpha(j) \right| \le \sum_{j=0}^{J} \left| \alpha_n(j) - \alpha(j) \right| + \sum_{j > J} 2\beta(j).$$

Now let $n \to \infty$. Then let $J \to \infty$. ∎

Proof of Theorem 4.3.3. $(i) \Rightarrow (ii)$: Since $p_{ij}^{(n+1)} = \sum_k p_{ik}^{(n)} p_{kj}$, Fatou's lemma gives
$\pi_j \ge \sum_k \pi_k p_{kj}$. Since $\sum_j p_{ij}^{(n)} = 1$, Fatou's lemma also gives $\sum_{j=0}^{\infty} \pi_j \le 1$. Thus $\sum_j \pi_j \ge$
1" $\sum_j \sum_k \pi_k p_{kj} = \sum_k \pi_k \sum_j p_{kj} = \sum_k \pi_k$, whence $\forall j\ \ \pi_j = \sum_k \pi_k p_{kj}$. ... Therefore, $p_i :=$
$\pi_i / \sum \pi_j$ form a stationary probability distribution, where the denominator is positive
since each $\pi_j > 0$.

$(ii) \Rightarrow (iii)$: If $\langle p_i \rangle$ is any stationary probability distribution, then $\forall n, j\ \ p_j = \sum_{i=0}^{\infty} p_i p_{ij}^{(n)}$,
2" ... whence by the LDCT, $p_j = \sum_{i=0}^{\infty} p_i \pi_j = \pi_j$. That is, $\langle \pi_j \rangle$ is a stationary probability
distribution and is unique.

$(iii) \Rightarrow (i)$ since some $\pi_j > 0$.

Finally, in the case of positive recurrence and $x_j = \sum_i x_i p_{ij}$, we have $\forall n\ \ x_j =$
2" $\sum_i x_i p_{ij}^{(n)}$ (as $[x_j]$ is a left eigenvector for P), ... so $x_j \ge \sum_i x_i \pi_j = c\pi_j$. This means
that $c < \infty$, whence the LDCT gives $x_j = c\pi_j$. ∎

In the irreducible positive recurrent periodic case, the unique stationary probability distribution is still $\pi_j := 1/\mu_{jj} = \lim_{n\to\infty} p_{jj}^{(nd)}/d$. See Exercise 4.17, p. 221 (solution in the back of the book).

Example (Simple Random Walk on $\mathbb{Z}$). This is null recurrent. If not, then every
2" solution to $\forall j\ \ x_j = \sum_i x_i p_{ij}$ satisfies $\sum_j x_j < \infty$. However, $x_j \equiv 1$ is a solution. ...

Example 4.4(a) (The Gambler's Ruin Problem). A gambler needs \$$N$ but has only \$$i$ ($1 \le i \le N-1$). He plays games that give him chance p of winning \$1 and $q := 1 - p$ of losing \$1 each time. When his fortune is either \$$N$ or 0, he stops. What is his chance of success?

↓

Solution. The gambler's fortune is a Markov chain on $\{0, 1, \ldots, N\}$. Let α_i be the probability of success. Then $\alpha_0 = 0$, $\alpha_N = 1$, and

$$\text{for } 1 \le i \le N-1 \qquad \alpha_i = p\alpha_{i+1} + q\alpha_{i-1},$$

which gives

$$\alpha_{i+1} - \alpha_i = \frac{q}{p}(\alpha_i - \alpha_{i-1}).$$

Therefore

$$\alpha_{i+1} - \alpha_i = \left(\frac{q}{p}\right)^i (\alpha_1 - \alpha_0) = \left(\frac{q}{p}\right)^i \alpha_1.$$

To determine α_1, add these up:

$$1 = \alpha_N = \sum_{i=0}^{N-1} (\alpha_{i+1} - \alpha_i) = \sum_{i=0}^{N-1} \left(\frac{q}{p}\right)^i \alpha_1,$$

so

$$\alpha_1 = 1 \Big/ \sum_{i=0}^{N-1} \left(\frac{q}{p}\right)^i.$$

By adding only the equations for $\alpha_1 - \alpha_0, \alpha_2 - \alpha_1, \ldots, \alpha_i - \alpha_{i-1}$, we get

$$\alpha_i = \sum_{j=0}^{i-1} \left(\frac{q}{p}\right)^j \Big/ \sum_{j=0}^{N-1} \left(\frac{q}{p}\right)^j.$$

6" ↑ For example, when $p = \frac{1}{2}$, we have $\alpha_i = i/N$.

For an application to statistics, see Exercise 4.30, p. 224 (with answer in the back).

§4.5. Branching Processes.

An individual has k children with probability p_k, where $\sum_{k=0}^{\infty} p_k = 1$. The children reproduce independently according to the same offspring distribution. Let Z_n be the size of the nth generation. Clearly Z_n is a Markov chain, called a ***Galton–Watson branching process***. The initial state Z_0 is usually assumed to be 1.

This was introduced in order to study British family names. It is also of interest in biology, chain reactions, electron multipliers, and analysis of various probabilistic processes. Many, many variations have been (and continue to be) studied.

Let L be a random variable with $P[L = k] = p_k$ and let $\langle L_i^{(n)} ;\ n, i \geq 1 \rangle$ be independent copies of L, so that $Z_{n+1} = \sum_{i=1}^{Z_n} L_i^{(n+1)}$. The ***probability generating function*** (p.g.f.) of L is

$$f(s) := E[s^L] = \sum_{k \geq 0} p_k s^k \quad (0 \leq s \leq 1).$$

Proposition. *The p.g.f. of Z_n is $f^{(n)} := \underbrace{f \circ \cdots \circ f}_{n\ \text{times}}$.*

Proof. $E[s^{Z_n}] = E\Big[E\big[s^{\sum_{i=1}^{Z_{n-1}} L_i^{(n)}} \mid Z_{n-1}\big]\Big] = E\Big[\prod_{i=1}^{Z_{n-1}} E[s^L]\Big] = E[f(s)^{Z_{n-1}}]$. Apply this n times. ∎

Write $q := P[Z_n \to 0] = P[\exists n \ \ Z_n = 0]$.

Corollary. $q = \lim_{n\to\infty} f^{(n)}(0)$.

Proof. $q = \lim_n P[Z_n = 0] = \lim_n f^{(n)}(0)$. ∎

3" Looking at a graph of f, which is increasing and convex, we see that ...

Proposition. *Suppose $p_1 \neq 1$. We have $q = 1 \iff f'(1) \leq 1$. Also, q is the smallest root of $f(s) = s$ in $[0, 1]$ — the only other possible root being 1.*

Note that $f'(1) = E[L] =: m$, the mean number of offspring per individual.

§4.6. Applications of Markov Chains.

§4.6.1. A Markov Chain Model of Algorithmic Efficiency.

Certain optimization algorithms move from a point to a better point repeatedly until reaching an optimal point. How many steps do they take? We model this very generally and very crudely as follows. There are a known number of points, N, and we start at the worst one. Each step chooses randomly uniformly among the better points, independently of the past points. Thus, we can call the jth best point state j; then we see a Markov chain on the states $\{1, 2, \ldots, N\}$ that starts at state N and ends at state 1. To end at state 1 means that the transition probability from state 1 to state 1 is 1. We will analyze T_N, the number of steps until state 1 is reached. Note that

$$T_N = \sum_{j=1}^{N-1} I_j,$$

where $I_j :=$ indicator of ever being in state j.

Lemma 4.6.1. *$I_1, \ldots, I_{N-1}$ are independent and $P[I_j = 1] = 1/j$.*

Proof. It suffices to show that

$$P\big[I_j = 1 \mid I_{j+1}, \ldots, I_{N-1}\big] = \frac{1}{j}.$$

This is clear for $j = N-1$. For $j < N-1$, given $I_{j+1}, \ldots, I_{N-1}$, let $K := \min\{k > j\,;\ I_k = 1\}$. Then on the event that $K = n$, we have

$$\begin{aligned} P[I_j = 1 \mid I_{j+1}, \ldots, I_{N-1}] &= P[I_j = 1 \mid K = n] \\ &= P[I_j = 1 \mid I_{j+1} = 0, \ldots, I_{n-1} = 0, I_n = 1] \\ &= \frac{P[I_j = 1, I_{j+1} = 0, \ldots, I_{n-1} = 0 \mid I_n = 1]}{P[I_{j+1} = 0, \ldots, I_{n-1} = 0 \mid I_n = 1]} \\ &= \frac{1/(n-1)}{j/(n-1)} = \frac{1}{j}. \end{aligned}$$ ∎

Proposition 4.6.2. *$E[T_N] = \sum_{j=1}^{N-1} 1/j$, $\mathrm{Var}(T_N) = \sum_{j=1}^{N-1} (1/j)(1 - 1/j)$, and*

$$\frac{T_N - \log N}{\sqrt{\log N}} \Rightarrow \mathrm{N}(0,1) \qquad \textit{as } n \to \infty.$$

2" *Proof.* Use Lindeberg's CLT: since the random variables are bounded, this … requires merely that the variance of the sum tend to infinity. ∎

(Note: the book says that $T_N \approx \mathrm{Pois}(\log N)$. This is also true. On the other hand, *Probability Models* gives the normal limit!)

§4.7. Time-Reversible Markov Chains.

The definition of the Markovian property shows that given a finite time N, the sequence

$$X_N, X_{N-1}, \ldots, X_0$$

arises from N steps of a possibly nonhomogeneous Markov chain. Now, the transition probabilities are

$$P\big[X_m = j \mid X_{m+1} = i\big] = \frac{P[X_m = j, X_{m+1} = i]}{P[X_{m+1} = i]} = \frac{P[X_{m+1} = i \mid X_m = j]P[X_m = j]}{P[X_{m+1} = i]}.$$

Therefore, if $\langle X_n \rangle$ is stationary with stationary probabilities $P[X_0 = i] = \pi_i$, then the transition probabilities are homogeneous and equal

$$P[X_m = j \mid X_{m+1} = i] = \frac{\pi_j p_{ji}}{\pi_i} =: p^*_{ij}.$$

Since $P[X_k = i] = \pi_i$, this time-reversed chain has the same stationary probabilities. If it happens that $\forall i, j \;\; p^*_{ij} = p_{ij}$, then the Markov chain is called ***reversible***. This can be written as

$$\forall\, i, j \quad \pi_i p_{ij} = \pi_j p_{ji}.$$

If the chain is irreducible and $\exists x_i \geq 0$ such that $\sum x_i = 1$ and $\forall i, j \;\; x_i p_{ij} = x_j p_{ji}$, then actually $x_i = \pi_i$ and so the chain is reversible since

$$\forall j \qquad \sum_i x_i p_{ij} = \sum_i x_j p_{ji} = x_j.$$

We can also write $w_{ij} := \pi_i p_{ij}$, so that $w_{ji} = w_{ij}$ and

$$\forall i \quad \sum_j w_{ij} = \sum_j \pi_i p_{ij} = \pi_i,$$

whence $p_{ij} = w_{ij} / \sum_k w_{ik}$. Conversely, if $\exists w_{ij} = w_{ji} \geq 0$ with $0 < w := \sum_{i,j} w_{ij} < \infty$ and $p_{ij} = w_{ij} / \sum_k w_{ik}$, then set $w_i := \sum_k w_{ik}$ and

$$\pi_i := \frac{w_i}{\sum_\ell w_\ell} = \frac{w_i}{w}.$$

We have

$$\pi_i p_{ij} = \frac{w_i}{w} \cdot \frac{w_{ij}}{w_i} = \frac{w_{ij}}{w} = \frac{w_{ji}}{w} = \pi_j p_{ji},$$

so the chain is reversible with these stationary probabilities.

3" Note that this is the same as a random walk on a graph with weighted edges. …

To find a condition for reversibility not requiring finding numbers x_i or w_{ij}, consider

$$\frac{p_{ij}}{p_{ji}} = \frac{\pi_j}{\pi_i}.$$

This means that around every cycle $i_0, i_1, \ldots, i_n, i_{n+1} = i_0$ where the successive transitions have positive probability, we have

$$\prod_{j=0}^{n} \frac{p_{i_j,i_{j+1}}}{p_{i_{j+1},i_j}} = \prod_{j=0}^{n} \frac{\pi_{i_{j+1}}}{\pi_{i_j}} = 1.$$

Conversely, if this holds and the chain is irreducible, then we may define numbers x_ℓ by making $x_0 > 0$ arbitrary and setting

$$x_\ell := x_0 \prod_{j=0}^{k-1} \frac{p_{i_j,i_{j+1}}}{p_{i_{j+1},i_j}}$$

2" for any path $0 = i_0, i_1, \ldots, i_k = \ell$ since any two paths give the same value. ... This implies
1" that $x_i p_{ij} = x_j p_{ji}$, ... so $x_j = \sum_i x_j p_{ji} = \sum_i x_i p_{ij}$. By our version of Theorem 4.3.3, if the chain is positive recurrent, this means $\sum x_j < \infty$ and $x_j = \pi_j \sum_i x_i$, so $\pi_i p_{ij} = \pi_j p_{ji}$. Thus, we have proved

THEOREM 4.7.1. *An irreducible stationary Markov chain is reversible iff for every cycle $i_0, i_1, \ldots, i_n, i_{n+1} = i_0$ where the successive transitions have positive probability, we have*

$$\prod_{j=0}^{n} \frac{p_{i_j,i_{j+1}}}{p_{i_{j+1},i_j}} = 1. \tag{N4}$$

In fact, we *extend* the notion of reversibility beyond positive recurrent chains to include all those Markov chains for which $\exists x_i > 0\ \forall i, j\ \ x_i p_{ij} = x_j p_{ji}$. We still have $p_{ij} = w_{ij}/x_i$ if $w_{ij} := x_i p_{ij} = w_{ji}$, but it may be that $\sum x_i = +\infty$. Likewise, if $w_{ij} = w_{ji}$ are given with $\forall i\ \ x_i := \sum_j w_{ij} < \infty$ and $p_{ij} = w_{ij}/x_i$, then the chain is reversible. Theorem 4.7.1 extends to say that any irreducible Markov chain is reversible iff (N4) holds for all cycles. Furthermore, the chain has a stationary probability distribution iff $\sum x_i < \infty$ (i.e., the
2" sum of the weights is finite) by Theorem 4.3.3. ...

EXAMPLE 4.7(A). Any nearest-neighbor random walk on $\mathbb{Z}$ or on a tree is reversible.

EXAMPLE: Simple random walk on any graph, such as the lattice $\mathbb{Z}^d$, is reversible. If the graph is infinite, then simple random walk is not positive recurrent.

`Suppose we toss a fair coin repeatedly. What is the expected number of tosses until the numbers of heads and tails are equal?`

We will later study reversibility in continuous time and see that certain diffusions, including Brownian motion, are reversible.

We now show that electrical networks are intimately connected to reversible Markov chains. The states i will now be vertices x and the weights w_{ij} of edges will now be conductances C_{xy}.

Let G be a finite connected graph, x a vertex of G, and A, Z disjoint subsets of vertices of G. Let T_A be the first time that the random walk visits ("hits") some vertex in A; if the random walk happens to start in A, then this is 0. Occasionally, we will use T_A^+, which is the first time after 0 that the walk visits A; this is different from T_A only when the walk starts in A. Usually A and Z will be singletons. Often, all the edge weights are equal; we call this case ***simple random walk***.

Consider the probability that the random walk visits A before it visits Z as a function of its starting point x:

$$F(x) := P_x[T_A < T_Z]. \tag{N5}$$

We use $\restriction$ to denote the restriction of a function to a set. Then clearly $F \restriction A \equiv 1$, $F \restriction Z \equiv 0$, and for $x \notin A \cup Z$,

$$\begin{aligned} F(x) &= \sum_y P_x[\text{first step is to } y]\, P_x[T_A < T_Z \mid \text{first step is to } y] \\ &= \sum_{x \sim y} p_{xy} F(y) = \frac{1}{C_x} \sum_{x \sim y} C_{x,y} F(y), \end{aligned}$$

where $x \sim y$ indicates that x, y are adjacent in G. In the special case of simple random walk, this equation becomes

$$F(x) = \frac{1}{\deg x} \sum_{x \sim y} F(y),$$

where $\deg x$ is the degree of x, i.e., the number of edges incident to x. That is, $F(x)$ is the average of the values of F at the neighbors of x. In general, this is still true, but the average is taken with weights. We say that F is ***harmonic*** at such a point. Now harmonic functions satisfy a maximum principle: For $H \subseteq G$, write $\overline{H}$ for the set of vertices that are either in H or are adjacent to some vertex in H. When we say that a function is defined on a graph, we mean that it is defined on its vertex set.

Maximum Principle. *If $H \subseteq G$, H is connected, $f\colon G \to \mathbb{R}$, f is harmonic on H, and $\max f{\restriction}H = \max f$, then $f{\restriction}\overline{H} \equiv \max f$.*

Proof. Let $K := \{y \in \overline{H}\,;\ f(y) = \max f\}$. Note that if $x \in H$, $x \sim y$, and $f(x) = \max f$, then $f(y) = \max f$ by harmonicity of f at x. Thus, $\overline{K} \cap \overline{H} = K$. Since H is connected, it follows that $K = \overline{H}$. ∎

This leads to the

Uniqueness Principle. *If $H \subsetneqq G$, $f, g\colon G \to \mathbb{R}$, f, g are harmonic on H, and $f{\restriction}(G \setminus H) = g{\restriction}(G \setminus H)$, then $f = g$.*

Proof. Let $h := f - g$. We claim that $h \le 0$. This suffices to establish the corollary since then $h \ge 0$ by symmetry, whence $h = 0$.

Now $h = 0$ off H, so if $h \not\le 0$ on H, then h is positive somewhere on H, whence $\max h{\restriction}H = \max h$. Therefore, according to the maximum principle, h is a positive constant on the closure of some component K of H. In particular, $h > 0$ on the non-empty set $\overline{K} \setminus K$. However, $\overline{K} \setminus K \subseteq G \setminus H$, whence $h = 0$ on $\overline{K} \setminus K$. This is a contradiction. ∎

Thus, the harmonicity of the function $x \mapsto P_x[T_A < T_Z]$ (together with its values where it is not harmonic) characterizes it.

Existence Principle. *If $H \subsetneqq G$ and $f_0 : G \setminus H \to \mathbb{R}$, then $\exists f\colon G \to \mathbb{R}$ such that $f{\restriction}(G \setminus H) = f_0$ and f is harmonic on H.*

Proof. Let X be the first vertex in $G \setminus H$ visited by the corresponding random walk. Set $f(x) := E_x[f_0(X)]$. ∎

This is the solution to the so-called Dirichlet problem. The function F of (N5) is the particular case $H = G \setminus (A \cup Z)$, $f_0{\restriction}A \equiv 1$, and $f_0{\restriction}Z \equiv 0$.

In fact, we could have immediately deduced existence from uniqueness or vice versa: The Dirichlet problem on a finite graph consists of a finite number of linear equations, one for each vertex in H. Since the number of unknowns is equal to the number of equations, we get the equivalence of uniqueness and existence.

In order to study the solution to the Dirichlet problem, especially for a sequence of subgraphs of an infinite graph, we will discover that electrical networks are useful. Electrical networks, of course, have a physical meaning whose intuition is useful to us, but also they can be used as a rigorous mathematical tool.

Mathematically, an electrical network is just a weighted graph. But now we call the weights of the edges ***conductances*** and write them as C_{xy}; their reciprocals are called ***resistances***, written R_{xy}. We hook up a battery or batteries (this is just intuition) between A and Z so that the ***voltage*** (or ***potential***) at every vertex in A is 1 and in Z is 0 (more generally, so that the voltages on $G \setminus H$ are given by f_0). ***Voltages*** V are then established at every vertex and ***current*** I runs through the edges. These functions are implicitly defined and uniquely determined, as we will see, by two "laws":

> ***Ohm's Law*** : If $x \sim y$, the current flowing from x to y satisfies $(V_x - V_y) = I_{xy} R_{xy}$.
>
> ***Kirchhoff's Node Law***: The sum of all currents flowing out of a given vertex is 0, provided the vertex is not connected to a battery.

Physically, Ohm's law, which is usually stated as $V = IR$ in engineering, is an empirical statement about linear response to voltage differences — certain components obey this law over a large range of voltage differences. Kirchhoff's node law expresses the fact that charge does not build up at a node (current being charge per unit time). If we add wires corresponding to the batteries, then the proviso in Kirchhoff's node law is unnecessary.

Mathematically, we'll take Ohm's law to be the definition of current in terms of voltage. In particular, $I_{xy} = -I_{yx}$. Then Kirchhoff's node law presents a constraint on what kind of function V can be. Indeed, it determines V uniquely: Current flows into G at A and out at Z. Thus, we may combine the two laws on $G \setminus (A \cup Z)$ to obtain

$$\forall x \notin A \cup Z \qquad 0 = \sum_{x \sim y} I_{xy} = \sum_{x \sim y} (V_x - V_y) C_{xy},$$

or

$$V_x = \frac{1}{C_x} \sum_{x \sim y} C_{xy} V_y,$$

where

$$C_x := \sum_{y \sim x} C_{xy}.$$

That is, $V_\bullet$ is harmonic on $G \setminus (A \cup Z)$. Since $V \restriction A \equiv 1$ and $V \restriction Z \equiv 0$, it follows that $V = F$; in particular, we have uniqueness and existence of voltages. The voltage function is just the solution to the Dirichlet problem.

Suppose that $A = \{a\}$ is a singleton. What is the chance that a random walk starting at a will hit Z before it returns to a? Write this as

$$P[a \to Z] := P_a[T_Z < T_{\{a\}}^+].$$

Impose a voltage of V_a at a and 0 on Z. Since $V_\bullet$ is linear in V_a, we have that $P_x[T_{\{a\}} < T_Z] = V_x/V_a$, whence

$$\begin{aligned} P[a \to Z] &= \sum_x p_{ax}\Big(1 - P_x[T_{\{a\}} < T_Z]\Big) = \sum_x \frac{C_{ax}}{C_a}(1 - V_x/V_a) \\ &= \frac{1}{V_a C_a}\sum_x C_{ax}(V_a - V_x) = \frac{1}{V_a C_a}\sum_x I_{ax}. \end{aligned}$$

In other words,

$$V_a = \frac{\sum_x I_{ax}}{C_a\, P[a \to Z]}.$$

Since $\sum_x I_{ax}$ is the total amount of current flowing into the circuit at a, we may regard the entire circuit between a and Z as a single conductor of net, or ***effective*****,** ***conductance***

$$\mathcal{C}_{\text{eff}} := C_a\, P[a \to Z] =: \mathcal{C}(a \leftrightarrow Z), \tag{N6}$$

where the last notation indicates the dependence on a and Z. We define the ***effective resistance*** $\mathcal{R}(a \leftrightarrow Z)$ to be its reciprocal. One answer to our question above is thus $P[a \to Z] = \mathcal{C}(a \leftrightarrow Z)/C_a$. Later, we will see some ways to compute effective conductance.

Now the number of visits to a before hitting Z is a geometric random variable with mean $P[a \to Z]^{-1} = C_a \mathcal{R}(a \leftrightarrow Z)$. This generalizes as follows. Let $G(a, x)$ be the expected number of visits to x strictly before hitting Z by a random walk started at a. Thus, $G(a, a) = C_a \mathcal{R}(a \leftrightarrow Z)$ and $G(a, x) = 0$ for $x \in Z$. The function $G(\cdot, \cdot)$ is the ***Green function*** for the random walk absorbed on Z.

Theorem (Green Function = Voltage). *When a voltage is imposed so that a unit current flows from a to Z, then $V_x = G(a, x)/C_x$ for all x.*

Proof. We have just shown that this is true for $x \in \{a\} \cup Z$, so it suffices to establish that $G(a, x)/C_x$ is harmonic elsewhere. But by Exercise –?– , we have that $G(a, x)/C_x = G(x, a)/C_a$ and the harmonicity of $G(x, a)$ is established just as for the function of (N5). ∎

Given that we have two probabilistic interpretations of voltage, we naturally wonder whether current has a probabilistic interpretation. We might guess one by the following unrealistic but simple model of electricity: positive particles enter the circuit at a, they do Brownian motion on G (taking longer to pass through small conductors) and, when they hit Z, they are removed. The net flow of particles across an edge would then be the current on that edge. It turns out that in our mathematical model, this is correct:

Proposition (Interpretation of Current). *Start a random walk at a and absorb it when it first visits Z. For $x \sim y$, let S_{xy} be the number of transitions from x to y. Then $E[S_{xy}] = G(a,x)p_{xy}$ and $E[S_{xy} - S_{yx}] = I_{xy}$, where I is the current in G when a potential is applied between a and Z in such a way that unit current flows in at a.*

Note that we count a transition from y to x when $y \notin Z$ but $x \in Z$, although we do not count this as a visit to x in computing $G(a,x)$.

Proof. We have

$$\begin{aligned} E[S_{xy}] &= E\left[\sum_{k=0}^{\infty} \mathbf{1}_{[X_k=x]} \mathbf{1}_{[X_{k+1}=y]}\right] = \sum_{k=0}^{\infty} P[X_k = x,\, X_{k+1} = y] \\ &= \sum_{k=0}^{\infty} P[X_k = x]\, p_{xy} = E\left[\sum_{k=0}^{\infty} \mathbf{1}_{[X_k=x]}\right] p_{xy} = G(a,x)p_{xy}. \end{aligned}$$

Hence by the preceding theorem, we have $\forall x, y$,

$$E[S_{xy} - S_{yx}] = G(a,x)p_{xy} - G(a,y)p_{yx} = \left(\frac{G(a,x)}{C_x} - \frac{G(a,y)}{C_y}\right) C_{xy} = (V_x - V_y)C_{xy} = I_{xy}.$$

∎

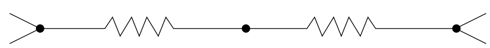
Effective conductance is a key quantity because of the following relationship to the question of transience and recurrence when G is infinite. For an infinite graph G, we assume that there are only a finite number of edges incident to each vertex. But we allow more than one edge between a given pair of vertices: each such edge has its own conductance. Loops are also allowed (edges with only one endpoint), but these may be ignored since they only delay the random walk. Strictly speaking, then, G may be a ***multigraph***, not a graph. However, we will ignore this distinction.

Let $\langle G_n \rangle$ be any sequence of finite subgraphs of G that ***exhaust*** G, i.e., $G_n \subseteq G_{n+1}$ and $G = \bigcup G_n$. Let Z_n be the set of vertices in $G \setminus G_n$. (Note that if Z_n is contracted to a point, the graph will have finitely many vertices but may have infinitely many edges.) Then for each $a \in G$, the limit $\lim_n P[a \to Z_n]$ is the probability of never returning to a in G — the escape probability from a. This is positive iff the random walk on G is transient. By (N6), $\lim_{n\to\infty} \mathcal{C}(a \leftrightarrow Z_n)$ has the same property. We call $\lim_{n\to\infty} \mathcal{C}(a \leftrightarrow Z_n)$ the ***effective conductance*** from a to ∞ in G and denote it by $\mathcal{C}(a \leftrightarrow \infty)$ or, if a is understood, by C_{eff}. Its reciprocal, ***effective resistance***, is denoted R_{eff}. We have shown:

THEOREM (TRANSIENCE EQUIVALENT TO FINITE EFFECTIVE CONDUCTANCE). *Random walk on a connected network is transient iff the effective conductance from any of its vertices to infinity is positive.*

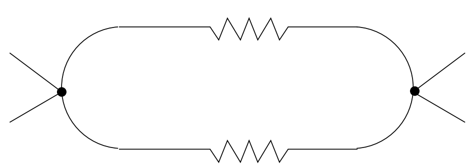
How do we calculate effective conductance of a network? Since we want to replace a network by an equivalent single conductor, it is natural to attempt this by replacing more and more of G through simple transformations. There are, in fact, three such simple transformations, series, parallel, and star-triangle, and it turns out that they suffice to reduce all finite planar networks by a theorem of Epifanov.

I. Series. Two resistors R_1 and R_2 in series are equivalent to a single resistor $R_1 + R_2$. In other words, if $v \in G \setminus (A \cup Z)$ is a node of degree 2 with neighbors u_1, u_2 and we replace the edges (u_i, v) by a single edge (u_1, u_2) having resistance $R_{u_1 v} + R_{v u_2}$, then all potentials and currents in $G \setminus \{v\}$ are unchanged and the current that flows from u_1 to u_2 equals $I_{u_1 v}$.



Proof. It suffices to check that Ohm's and Kirchhoff's laws are satisfied on the new network for the voltages and currents given. This is easy. ∎

II. Parallel. Two conductors C_1 and C_2 in parallel are equivalent to one conductor $C_1 + C_2$. In other words, if two edges e_1 and e_2 that both join vertices $v_1, v_2 \in G$ are replaced by a single edge e joining v_1, v_2 of conductance $C_e := C_{e_1} + C_{e_2}$, then all voltages and currents in $G \setminus \{e_1, e_2\}$ are unchanged and the current I_e equals $I_{e_1} + I_{e_2}$. The same is true for an infinite number of edges in parallel.

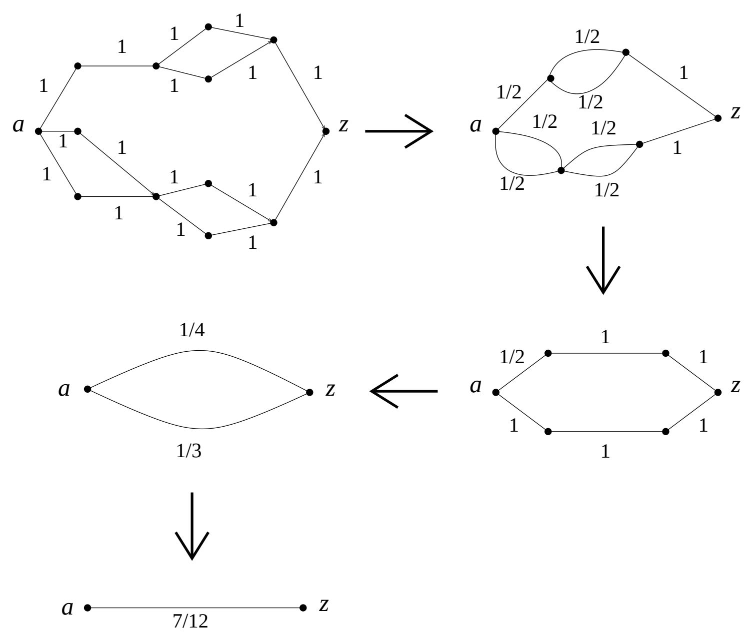


Proof. Check Ohm's and Kirchhoff's laws with $I_e := I_{e_1} + I_{e_2}$. ∎

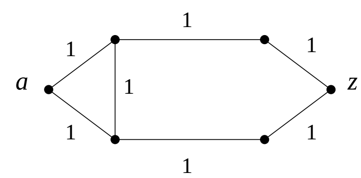
EXAMPLE (GAMBLER'S RUIN). Consider simple random walk on $\mathbb{Z}$. Let $0 \le k \le n$. What is $P_k[T_0 < T_n]$? It is the voltage at k when there is a unit voltage imposed at 0 with 0 voltage at n. If we replace the resistors in series from 0 to k by a single resistor with resistance k and the resistors from k to n by a single resistor of resistance $n-k$, then the voltage at k does not change. But now this voltage is simply the probability of taking a step to 0, which is thus $(n-k)/n$.

EXAMPLE: Suppose that each edge in the following network has equal conductance. What is $P[a \to z]$? Following the transformations indicated in the figure, we obtain $\mathcal{C}(a \leftrightarrow z) = 7/12$, so that

$$P[a \to z] = \frac{\mathcal{C}(a \leftrightarrow z)}{C_a} = \frac{7/12}{3} = \frac{7}{36}\,.$$



Example: What is $P[a \to z]$ in the following network?



There are 2 ways to deal with the vertical edge:

(1) Remove it: by symmetry, the voltages at its endpoints are equal, whence no current flows on it.

(2) Contract it, i.e., remove it but combine its endpoints into one vertex (we could also combine the other two unlabelled vertices with each other): the voltages are the same, so they may be combined.

In either case, we get $\mathcal{C}(a \leftrightarrow z) = 2/3$, whence $P[a \to z] = 1/3$.

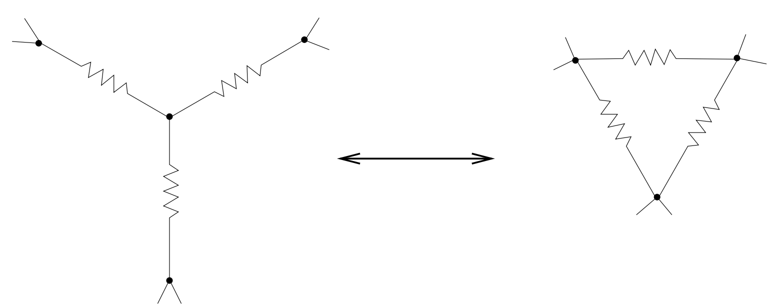
III. Star-triangle. The configurations below are equivalent when

$$\forall i \in \{1, 2, 3\} \qquad C_{uv_i} C_{v_{i-1} v_{i+1}} = \gamma,$$

where indices are taken mod 3 and

$$\gamma := \frac{\prod_i C_{uv_i}}{\sum_i C_{uv_i}} = \frac{\sum_i R_{v_{i-1} v_{i+1}}}{\prod_i R_{v_{i-1} v_{i+1}}}.$$

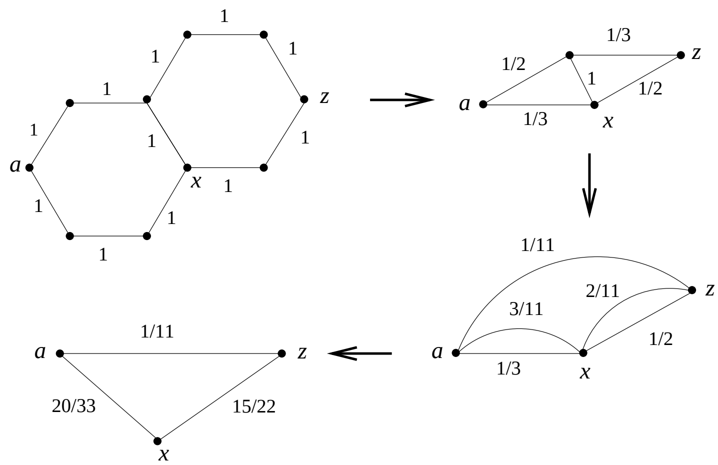
We won't prove this equivalence.



Actually, there is a fourth trivial transformation: we may prune (or add) vertices of degree 1 (and attendant edges) as well as loops.

Either of the transformations star-triangle or triangle-star can also be used to reduce the network in the preceding example.

Example: What is $P_x[\tau_a < \tau_z]$ in the following network?



Following the transformations indicated in the figure, we obtain

$$P_x[\tau_a < \tau_z] = \frac{10/33}{10/33 + 15/22} = \frac{4}{13}\,.$$

THEOREM. *For any positive recurrent Markov chain and any states $a \neq z$,*

$$E_a[\textit{time to first return to } a \textit{ that occurs after } T_z] = E_a T_z + E_z T_a = \frac{1}{\pi_a P_a[T_z < T_a^+]}.$$

If the chain is reversible, this equals $2\gamma/\mathcal{C}(a \leftrightarrow z)$ [where $\gamma = \sum_{x \sim y} C_{xy}$].

Proof. $P_a\Big[T_z < T_a^+\Big] =$ rate of commutes to z among excursions from a

$$= \frac{\text{rate of commutes to } z \text{ among steps}}{\text{rate of excursions from } a \text{ among steps}} = \frac{1/\text{expected commute time}}{\pi_a}.$$

In the reversible case, use (N6) and the fact that $\pi_a = C_a/(2\gamma)$. ∎

Another important concept concerns energy, but we omit it in favor of simply reciting some of its consequences.

RAYLEIGH'S MONOTONICITY LAW. *If C and C' are two assignments of conductances on the same graph and $C \le C'$ on each edge, then $\mathcal{C}(a \leftrightarrow Z) \le \mathcal{C}'(a \leftrightarrow Z)$ for any a, Z.*

COROLLARY. *If $C \asymp C'$ (i.e., $\exists k_1, k_2\ \ k_1 C \le C' \le k_2 C$ on each edge), then the corresponding random walks are both transient or both recurrent.*

To generalize this, given two networks G, G' with conductance C, C', call a map ϕ from the vertices of G to those of G' ***bounded*** if $\exists k < \infty\ \exists$ map Φ on edges of G such that

(i) $\forall$ edge $(v, w) \in G$, $\Phi(v, w)$ is a path of edges joining $\phi(v)$ and $\phi(w)$ with

$$\sum_{e' \in \Phi(v,w)} C'(e')^{-1} \le kC(v, w)^{-1}; \quad \text{and}$$

(ii) $\forall$ edge $e' \in G'$, there are $\le k$ edges in G whose image under Φ contains e'.
[Think of resistances as lengths of edges.]

EXAMPLE: $G \le G'$, $\phi =$ inclusion, $C \le kC'$. We call two networks ***roughly equivalent*** if there are bounded maps in both directions.

THEOREM (KANAI). *Two roughly equivalent networks are both transient or both recurrent. In fact, if there is a bounded map from G to G' and G is transient, then G' is transient.*

New proof of Pólya's Theorem in $\mathbb{Z}^2$. In $\mathbb{Z}^2$, short together all (x, y) with constant $|x| \vee |y|$. This is recurrent since $\sum \frac{1}{n} = \infty$. ∎

`Give idea for` $\mathbb{Z}^3$`. Talk about continuous case (spherical symmetry helps). Give spherically symmetric tree examples.`

Chapter 5

Continuous-Time Markov Chains

We will do only §§2–4.

§5.2. Continuous-Time Markov Chains.

This section consists of definitions.

A stochastic process with time being an interval in $\mathbb{R}$ is called ***Markov*** if the future and past are independent given the present: $\forall t\ \langle X(s)\,;\ s > t\rangle$ and $\langle X(s)\,;\ s < t\rangle$ are independent given $X(t)$. If the number of states is countable, the process is called a ***chain***. We then identify the states with $\mathbb{N}$. We will deal only with Markov chains that do not have instantaneous states and are right-continuous, i.e., with probability 1 $\forall i \in \mathbb{N}\ \forall t\ \ X(t) = i \Rightarrow \exists \varepsilon > 0\ \ \forall s \in (0, \varepsilon)\ \ X(t+s) = i$. We also assume homogeneous transition probabilities $p_{ij}(s) := P\big[X(t+s) = j \mid X(t) = i\big]$.

By the Markov property, the time spent in each state is a memoryless random variable, hence is an exponential random variable; call the rate ν_i when in state i. Let the probability distribution of the next state visited be P_{ij}; the next state is independent of the time spent in i by the Markov property again. This gives a constructive view of a continuous time Markov chain: use a timer at each state; when it rings, move according to a discrete-time Markov chain. However, there is a difficulty: what if we make an infinite number of transitions in a finite time period? Example: $P_{i,i+1} = 1$, $\nu_i = i^2$. If τ_i is the time spent in i, then

$$E[\tau_i] = \frac{1}{i^2}, \quad \text{so} \quad E\Big[\sum \tau_i\Big] < \infty,$$

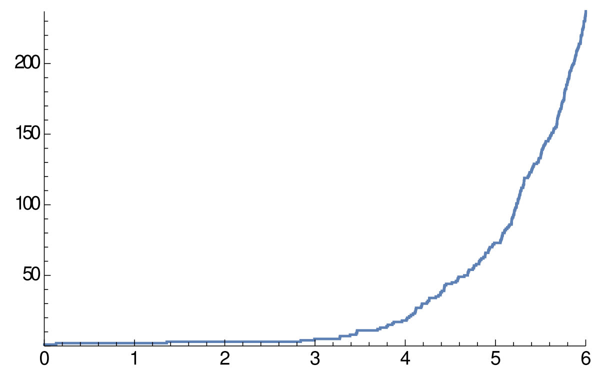
so $\sum \tau_i < \infty$ a.s. `The paradox of Lincoln's penny.` We will not treat such chains, only ***regular*** ones, i.e., ones defined on $[0, \infty)$ that with probability 1 make only a finite number of transitions in $[0, N)$ for every $N < \infty$.

Another construction of continuous-time Markov chains is as follows. Let $q_{ij} := \nu_i P_{ij}$
1" for $i \neq j$. This is the ***transition rate*** from i to j. We could have at state i timers . . .
with rates q_{ij}; the first to ring determines the next state: from Problem 1.34, p. 53, the
1" probability of being the first to ring is proportional to the rate. . . .

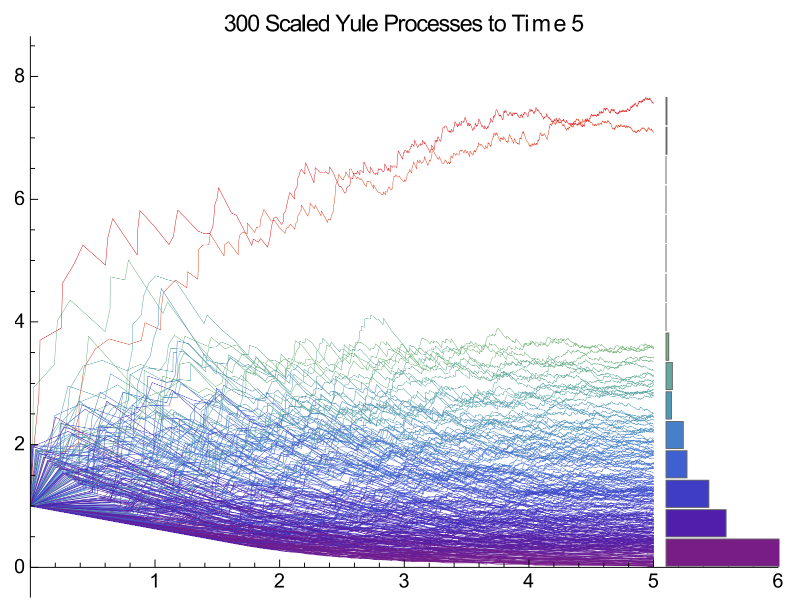
§5.3. Birth and Death Processes.

In case $P_{ij} = 0$ for $|i - j| > 1$, the chain is called a ***birth and death process***. We think of the state as representing the size of a population. Let the ***birth rates*** be $\lambda_i := q_{i,i+1}$ and the ***death rates*** be $\mu_i := q_{i,i-1}$.

If there are no deaths, the process is called a ***pure birth process***. The Poisson process, $\lambda_n = \lambda$ for $n \geq 0$, is such a process. Another is the ***Yule process***, where $\lambda_n = n\lambda$ for $n \geq 0$.



A sample of a Yule process.



A sample of 300 scaled Yule processes, dividing by e^t.

1" . . . The Yule process is regular: as shown on p. 235 of the book, if τ_i denotes the time spent in state i, then for every $k \ge 1$ and $t \in (0, \infty)$, we have (if the chain starts in state 1)

$$P\Big[\sum_{i=1}^{k} \tau_i \le t\Big] = \big(1 - e^{-\lambda t}\big)^k,$$

1" whence $P\Big[\sum_{i=1}^{\infty} \tau_i \le t\Big] = 0$. . . .

Another way to see this result, which says that $P\big[X(t) > k\big] = \big(1 - e^{-\lambda t}\big)^k$, or that $X(t) \sim \mathrm{Geom}\big(e^{-\lambda t}\big)$, is the following. The time $\tau_i \sim \mathrm{Exp}(\lambda i)$, which is also the distribution of the minimum of i independent $\mathrm{Exp}(\lambda)$ random variables. Thus, $\sum_{i=1}^{k} \tau_i$ has the same distribution as $\max_{1 \le i \le k} Z_i$, where $Z_i \sim \mathrm{Exp}(\lambda)$ are independent: τ_k has the same distribution as $\min_{1 \le i \le k} Z_i$, then τ_{k-1} the same as the minimum of the time from the minimum of Z_i to the next smallest, etc. But the cdf of $\max_{1 \le i \le k} Z_i$ is easy to calculate.

Example 5.3(a).

(i) Let $X(t)$ be the number of people in the system of an M/M/s queue, where arrivals have rate λ and service has rate μ. Then $\lambda_n = \lambda$ for $n \ge 0$, $\mu_n = n\mu$ for $1 \le n \le s$, and $\mu_n = s\mu$ for $n > s$.

(ii) A linear growth process with immigration assumes that each individual in the population gives birth at exponential rate λ and dies at rate μ, while there is also immigration at rate θ. Thus $\lambda_n = n\lambda + \theta$ for $n \ge 0$ and $\mu_n = n\mu$ for $n \ge 1$. This can be shown to be regular by a more general method than the one we used for a Yule process. Namely, note that $M_n := E\Big[\sum_{i=1}^{n} \tau_i\Big] \to \infty$ as $n \to \infty$ and show that
2" $P\Big[M_n^{-1} \sum_{i=1}^{n} \tau_i \to 1\Big] = 1$ by calculating the variance of $\sum_{i=1}^{n} \tau_i$. . . .

§5.4. The Kolmogorov Differential Equations.

A pure birth process is the easiest to analyze, since it can always be reduced to a sum of independent (though not necessarily identically distributed) exponential random
2" variables. ... For other processes, we require new tools.

Recall that $p_{ij}(t) = P\big[X(s+t) = j \mid X(s) = i\big]$. There are two sets of differential equations that these functions satisfy, obtained by conditioning on intermediate states.

Theorem 5.4.3 (Kolmogorov's Backward Equations). $\forall i, j, t$

$$p'_{ij}(t) = \sum_{k \neq i} q_{ik} p_{kj}(t) - \nu_i p_{ij}(t).$$

This can be written with matrices as

$$P'(t) = QP(t),$$

where $P(t) := \big(p_{ij}(t)\big)_{i,j}$, $Q := (q_{ij})_{i,j}$, and

$$q_{ii} := -\nu_i = \text{ the negative of the rate of transition out of } i.$$

Proof. First we write an integral equation for $p_{ij}(t)$. Either the chain has jumped by time t or not; if it has, then its first jump is to some state $k \neq i$, from which it eventually reaches state j. Let τ be the time of the first jump. Thus, we have

$$\begin{aligned}
p_{ij}(t) &= E_i\big[P_i[X(t) = j \mid \tau]\big] = E_i\big[\delta_{ij}\mathbf{1}_{[\tau > t]} + P_i[X(t) = j \mid \tau]\mathbf{1}_{[\tau \le t]}\big] \\
&= \delta_{ij} e^{-\nu_i t} + E_i\Big[\sum_{k \neq i} P_{ik} p_{kj}(t - \tau)\mathbf{1}_{[\tau \le t]}\Big] \\
&= \delta_{ij} e^{-\nu_i t} + \int_0^t \sum_{k \neq i} P_{ik} p_{kj}(t - s)\nu_i e^{-\nu_i s}\, ds \\
&= \delta_{ij} e^{-\nu_i t} + \int_0^t \sum_{k \neq i} q_{ik} p_{kj}(u) e^{-\nu_i (t-u)}\, du \\
&= e^{-\nu_i t} H_{ij}(t),
\end{aligned}$$

where

$$H_{ij}(t) := \delta_{ij} + \int_0^t \sum_{k \neq i} q_{ik} p_{kj}(u) e^{\nu_i u}\, du.$$

1" Observe that the integrand is bounded by $\nu_i e^{\nu_i t}$. ... Therefore $H_{ij}(\cdot)$ is continuous, and thus so is $p_{ij}(\cdot)$. But this means that the integrand is continuous by the LDCT for
2" series, ... whence $H_{ij}(\cdot)$ and $p_{ij}(\cdot)$ are differentiable and we can apply the Fundamental Theorem of Calculus to derive

$$p'_{ij}(t) = -\nu_i p_{ij}(t) + e^{-\nu_i t} \sum_{k \neq i} q_{ik} p_{kj}(t) e^{\nu_i t},$$

as desired. ∎

If we use $t := 0$, then we get that $p'_{ij}(0) = q_{ij}$.

The name "backward equations" arises because we conditioned all the way back to the time of the first jump. The forward equations come from a more natural conditioning, yet are more difficult to establish—indeed, they do not always hold. The forward equations arise as follows. Since

$$p_{ij}(t+h) = \sum_k p_{ik}(t)p_{kj}(h) \quad (h > 0),$$

we have

$$\frac{p_{ij}(t+h) - p_{ij}(t)}{h} = \sum_{k \neq j} p_{ik}(t)\frac{p_{kj}(h)}{h} - p_{ij}(t)\frac{1 - p_{jj}(h)}{h}.$$

If we could interchange $\lim_{h \to 0^+}$ with $\sum_{k \neq j}$, e.g., if there are only finitely many states, then we would get

$$p'_{ij}(t) = \sum_{k \neq j} p_{ik}(t)q_{kj} - p_{ij}(t)\nu_j,$$

or

$$P'(t) = P(t)Q$$

in matrix notation. (Note that the continuity of $p_{ik}(\cdot)$ allows a similar argument for the left-hand derivative.)

Example 5.4(a) (The Two-State Chain). Let $q_{01} = \lambda$ and $q_{10} = \mu$. Then $\nu_0 = \lambda$, $\nu_1 = \mu$, and

$$Q = \begin{pmatrix} -\lambda & \lambda \\ \mu & -\mu \end{pmatrix}.$$

Since the matrices are finite, the solution to $P'(t) = QP(t)$ is

$$P(t) = e^{Qt} \quad \text{(the multiplicative constant } = 1 \text{ since } P(0) = I).$$

This is intuitive from the following calculation:

$$P(t) = P(t/n)^n = \big(I + [P(t/n) - I]\big)^n = \big(I + Qt/n + o(1/n)\big)^n = e^{Qt}.$$

Exponentiation is calculated via diagonalization: If $Q = ADA^{-1}$, then

$$e^{Qt} = Ae^{Dt}A^{-1}.$$

Here, it is easily calculated that the eigenvalues of Q are 0 and $-(\lambda+\mu)$, with corresponding eigenvectors $\binom{1}{1}$ and $\binom{\lambda}{-\mu}$. Thus, we use

$$D := \begin{pmatrix} 0 & 0 \\ 0 & -(\lambda+\mu) \end{pmatrix}, \quad A := \begin{pmatrix} 1 & \lambda \\ 1 & -\mu \end{pmatrix}, \quad A^{-1} = \frac{1}{\lambda+\mu} \begin{pmatrix} \mu & \lambda \\ 1 & -1 \end{pmatrix},$$

so

$$\begin{aligned} e^{Qt} &= \frac{1}{\lambda+\mu} \begin{pmatrix} 1 & \lambda \\ 1 & -\mu \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & e^{-(\lambda+\mu)t} \end{pmatrix} \begin{pmatrix} \mu & \lambda \\ 1 & -1 \end{pmatrix} \\ &= \frac{1}{\lambda+\mu} \begin{pmatrix} \mu+\lambda e^{-(\lambda+\mu)t} & \lambda-\lambda e^{-(\lambda+\mu)t} \\ \mu-\mu e^{-(\lambda+\mu)t} & \lambda+\mu e^{-(\lambda+\mu)t} \end{pmatrix}. \end{aligned}$$

E.g., $p_{00}(t) = (\mu+\lambda e^{-(\lambda+\mu)t})/(\lambda+\mu)$.

Chapter 6

Martingales

§6.1. Martingales.

Recall

Theorem 3.3.2 (Wald's Equation). *Let X_n be random variables all with the same mean μ. Suppose that N is an $\mathbb{N}$-valued random variable such that $\forall n \ge 0\ \ \forall i \ge 1\ \ [N = n]$ is independent of X_{n+i}. If either*

(a) *all $X_n \ge 0$ or*

(b) *$E[N] < \infty$ and $\sup_n E|X_n| < \infty$,*

then

$$E\Big[\sum_{n=1}^{N} X_n\Big] = \mu \cdot E[N].$$

A modification of Wald's equation is:

Theorem (Extension of Wald's Equation). *Let X_k be random variables for $k \ge 1$, N an $\mathbb{N}$-valued random variable, $\mu \in \mathbb{R}$,*

(i) *$\forall k\ \ E[X_k \mathbf{1}_{[N \ge k]}] = \mu P[N \ge k]$, and*

(ii) *either*

  (a) *$\forall k\ \ X_k \ge 0$ or*

  (b) *$E\Big[\big|\sum_{k=1}^{N} X_k\big|\Big] < \infty$ and $\lim_{n\to\infty} E\Big[\sum_{k=1}^{n} X_k \mathbf{1}_{[N>n]}\Big] = 0$.*

Then

$$E\Big[\sum_{k=1}^{N} X_k\Big] = \begin{cases} \mu E[N] & \text{if } \mu \ne 0, \\ 0 & \text{if } \mu = 0. \end{cases}$$

Proof. The case (ii)(a) is as before, since this is what we really used in the proof of that case. However, we won't use it, so assume (ii)(b). Write $Z_n := \sum_{k=1}^{n} X_k$. By the first part of (ii)(b) and the LDCT, $E[Z_N] = \lim_{n\to\infty} E[Z_N \mathbf{1}_{[N \le n]}]$. Now

$$E\Big[Z_N \mathbf{1}_{[N \le n]}\Big] = E\Big[\sum_{k=1}^{n} X_k \mathbf{1}_{[k \le N \le n]}\Big] = \sum_{k=1}^{n} E\Big[X_k\big(\mathbf{1}_{[N \ge k]} - \mathbf{1}_{[N > n]}\big)\Big],$$

which, by (i),

$$= \sum_{k=1}^{n} \Big\{ \mu P[N \ge k] - E[X_k \mathbf{1}_{[N>n]}] \Big\} = \mu \sum_{k=1}^{n} P[N \ge k] - E[Z_n \mathbf{1}_{[N>n]}].$$

1" Consider separately the cases $\mu = 0$ and $\mu \ne 0$ and apply the second part of (ii)(b). ... ∎

Definition. We call $\langle Z_n\,;\ n \ge 0\rangle$ a ***martingale*** if
(i) $\forall n\ \ E\big[|Z_n|\big] < \infty$ and
(ii) $\forall n \ge 1\ \ E[Z_n \mid Z_0, Z_1, \ldots, Z_{n-1}] = Z_{n-1}$.
In particular, $E[Z_n]$ does not depend on n by the tower property (1.5.1).

▷ `Read pp. 296--297 in the book.`

We say that an $\mathbb{N} \cup \{\infty\}$-valued random variable N is a ***stopping time*** with respect to $\langle Z_n\,;\ n \ge 0\rangle$ if $\forall n\ \ \mathbf{1}_{[N=n]}$ is a function of $Z_0, \ldots, Z_n$. This is equivalent to the condition that $\forall n\ \ \mathbf{1}_{[N \le n]}$ is a function of $Z_0, \ldots, Z_n$.

Corollary (The Optional Stopping Theorem). *Let $\langle Z_n\,;\ n \ge 0\rangle$ be a martingale and N be an a.s. finite stopping time. If $E|Z_N| < \infty$ and $\lim_{n\to\infty} E[Z_n \mathbf{1}_{[N>n]}] = 0$, then*

$$E[Z_N] = E[Z_0].$$

Proof. Apply the extension of Wald's equation with $X_n := Z_n - Z_{n-1}$ for $n \ge 1$ and
↓ $\mu := 0$. Note that $\mathbf{1}_{[N \ge k]} = \mathbf{1} - \mathbf{1}_{[N \le k-1]}$ is a function of $Z_0, \ldots, Z_{k-1}$, so we can compute
2" ↑ $E[X_k \mid N \ge k]$ via the tower property.

∎

In fact, the same proof works for N that is a finite stopping time with respect to a stochastic process of the form $\langle (Z_n, W_n)\,;\ n \ge 0\rangle$: we simply use the tower property a bit more. Typically, this is used when Z_n is a function of W_n.

For example, the hypotheses of this corollary hold if N is a bounded stopping time,
↓ but not if you gamble on fair games until you are ahead. If $N \le A$ a.s., then $|Z_N| \le$
2" ↑ $\max_{n \le A} |Z_n| \le \sum_{n=0}^{A} |Z_n|$ and $Z_n \mathbf{1}_{[N>n]} = 0$ for large enough n.

↓ The hypotheses also hold if $Z_n \mathbf{1}_{[N \ge n]}$ is bounded uniformly in n. If $|Z_n \mathbf{1}_{[N \ge n]}| \le B$,
2" ↑ then $|Z_N| = |Z_n \mathbf{1}_{[N=n]}| \le B$ and $E[Z_n \mathbf{1}_{[N>n]}] \to 0$ by the BCT.

More generally, it suffices that there is some nonnegative random variable W with $E[W] < \infty$ such that for all $n \geq 0$, we have $|Z_{N \wedge n}| \leq W$. Use the LDCT and note that $Z_n \mathbf{1}_{[N>n]} = Z_{N \wedge n} \mathbf{1}_{[N>n]}$.

For some nice applications of the optional stopping theorem, note that if N is a
1" stopping time and $n \in \mathbb{N}$, then $N \wedge n$ is a (bounded) stopping time: ...

Example 6.2(a) and more of 3.5(a) (Computing Mean Time to Occurrence of a Pattern). Suppose Y_n are i.i.d. for $n \geq 1$, with values $0, 1, 2$ that have corresponding probabilities $\frac{1}{2}$, $\frac{1}{3}$, $\frac{1}{6}$. Let N be the first time we see the pattern 020. What is $E[N]$?

Suppose that at each time n, gambler number n begins betting the pattern will occur starting then. More precisely, at time n, gambler n pays us 1. If $Y_n \neq 0$, he gets nothing and quits. If $Y_n = 0$, then we pay him 2 (to be fair) and gambler n bets 2 on $[Y_{n+1} = 2]$. If $Y_{n+1} \neq 2$, then we keep the 2 of gambler n, but if $Y_{n+1} = 2$, then we pay him back $6 \times 2 = 12$ and gambler n bets 12 on $Y_{n+2} = 0$. If he loses, then we keep his 12. If he wins, we pay back $2 \times 12 = 24$. At this point, gambler n quits.

Let X_n be *our* net gain after seeing $Y_1, \ldots, Y_n$ and write $R_n := n - X_n$. At each time, one gambler pays us 1. Sometimes a gambler pays us some other amount to place a new bet, but at those times, the amount equals what we paid the gambler on the previous (winning) bet. Thus, $X_n = n - R_n$ and R_n is the total we have paid all the gamblers up through time n. If a gambler has lost a bet before time n, then we will have paid that gambler nothing (while we have still collected the initial 1 from that gambler). Thus, at time N, every gambler who started before play $N-2$ has lost, whereas we have paid 24 to player $N-2$, nothing to player $N-1$, and 2 to player N. Thus $R_N = 24 + 2 = 26$.
.5" ... Since $\langle X_n \rangle$ is a martingale (all bets being fair), it follows (since N is a stopping time) that $\forall n$

$$0 = E[X_{N \wedge n}] = E[N \wedge n] - E[R_{N \wedge n}].$$

Since $0 \leq R_{N \wedge n} \leq 26$, it follows that $E[N \wedge n]$ is bounded, whence $E[N] < \infty$ by the MCT. In particular, $N < \infty$ a.s., so we may deduce that $R_{N \wedge n} \to R_N = 26$ as $n \to \infty$ by
.5" the BCT. It follows that $E[N] = 26$. ...

Similarly, the mean time until HHTTHH if $P(\mathrm{H}) = p$ is equal to the payback at the
1" corresponding time N, i.e., $p^{-4}q^{-2} + p^{-2} + p^{-1}$, where $q := 1 - p$. ...

We now compute the chance that one pattern occurs before another, e.g., $A := \langle 0, 2, 0 \rangle$ before $B := \langle 1, 0, 0, 2 \rangle$ in the first game above. The key is the use of the following relations. Let N_A, N_B be the first time A, B occur, respectively, $M := N_A \wedge N_B$, and P_A be the probability that A occurs before B. Let $N_{A|B}$ be the number of trials after B occurs until A occurs, and define $N_{B|A}$ likewise. Then

$$\begin{aligned}
E[N_A] &= E[M + (N_A - M)] = E[M] + E[N_A - M \mid N_B < N_A](1 - P_A) \\
&= E[M] + (1 - P_A)E[N_{A|B}], \\
E[N_B] &= E[M] + P_A E[N_{B|A}].
\end{aligned}$$

1" . . . Solving for P_A and $E[M]$ gives

$$P_A = \frac{E[N_B] + E[N_{A|B}] - E[N_A]}{E[N_{B|A}] + E[N_{A|B}]}$$

and

$$E[M] = E[N_B] - P_A E[N_{B|A}].$$

In the case at hand, we already have $E[N_A] = 26$. Similar reasoning gives $E[N_B] = 72$. Clearly $E[N_{B|A}] = 72$ as well. To calculate $E[N_{A|B}]$, we could use the same scheme as before, with gamblers starting to bet at each trial hoping to get A, but now we simply assume that B occurs immediately, i.e., the first 4 trials are $1, 0, 0, 2$: what occurs after this is still a martingale, but X_4, our net gain after these 4 trials, no longer has expectation 0. However, this is a little confusing, so instead, we do the following. We have that
.5" $N_{A|B} = N_{A|\langle 0,2\rangle}$ and $N_A = N_{\langle 0,2\rangle} + N_{A|\langle 0,2\rangle}$, . . . whence $E\big[N_{A|B}\big] = E[N_A] - E\big[N_{\langle 0,2\rangle}\big] = 26 - 12 = 14$. Substitution into our formulas gives

$$P_A = \frac{30}{43} \text{ and } E[M] = \frac{936}{43}.$$

Example (Problem 6.11). Let $\langle Z_n\rangle$ be simple random walk on $\mathbb{Z}$ starting from $Z_0 = 0$; this means that $Z_n = \sum_{k=1}^n X_k$, where X_k are i.i.d. for $k \ge 1$ and equal ± 1 with equal probability. Let $\tau_a := \inf\{n;\ Z_n = a\}$. Set $\tau := \tau_a \wedge \tau_{-b}$ for $a, b > 0$. The process
2" $\big\langle (a - Z_n)(Z_n + b) + n\big\rangle$ is a martingale: . . . Fix $N \in \mathbb{N}$. Then we may apply the optional stopping theorem to get

$$ab = (a - Z_0)(Z_0 + b) + 0 = E\big[(a - Z_{\tau\wedge N})(Z_{\tau\wedge N} + b) + \tau \wedge N\big].$$

Since $Z_{\tau\wedge N}$ is bounded, it follows that so is $E[\tau \wedge N]$, whence $E[\tau] < \infty$ by the MCT.
1" This implies that $\tau < \infty$ a.s., whence letting $N \to \infty$ gives $E[\tau] = ab$. . . .
.5" Now $\langle Z_n\rangle$ is itself a martingale, whence . . .

$$0 = E[Z_\tau] = a\,P[\tau_a < \tau_{-b}] - b\,P[\tau_{-b} < \tau_a],$$

giving $P[\tau_a < \tau_{-b}] = b/(a+b)$.

If we let $b \to \infty$, we obtain that $\tau_a < \infty$ a.s. Thus, simple random walk is recurrent, i.e., it visits every integer a.s., whence it visits every integer infinitely many times a.s. In addition, the conclusion of Wald's equation Theorem 3.3.2 does *not* hold for the stopping time τ_a, whence $E[\tau_a] = \infty$.

Example 6.2(c) and Problem 6.12. Consider three players, A, B, and C, who play the following game: They begin with fortunes $a, b, c \in \mathbb{N}^+$, respectively. At each time, a pair of the players whose fortunes are strictly positive is chosen at random and a random one of that pair gives one unit to the other. Let X_n, Y_n, Z_n be their respective fortunes after n plays. What is the expected time until some player's fortune is zero? What is the expected time until two players' fortunes are zero (at which time the game ends)?

Solution. The previous example shows that with only two players, A and B, the process $\langle X_n Y_n + n \rangle$ is a martingale. It follows that $\langle X_n Y_n + Y_n Z_n + Z_n X_n + n \rangle$ is a martingale. Because $X_n Y_n + Y_n Z_n + Z_n X_n$ is bounded, we get as above that the expected time until two players' fortunes are zero equals $ab + bc + ca$.

For the first question, we get the idea of somehow using $X_n Y_n Z_n$. Indeed, calculating with it, we find that $\langle X_n Y_n Z_n + n(X_n + Y_n + Z_n)/3 \rangle$ is a martingale. Because $X_n Y_n Z_n$, again, is bounded, we obtain that the expected time until some player's fortune is zero equals $3abc/(a+b+c)$.

Example: Let $\langle Z_n \rangle$ be asymmetric random walk on $\mathbb{Z}$ starting from $Z_0 = 0$; this means that $Z_n = \sum_{k=1}^n X_k$, where X_k are i.i.d. for $k \ge 1$ and equal ± 1 with probabilities p and $q := 1 - p$, respectively. For which $s, r > 0$ is $\langle s^{Z_n} r^{-n} \rangle$ a martingale? Choose s and r to deduce that $P[\tau_a \ge n] \le (q/p)^{-a/2} (2\sqrt{pq})^n$ for $a \ge 1$ and $p > q$. Thus, $E[\tau_a] < \infty$ when $p > q$ (due to the inequality $(a+b)/2 \ge \sqrt{ab}$ with equality iff $a = b$). Make another choice of s and r to calculate $E[\tau_a \wedge \tau_{-b}]$ when $p \ne q$.

Solution. We get the condition $ps + qs^{-1} = r$. Note that $r \ge 2\sqrt{pq}$ with equality iff $s = \sqrt{q/p}$. We make that choice. Then

$$1 = E\left[s^{Z_{\tau_a \wedge n}} r^{-(\tau_a \wedge n)}\right] \ge E[s^a r^{-n} \,;\, \tau_a \ge n]$$

because $s \le 1$. Thus, $P[\tau_a \ge n] \le s^{-a} r^n$.

Next choose $r = 1$ and $s = q/p$ to obtain that $\langle (q/p)^{X_n} \rangle$ is a martingale. Write $\tau := \tau_a \wedge \tau_{-b}$. This martingale is bounded up to time τ, whence the optional stopping theorem applies:

$$1 = E\big[(q/p)^{X_0}\big] = E\big[(q/p)^{X_\tau}\big] = (q/p)^a \, P[\tau_a < \tau_{-b}] + (q/p)^{-b} \, P[\tau_{-b} < \tau_a].$$

5" ↑ We can solve this to derive the probabilities in gambler's ruin when $p \neq q$.

EXAMPLE 6.2(B). For asymmetric random walk, what is $E[\tau_a]$ when $p > q$, $a \geq 1$, and $Z_0 = 0$?

↓

Solution. We use another martingale: $\langle Z_n - (p-q)n \rangle$. Fix $N \in \mathbb{N}$ and define $\tau := \tau_a \wedge N$. Then τ is bounded, so $0 = E[Z_0 - (p-q)0] = E[Z_\tau - (p-q)\tau]$, i.e., $(p-q)E[\tau_a \wedge N] = E[Z_{\tau_a \wedge N}]$. By the MCT, $E[\tau_a \wedge N] \to E[\tau_a]$ as $N \to \infty$. Since $|Z_{\tau_a \wedge N}| \leq \tau_a \wedge N \leq \tau_a$ and $E[\tau_a] < \infty$ by the previous example (the book neglected to verify this), the LDCT gives $E[Z_{\tau_a \wedge N}] \to E[Z_{\tau_a}] = a$ as $N \to \infty$. In conclusion,

$$E[\tau_a] = \frac{a}{p-q}.$$

4" ↑ We could also have applied Wald's equation Theorem 3.3.2.

Chapter 7

Random Walks

§7.1. Duality in Random Walks.

If $\langle X_i \rangle$ are i.i.d. and $S_n := \sum_{i=1}^n X_i$ is the corresponding random walk, then a useful observation is the "duality" property that $\langle X_1, X_2, \ldots, X_n \rangle \overset{\mathscr{D}}{=} \langle X_n, X_{n-1}, \ldots, X_1 \rangle$. We give two applications. Let

$$\begin{aligned} N &:= \min\{n\,;\ S_n > 0\}, \\ M &:= \text{number of new minima of } \langle S_n \rangle \\ &= |\{n\,;\ \forall k \in [0,n) \quad S_n \le S_k\}|, \\ R_n &:= \text{number of distinct values of } \langle S_k\,;\ 0 \le k \le n \rangle \\ &= |\{S_0, S_1, \ldots, S_n\}| \\ &= |\{k \in [0,n]\,;\ \forall j \in [0,k) \quad S_k \ne S_j\}| \\ &= |\{k \in [0,n]\,;\ \forall j \in (k,n] \quad S_k \ne S_j\}|. \end{aligned}$$

R_n is called (the size of) the ***range*** of $\langle S_k\,;\ 0 \le k \le n \rangle$. If $E[X]$ exists and is positive, then by the SLLN, $S_n \to \infty$ a.s., whence $N < \infty$ a.s. and $M < \infty$ a.s. Always $R_n \to \infty$ a.s. (except if $X = 0$ a.s.).

Proposition 7.1.1. *If $\mu := E[X] > 0$, then $E[N] = E[M] < \infty$ and $E[S_N] = \mu E[N]$.*

Proposition 7.1.2. *Without any assumption on $E[X]$,*

$$\lim_{n\to\infty} \frac{E[R_n]}{n} = P[\textit{no return to } 0] = P[\forall n > 0\ \ S_n \ne 0].$$

Hence $\lim E[R_n]/n = 0 \iff$ the random walk is recurrent.

Proof of Proposition 7.1.1. We have

$$E[N] = \sum_{n=0}^{\infty} P[N > n] = \sum_{n \ge 0} P[\forall k \le n\ \ S_k \le 0].$$

1" By duality, this ...

$$= \sum_{n \ge 0} P[\forall k \le n \ \ S_n - S_{n-k} \le 0] = \sum_{n \ge 0} P[S_n \text{ is a new minimum}] = E[M].$$

Now the sequence of times at which minima occur form a renewal sequence with interarrival times allowed to be infinite. Indeed, there are only a finite number of renewals since $\mu > 0$; their number is a geometric random variable, so $E[M] < \infty$. Since this means that $E[N] < \infty$ also, we may apply Wald's equation (not its extension) to conclude when $\mu < \infty$. If $\mu = \infty$, then $E[S_N] \ge E[S_N \mathbf{1}_{[N=1]}] = E[X_1 \mathbf{1}_{[X_1 > 0]}] \ge E[X_1] = \infty$, so the result holds still. ∎

Proof of Proposition 7.1.2. We have

$$E[R_n] = \sum_{k=0}^{n} P[\forall j \in (k, n] \ \ S_k \ne S_j]$$

1" ...

$$= \sum_{k=0}^{n} P[\forall i \in (0, n-k] \ \ S_0 \ne S_i]$$

$$= \sum_{k=0}^{n} P[\forall i \in (0, k] \ \ S_i \ne 0].$$

Since the summands $\to P[\text{no return to } 0]$, the result follows. ∎

REMARK. For the proof of Proposition 7.1.2, we used stationarity and reversed counting, rather than duality (which the book uses). Thus, the result is more general: it applies to stationary $\langle X_i \rangle$. We also don't need the values of X to lie in $\mathbb{R}$. It is also true that $R_n/n \to P[\text{no return}]$ a.s. (Kesten, Spitzer, and Whitman; use the Kingman subadditive ergodic theorem to prove this).

EXAMPLE 7.1(A). Suppose that $X = \pm 1$ with probability $\frac{1}{2} \pm \alpha$. Then $P[\text{no return}] = 1 - f_{00} = 2|\alpha|$ (see the calculation of the Green function or the gambler's ruin problem).

1" ...

▷ `Read Proposition 7.1.3 in the book.`

Chapter 8

Brownian Motion and Other Markov Processes

§8.1. Introduction and Preliminaries.

The motion of a particle floating on water seems random. Model one coordinate $\langle X(t)\rangle$ as follows: $\langle X(t)\rangle$ is a stochastic process with independent, stationary increments and continuous paths. Because of momentum, the independence of increments is not a great assumption, but we will not study a better model.

For $0 \le s \le t$, define

$$M(s,t) := \max\big\{|X(u) - X(v)|\,;\ s \le u \le v \le t\big\}.$$

Now, because $\langle X(t)\rangle$ is uniformly continuous on $[0,1]$, we have

$$J_n := \max_{\substack{|s-t|=1/n \\ s,t\in[0,1]}} M(s,t) \to 0 \text{ as } n \to \infty,$$

.5" whence $\forall\delta > 0\quad P[J_n \ge \delta] \to 0$. ... `These events are decreasing to ∅, but the BCT would give this even if not.` In particular, if

$$H_n := \max_{1\le k\le n} M\big((k-1)/n, k/n\big),$$

then $\forall\delta > 0\quad P[H_n \ge \delta] \to 0$ since $H_n \le J_n$. Now

$$\begin{aligned} P[H_n \ge \delta] = 1 - P[H_n < \delta] &= 1 - \prod_{k=1}^{n} P\big[M\big((k-1)/n, k/n\big) < \delta\big] \\ &= 1 - P\big[M(0,1/n) < \delta\big]^n \\ &= 1 - \Big[1 - P\big[M(0,1/n) \ge \delta\big]\Big]^n \\ &\ge 1 - e^{-nP[M(0,1/n)\ge\delta]} \end{aligned}$$

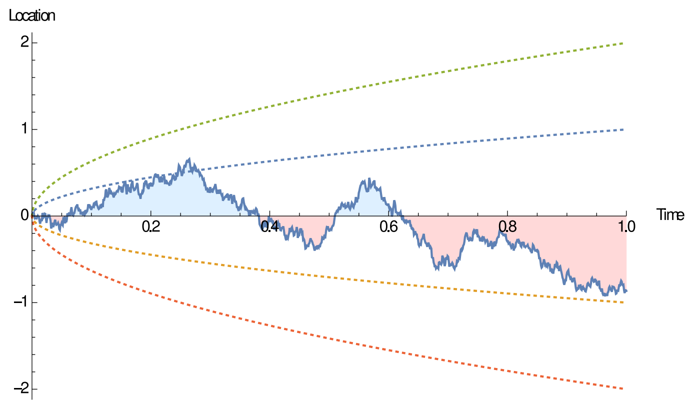
1" since $1 - x \le e^{-x}$ for all $x \in \mathbb{R}$. Therefore, $nP\big[M(0, 1/n) \ge \delta\big] \to 0$. ... By considering the largest n for which $h \le 1/n$, it follows that

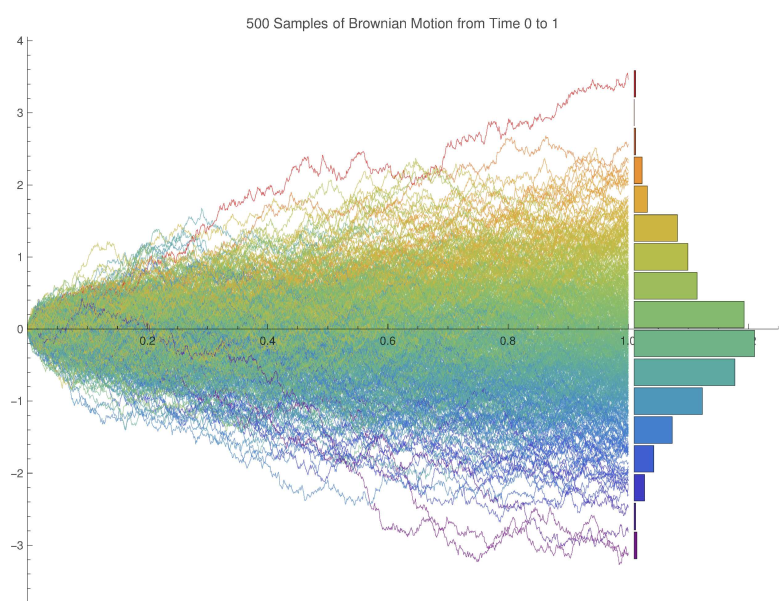
$$\forall \delta > 0 \quad \lim_{h \to 0^+} \frac{P\big[M(0,h) \ge \delta\big]}{h} = 0. \tag{N1}$$

1" `... Compare to a Poisson process; consider` $\delta \le 1$ `or` $\delta > 1$`.` Using this and the CLT, one can show (see Breiman's *Probability*, Proposition 12.4) that

$$\exists \mu \in \mathbb{R} \ \exists \sigma \ge 0 \ \forall t \ge 0 \quad X(t) - X(0) \sim \mathrm{N}(\mu t, \sigma^2 t). \tag{N2}$$

`Demos of B.M.`





We call such a process a ***Brownian motion*** (B.M.) if $X(0)$ is independent of $\langle X(t) - X(0)\,;\ t > 0\rangle$. If $X(0) \equiv 0$, $\mu = 0$, and $\sigma = 1$, it is a ***standard Brownian motion*** or, simply, ***Brownian motion***. In general, μ is called the ***drift*** and σ^2 is the ***variance parameter***. We always assume that $\sigma \neq 0$. Note that

$$\frac{X(t) - X(0) - \mu t}{\sigma}$$

is a standard Brownian motion and, if $B(t)$ is a standard Brownian motion, then $X(t) := a + \mu t + \sigma B(t)$ is a Brownian motion starting at a with drift μ and variance parameter σ^2.
1" … Also, if X is a Brownian motion, then $-X$ is a Brownian motion. By the independent increments property, a Brownian motion is a ***Markov process***, i.e., the future and the past are independent given the present: for all $t > 0$, the two collections of random variables $\langle X(s)\,;\ s > t\rangle$ and $\langle X(s)\,;\ s < t\rangle$ are independent given $X(t)$.

A stochastic process with independent, stationary increments satisfying (N2) need
1" not be continuous, even a.s., despite the book's claim (p. 358). … Adding (N1) to (N2) does imply a.s. continuity. However, given a process $X(\cdot)$ with independent, stationary increments satisfying (N2), there is a Brownian motion $\widetilde{X}(\cdot)$ such that $\forall t\ \ P\big[X(t) = \widetilde{X}(t)\big] = 1$. This is not easy to show. A strengthening follows from its proof:

THEOREM. *Let $X(\cdot)$ be a stochastic process on an interval $I \subseteq [0, \infty)$ that is a.s. continuous and has the same finite-dimensional marginals as Brownian motion does on I. Then $X(\cdot)$ is the restriction to I of a Brownian motion.*

Why should we believe that a Brownian motion exists? We can get it as a limit of random walks that take small steps very quickly: let Y_n be i.i.d. ± 1 steps with $E[Y_n] = 0$, $\Delta x > 0$, and $\Delta t > 0$. If we want a step of size Δx to take place in time Δt, we can let

$$D(t) := \sum_{k=1}^{\lfloor t/\Delta t\rfloor} \Delta x \cdot Y_k.$$

How should Δx and Δt be related? We have

$$\operatorname{Var} D(t) = \lfloor t/\Delta t\rfloor \cdot (\Delta x)^2 \cdot 1,$$

so if $D(t)$ converges to standard Brownian motion $X(t)$, we should have $(\Delta x)^2/\Delta t$ converging to 1. So take $\Delta t := 1/n$, $\Delta x = 1/\sqrt{n}$, and let $D_n(t)$ be the corresponding process. By the CLT, $D_n(t) \Rightarrow \mathrm{N}(0, t)$ for each t. In fact, the finite dimensional marginals of $D_n(t)$
2" converge to those of standard Brownian motion. … This makes quite plausible the existence of Brownian motion and shows how its study extends the study of sums of i.i.d. random variables. Note that there was little special about Y_n being ± 1.

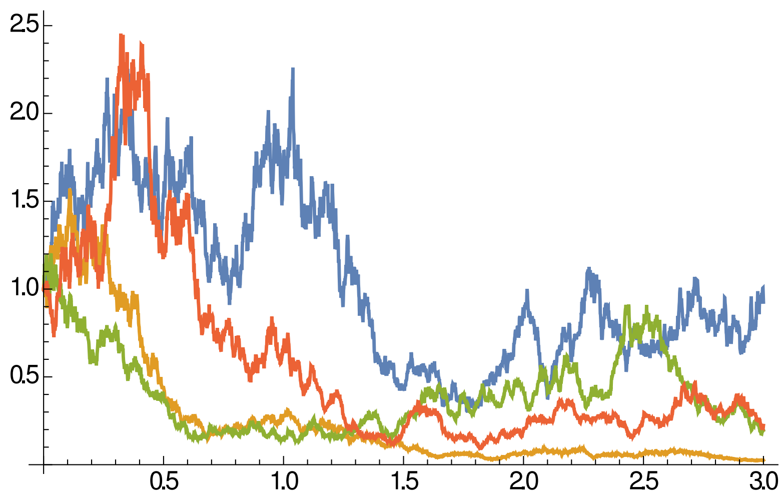
`Demo of random walk.`

§8.3.3. Geometric Brownian Motion.

To model stock prices, say, one needs a stochastic process that is ≥ 0. Furthermore, it might be reasonable that % changes over time intervals of the same length are i.i.d., that is, that the distribution of $[X(t+\Delta t)-X(t)]/X(t)$ should depend on Δt but not on t, and that these quotients should be independent when disjoint intervals $[t, t+\Delta t]$ are considered. This is the same as having $X(t+\Delta t)/X(t)$ be i.i.d., or of having $\log X(t+\Delta t) - \log X(t)$ be i.i.d. In other words, if prices are continuous, then this is the same as $Y(t) := \log X(t)$ being a Brownian motion, i.e.,

$$X(t) = e^{Y(t)}, \qquad Y \text{ a Brownian motion.}$$

Such an X is called a ***geometric Brownian motion***. Actually, for such modeling, X should have units of currency, whereas Y cannot have units since it is exponentiated. In the quotient $X(t+\Delta t)/X(t)$, such units cancel. Thus, it makes more sense to use $X(t) = X(0)e^{Y(t)}$ with $Y(0) \equiv 0$ and $X(0)$ having currency units.



Four samples of a geometric Brownian motion.

Recall that the m.g.f. of a normal random variable W is

$$E\big[e^{aW}\big] = e^{aE[W]+a^2\operatorname{Var}(W)/2}.$$

Thus, for X as above with Y having drift μ and variance parameter σ^2, and for $s < t$,

$$E\big[X(t)/X(s)\big] = E\big[e^{Y(t)-Y(s)}\big] = e^{\mu(t-s)+\sigma^2(t-s)/2}.$$

Note that

$$\begin{aligned}E\big[X(t) \mid \langle X(u)\,;\ 0 \le u \le s\rangle\big] &= X(s)E\Big[\frac{X(t)}{X(s)} \;\Big|\; \langle X(u)\,;\ 0 \le u \le s\rangle\Big] \\ &= X(s)E\Big[\frac{X(t)}{X(s)}\Big] = X(s)e^{\mu(t-s)+\sigma^2(t-s)/2}.\end{aligned}$$

Therefore,

$$E\big[e^{-\alpha t}X(t) \mid \langle X(u)\,;\ 0 \le u \le s\rangle\big] = e^{-\alpha s}X(s) \tag{N3}$$

when

$$\alpha = \mu + \sigma^2/2. \tag{N4}$$

Later, we will see that this means that $\langle e^{-\alpha t}X(t)\rangle$ is a continuous-time martingale. We call α the ***drift parameter*** of X and σ^2 the ***variance parameter*** of X. Note that μt and $\sigma^2 t$ are unitless, whence so is αt.

Note that if α is the (continuously compounded, risk-free, constant) interest rate, then the "theoretical" future price of the stock discounted to present value should be a martingale: if the $=$ in (N3) didn't hold, either no one would buy, so the price would fall, or there would be an infinite demand, so the price would rise. Thus, if geometric Brownian motion is to be a good model for stock prices and these other ideal assumptions hold, then we would certainly need (N3) and (N4), where α is the interest rate.

Now various kinds of options are also available on the stock. For example, for cost c, you can purchase a "call" option that gives you the right (not obligation) to buy a share of the stock at a fixed time T for a fixed price K. What should c be as a function of T and K? At time T, this option is worth $\big(X(T)-K\big)^+$, so now, at time 0, it is worth $e^{-\alpha T}\big(X(T)-K\big)^+$. This means that we should have the theoretical price

$$c = E\Big[e^{-\alpha T}\big(X(T)-K\big)^+\Big].$$

When computed, this gives the ***Black–Scholes formula***. Briefly, this goes as follows: Denote $\kappa := K/X(0)$, which is unitless. Since $Y(T)-Y(0) \sim \mathrm{N}(\mu T, \sigma^2 T)$, we have

$$\begin{aligned}\frac{c}{X(0)} &= E\Big[e^{-\alpha T}\big(e^{Y(T)-Y(0)}-\kappa\big)^+\Big] \\ &= e^{-\alpha T}\int_{-\infty}^{\infty}\big(e^y-\kappa\big)^+\frac{1}{\sqrt{2\pi\sigma^2 T}}e^{-(y-\mu T)^2/(2\sigma^2 T)}\,dy \\ &= \frac{e^{-\alpha T}}{\sqrt{2\pi\sigma^2 T}}\int_{\log\kappa}^{\infty}\big(e^y-\kappa\big)e^{-(y-\mu T)^2/(2\sigma^2 T)}\,dy \\ &= \Phi(\sigma\sqrt{T}+b) - \kappa e^{-\alpha T}\Phi(b),\end{aligned}$$

where

$$b := \frac{\alpha T - \sigma^2 T/2 - \log \kappa}{\sigma \sqrt{T}} \qquad \text{and} \quad \Phi := \text{c.d.f. of N}(0,1).$$

This uses (N4) to eliminate μ in favor of α and σ^2. Note that if we look at the currency unit, then we see that $c/X(0)$ is a function of the unitless κ, as it ought to be.

§8.2. Hitting Times, Maximum Value, and Arc Sine Laws.

How long does it take (standard) Brownian motion to hit $a \neq 0$? By symmetry, we may consider only $a > 0$. Let $T_a := \inf\{t \geq 0\,;\ X(t) = a\}$ be the hitting time of a. Note that

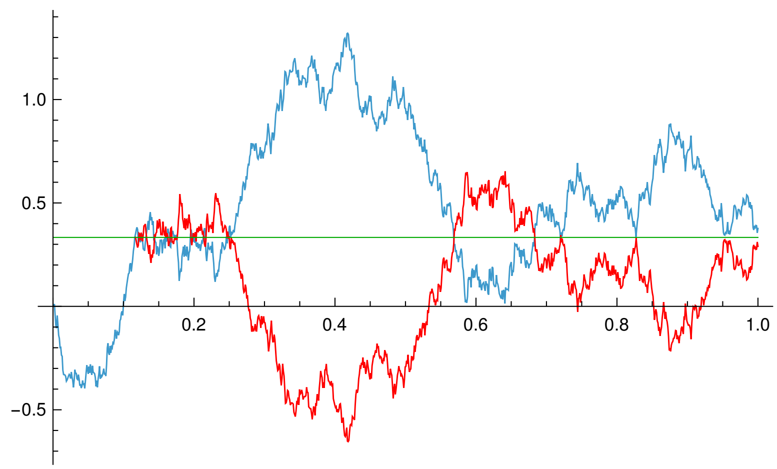


Illustration of reflection of Brownian motion at level 1/3.

$$P\big[X(t) > a \,\big|\, T_a < t\big] = \frac{1}{2}$$

2" . . . by symmetry and the strong Markov property.

This property was stated in Chap. 2: Suppose τ is a random variable with values in $[0, \infty)$ such that for all s, the event that $\tau \leq s$ depends only on $X(t)$ for $t \leq s$. Then if an event A is defined in terms of $X(t)$ for $t \leq \tau$ and another event B is defined in terms of $X(\tau + t) - X(\tau)$ for $t \geq 0$, then A and B are independent. Furthermore, the law of $\langle X(\tau + t) - X(\tau)\,;\ t \geq 0\rangle$ is the same as the law of $\langle X(t)\,;\ t \geq 0\rangle$.

This is called the ***reflection principle***. Note that $P\big[X(t) = a\big] = 0$ and $[T_a = t] \subseteq [X(t) = a]$, so also $P[T_a = t] = 0$. Therefore, whether we have strict inequalities or not in the reflection principle makes no difference. Since $X(t) \geq a \Rightarrow T_a \leq t$, the reflection principle is the same as

$$P[T_a \leq t] = 2\,P\big[X(t) \geq a\big] = 2\,P\big[X(t)/\sqrt{t} \geq a/\sqrt{t}\,\big] = \sqrt{\frac{2}{\pi}} \int_{a/\sqrt{t}}^{\infty} e^{-y^2/2}\, dy. \qquad \text{(N5)}$$

1" ... In particular, $P[T_a < \infty] = 1$, ... so Brownian motion is ***recurrent***: a.s., it visits every real number. (In the random walk demo, if we regard the scaling for the random walk to time n not as $\Delta t = 1/n$ but $\Delta t = 1/\sqrt{n}$ (and so $\Delta x = 1/n^{1/4}$), then we see an approximation to B.M. up to time $\sqrt{n}$.) Is it ***positive recurrent***, by which we mean: is $E[T_a] < \infty$? Note that

$$P[T_a > t] = \sqrt{\frac{2}{\pi}} \int_0^{a/\sqrt{t}} e^{-y^2/2}\,dy \sim \sqrt{\frac{2}{\pi}}\frac{a}{\sqrt{t}} \quad \text{as } t \to \infty,$$

whence $E[T_a] = \int_0^\infty P[T_a > t]\,dt = \infty$. Thus, Brownian motion is ***null recurrent***.

This could also have been derived from the null recurrence of simple random walk on $\mathbb{Z}$: Let $\tau_0 := \mathbf{0}$, $\tau_1 := T_1 \wedge T_{-1}$, and, in general, let τ_{n+1} be the first time t after τ_n that $X(t)$ equals $X(\tau_n) \pm 1$. Then $\langle X(\tau_n) \rangle_n$ is simple random walk with $E[\tau_{n+1} - \tau_n] = E[\tau_1] = 1$ (to be proved later). Let $N := \min\{n\,;\ X(\tau_n) = 1\}$. Then

$$T_1 = \sum_{n=1}^{N} (\tau_n - \tau_{n-1}),$$

1" so Wald's equation gives $E[T_1] = E[N] = \infty$. ...

Note that we easily get the distribution of another random variable: for $a > 0$,

$$P\Big[\max_{0 \le s \le t} X(s) < a\Big] = P\big[T_a > t\big] = \sqrt{\frac{2}{\pi}} \int_0^{a/\sqrt{t}} e^{-y^2/2}\,dy.$$

§8.3.1. Brownian Motion Absorbed at a Value.

Define ***Brownian motion absorbed at*** a by

$$Z(t) := \begin{cases} X(t), & \text{if } t < T_a, \\ a, & \text{if } t \ge T_a. \end{cases}$$

If $a > 0$, then

$$P\big[Z(t) = a\big] = P[T_a \le t] = \sqrt{\frac{2}{\pi}} \int_{a/\sqrt{t}}^{\infty} e^{-y^2/2}\,dy.$$

What is more interesting is the rest of the distribution of $Z(t)$: for $x < a$,

$$P\big[Z(t) \le x\big] = \frac{1}{\sqrt{2\pi t}} \int_{x-2a}^{x} e^{-y^2/(2t)}\,dy.$$

This is shown by clever use of the reflection principle: read §8.3.1.

§8.3.2. Brownian Motion Reflected at the Origin.

How do we reflect Brownian motion at, say, 0? We simply define

$$Z(t) := |X(t)|.$$

The c.d.f. of $Z(t)$ is easily computed: for $y \geq 0$, we have

$$P\big[Z(t) \leq y\big] = P\big[X(t) \leq y\big] - P\big[X(t) < -y\big] = \Phi(y/\sqrt{t}\,) - \Phi(-y/\sqrt{t}\,) = 2\Phi(y/\sqrt{t}\,) - 1,$$

1" . . . where Φ is the c.d.f. of $\mathrm{N}(0,1)$.

If we wished to reflect at $a > 0$, we would define

$$Z(t) := \begin{cases} X(t) & \text{if } X(t) \leq a, \\ 2a - X(t) & \text{if } X(t) \geq a. \end{cases}$$

§8.4. Brownian Motion with Drift.

If $\langle X(t)\,;\ t \geq 0\rangle$ is a Markov process, we write P_x for probability and E_x for expectation when the process starts at x, i.e., $X(0) = x$. If f is a real-valued function on the state space for which

$$(Lf)(x) := \lim_{t \to 0^+} E_x\Big[\frac{f\big(X(t)\big) - f(x)}{t}\Big]$$

exists for every initial state x, then we write $f \in \mathscr{D}_L$. More generally, $\mathscr{D}_L(x)$ denotes the set of functions f for which $(Lf)(x)$ exists. Thus, $\mathscr{D}_L = \bigcap_x \mathscr{D}_L(x)$. The functional L defined on $\mathscr{D}_L$ is called the ***infinitesimal generator*** of the process.

For Brownian motion, the above notation is interpreted as $X(t) = B(t) + x$, where $B(\cdot)$ is a Brownian motion with the same drift and variance parameter as X, but $B(0) \equiv 0$.

THEOREM. *If $\langle X(t)\rangle$ is a Brownian motion, then $\mathscr{D}_L \supseteq C^2_{\mathrm{b}}(\mathbb{R})$ (the space of bounded functions on $\mathbb{R}$ with a continuous second derivative) and*

$$Lf = \mu f' + \frac{\sigma^2}{2} f'' \quad \textit{for } f \in C^2_{\mathrm{b}}(\mathbb{R}).$$

More generally, if f is bounded on $\mathbb{R}$ and has a continuous second derivative in a neighborhood of x, then $f \in \mathscr{D}_L(x)$ and

$$(Lf)(x) = \mu f'(x) + \frac{\sigma^2}{2} f''(x).$$

Proof. Fix $x \in \mathbb{R}$. If f has a continuous second derivative near x, then

$$f(y) = f(x) + f'(x)[y-x] + \frac{1}{2} f''(x)[y-x]^2 + o\big([y-x]^2\big). \tag{N6}$$

Therefore

$$E_x\big[f\big(X(t)\big) - f(x)\big] = f'(x)\mu t + f''(x)\frac{\sigma^2 t + (\mu t)^2}{2} + o(t). \tag{N7}$$

Interchanging E_x and $o(\cdot)$ to derive (N7) from (N6) requires justification. We will merely describe how to do this: Let f_0 be a function in $C^2_{\mathrm{b}}(\mathbb{R})$ that has a bounded second derivative and equals f near x. The chance that $X(t)$ is not near x is $o(t)$ by (N1), so we may replace f by f_0. Now use $f_0(y) = f(x) + f'(x)[y-x] + (1/2)f''(x)[y-x]^2 + g(y)[y-x]^2$ in place of (N6), where $|g(y)| \le \max\big\{|f_0''(z) - f''(x)|/2\,;\ |z-x| \le |y-x|\big\}$. Then use the LDCT to get (N7).

Thus,

$$E_x\Big[\frac{f\big(X(t)\big) - f(x)}{t}\Big] = f'(x)\mu + f''(x)\frac{\sigma^2 + \mu^2 t}{2} + o(1). \qquad ∎$$

To apply this result, let $a, b > 0$ and let

$$f(x) := P_x[T_a < T_{-b}].$$

We claim that $Lf = 0$ on $(-b, a)$, i.e., that $E_x\big[f\big(X(t)\big)\big] = f(x) + o(t)$ as $t \to 0$ for
.5" $-b < x < a$. ... Note first that the Markov property implies that

$$\Big|f\big(X(t)\big) - P_x\big[T_a < T_{-b} \mid X(s)\ (s \le t)\big]\Big| \le \mathbf{1}_{[T_a \wedge T_{-b} < t]}.$$

1" ... Therefore,

$$\begin{aligned}
\Big|E_x\big[f\big(X(t)\big)\big] - f(x)\Big| &= \Big|E_x\big[f\big(X(t)\big)\big] - E_x\big[P_x[T_a < T_{-b} \mid X(s)\ (s \le t)]\big]\Big| \\
&\le E_x\Big[\Big|f\big(X(t)\big) - P_x\big[T_a < T_{-b} \mid X(s)\ (s \le t)\big]\Big|\Big] \\
&\le E_x\big[\mathbf{1}_{[T_a \wedge T_{-b} < t]}\big] = P_x[T_a \wedge T_{-b} < t].
\end{aligned}$$

Now

$$P_x[T_a \wedge T_{-b} < t] \le P_x\Big[\max_{0 \le s \le t} |X(s) - x| \ge |x-a| \wedge |x+b|\Big] = o(t)$$

by (N1). Therefore $f \in \mathscr{D}_L$ and $Lf = \mathbf{0}$, as claimed. We will not prove that $f \in C^2(-b, a) \cap C[-b, a]$, which is true, but assume it. Consider first the case where $\mu = 0$. Then $f'' = 0$, so f is linear. Since $f(a) = 1$ and $f(-b) = 0$, we get

$$f(x) = \frac{x+b}{a+b} \qquad (-b \le x \le a)\,;$$

in particular,

$$f(0) = b/(a+b).$$

`Recall resistances, simple random walk.`

Now let $\mu \neq 0$. The equation $Lf = \mathbf{0}$ is

$$\mu f'(x) + \frac{\sigma^2}{2} f''(x) = 0.$$

Integration gives

$$\mu f(x) + \frac{\sigma^2}{2} f'(x) = C,$$

whence

$$\frac{d}{dx}\left\{\frac{\sigma^2}{2} e^{2\mu x/\sigma^2} f(x)\right\} = C e^{2\mu x/\sigma^2},$$

1" ... so that

$$f(x) = C_1 + C_2 e^{-2\mu x/\sigma^2}.$$

1" ... [We could also have used separation of variables to solve the differential equation.] Since $f(a) = 1$ and $f(-b) = 0$, we get

$$f(x) = \frac{e^{2\mu b/\sigma^2} - e^{-2\mu x/\sigma^2}}{e^{2\mu b/\sigma^2} - e^{-2\mu a/\sigma^2}}.$$

In particular,

$$f(0) = \frac{e^{2\mu b/\sigma^2} - 1}{e^{2\mu b/\sigma^2} - e^{-2\mu a/\sigma^2}}.$$

This corresponds to a conductivity at x of $e^{2\mu x/\sigma^2}$, since for a variable conductivity $C(x)$, we have

$$\begin{aligned} f(x) &= \frac{\mathcal{C}(x \leftrightarrow a)}{\mathcal{C}(x \leftrightarrow a) + \mathcal{C}(x \leftrightarrow -b)} = \frac{\mathcal{R}(x \leftrightarrow -b)}{\mathcal{R}(-b \leftrightarrow a)} \\ &= \int_{-b}^{x} C(s)^{-1} ds \Big/ \int_{-b}^{a} C(s)^{-1}\, ds \\ &= \frac{-e^{2\mu s/\sigma^2}\big|_{-b}^{x}}{-e^{-2\mu s/\sigma^2}\big|_{-b}^{a}}. \end{aligned}$$

Note that $T_{-b_1} \le T_{-b_2}$ when $X(0) = 0$ and $0 < b_1 < b_2$, (we have equality when $T_{-b_1} = \infty$) so that $[T_a < T_{-b_1}] \subseteq [T_a < T_{-b_2}]$. Also, $\bigcup_{b>0}[T_a < T_{-b}] = [T_a < \infty]$ (we
1" allow $T_{-b} = \infty$). ... Thus, letting $b \to +\infty$ gives

$$P_0[T_a < \infty] = P_0\Big[\max_{t\ge 0} X(t) \ge a\Big] = \begin{cases} 1 & \text{if } \mu \ge 0, \\ e^{2\mu a/\sigma^2} & \text{if } \mu < 0. \end{cases}$$

1" ... Thus, if $\mu < 0$, we find that $\max_{t\ge 0} X(t) \sim \mathrm{Exp}(-2\mu/\sigma^2)$. That $\max_{t\ge 0} X(t)$ has an
2" exponential distribution follows also from the strong Markov property. ...

By symmetry, it also follows that $P_0[T_{-b} < \infty] = 1$. Hence a Brownian motion with negative drift visits every negative real number with probability 1. We claim that, in fact, $\lim_{t\to\infty} X(t) = -\infty$ a.s. To see this, let A be the event that the set of times t where $X(t) = 0$ is unbounded. Then $A \subseteq [\exists t > T_{-n}\ \ X(t) = 0] \cup [T_{-n} = \infty]$ for each $n \in \mathbb{N}^+$. That last set has P_0-probability 0 by what we just proved. On the other hand,
.5" $P_{-n}[T_0 < \infty] = e^{2\mu n/\sigma^2}$. ... Thus, $P_0(A) = 0$, so 0 is a.s. not visited after some random, finite time. Similarly, let A_k be the event that the set of times t where $X(t) = -k$ is unbounded. Then $A_k \subseteq [\exists t > T_{-k-n}\ \ X(t) = -k] \cup [T_{-k-n} = \infty]$ for each $n \in \mathbb{N}^+$. Again, this gives us that

$$\begin{aligned} P_0(A_k) &\le P_0[\exists t > T_{-k-n}\ \ X(t) = -k] = E_0\big[P_0[\exists t > T_{-k-n}\ \ X(t) = -k \mid T_{-k-n}]\big] \\ &= E_0\big[P_{-k-n}[\exists t > 0\ \ X(t) = -k]\big] = e^{2\mu n/\sigma^2} \end{aligned}$$

by the strong Markov property, whence $P_0(A_k) = 0$. Therefore, $P_0\big(\bigcap_k A_k^{\mathrm{c}}\big) = 1$, which combines with $P_0\big[\liminf_t X(t) = -\infty\big] = 1$ to yield $P_0\big[\lim_t X(t) = -\infty\big] = 1$.

Here's an interesting game: Brownian motion $X(t)$ with parameter (μ, σ^2), $\mu < 0$, $X(0) \equiv 0$, is run for all $t \ge 0$ "quickly"; e.g., we may observe $X\big(s/(1-s)\big)$ for $0 \le s < 1$. You may stop it at any time depending on only the values of X before that time (e.g., you are not allowed to know $\max_{t\ge 0} X(t)$); if you stop it at time t, then you collect $X(t)$. You are not required to stop it. How much should you pay to play?

Suppose that your rule is to fix x and stop if and when $X(t) = x$. Then your expected gain would be $x \cdot P\big[\max_{t\ge 0} X(t) \ge x\big] = x e^{2\mu x/\sigma^2}$, which is maximal at $x = -\sigma^2/(2\mu)$, so the value is $-\sigma^2/(2\mu e)$. The clairvoyant value of this game is $-\sigma^2/(2\mu)$, so we lose only a factor of e. Best possible strategy, as we'll see later using martingales.

§8.1. Introduction and Preliminaries (again).

For some additional topics, we need to study the multivariate normal distribution, Example 1.4(b), since $\langle X(t_1), \ldots, X(t_n) \rangle$ has this distribution. Recall again that the m.g.f. of $\mathrm{N}(\mu, \sigma^2)$ is $t \mapsto e^{\mu t + \sigma^2 t^2/2}$. If $X_i \sim \mathrm{N}(\mu_i, \sigma_i^2)$ $(i = 1, 2)$ are independent, then the m.g.f. of $X_1 + X_2$ is

$$t \mapsto E\big[e^{t(X_1+X_2)}\big] = E\big[e^{tX_1}\big]E\big[e^{tX_2}\big] = \exp\big\{(\mu_1 + \mu_2)t + (\sigma_1^2 + \sigma_2^2)t^2/2\big\},$$

whence, by uniqueness of the m.g.f. (which requires some assumptions—it suffices that the m.g.f. exists on the entire real line—and which we did not prove), $X_1 + X_2 \sim \mathrm{N}(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$.

Define the ***joint m.g.f.*** of random variables $X_1, \ldots, X_n$ by

$$(t_1, \ldots, t_n) \mapsto E\big[e^{\sum_{i=1}^n t_i X_i}\big].$$

This uniquely determines the joint distribution of $\langle X_1, \ldots, X_n \rangle$ when it is finite for all $(t_1, \ldots, t_n)$. The notation is simpler if we use vectors: $\mathbf{t} := (t_1, \ldots, t_n)$, $\mathbf{X} := (X_1, \ldots, X_n)$, $\mathbf{t} \mapsto E[e^{\mathbf{t}\cdot\mathbf{X}}]$.

Now let $Z_1, \ldots, Z_n$ be independent normal random variables, $\mu_i, a_{ij} \in \mathbb{R}$ $(1 \le i \le m$, $1 \le j \le n)$, and

$$X_i = \sum_{j=1}^n a_{ij} Z_j + \mu_i \quad (1 \le i \le m).$$

Then we say that $\langle X_1, \ldots, X_m \rangle$ has a ***multivariate normal distribution***, or that they are ***jointly normal random variables***. Note that by the preceding result, each X_i is normal. In addition, it follows from the definition that random variables that are linear combinations of the coordinates of a multivariate normal random vector also have a multivariate normal distribution.

What is the joint m.g.f. $\mathbf{t} \mapsto E[e^{\mathbf{t}\cdot\mathbf{X}}]$ of a multivariate normal random variable, $\mathbf{X}$? Since $\mathbf{t} \cdot \mathbf{X}$ is normal for a given $\mathbf{t}$, we need merely find $E[\mathbf{t} \cdot \mathbf{X}]$ and $\mathrm{Var}(\mathbf{t} \cdot \mathbf{X})$ in order to determine $E[e^{\mathbf{t}\cdot\mathbf{X}}]$. These functions of $\mathbf{t}$, in turn, are easily specified via $E[X_i]$ and $\mathrm{Cov}(X_i, X_j)$. We conclude that the joint distribution of $\langle X_1, \ldots, X_m \rangle$ is uniquely determined by their individual expectations and their pairwise covariances.

Let's be more explicit. We have

$$E[e^{\mathbf{t}\cdot\mathbf{X}}] = \exp\big\{E[\mathbf{t} \cdot \mathbf{X}] + \mathrm{Var}(\mathbf{t} \cdot \mathbf{X})/2\big\} = \exp\Big\{\sum_i t_i E[X_i] + \sum_{i,j} t_i t_j \mathrm{Cov}(X_i, X_j)/2\Big\}.$$

Now consider the case that for some index subset I, we have $\mathrm{Cov}(X_k, X_\ell) = 0$ whenever $k \in I$ and $\ell \notin I$. Then the above can be factored as

$$E[e^{\mathbf{t}\cdot\mathbf{X}}] = \exp\Big\{\sum_{i\in I} t_i E[X_i] + \sum_{i,j\in I} t_i t_j \mathrm{Cov}(X_i, X_j)/2\Big\}\times \\ \exp\Big\{\sum_{i\notin I} t_i E[X_i] + \sum_{i,j\notin I} t_i t_j \mathrm{Cov}(X_i, X_j)/2\Big\}.$$

1" ... Since the m.g.f. determines the joint distribution, it follows that $\langle X_k\,;\ k \in I\rangle$ is
1" independent of $\langle X_\ell\,;\ \ell \notin I\rangle$. ... In other words, for jointly normal random variables, pairwise vanishing covariances is the same as pairwise independence and is the same as mutual independence.

Definition. A ***Gaussian process*** $\langle X(t)\,;\ t \geq 0\rangle$ is a stochastic process such that $\forall t_1, \ldots, t_n \quad \langle X(t_1), \ldots, X(t_n)\rangle$ has a multivariate normal distribution.

By the preceding, we see that a Gaussian process has all its finite-dimensional marginals uniquely determined by the means $E\big[X(t)\big]$ and the covariances $\mathrm{Cov}\big(X(s), X(t)\big)$. This determines the process as a whole in the usual sense of distribution; continuous sample paths are a different matter.

For example, every Brownian motion X with $X(0)$ being constant is a Gaussian
1" process. ... The means are $E\big[X(t)\big] = X(0) + \mu t$. If $s \leq t$, then

$$\begin{aligned}\mathrm{Cov}\big(X(s), X(t)\big) &= \mathrm{Cov}\big(X(s), X(t) - X(s) + X(s)\big) \\ &= \mathrm{Var}\,\big(X(s)\big) = \sigma^2 s\,.\end{aligned}$$

Thus, for general s and t, we have $\mathrm{Cov}\big(X(s), X(t)\big) = \sigma^2(s \wedge t)$.

Suppose that a Gaussian process X with $X(0) \equiv 0$ is at A at time t. How did it get there? I.e., what is the distribution of the process $\langle X(s)\rangle_{0\leq s\leq t}$ given $X(t) = A$? Let $0 < s < t$. Forget the conditioning for the moment. Write $Q(s,t) := \mathrm{Cov}\big(X(s), X(t)\big)$ and

$$Y(s) := X(s) - \frac{Q(s,t)}{Q(t,t)}X(t).$$

Write this as

$$X(s) = Y(s) + \frac{Q(s,t)}{Q(t,t)}X(t)\,;$$

1" the first part, $Y(s)$, has covariance 0 with the second part, ... whence the two parts are independent. That is, the distribution of $X(s)$ given $X(t) = A$ is the same as the
1" unconditional distribution of $Y(s) + \frac{Q(s,t)}{Q(t,t)}A$. ... Because pairwise independence implies

mutual independence for jointly normal random variables, it follows that given $0 \le s_1 < s_2 < \cdots < s_n < t$, the random variables $\langle Y(s_k);\ 1 \le k \le n\rangle$ are independent of $X(t)$, and so the conditional distribution of $\langle X(s_k);\ 1 \le k \le n\rangle$ given $X(t) = A$ equals the unconditional distribution of $\langle Y(s_k) + \frac{Q(s_k,t)}{Q(t,t)}A;\ 1 \le k \le n\rangle$. These random variables are linear combinations of those in X (plus constants), whence have a multivariate normal distribution. That is, $\langle Y(s) + \frac{Q(s,t)}{Q(t,t)}A;\ 0 \le s \le t\rangle$ is a Gaussian process. In other words, the process $\langle X(s);\ 0 \le s \le t\rangle$ given $X(t) = A$ is a Gaussian process. (In fact, the argument above works for all $s > 0$.)

Consider the special case of a Brownian motion with $X(0) \equiv 0$ and parameter (μ, σ^2). Conditional that $X(t) = A$, this is called ***Brownian bridge*** $Z(\cdot)$ (from 0 to A on $[0, t]$). Using our formula for general Gaussian processes, we get that the distribution of $Z(s)$
1.5" equals the (unconditional) distribution of $X(s) - (s/t)\big(X(t) - A\big)$. ... Since this is a Gaussian process, we can characterize it by its means and covariances. These are, for $0 \le s \le s' \le t$,

$$E\big[Z(s)\big] = E\big[X(s) - (s/t)\big(X(t) - A\big)\big] = \mu s - (s/t)(\mu t - A) = As/t$$

and

$$\begin{aligned}\operatorname{Cov}\big(Z(s), Z(s')\big) &= \operatorname{Cov}\big(X(s) - (s/t)X(t), X(s') - (s'/t)X(t)\big) \\ &= \sigma^2\big(s - ss'/t - ss'/t + ss'/t\big) = \sigma^2 s(t - s')/t.\end{aligned}$$

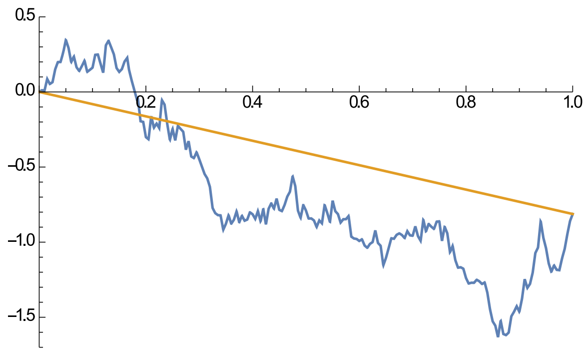
For example, the conditional distribution of $X(s)$ given $X(t) = A$ is

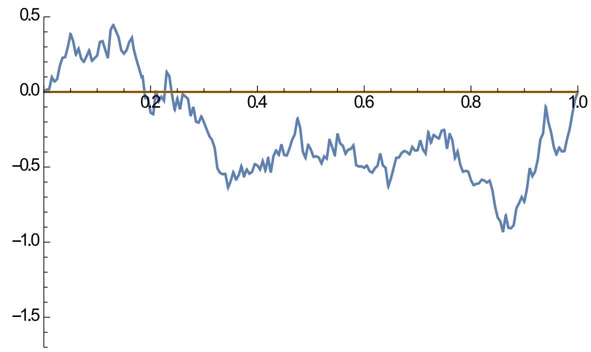
$$\big(X(s) \mid X(t) = A\big) \sim \mathrm{N}(As/t, \sigma^2 s(t - s)/t).$$

Also note that $E[X(s) \mid X(t) = A]$ is linear in s and the conditional variance $\to 0$ as $s \to 0$ and as $s \to t$.

We also notice that the distribution of the Brownian bridge process is independent of μ, which is quite surprising at first. Now the unconditional law of $\langle X(s)\rangle_{0\le s\le t}$ is a mixture of these conditional laws, where $A = X(t) \sim \mathrm{N}(\mu t, \sigma^2 t)$ does depend on μ. In other words, first sample $X(t) \sim \mathrm{N}(\mu t, \sigma^2 t)$, and then sample the Brownian bridge from 0 to this value of $X(t)$. Thus, the drift μ cannot be estimated more precisely by knowing $\langle X(s);\ s \le t\rangle$ than by knowing $X(t)$ alone. We can also note that the law of $\langle X(s) - As/t;\ 0 \le s \le t\rangle$ given $X(t) = A$ is the same as the law of $\langle X(s);\ 0 \le s \le t\rangle$ given $X(t) = 0$. Thus, when we sample the Brownian bridge from 0 to A, we could just sample a Brownian bridge from 0 to 0 and add to it the linear function $s \mapsto As$ to get a bridge from 0 to A.

The ***standard Brownian bridge*** has $A = 0$, $t = 1$, and $\sigma = 1$. In fact, for simplicity, we now take $t = 1$ for the following summary:





Proposition 8.1.1. *If $\langle X(t)\,;\ t \ge 0\rangle$ is a Brownian motion with $X(0) \equiv 0$ and $Z(t) := X(t) - X(1)t + At$, then $\langle Z(t)\,;\ 0 \le t \le 1\rangle$ is a Brownian bridge from 0 to A.*

These considerations also tell us one way to simulate (or even construct) Brownian motion: For instance, suppose we want to simulate Brownian motion $X(\cdot)$ on $[0,1]$. First, choose $X(1) \sim \mathrm{N}(0,1)$. Then choose $X(1/2)$ with the appropriate distribution for a bridge from 0 to $X(1)$, namely, $X(1/2) \sim \mathrm{N}\big(X(1)/2, 1/4\big)$. Similarly, choose $X(1/4) \sim \mathrm{N}\big(X(1/2)/2, 1/8\big)$ and $X(3/4) \sim \mathrm{N}\big((X(1/2) + X(1))/2, 1/8\big)$: see Exercise 93. Continue to the desired precision and finally linearly interpolate between successive values. This method is due to Paul Lévy.

▷ `Read pp. 361--363 in the book.`

§8.4.1. Using Martingales to Analyze Brownian Motion.

Generalizing from the case of discrete time, we call $\langle Z(t)\,;\ t \ge 0\rangle$ a ***martingale*** if

(i) $\forall t\ \ E|Z(t)| < \infty$ and

(ii) $\forall s < t\ \ E\big[Z(t) \mid \langle Z(u)\,;\ 0 \le u \le s\rangle\big] = Z(s)$.

We call a $[0,\infty]$-valued random variable τ a ***stopping time*** (for $Z(\cdot)$) if $\forall t\ \mathbf{1}_{[\tau \le t]}$ is a function of $\langle Z(s)\,;\ 0 \le s \le t\rangle$. If τ is a stopping time, then so is $\tau \wedge r$ for every $r \ge 0$. More generally, if τ_1 and τ_2 are both stopping times, then so is $\tau_1 \wedge \tau_2$. Indeed, $\tau_1 \wedge \tau_2 \le t$ iff $\tau_1 \le t$ or $\tau_2 \le t$. Hence, $\mathbf{1}_{[\tau_1 \wedge \tau_2 \le t]} = \mathbf{1}_{[\tau_1 \le t]} \vee \mathbf{1}_{[\tau_2 \le t]}$.

The Optional Stopping Theorem. *Let $\langle Z(t)\,;\ t \ge 0\rangle$ be a martingale with right-continuous sample paths a.s. and let τ be an a.s. finite stopping time. If*

(i) $E|Z(\tau)| < \infty$ *and*

(ii) $\lim_{t\to\infty} E\big[Z(t)\mathbf{1}_{[\tau > t]}\big] = 0$,

then

$$E\big[Z(\tau)\big] = E\big[Z(0)\big].$$

The hypotheses (i) and (ii) hold if τ is bounded.

`To see the last sentence, approximate` τ `by finite-valued stopping times and use that` $|Z(t)|$ `is a submartingale, so` $E|Z(\tau)| \le E|Z(t)|$ `for any fixed` $t > \|\tau\|_\infty$`. To prove this theorem, apply Durrett, Theorem 7.5.1, to` $\tau \wedge t$ `and let` $t \to \infty$`.`

In fact, the same holds for τ that is an a.s. finite stopping time with respect to a stochastic process of the form $\langle (Z_t, W_t)\,;\ t \ge 0\rangle$. Typically, this is used when Z_t is a function of W_t.

To see that some hypothesis is needed, note that standard Brownian motion $X(\cdot)$ is a martingale, but if $\tau = T_1$ is the hitting time of 1, then $E\big[X(T_1)\big] = 1 \ne 0 = E\big[X(0)\big]$. Here, T_1 clearly satisfies (i), so it must be that T_1 does not satisfies (ii). The conclusion, then, is that $X(t)$ might be very negative when $t < T_1$.

A sufficient condition for (i) and (ii) is that there is some nonnegative random variable W with $E[W] < \infty$ such that for all $t \ge 0$, we have $|Z(\tau \wedge t)| \le W$ (e.g., W could be bounded). `Use the LDCT and note that` $Z(t)\mathbf{1}_{[\tau>t]} = Z(\tau \wedge t)\mathbf{1}_{[\tau>t]}$`.`

Example: If $\langle X(t)\rangle$ is standard Brownian motion, then $\langle X(t)^2 - t\rangle$ is a martingale. Indeed, for $s < t$ we have

$$\begin{aligned}
&E\Big[X(t)^2 - t \;\Big|\; \langle X(u)^2 - u\,;\ 0 \le u \le s\rangle\Big] \\
&\quad = E\Big[E\big[X(t)^2 - t \mid \langle X(u)\,;\ 0 \le u \le s\rangle\big] \;\Big|\; \langle X(u)^2 - u\,;\ 0 \le u \le s\rangle\Big] \\
&\quad = E\Big[E\big[(\underbrace{\underbrace{X(t) - X(s)}_{\mathrm{N}(0,t-s)} + X(s)}_{\mathrm{N}(X(s),t-s)})^2 - t \mid \langle X(u)\,;\ 0 \le u \le s\rangle\big] \;\Big|\; \langle X(u)^2 - u\,;\ 0 \le u \le s\rangle\Big] \\
&\quad = E\Big[(t-s) + X(s)^2 - t \;\Big|\; \langle X(u)^2\,;\ 0 \le u \le s\rangle\Big]
\end{aligned}$$

.5" $\quad\cdots$

$$= X(s)^2 - s.$$

Corollary. *For standard Brownian motion, $E[T_1 \wedge T_{-1}] = 1$.*

Proof. Let $\tau := T_1 \wedge T_{-1}$, which is the hitting time of the set $\{1, -1\}$. Then $\forall t$, we may apply the optional stopping theorem to the bounded stopping time $\tau \wedge t$ to obtain

$$E\big[X(\tau \wedge t)^2 - \tau \wedge t\big] = E\big[X(0)^2 - 0\big] = 0,$$

i.e.,

$$E[\tau \wedge t] = E\big[X(\tau \wedge t)^2\big].$$

By the MCT, the left-hand side tends to $E[\tau]$ as $t \to \infty$. Because the right-hand side is at most 1, it follows that $E[\tau] < \infty$ and so $\tau < \infty$ a.s. Therefore, we can apply the BCT to the right-hand side and obtain the limit $E\big[X(\tau)^2\big] = E[\mathbf{1}] = 1$. This is then the value of $E[\tau]$. ∎

We can get easily a new proof that for standard Brownian motion, $P[T_a < T_{-b}] = b/(a+b)$ $(a, b > 0)$. First note that a proof similar to the above corollary shows that $E[T_a \wedge T_{-b}] < \infty$, so, in particular, $T_a \wedge T_{-b} < \infty$ a.s. Therefore, if $p := P[T_a < T_{-b}]$, we have

$$0 = E\big[X(T_a \wedge T_{-b})\big] = ap - b(1-p).$$

There are special extensions of the optional stopping theorem for Brownian martingales:

Wald's First Identity. *Let $X(\cdot)$ be standard Brownian motion. If τ is a stopping time with $E[\tau] < \infty$, then $E\big[X(\tau)\big] = 0$.*

Proof. We will show that the hypotheses of the optional stopping theorem as we stated it above hold. First, consider $t < \infty$ and the stopping time $\tau \wedge n$. We have

$$E\big[|X(\tau \wedge n)|\big]^2 \le E\big[X(\tau \wedge n)^2\big] = E[\tau \wedge n] \le E[\tau], \tag{N8}$$

because the square of the first moment of a random variable is at most its second moment. Hence, Fatou's lemma gives

$$E\big[|X(\tau)|\big] \le \liminf_{n \to \infty} E\big[|X(\tau \wedge n)|\big] \le E[\tau]^{1/2} < \infty.$$

This establishes condition (i) of the optional stopping theorem. Next, use the Cauchy–Schwarz inequality (Exercise 8.28) to see that

$$E\big[X(t)\mathbf{1}_{[\tau > t]}\big]^2 = E\big[X(\tau \wedge t)\mathbf{1}_{[\tau > t]}\big]^2 \le E\big[X(\tau \wedge t)^2\big] E[\mathbf{1}_{[\tau > t]}^2] \le E[\tau] P[\tau > t].$$

Since $\tau < \infty$ a.s., this tends to 0 as $t \to \infty$, yielding condition (ii). ∎

Wald's Second Identity. *Let $X(\cdot)$ be standard Brownian motion. If τ is a stopping time with $E[\tau] < \infty$, then $E\big[X(\tau)^2\big] = E[\tau]$.*

Proof. As above, Fatou's lemma and (N8) yield $E\big[X(\tau)^2\big] \le E[\tau]$, so it remains to establish the reverse inequality. Write $X(\tau) = X(\tau \wedge n) + \big[X(\tau) - X(\tau \wedge n)\big]$, square, and take the expectation:

$$E\big[X(\tau)^2\big] = E\big[X(\tau \wedge n)^2\big] + E\Big[\big[X(\tau) - X(\tau \wedge n)\big]^2\Big] + 2\,E\Big[X(\tau \wedge n)\big[X(\tau) - X(\tau \wedge n)\big]\Big].$$

By the strong Markov property and independent increments, we have that the last term equals $2\,E\big[X(\tau \wedge n)\big]\,E\big[X(\tau) - X(\tau \wedge n)\big]$, which is 0 by Wald's first identity. Therefore, $E\big[X(\tau)^2\big] \ge E\big[X(\tau \wedge n)^2\big] = E[\tau \wedge n]$ by (N8). Taking $n \to \infty$ and using the MCT gives the desired result. ∎

Other techniques show that both identities hold assuming only that $E\big[\sqrt{\tau}\,\big] < \infty$. The choice $\tau = T_1$ gives counterexamples when $E\big[\tau^c\big] < \infty$ for each $c < 1/2$.

EXAMPLE: If $\langle X(t)\rangle$ is a Brownian motion with drift $\mu \neq 0$ and $X(0) \equiv 0$, then $\langle X(t) - \mu t\rangle$
1" is a martingale. … Thus, if τ satisfies the conditions of the optional stopping theorem, then

$$E\big[X(\tau) - \mu\tau\big] = 0. \tag{N9}$$

In particular, for every bounded stopping time, τ, we have

$$E\big[X(\tau)\big] = \mu\, E[\tau]. \tag{N10}$$

For example, if $\mu < 0$ and we are forced to choose such a stopping time and collect $X(\tau)$, then to maximize our expected gain, we should take $\tau = \mathbf{0}$.

It turns out that the same equation (N10) holds without any assumptions on τ other than that τ is finite a.s., in contrast to the case where $\mu = 0$ and also in contrast to the game we analyzed earlier, which did not force us to stop at a finite time. This may be a surprise since a.s. there exists an open interval of time where $X > 0$. What makes it impossible to stop a.s. at such a time, however, is that there is no time $t > 0$ such that a.s. there is some $s \geq t$ with $X(s) > 0$, nor is there any $x > 0$ such that a.s. there is some time t with $X(t) \geq x$.

Indeed, let τ be any a.s. finite stopping time and $\mu > 0$. Write $B(t) := X(t) - \mu t$, which is Brownian motion with drift 0 and variance parameter σ^2. By dividing by σ and changing μ to μ/σ, we may assume that $\sigma = 1$ without loss of generality. By Wald's first identity, if $E[\tau] < \infty$, then (N9) holds, which implies that (N10) holds. Consequently, it remains to show that $E\big[X(\tau)\big] = \infty$ when $E[\tau] = \infty$. Thus, assume that $\int_0^\infty P[\tau > t]\,dt = \infty$. We have that

$$\int_0^\infty P\big[\exists s \geq t\ \ B(s) \geq \mu s/2\big]\,dt < \infty.$$

(Use time inversion: with $Y(u) := uB(1/u)$, we have

$$P[\exists s \geq t\ \ B(s) \geq \mu s/2] = P[\exists u \leq 1/t\ \ Y(u) \geq \mu/2] = 2\Phi(\sqrt{t}\mu/2) - 2 \leq 2e^{-t\mu^2/8}$$

by (N5) for the second equality and Proposition 1.7.2 for the inequality.) Now

$$\begin{aligned} P\big[X(\tau) > \mu t\big] &\geq P\big[X(\tau) > \mu\tau/2, \tau > 2t\big] = P[\tau > 2t] - P\big[\tau > 2t, B(\tau) \leq -\mu\tau/2\big] \\ &\geq P[\tau > 2t] - P\big[\exists s \geq 2t\ \ B(s) \geq \mu s/2\big]. \end{aligned}$$

Therefore, $\int_0^\infty P\big[X(\tau)/\mu > t\big]\,dt = \infty$. Combining this with the fact that $-\min_{s\geq 0} X(s)$ has an exponential distribution, we obtain that $E\big[X(\tau)\big] = \infty$ as desired. The upshot is that the longer you wait to stop, the more you will lose in expectation.

Now we prove that the strategy we used to stop a Brownian motion $X(\cdot)$ with negative drift μ at a time T in order to maximize $R := E_0\big[X(T)\,;\ T < \infty\big]$ is indeed the best one. Recall that we used $T := T_\lambda$, where $\lambda := -\sigma^2/(2\mu)$, for which we had $R = \lambda/e$. Recall also that $\max X(t) \sim \mathrm{Exp}(1/\lambda)$. Let T be any stopping time (with $X(T) \ge 0$ on the event $T < \infty$). Given $a > 0$, define $\tau := T \wedge T_a$, $\ p_a := P_0[\tau < \infty]$, and $R_a := E_0\big[X(\tau)\,;\ \tau < \infty\big]$. A proof similar to the solution of Exercise 95 tells us that

$$E_0\big[\exp\big\{X(\tau \wedge T_{-b})/\lambda\big\}\big] = 1$$

1" for every $b > 0$. The BCT allows us to take $b \to \infty$ ... and obtain

$$E_0\big[\exp\big\{X(\tau)/\lambda\big\}\,;\ \tau < \infty\big] = 1,$$

i.e.,

$$E_0\big[\exp\big\{X(\tau)/\lambda\big\} \bigm| \tau < \infty\big]p_a = 1.$$

2" Now convexity of the exponential (Jensen's inequality, Proposition 1.7.3) ... yields

$$E_0\big[\exp\big\{X(\tau)/\lambda\big\} \bigm| \tau < \infty\big] \ge \exp\Big\{E_0\big[X(\tau)/\lambda \bigm| \tau < \infty\big]\Big\} = e^{R_a/(p_a\lambda)}.$$

Putting these together, we arrive at

$$1 \ge p_a e^{R_a/(p_a\lambda)},$$

whence

$$\frac{R_a}{p_a} e^{-R_a/(p_a\lambda)} \ge R_a.$$

Since $\max_{x \ge 0} xe^{-x} = 1/e$, we arrive at the inequality $R_a \le \lambda/e$. Since $0 \le X(T) \le \max_t X(t)$ on the event $T < \infty$, and $\max_t X(t)$ has finite expectation, the LDCT shows that $\lim_{a\to\infty} R_a =$
2" R, ... whence also $R \le \lambda/e$, as desired.

Furthermore, examination of the equality condition in Jensen's inequality shows that the only stopping time that achieves $R = \lambda/e$ is the one we used, T_λ. We sketch the proof. We first show that if $R = \lambda/e$, then there is a constant c such that $X(T) = c$ a.s. given $T < \infty$. Suppose T does not have this form. Then

$$P\big[|X(T) - E[X(T)]| > 2\lambda\epsilon' \bigm| T < \infty\big] > 2\epsilon$$

for some $\epsilon, \epsilon' > 0$, whence $P\big[|X(\tau) - E[X(\tau)]| > \lambda\epsilon' \bigm| \tau < \infty\big] > \epsilon$ for all large a (as before, $\tau := T \wedge T_a$). We now look at a more refined version of Jensen's inequality. There is some $\delta > 0$ such that $e^t \ge 1 + t + \delta\mathbf{1}_{\{|t|>\epsilon'\}}$, whence for every random variable Y with mean 0, we have

$$E\big[e^Y\big] \ge 1 + \delta\, P\big[|Y| > \epsilon'\big].$$

Use $Y := \big(X(\tau) - E_0\big[X(\tau) \bigm| \tau < \infty\big]\big)/\lambda$ conditional on $\tau < \infty$ to get

$$E_0\Big[\exp\big\{X(\tau)/\lambda\big\} \bigm| \tau < \infty\Big] \ge \exp\Big\{E_0\big[X(\tau)/\lambda \bigm| \tau < \infty\big]\Big\}(1 + \delta\epsilon) = e^{R_a/(p_a\lambda)}(1 + \delta\epsilon).$$

Then proceeding as before yields $R_a \le \lambda/\big(e(1 + \delta\epsilon)\big)$, whence the same holds for R, a contradiction.

Second, we claim that this means $T = T_c$ a.s., which implies that $c = \lambda$ by our earlier calculation. Indeed, if not, then $T > T_c$ with positive probability. Since $X(T) = c$ a.s. given that $T < \infty$, it follows that to obtain the same reward as T_c, it must be that $T < \infty$ a.s. given that $T > T_c$. However, this contradicts (N10) applied to T and X starting from time T_c given that $T_c < \infty$.

Chapter 9

Stochastic Order Relations

§9.1. Stochastically Larger.

Given two random variables X and Y, we say that X is ***stochastically larger than*** or ***stochastically dominates*** Y, written $X \succcurlyeq Y$, if $\overline{F}_X \ge \overline{F}_Y$. We have seen this in connection with renewal processes (Exercise 38). Also, $X \succcurlyeq Y$ iff $\forall a\ P[X \ge a] \ge P[Y \ge a]$ (Exercise 97). If we decompose X and Y into their positive and negative parts, $X = X^+ - X^-$, $Y = Y^+ - Y^-$, then

$$X \succcurlyeq Y \iff X^+ \succcurlyeq Y^+ \text{ and } X^- \preccurlyeq Y^-$$

(Exercise 97).

LEMMA 9.1.1. *If $X \succcurlyeq Y$, then $E[X] \ge E[Y]$ when both $E[X]$ and $E[Y]$ are defined in $[-\infty, \infty]$.*

Proof. If $X, Y \ge 0$, then $E[X] = \int_0^\infty \overline{F}_X(a)\, da \ge \int_0^\infty \overline{F}_Y(a)\, da = E[Y]$. Thus, $E[X^+] \ge E[Y^+]$ and $E[X^-] \le E[Y^-]$, so $E[X] = E[X^+] - E[X^-] \ge E[Y^+] = E[Y^-] = E[Y]$. ∎

PROPOSITION 9.1.2. *$X \succcurlyeq Y$ iff for all increasing $f \colon \mathbb{R} \to \mathbb{R}$, we have $E\big[f(X)\big] \ge E\big[f(Y)\big]$ when both expectations are defined in $[-\infty, \infty]$.*

Proof. $\Longleftarrow$: Let $f := \mathbf{1}_{(a,\infty)}$, $a \in \mathbb{R}$.

$\Longrightarrow$: (The proof in the book is incorrect unless f is continuous.) By the lemma, it suffices to show that $f(X) \succcurlyeq f(Y)$. Set

$$I_f(a) := \{x\,;\ f(x) > a\}.$$

Then $I_f(a)$ is an interval of the form (s, ∞) or $[s, \infty)$ and $f(X) > a \iff X \in I_f(a)$, whence

$$P\big[f(X) > a\big] = P\big[X \in I_f(a)\big] \ge P\big[Y \in I_f(a)\big] = P[f(Y) > a].$$ ∎

§9.2. Coupling.

Coupling refers generally to creating a new pair of random variables (X^*, Y^*) out of given random variables X and Y such that X^* and Y^* are defined on the same probability space as each other, yet still $X^* \overset{\mathscr{D}}{=} X$ and $Y^* \overset{\mathscr{D}}{=} Y$. Usually one wants to get the new pair to have some special properties. Often, this is used to convert distributional properties to more direct comparisons of random variables. We will use the following kind of inverse to a c.d.f., F:

$$F^{-1}(s) := \inf\{x\,;\ F(x) \geq s\} = \min\{x\,;\ F(x) \geq s\}.$$

2" ... Note that $F^{-1}(s) \leq y$ iff $F(y) \geq s$. ...

PROPOSITION. *Let F be a c.d.f. If $U \sim \mathrm{Unif}[0,1]$, then $F^{-1}(U) \sim F$.*

Proof. For every $y \in \mathbb{R}$, we have

$$P\big[F^{-1}(U) \leq y\big] = P\big[F(y) \geq U\big] = F(y). \qquad \blacksquare$$

PROPOSITION 9.2.2. *$X \succcurlyeq Y$ iff there exist random variables X^* and Y^* with $X^* \overset{\mathscr{D}}{=} X$, $Y^* \overset{\mathscr{D}}{=} Y$, and $X^* \geq Y^*$.*

This makes Proposition 9.1.2 obvious.

Proof. $\Longleftarrow$: We have for all a,

$$P[X > a] = P[X^* > a] \geq P[Y^* > a] = P[Y > a].$$

$\Longrightarrow$: Let $U \sim \mathrm{Unif}[0,1]$. Define $X^* := F_X^{-1}(U)$ and $Y^* := F_Y^{-1}(U)$. By the preceding
1" proposition, $X^* \overset{\mathscr{D}}{=} X$ and $Y^* \overset{\mathscr{D}}{=} Y$. Since $F_X \leq F_Y$, we have $F_X^{-1} \geq F_Y^{-1}$, ... which gives the result. $\blacksquare$

2" EXAMPLE: If $n_1 \geq n_2$ and $p_1 \geq p_2$, then $\mathrm{Bin}(n_1, p_1) \succcurlyeq \mathrm{Bin}(n_2, p_2)$. ...

EXAMPLE 9.2(B). $\mathrm{Pois}(\lambda)$ is stochastically increasing in λ: This is not hard to show by analytic calculation, but here is a coupling proof: Let $\lambda < \mu$, $X \sim \mathrm{Pois}(\mu)$, and $Y \sim \mathrm{Bin}(X, \lambda/\mu)$. Then $Y \leq X$ and $Y \sim \mathrm{Pois}(\mu \cdot \frac{\lambda}{\mu})$.

§9.2.2. Exponential Convergence in Markov Chains.

Recall that if $\langle X_n\,;\ n \ge 0\rangle$ is a finite-state irreducible aperiodic Markov chain, then for all i and j, we have $p_{ij}^{(n)} \to \pi_j$, the stationary probabilities. How fast is the convergence? We show that it is exponential; an active area of research involves estimating the precise rate for various Markov chains on large state spaces.

Theorem. *Let $\langle X_n\,;\ n \ge 0\rangle$ be a finite-state irreducible aperiodic Markov chain and π_j be the stationary probabilities. Then $\exists c > 0,\ \beta < 1$ such that $\forall n, i, j\ \ |p_{ij}^{(n)} - \pi_j| \le c\beta^n$.*

Proof. By Exercise 4.14, $\pi_j > 0$ for all j. Thus, by our recollection above, $\exists N\ \ \forall i, j\ \ p_{ij}^{(N)} > 0$. Let $\varepsilon := \min_{i,j} p_{ij}^{(N)}$. Let $\langle X'_n\,;\ n \ge 0\rangle$ be an independent Markov chain with the same transition probabilities, but with the stationary distribution used as the initial distribution. Let's take $X_0 \equiv i$. Define

$$T := \inf\{n\,;\ X_n = X'_n\}$$

and set

$$\overline{X}_n := \begin{cases} X_n & \text{if } n \le T, \\ X'_n & \text{if } n \ge T. \end{cases}$$

Then $\langle \overline{X}_n\rangle$ is a Markov chain with the same distribution as $\langle X_n\rangle$, as we'll verify. Since $[\overline{X}_n \ne X'_n] \subseteq [T > n]$, we have

$$|p_{ij}^{(n)} - \pi_j| = |P[\overline{X}_n = j] - P[X'_n = j]| \le P[\overline{X}_n \ne X'_n] \le P[T > n].$$

1" … Thus, it remains to bound $P[T > n]$.

Now by choice of N, we have $P[X_N = X'_N] \ge P[X_N = j = X'_N] \ge \varepsilon^2$ (for any j) and, in fact,

$$P[X_{kN} = X'_{kN} \mid X_0, X'_0, X_N, X'_N, \ldots, X_{(k-1)N}, X'_{(k-1)N}] \ge \varepsilon^2,$$

1" whence $P[T > kN] \le (1 - \varepsilon^2)^k$. … This proves the result. …

It remains to verify that $\langle \overline{X}_n\rangle$ is a Markov chain with the same distribution as $\langle X_n\rangle$. Consider $Z_n := (X_n, X'_n)$. This is clearly a Markov chain. Furthermore, T is a stopping time for it. If we define $Z'_n := (X'_n, X_n)$, then the Markov property implies that given $T = t$ and $Z_T = (j, j)$, the distribution of $\langle Z_{T+k}\,;\ k \ge 0\rangle$ equals the distribution of $\langle Z'_{T+k}\,;\ k \ge 0\rangle$. Hence the same is true given $T = t$, which means that if we define

$$\overline{Z}_n := \begin{cases} Z_n & \text{if } n \le T, \\ Z'_n & \text{if } n \ge T, \end{cases}$$

then $\langle \overline{Z}_n\rangle$ has the same distribution as $\langle Z_n\rangle$. Looking at the first coordinates of these processes gives what we wanted. ∎

Homework Problems

Note that while problems from the book are reproduced here, there is often information in the back of the book that is not reproduced here.

Exercise not to Hand In: Based on experience from similar oil fields, an oil executive has determined that the probability that a certain oil field contains a significant quantity of oil is 0.6. Before drilling, she orders a seismological test for further information. This test is not 100% accurate; if there is a significant quantity of oil, then the test confirms this with probability 0.9, but if there is not a significant quantity, then it confirms that with probability 0.8. Suppose that the seismological test does say that there is a significant quantity of oil. What should the executive now estimate as the probability of a significant quantity of oil?

Exercise not to Hand In: Five communication towers are erected in a straight line, each exactly 8 miles from its neighbors. The signal from each tower travels 16.6 miles. Assume that on a given day, the communication equipment in each tower is broken with probability 0.002, independently of each other. If it is not broken, then it transmits each signal that it receives. What is the probability that a signal from the first tower reaches the fifth tower?

Exercise not to Hand In: What is the median of an $\mathrm{Exp}(\lambda)$ distribution? Why is this called "half-life" in radioactive decay?

Exercise not to Hand In: Suppose that the joint density of X, Y is

$$f(x,y) = \begin{cases} cx^2y^2 & \text{if } x \geq 0, y \geq 0, x+y \leq 1, \\ 0 & \text{otherwise.} \end{cases}$$

What is the value of c? What is $E[X]$? What is $E[XY]$?

1. p. 48, 1.11: If X is a nonnegative integer-valued random variable, then the function $P(z)$, defined for $|z| \le 1$ by

$$P(z) = E\big[z^X\big] = \sum_{j=0}^{\infty} z^j P[X = j],$$

is called the ***probability generating function*** of X.

(a) Show that

$$\frac{d^k}{dz^k} P(z)\Big|_{z=0} = k! P[X = k].$$

(b) Show that

$$P[X \text{ is even}] = \frac{P(-1) + P(1)}{2}.$$

(c) If X is binomial with parameters n and p, show that

$$P[X \text{ is even}] = \frac{1 + (1 - 2p)^n}{2}.$$

(d) If X is Poisson with mean λ, show that

$$P[X \text{ is even}] = \frac{1 + e^{-2\lambda}}{2}.$$

(e) If X is geometric with parameter p, show that

$$P[X \text{ is even}] = \frac{1 - p}{2 - p}.$$

(f) If X is a negative binomial random variable with parameters r and p, show that

$$P[X \text{ is even}] = \frac{1}{2}\left[1 + (-1)^r \left(\frac{p}{2 - p}\right)^r\right].$$

2. Show that if X, Y have a joint density $f_{X,Y}$, then X has the density

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x, y)\, dy.$$

3. Give an example of independent random variables X and Y and a constant $a \in \mathbb{R}$ such that if we denote $Z_a := X + aY$, then Z_1 and Z_a are not independent yet $\mathrm{Cov}(Z_1, Z_a) = 0$.

4. Let $A_1, \ldots, A_n$ be independent events with respective probabilities $p_1, \ldots, p_n$. Let N be the number of these events that occur. Show that the moment generating function of N is $t \mapsto \prod_{k=1}^{n} \big(1 - p_k + p_k e^t\big)$.

5. p. 47, 1.8: Let X_1 and X_2 be independent Poisson random variables with means λ_1 and λ_2.

(a) Find the distribution of $X_1 + X_2$.

(b) Compute the conditional distribution of X_1 given that $X_1 + X_2 = n$.

(Note that both answers involve named distributions.)

6. Verify the conditional change-of-variable formula (N2) when X and Y are both discrete.

7. Find the mean and standard deviation of the time until the miner reaches safety in Example 1.5(b) (p. 23 of the book).

8. Verify the values given in Table 1.4.1 (p. 16) for the moment generating functions, means, and variances. Your verification should use the conceptual definitions of the distributions as much as possible, rather than the formulas for the distributions.

9. Show using the tower property, (1.5.1), and the moment generating function of an exponential random variable that if $X_i \sim \mathrm{Exp}(\lambda_i)$ for $i = 1, 2$ are independent, then $P[X_1 < X_2] = \lambda_1/(\lambda_1 + \lambda_2)$.

10. Let X be a random variable and A be an event. Show directly from our definition that $E[X \mid A] = E[X^+ \mid A] - E[X^- \mid A]$. (You may not use linearity of conditional expectation nor (N3), because we used this exercise to prove those. Of course, you may use linearity of ordinary expectation.)

11. p. 46, 1.1: Let N denote a nonnegative, integer-valued, random variable. Show that

$$E[N] = \sum_{k=1}^{\infty} P[N \geq k] = \sum_{k=0}^{\infty} P[N > k].$$

In general, show that if X is nonnegative with distribution F, then

$$E[X] = \int_0^\infty \overline{F}(x)\,dx$$

and

$$E[X^n] = \int_0^\infty n x^{n-1} \overline{F}(x)\,dx.$$

12. Suppose that U is a Unif$[0,1]$ random variable. Given the value of U, say, $U = u$, another random experiment is made to get the value of X so that X has a Exp(u) distribution (i.e., an exponential random variable with parameter u; note that it is *not* true that $X = e^U$). Prove that X has a density and compute it.

13. Let X be a nonnegative random variable with c.d.f. F and c a positive constant. Show that

$$E\big[\min\{X, c\}\big] = \int_0^c \overline{F}(x)\,dx$$

and, if $F(c) > 0$,

$$E[X \mid X \le c] = \int_0^c \left[1 - \frac{F(x)}{F(c)}\right] dx.$$

14. p. 46, 1.3: Let X_n denote a binomial random variable with parameters (n, p_n), $n \ge 1$. If $np_n \to \lambda$ as $n \to \infty$, show that

$$P[X_n = i] \to e^{-\lambda}\lambda^i/i! \qquad \text{as } n \to \infty.$$

Here, $\lambda < \infty$. Show that if $\lambda = \infty$, then $P[X_n = i] \to 0$ as $n \to \infty$, perhaps by comparing to the case of finite λ.

(The last part was added. Show all this directly, not by using the Poisson convergence theorem.)

15. p. 50, 1.18: A coin, which lands on heads with probability p, is continually flipped. Compute the expected number of flips that are made until a string of r heads in a row is obtained. *Hint:* Condition on the number of flips until the first tail appears.

Exercise not to Hand In: p. 51, 1.22

16. p. 53, 1.30: In Example 1.6(A) if server i serves at an exponential rate λ_i, $i = 1, 2$, compute the probability that Mr. A is the last one out.

17. p. 53, 1.31: If X and Y are independent exponential random variables with respective means $1/\lambda_1$ and $1/\lambda_2$, compute the distribution of $Z = \min(X, Y)$. What is the conditional distribution of Z given $Z = X$? (For a fun challenge, see whether you can solve both parts of this problem with virtually no calculation. Note that there are 3 natural ways to express that two events, A and B, are independent: $P(A \mid B) = P(A)$, $P(B \mid A) = P(B)$, and $P(AB) = P(A)P(B)$. In fact, there are 12 ways because each of A and B could be replaced by their complements.)

Exercise not to Hand In: p. 53, 1.34

18. Let X and Y be independent $\mathrm{Exp}(\lambda)$ random variables. Find the distribution of $|X-Y|$. (Hint: there is a short solution that needs hardly any calculation.)

19. Consider again the two teams of gladiators. At each round, each team sends one of its gladiators to battle. When two fight, the probability of winning is proportional to the strength of a gladiator. The loser never plays again. In this variant, however, the winning gladiator inherits the strength of the loser. Show that each team wins with probability proportional to its total strength, regardless of order. (Hint: induct on the total number of gladiators.)

20. Suppose that the lifetime of a machine is an $\mathrm{Exp}(\lambda)$ random variable. The machine is checked to see whether it is operating at regular intervals, namely, at times s, $2s$, $3s$, etc., for some fixed $s>0$. Eventually, of course, the machine is discovered to be down. In terms of λ and s, what is the expected duration of the time that the machine is actually down before it is discovered to be down?

21. A component of a machine has an exponentially distributed lifetime with mean 750 hours. When it fails, it is replaced in the machine by a new component with an independent lifetime of the same distribution. What is the smallest number of spare components that should be provided in order that the machine last for 2000 hours (using the original component and these spares only) with probability at least 95%?

22. p. 89, 2.5: Suppose that $\{N_1(t), t\ge 0\}$ and $\{N_2(t), t\ge 0\}$ are independent Poisson processes with rates λ_1 and λ_2. Show that the combined process $\{N_1(t)+N_2(t), t\ge 0\}$ is a Poisson process with rate $\lambda_1+\lambda_2$. Also, show that the probability that the first event of the combined process comes from $\{N_1(t), t\ge 0\}$ is $\lambda_1/(\lambda_1+\lambda_2)$, independently of the time of the event. (Note that this last statement means that the probability that the first event comes from $N_1(\cdot)$ given that the time of the first event is t is equal to $\lambda_1/(\lambda_1+\lambda_2)$ for every $t>0$.)

State and prove the analogue for any finite number of independent Poisson processes, not just for two processes.

23. You want to cross a road at a spot where cars pass according to a Poisson process with rate λ. You begin to cross as soon as you see there will not be any cars passing for the next c time units. Let $N :=$ number of cars that pass before you cross, $T :=$ time you begin to cross.

(a) What is $E[N]$?

(b) Find $E[T]$, for example, by conditioning on N.

Exercise not to Hand In: p. 90, 2.8

24. p. 91, 2.14: Consider an elevator that starts in the basement and travels upward. Let N_i denote the number of people that get in the elevator at floor i. Assume the N_i are independent and that N_i is Poisson with mean λ_i. Each person entering at i will, independently of everything else, get off at j with probability P_{ij}, where $\sum_{j>i} P_{ij} = 1$. Let $O_j =$ number of people getting off the elevator at floor j.

(a) Compute $E[O_j]$.

(b) What is the distribution of O_j?

(c) What is the joint distribution of O_j and O_k for $j \neq k$?

(The solution needs more explanation than what appears in the back of the book!)

25. Let $T \sim \operatorname{Exp}(\lambda)$ and $M \sim \operatorname{Pois}(\mu T)$ given T. Use classified Poisson processes to calculate the unconditional distribution of M.

26. There are n radioactive particles in a substance at time 0. Their lifetimes are i.i.d. $\operatorname{Exp}(\lambda)$. Let $X(t)$ be the number of particles that have decayed by time t.

(a) What is $P\big[X(t) = k\big]$?

(b) Let T be the first time t that $X(t) = k$. What is $E[T]$?

27. Customers arrive at a store according to a Poisson process of rate λ/hr. Each customer spends a time in the store that is a random variable with $\operatorname{Exp}(\alpha/\text{hr})$ distribution, independent of other customer times, and then departs. Given that exactly five customers arrive in the first hour, what is the probability that the store is empty of customers at the end of this first hour?

28. A machine needs 2 types of parts to work, type A and type B. It has one part of each type to begin, and there are also 2 spare A parts and 1 spare B part. When a part fails, it is replaced by a spare part of the same type, if available, instantaneously. Suppose that the lifetimes (time in service) of all parts are independent; parts of type A are $\mathrm{Exp}(\lambda)$-distributed, while parts of type B are $\mathrm{Exp}(\mu)$-distributed. What is the expected time until the machine fails for lack of a needed part?

29. Let $N_i(\cdot)$ $(i = 1, 2)$ be independent Poisson processes with rates λ_i. Let X be the result of a fair coin flip that is independent of both processes $N_i(\cdot)$. Suppose that $N(t) = N_1(t)$ for all t if $X = \mathrm{H}$, while $N(t) = N_2(t)$ for all t if $X = \mathrm{T}$. (There is only one coin flip.)
(a) Does $N(\cdot)$ have stationary increments?
(b) Does $N(\cdot)$ have independent increments?
(c) Is $N(\cdot)$ a simple counting process?
(d) Is $N(\cdot)$ a Poisson process?

30. p. 94, 2.30: Let $T_1, T_2, \ldots$ denote the interarrival times of events of a nonhomogeneous Poisson process having intensity function $\lambda(t)$. Assume that $\int_0^\infty \lambda(t)\,dt = \infty$.
(a) Are the T_i independent?
(b) Are the T_i identically distributed?
(c) Find the distribution of T_1.
(d) Find the distribution of T_2.
(e) What are the failure rates of T_1 and T_2?
Note: part (e) was added.

31. p. 95, 2.33: Consider a two-dimensional Poisson point process with intensity λ. Given a fixed point, let X denote the distance from that point to its nearest event, where distance is measured in the usual Euclidean manner. Let R_i, $i \geq 1$, denote the distance from that point to the ith closest event to it. Put $R_0 := 0$. Show that:
(a) $P[X > t] = e^{-\lambda\pi t^2}$.
(b) $E[X] = 1/(2\sqrt{\lambda})$.
(c) $\pi R_i^2 - \pi R_{i-1}^2, i \geq 1$, are independent exponential random variables, each with rate λ.
(d) Let $N(r)$ be the number of points of the Poisson point process in $\mathbb{R}^2$ that are within distance r of the origin. Describe the law of the counting process $N(\cdot)$ on $[0, \infty)$.
Note: part (d) was added. Also, $X = R_1$.

32. p. 153, 3.1: Is it true that:

(a) $N(t) < n$ if and only if $S_n > t$?

(b) $N(t) \le n$ if and only if $S_n \ge t$?

(c) $N(t) > n$ if and only if $S_n < t$?

Note: $N(\cdot)$ is a renewal process with arrival times S_n.

33. Suppose that the interarrival distribution for a renewal process is Pois(μ) (so the interarrival times are discrete, but the renewal process is defined for continuous time).

(a) Find the distribution of S_n for each n.

(b) Find the distribution of $N(t)$ for all t.

34. Let $N(\cdot)$ be a renewal process with finite-mean interarrival time. Show that

$$\lim_{t \to \infty} S_{N(t)}/t = 1 \text{ a.s.}$$

35. Betsy is a consultant. Each time she gets a job to do, it lasts 3 months on average. The time between jobs is exponentially distributed with mean 2 weeks. At what rate does Betsy start new jobs in the long run?

36. A fair, 6-sided die is rolled repeatedly until the first of the following occurs:

(A) 3 consecutive rolls lie in $\{1, 2, 3\}$;

(B) 2 consecutive rolls lie in $\{4, 5\}$;

(C) a 6 appears.

What is the expected number of die rolls? Which of these three possible endings is *least* likely, A, B, or C? For example, if the rolls are 143254, then it ends on roll 6 with the outcome B.

37. Let $U_k \sim \text{Unif}[0, 1]$ be independent random variables. Define $N := \min\big\{n\,;\ \sum_{k=1}^{n} U_k > 1\big\}$. What is $E[N]$?

38. Let $N(\cdot)$ be a renewal process. Show that $\forall x, t \ge 0\ \ P\big[X_{N(t)+1} > x\big] \ge \overline{F}_X(x)$. Compute both sides exactly when $N(\cdot)$ is a Poisson process of rate λ. (Hint: Condition on $N(t)$ and $S_{N(t)}$.)

39. Consider a renewal process $N(\cdot)$ with nonconstant interrarival times having finite mean μ. Show that $E\big[S_{N(t)}\big] < \mu \cdot E\big[N(t)\big]$ for some $t > 0$.

40. Let $N_1(\cdot)$ and $N_2(\cdot)$ be independent renewal processes, both with interarrival times $\mathrm{Unif}[0,1]$. Define $N(t) := N_1(t) + N_2(t)$ for all t.

(a) Are the interarrival times of $N(\cdot)$ independent?

(b) Are the interarrival times of $N(\cdot)$ identically distributed?

(c) Is $N(\cdot)$ a renewal process?

41. p. 154, 3.6: Let $N(\cdot)$ be a renewal process and suppose that for all n and t, conditional on the event that $N(t) = n$, the event times $S_1, \ldots, S_n$ are distributed as the order statistics of a set of independent uniform $(0, t)$ random variables. Show that $N(\cdot)$ is a Poisson process.

Note: do not use the hint given in the book because that requires solving the preceding problem in the book. Besides, it will be more interesting if you do it directly. Think about where Poisson *distributions* arise.

42. A battery has a lifetime that is $\mathrm{Unif}[30, 60]$ in hours. If a battery is replaced as soon as it fails, what is the approximate distribution of the number of batteries that are replaced in 4500 hours after the installation of a new battery?

43. Events occur according to a Poisson process with rate λ. Any event that occurs within a time d of the event that immediately preceded it is called a d-event. For instance, if $d = 1$ and events occur at times 2, 2.8, 4, 6, 6.6, $\ldots$, then the events at times 2.8 and 6.6 would be d-events.

(a) At what (long-run) rate do d-events occur?

(b) What (long-run) proportion of all events are d-events?

44. Show that for a renewal process with $X \sim \mathrm{Exp}(\lambda)$, we have for every $t > 0$ that $A(t) \overset{\mathscr{D}}{=} X \wedge t$ by considering the event times in $[0, t]$ to be uniform i.i.d. given how many there are.

45. p. 156, 3.14: Let $A(t)$ and $Y(t)$ denote the age and excess at t of a renewal process. Fill in the missing terms:

(a) For $x \le t$, $A(t) \ge x \leftrightarrow 0$ events in the interval ________.

(b) $Y(t) > x \leftrightarrow 0$ events in the interval ________.

(c) $P[Y(t) > x] = P[A(\quad) \ge \quad]$.

(d) Compute the joint distribution of $A(t)$ and $Y(t)$ for a Poisson process.

Note: $\leftrightarrow$ means "if and only if". In (a) and (c), I changed each $>$ associated to $A(\cdot)$ to $\ge$. Of course, you must explain your answers. All blanks should be filled in with deterministic numbers, not anything random.

46. For a random variable X, write $\widehat{X}$ for a size-biased random variable corresponding to X. Show that if $X \sim \mathrm{Bin}(n, p)$, then $\widehat{X} - 1 \sim \mathrm{Bin}(n-1, p)$, while if $X \sim \mathrm{Pois}(\lambda)$, then $\widehat{X} - 1 \sim \mathrm{Pois}(\lambda)$.

47. Consider a nonlattice renewal process $N(\cdot)$ with interarrival times $X_k \sim F$ having finite mean, μ. We argued heuristically that $\lim_{t\to\infty} P[X_{N(t)+1} \le x] = P[\widehat{X} \le x]$ for all $x \ge 0$, among other things, to explain the key renewal theorem, where $\widehat{X}$ has the size-biased distribution corresponding to F. Use the key renewal theorem to prove this limit statement. (Hint: revisit Exercise 38.)

48. Consider a nonlattice renewal process $N(\cdot)$ with interarrival times $X_k \sim F$ having finite mean, μ, and finite variance, σ^2. Show that $\lim_{t\to\infty}\bigl(m(t) - t/\mu\bigr) = \sigma^2/(2\mu^2) - 1/2$. (Hint: Use (3.3.3).)

49.

(a) Suppose that X and Y are i.i.d. (real) random variables and g is a nondecreasing function. Assume that $E\bigl[|X|+|g(X)|\bigr] < \infty$. Show that $E\bigl[(X-Y)\bigl(g(X)-g(Y)\bigr)\bigr] \ge 0$. Deduce that $E\bigl[Xg(X)\bigr] \ge E[X]\,E\bigl[g(X)\bigr]$.

(b) Let $X \sim F$ be a nonnegative random variable with finite, positive mean. It follows from Exercise 38 and Exercise 47 that the c.d.f. of $\widehat{X}$ is at most F. Prove this directly from part (a).

50. p. 157, 3.20: Consider successive flips of a fair coin.

(a) Compute the mean number of flips until the pattern HHTHHTT appears.

(b) Which pattern requires a larger expected time to occur: HHTT or HTHT?

(c) Let A and B be two patterns of H and T of the same length, where A has no overlaps and B has at least one overlap. Give an intuitive explanation explanation why the expected time to the first occurrence of A is not the same as of B and which one is larger. (*Hint:* Count only occurrences of B that do not overlap in an infinite sequence of tosses.)

Note: part (c) was added.

51. A coin has probability p of H. What is E[time to THTHTHTHT]?

52. p. 154, 3.9(c): Consider a single-server bank in which potential customers arrive at a Poisson rate λ. However, an arrival enters the bank only if the server is free when he or she arrives. Let G denote the service distribution. Note: the system capacity is 1, so any customer who arrives while the server is busy is lost. What fraction of time is the server busy?

53. A particular ski slope has n skiers continually and independently climbing up and skiing down. (These skiers are inexhaustible.) The times it takes skiers to climb up or ski down are independent of each other and nonlattice but not identically distributed. In fact, the time it takes the ith skier to climb up has distribution F_i each time and the time it takes her to ski down has distribution G_i each time. All F_i and G_i have finite means.

(a) If $N(t)$ denotes the total number of times that the members of this ski group have skied down the slope by time t, summed over all n members, what are $\lim_{t\to\infty} N(t)/t$ a.s. and $\lim_{t\to\infty} E\big[N(t)\big]/t$?

(b) If $U(t)$ denotes the number of skiers that are climbing up the hill *at* time t, what is $\lim_{t\to\infty} E\big[U(t)\big]$?

54. p. 160, 3.31: A system consisting of four components is said to work whenever both at least one of components 1 and 2 works and at least one of components 3 and 4 works. Suppose that component i alternates between working and being failed in accordance with a nonlattice alternating renewal process with distributions F_i and G_i, $i = 1, 2, 3, 4$. If these alternating renewal processes are independent, find $\lim_{t\to\infty} P[\text{system is working at time } t]$.

55. Consider an alternating renewal process where $Z+Y$ has finite mean. Let $I(t)$ denote the indicator that the system is on at time t. Show that

$$\lim_{t\to\infty} \frac{1}{t}\int_0^t I(s)\,ds = \frac{E[Z]}{E[Z]+E[Y]} \text{ a.s.}$$

56. Consider an alternating renewal process where $Z+Y$ has finite mean. Suppose that $Z+Y$ has a lattice distribution with period d.

(a) For $0 \le u < d$, show that

$$\lim_{n\to\infty} P[\text{system is on at time } u+nd] = \frac{d\sum_{n=0}^{\infty} P[Z > u+nd]}{E[Z]+E[Y]}\,.$$

Hint: Follow the proof of Theorem 3.4.4 and use $g(s) := \overline{F}_Z(u+s)/\overline{F}_{Z+Y}(u+s)$.

(b) With $I(t)$ denoting the indicator that the system is on at time t, show that

$$\lim_{t\to\infty} E\left[\frac{1}{t}\int_0^t I(s)\,ds\right] = \frac{E[Z]}{E[Z]+E[Y]}\,.$$

57. Show that the long-run expected proportion of time that $X_{N(t)+1} \le c$ is $\int_0^c [F(c) - F(x)]\,dx/E[X]$ for a renewal process with $X_n \sim F$, where F has finite mean. Be sure to treat both the nonlattice and lattice cases.

58. In Example 3.4(a), find the long-run rate of restocking.

59. Suppose that G is uniform on $[0, S]$ in Example 3.4(a). Find explicitly the limiting distribution of the inventory level.

60. Suppose that $X_1, X_2, \ldots$ are the inter-renewal times of an equilibrium renewal process, where $X_2, X_3, \ldots$ each are equal to 1 or $\sqrt{2}$ with equal probability. Does X_1 have a density? If so, what is it? If not, why not?

61. Let X_n be the lifetimes of items assumed i.i.d. with c.d.f. F. Items are replaced when they reach age T if they have not yet failed. Find an F and T such that the actual *failure* rate when planned replacements are made is greater than that without planned replacement.

62. A warehouse stores and sells items. Customers arrive according to a Poisson process of rate 2 per day. Each customer demands exactly one item. The warehouse gives an item to a customer when it has one, but turns away the customer otherwise. The warehouse orders A more items from the supplier when the warehouse becomes empty, but it takes a random amount of time for the order to arrive; the order time has a mean of 3 days. Each such order costs the warehouse \$50 (regardless of the size of the order). Each item costs the warehouse \$1 per day to store. The supplier charges \$70 per item, but the warehouse sells each item for \$80.

(a) What is the long-run profit per day made by the warehouse?

(b) What value of A maximizes the long-run profit per day?

(A more realistic scenario would involve ordering more items before the warehouse becomes empty, but this is much harder to analyze.)

63. Let $Q(t)$ denote the number of customers in the system of an M/G/1/2 queue. Assume that G has finite mean and the arrival rate is λ. Show that in the long run, the proportion of time that $Q(t) = 1$ is

$$\frac{\int_0^\infty e^{-\lambda x}\overline{G}(x)\,dx}{\int_0^\infty \left(e^{-\lambda x}G(x) + \overline{G}(x)\right)\,dx.}$$

64. Consider an M/G/1/2 queue. Let $X_n :=$ the number of customers in the system when the nth customer leaves the system, $X_0 := 0$. What are the transition probabilities of this Markov chain?

65. p. 219, 4.4: Show that

$$p_{ij}^{(n)} = \sum_{k=0}^{n} f_{ij}^k p_{jj}^{(n-k)}.$$

66. p. 221, 4.12: A transition probability matrix P is said to be doubly stochastic if

$$\sum_i P_{ij} = 1 \qquad \text{for all } j.$$

That is, the column sums all equal 1. If a doubly stochastic chain has n states and is ergodic, calculate its limiting probabilities. (Note: "ergodic" means that the Markov chain with this transition matrix is irreducible, aperiodic, and positive recurrent.)

Exercise not to Hand In: p. 221, 4.16, 4.17

67. Suppose you have a deck of n cards. You shuffle them in the following simple manner: A card is chosen at random and put on the top. This is repeated many times, where the card chosen each time is independent of the preceding choices. Show that in the long run, the deck is perfectly shuffled in the sense that all $n!$ orderings are equally likely.

68. You have 3 coins, each with different probability of H. Namely, coin k has chance $(k+1)/(k+2)$ of coming up H ($k = 0, 1, 2$). The coins are tossed repeatedly in the following fashion. Coin 0 is tossed the first two times, but thereafter, coin k is tossed at time n when the number of H in tosses $n-1$ and $n-2$ is equal to k. What is the limiting probability that the nth coin comes up H as $n \to \infty$? *Hint:* Use a 4-state Markov chain.

69. We want to decide whether Brand A special high-intensity lightbulbs last longer than Brand B special high-intensity lightbulbs by making the following test. We turn on one bulb of each brand simultaneously and wait until one of the two fails. The brand whose bulb lasts longer gets a point. Then we repeat with new lightbulbs from each brand. We continue this test until one brand has accumulated 5 points more than the other brand. What is the chance that the test picks the better brand given that in fact the lifetimes of bulbs of Brand A are exponential with mean 25 hours while those of Brand B are exponential with mean 30 hours? *Hint:* Consider the difference between the numbers of points that each brand has accumulated.

70. p. 224, 4.27: Consider a particle that moves along a set of $m+1$ nodes, labeled 0, 1, $\ldots, m$. At each move it either goes one step in the clockwise direction with probability p or one step in the counterclockwise direction with probability $1-p$. It continues moving until all the nodes 1, 2, $\ldots, m$ have been visited at least once. Starting at node 0, find the probability that node i is the last node visited, $i = 1, \ldots, m$.

71. p. 223, 4.23: In the gambler's ruin problem show that

$$P[\text{she wins the next gamble} \mid \text{present fortune is } i, \text{ she eventually reaches } N]$$
$$= \begin{cases} p[1-(q/p)^{i+1}]/[1-(q/p)^i] & \text{if } p \neq \frac{1}{2}, \\ (i+1)/2i & \text{if } p = \frac{1}{2}. \end{cases}$$

72. p. 226, 4.33: Given that $\{X_n, n \geq 0\}$ is a branching process:

(a) Argue that either X_n converges to 0 or to infinity.

(b) Show that

$$\operatorname{Var}(X_n \mid X_0 = 1) = \begin{cases} \sigma^2 \mu^{n-1} \frac{\mu^n - 1}{\mu - 1} & \text{if } \mu \neq 1, \\ n\sigma^2 & \text{if } \mu = 1. \end{cases}$$

Note: assume that $p_1 \neq 1$. Also, part (a) is asking about the a.s. behavior of X_n.

73. Suppose that in a branching process, the number of offspring per individual has a binomial distribution with parameters $(2, p)$, where $0 < p < 1$. If the process starts with a single individual (generation 0), calculate:

(a) the probability of eventual extinction;

(b) the probability that the population becomes extinct for the first time in the second generation (i.e., the second generation is the earliest generation that has no individuals).

74. Consider a random walk on a network G that is either transient or is stopped on the first visit to a set of vertices Z. Let $G(x, y)$ be the corresponding Green function, i.e., the expected number of visits to y for a random walk started at x; if the walk is stopped at Z, we count only those visits that occur strictly before visiting Z. Show that for every pair of vertices x and y,

$$C_x G(x, y) = C_y G(y, x).$$

75. Consider an electrical network G with a vertex a and a set of vertices Z that does not include a. When a voltage is imposed so that a unit current flows from a to Z, show that the expected total number of times an edge (x, y) is crossed by a random walk starting at a and absorbed at Z equals $C_{xy}(V_x + V_y)$.

76. Consider an electrical network G with a vertex a and a set of vertices Z that does not include a. Show that $E_a[T_Z] = \sum_{x \in \mathsf{V}} C(x) V_x$ when a voltage is imposed so that a unit current flows from a to Z.

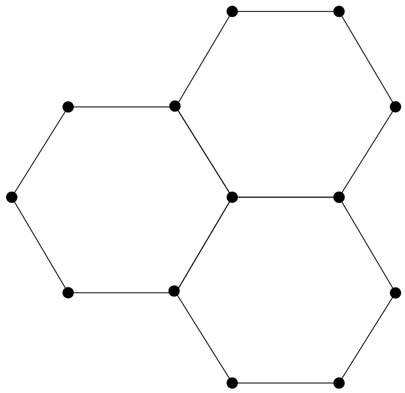
77. Let G be a network such that $\gamma := \sum_{e \in G} C_e < \infty$ (for example, G could be finite). For every vertex $a \in G$, show that the expected time for a random walk started at a to return to a is $2\gamma / C_a$.

78. Consider an electrical network G with a vertex a. Show that $\lim_n \mathcal{C}(a \leftrightarrow Z_n)$ is the same for every sequence $\langle G_n \rangle$ that exhausts G.

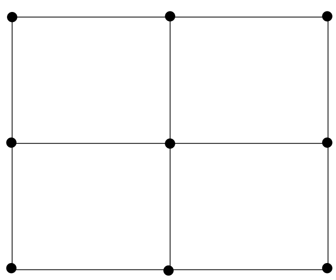
79. Let G be a network such that $\gamma := \sum_{e \in G} C_e < \infty$ and let a and z be two vertices of G. Let $x \sim y$ in G. Show that the expected number of transitions from x to y for a random walk started at a and stopped at the first return to a that occurs after visiting z is $C_{xy} \mathcal{R}(a \leftrightarrow z)$. This is, of course, invariant under multiplication of the edge conductances by a constant.

80. Let G be a network such that $\gamma := \sum_{e \in G} C_e < \infty$ and let a and z be two vertices of G. Show that the expected time for a random walk started at a to visit z and then return to a, the so-called "commute time" between a and z, is $2\gamma \mathcal{R}(a \leftrightarrow z)$.

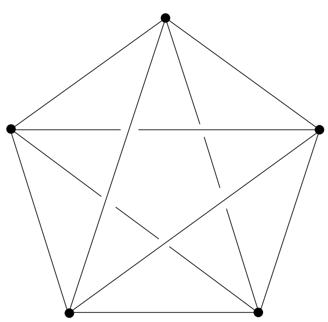
81. In the following networks, each edge has unit conductance.



What are $P_x[T_a < T_z]$, $P_a[T_x < T_z]$, and $P_z[T_x < T_a]$?



What is $\mathcal{C}(a \leftrightarrow z)$? (Or: show a sequence of transformations that could be used to calculate $\mathcal{C}(a \leftrightarrow z)$.)



What is $\mathcal{C}(a \leftrightarrow z)$? (Or: show a sequence of transformations that could be used to calculate $\mathcal{C}(a \leftrightarrow z)$.)

82. Find a (finite) graph that can't be reduced to a single edge by the four network transformations.

83. Let G be a connected graph with N edges and two vertices a and z of degree one with the same neighbor. Show that for simple random walk on G, $E_a[T_z] = 2N$.

84. p. 287, 5.3(b): Show that a continuous-time Markov chain is regular, given (a) that $\nu_i < M < \infty$ for all i or (b) that the discrete-time Markov chain with transition probabilities P_{ij} is irreducible and recurrent.

Note: do (a), but hand in only (b).

85. p. 287, 5.6: Verify the formula

$$A(t) = a_0 + \int_0^t X(s)ds,$$

given in Example 5.3(B).

86. p. 286, 5.1: A population of organisms consists of both male and female members. In a small colony each particular male is likely to mate with each particular female in each time interval of length h, with probability $\lambda h + o(h)$. Each mating immediately produces one offspring, equally likely to be male or female. Let $N_1(t)$ and $N_2(t)$ denote the number of males and females in the population at t. Derive the parameters of the continuous-time Markov chain $\{N_1(t), N_2(t)\}$.

87. Consider a Poisson process of rate 1 such that each event is classified as type i with probability α_i ($i = 0, 1$ and $\alpha_0 + \alpha_1 = 1$). Suppose that we put an event at time 0 that is classified as type i with probability p_i. Show that the two-state Markov chain with state i equal to type i and that changes state when the events of the Poisson process *change* type has transition rates $q_{01} = \alpha_1$ and $q_{10} = \alpha_0$. Use this connection to show that

$$P[\text{the chain is in state 0 at time } t] = p_0 e^{-t} + (1 - e^{-t})\alpha_0.$$

(This last equation could also be proved by using Example 5.4(A), but that is not what you are being asked to do here.)

88. p. 322, 6.3: Verify that X_n/m^n, $n \geq 1$, is a martingale when X_n is the size of the nth generation of a branching process whose mean number of offspring per individual is m.

Exercise not to Hand In: p. 323, 6.7: Let $X_1, \ldots$ be a sequence of independent and identically distributed random variables with mean 0 and variance σ^2. Let $S_n = \sum_{i=1}^n X_i$ and show that $\{Z_n, n \geq 1\}$ is a martingale when

$$Z_n = S_n^2 - n\sigma^2.$$

89. p. 323, 6.10: Consider successive flips of a coin having probability p of landing heads. Use a martingale argument to compute the expected number of flips until the following sequences appear:

(a) HHTTHHT

(b) HTHTHTH

90. Consider four players, A, B, C, and D, who play the following game: They begin with fortunes $a, b, c, d \in \mathbb{N}^+$, respectively. At each time, a pair of the players whose fortunes are strictly positive is chosen at random and a random one of that pair gives one unit to the other. Let X_n, Y_n, Z_n, W_n be their respective fortunes after n plays. What is the expected time until only one player's fortune is nonzero?

91. Let $\langle Z_n \rangle$ be asymmetric random walk on $\mathbb{Z}$ starting from $Z_0 = 0$; this means that $Z_n = \sum_{k=1}^{n} X_k$, where X_k are i.i.d. for $k \geq 1$ and equal ± 1 with probabilities p and $q := 1 - p$, respectively. Calculate the expected time until $Z_n \in \{a, -b\}$ for $a, b \in \mathbb{N}^+$ and $p > q$.

92. Let $\langle Z_n \rangle$ be asymmetric random walk on $\mathbb{Z}$ starting from $Z_0 = 0$. Calculate the variance of the time until $Z_n \in \{a, -b\}$ for $a, b \in \mathbb{N}^+$ and $p > q$. *Hint:* Use Problem 6.7.

93. Suppose $P[\mathrm{H}] = p$. Calculate the chance that HTHT appears before THTT.

94. Let X be 2 with probability p and -1 with probability $1 - p$, where $p > 1/3$. Let $\langle S_n \rangle$ be the corresponding random walk, $S_n := \sum_{i=1}^{n} X_i$, and N be the first time that the random walk is positive. Find the distribution of S_N, $E[S_N]$, and $E[N]$.

95. Let $\sigma > 0$ and $\mu \in \mathbb{R}$. In the following, we consider convergence only in the sense of finite-dimensional distributions.

(a) Show that if the steps of a random walk are of size $\sigma/\sqrt{n}$ taken each $1/n$ unit of time and are $\pm\sigma/\sqrt{n}$ with probability $\frac{1}{2} \pm \frac{\mu}{2\sigma\sqrt{n}}$ (for large n), then the random walk converges to Brownian motion with drift μ and variance parameter σ^2.

(b) Show that if the steps of a random walk are $\pm\sigma/\sqrt{n} + \mu/n$ taken each $1/n$ unit of time with probability $\frac{1}{2}$ each, then the random walk converges to Brownian motion with drift μ and variance parameter σ^2.

96. p. 399, 8.2: Let $W(t) = X(a^2 t)/a$ for $a > 0$. Verify that $W(t)$ is also Brownian motion. (Note: $X(t)$ is standard Brownian motion.)

97. Let $\langle X(t)\,;\ t \geq 0 \rangle$ be standard Brownian motion. Fix $r > 0$ and define $W(u) := X(r - u) - X(r)$ for $u \in [0, r]$. Argue that $\langle W(u)\,;\ 0 \leq u \leq r \rangle$ has the same distribution as $\langle X(t)\,;\ 0 \leq t \leq r \rangle$.

98. Verify the last step of the derivation of the Black–Scholes equation.

99. Suppose that $X(\cdot)$ is a geometric Brownian motion with drift parameter $\alpha > 0$. Show that $\lim_{t\to\infty} E\bigl[X(t)\bigr] = \infty$. Show that if the variance parameter $\sigma^2 > 2\alpha$, then $\lim_{\mathbb{N}\ni t\to\infty} X(t) = 0$ a.s. (Later, we will show this latter limit holds without the restriction that $t \in \mathbb{Z}$.)

100. Let $X(\cdot)$ be Brownian motion and T_a be the hitting time of a. Prove that $E[\sqrt{T_a}\,] = \infty$ for $a \ne 0$.

101. p. 400, 8.7: Let $\{X(t), t \ge 0\}$ denote Brownian motion. Find the density of:
(a) $|X(t)|$.
(b) $\Bigl|\min\limits_{0\le s\le t} X(s)\Bigr|$.
(c) $\max\limits_{0\le s\le t} X(s) - X(t)$.
(Note: I changed "distribution" to "density".)

102. p. 401, 8.14: Let T_x denote the time until Brownian motion hits x. Compute $P[T_1 < T_{-1} < T_2]$.

103. Let $\langle X(t)\,;\ t \ge 0\rangle$ be standard Brownian motion, $a, b > 0$, $\mu \in \mathbb{R}$. Let L_y be the line through $(0, y)$ with slope μ. Let τ_y be the first time $\bigl\langle \bigl(t, X(t)\bigr)\,;\ t \ge 0 \bigr\rangle$ hits L_y, i.e.,

$$\tau_y := \inf\bigl\{t\,;\ \bigl(t, X(t)\bigr) \in L_y\bigr\}.$$

Prove that $\tau_a \wedge \tau_{-b} < \infty$ a.s. and calculate $P_0[\tau_a < \tau_{-b}]$.

104. Let $X(\cdot)$ be a Brownian motion with parameters $(0, \sigma^2)$ and $X(0) \equiv 0$. Let $s, t, u, w > 0$. Compute $E\bigl[X(s)X(s+t)X(s+t+u)X(s+t+u+w)\bigr]$.

105. Suppose that $X_1, \ldots, X_n$ are random variables. Show that the following conditions are equivalent:
(a) $\forall a_1, \ldots, a_n \in \mathbb{R}\ \ \sum_{k=1}^n a_k X_k$ is a normal random variable;
(b) $\langle X_k\,;\ 1 \le k \le n\rangle$ has a multivariate normal distribution.

106. p. 399, 8.1: Let $Y(t) = tX(1/t)$.

(a) What is the distribution of $Y(t)$?

(b) Compute $\mathrm{Cov}(Y(s), Y(t))$.

(c) Argue that $\{Y(t), t \geq 0\}$ is also Brownian motion.

(d) Let

$$T = \inf\{t > 0\,;\ X(t) = 0\}.$$

Using (c) present an argument that

$$P[T = 0] = 1.$$

107. Prove that for Brownian motion $X(\cdot)$ with drift μ, we have $\lim_{t\to\infty} X(t)/t = \mu$ a.s.

Exercise not to Hand In: p. 399, 8.3: Compute the conditional distribution of $X(s)$ given that $X(t_1) = A$ and $X(t_2) = B$, where $t_1 < s < t_2$.

108. Let $X(\cdot)$ be standard Brownian motion. Define $Z(s) := (1-s)X\big(s/(1-s)\big)$ for $0 \leq s < 1$ and $Z(1) := 0$. Show that $Z(\cdot)$ is a standard Brownian bridge.

109. Let $\langle X(t)\,;\ t \in I\rangle$ be a Gaussian process on an interval, I. Write

$$Q(s,t) := \mathrm{Cov}\big(X(s), X(t)\big).$$

Show that $X(\cdot)$ has the Markov property iff for all $s_1 < t < s_2$ in I, we have

$$Q(s_1, s_2)Q(t,t) = Q(s_1, t)Q(s_2, t).$$

110. Let $\langle X(s)\,;\ 0 \leq s \leq t\rangle$ be a Brownian bridge from $X(0) = A$ to $X(t) = B$. Fix $t' \in (0,t)$. Given that $X(t') = C$, show that $\langle X(s)\,;\ 0 \leq s \leq t'\rangle$ is a Brownian bridge from A to C, that $\langle X(s)\,;\ t' \leq s \leq t\rangle$ is a Brownian bridge from C to B, and that these two bridges are independent.

111. p. 402, 8.21: Verify that if $\{B(t), t \geq 0\}$ is standard Brownian motion, then $\{Y(t), t \geq 0\}$ is a martingale with mean 1, when $Y(t) = \exp\{cB(t) - c^2t/2\}$. (Here, $c \in \mathbb{R}$.)

112. Let $\langle X(t)\rangle$ be a Brownian motion with parameter (μ, σ^2), where $\mu \neq 0$, $X(0) \equiv 0$, and let $a, b > 0$. Use Exercise 111 with $c := -2\mu/\sigma$ to show that

(a) $E\Big[\exp\big\{-2\mu X(T_a \wedge T_{-b})/\sigma^2\big\}\Big] = 1$ and

(b) $\max_{t\geq 0} X(t) \sim \mathrm{Exp}(-2\mu/\sigma)$ if $\mu < 0$.

(c) Use $c := -\mu/\sigma$ to show that $P[T_{-b} \geq t] \leq e^{-\mu b - \mu^2 t/2}$ if $\mu < 0$.

(d) Use (a) to give a new calculation of $P[T_a < T_{-b}]$.

113. Use the martingale $\langle X(t) - \mu t \rangle$ to calculate $E[T_a \wedge T_{-b}]$ for a Brownian motion starting at 0 with drift $\mu \neq 0$ and $a, b > 0$. Use a martingale to show that for standard Brownian motion, $E[T_a \wedge T_{-b}] = ab$.

114. Show that $X \succcurlyeq Y \iff \forall a \;\; P[X \geq a] \geq P[Y \geq a] \iff X^+ \succcurlyeq Y^+$ and $X^- \preccurlyeq Y^-$.