ST202 Stochastic Processes

Lecturer: Saul D. Jacka
Author: Wilfrid S. Kendall
Department of Statistics, University of Warwick

Outline notes Autumn Term 2005-2006

IMPORTANT: Lectures / syllabus for later years may vary from these older notes!

June 3, 2006

Contents

Administrative details

We begin with various administrative details of the module, also to be found on the handout distributed during the first lecture. It is assumed you have read these details carefully: scarce lecture time will not be spent on going through them in detail.

Aims and objectives

This module *aims* to introduce the idea of a stochastic process, and to show how simple probability and matrix theory can be used to build this notion into a beautiful and useful piece of applied mathematics. Its *objectives* are as follows: at the end of the module students will:

(a) understand the notion of a Markov chain, and how simple ideas of conditional probability and matrices can be used to give a thorough and effective account of discrete-time Markov chains;

(b) understand notions of long-time behaviour including transience, recurrence, and equilibrium;

(c) know how to apply these ideas to answer basic questions in several applied situations including genetics, branching processes and random walks.

Module assessments

We employ a successful system of assessment originally introduced by Prof. Saul Jacka for ST213. The assessed component will be conducted as follows: exercise sheets will be handed out approximately every week, totalling 8 sheets. In the middle of the lecture one week after an even-numbered sheet has been handed out there will be a brief test: you produce an answer to a modest variation on an aspect of one of the questions on the previous two sheets, specified at the start of the test. The tests will be marked[1], and the assessed component will be based on the best 3 out of 4 of your answers. A model answers to the assessed question will be attached to your marked answers when they are returned to you. Spare handouts will be placed by the Statistics undergraduate pigeonholes.
Note also that some lectures will depend on your understanding of results which you are asked to investigate as part of these exercises.

[1] **Note:** as in all Warwick Statistics examinations, full marks cannot be obtained unless you *justify your answer!*

This method helps you learn *during* the lecture module and to keep up to date with the lectures, so should

- improve your exam marks;
- increase your enjoyment of the module;
- cost less time than end-of-term assessment.

WARWICK
Home Page
Title Page
Contents
Page 5 of 237
Go Back
Full Screen
Close
Quit

Learning Resources

- *Lectures, also an examples class about once a fortnight;*
  It is obvious, but important to note, you should expect to think about and to work on the module material at other times than simply in the lectures: it may help to bear in mind that 1 CAT of credit is supposed to correspond loosely to 10 hours of study. Thus a 12 CAT module corresponds to 120 hours of study. Lectures account for just 30 of these hours, with a further 45 (say) for revision *etc*, so as a rule of thumb each 1 lecturing hour in a week should be balanced by on the order of 1.5 hours work on your own (but your mileage will vary . . .).

- *When possible, I allow time at end of lectures for you to come and ask questions;*
  I try when possible to allow some time at the end of each lecture for people to come and ask me questions; often this is the best way of quickly clearing up a point of confusion.

- *Office hours as given above;*
  occasionally I have to be absent but then will post a notice of my absence on the noticeboard by my door, if at all possible at least one day before.

WARWICK
Home Page
Title Page
Contents
Page 7 of 237
Go Back
Full Screen
Close
Quit

- *Module notes.* Available *via* links from home page as above. Released on the web in sections throughout the module; final printed version placed in Library SRC after the end of module; It is my firm and established policy *not* to make the entire set of notes available at the start of the module. The primary reason for this is that it is helpful for you to make your own notes as you attend the lectures, in order properly to master the concepts involved. (The secondary reason is that I update the notes throughout the module every year: an up-to-date version *will not exist* before the end of term!) The online notes should be used in the revision process, to give you a second view on what has been said, and they may also be useful in checking dubious points in your personal notes. The desperate, or the IT-challenged, can look for the printed version in the library SRC after the end of the term in which the lectures have taken place.

- I have set up a collection of *online quizzes* at

  http://www.warwick.ac.uk/go/st202/st202q.pdf

  to help you check your understanding of various sections. It is entirely up to you to make use of this resource! In due course I expect also to experiment with animations . . . .

- *Previous years' Statistics examination papers, often together with (rough) outline solutions,* can be obtained from the web (youĺl need to sign in using your `userid` with password):

  http://www2.warwick.ac.uk/elearning/projects/exampapers/

Books

See the bibliography at the end of the online notes for details of books which may help you understand the concepts described in this module.

The book by Grimmett and Stirzaker (1982) on probability and random processes is a good general reference which would serve you well not just for this module but also for the third-year module ST333 *Applied Stochastic Processes*. Norris (1997) specializes in Markov chains rather than applications, and goes rather further into the theory than this module does. Häggström (2002) gives a concise introduction oriented towards the exciting new field of randomized algorithms. Ross (1997) is a good broad introduction to applied probability. Jones and Smith (2001) gives a straightforward treatment of many of the topics in this module and of the module ST333 *Applied Stochastic Processes*.

Technical details

The notes are in Acrobat pdf format: this is a good way to disseminate information which includes mathematical formulae, and can be read using widely available *free* software (*Adobe Acrobat Reader*: follow links from Adobe Acrobat Reader downloads homepage; it is also on numerous CD-ROMS accompanying the more reputable computer magazines!). Acrobat pdf format allows me to include numerous hypertext references and I have made full use of this. As a rule of thumb, clicking on coloured text is quite likely to:

- move you to some other relevant text, either in the current document such as here: Aims and objectives, or occasionally in supporting documents or other module notes;
- or launch a Worldwide Web browser as in

  `http://www.warwick.ac.uk/go/wskendall/`

  (assuming your system has configured a browser appropriately).

You should notice that *Adobe Acrobat Reader* includes a facility for going *back* from the current page to the previously visited page: this can be accessed by the "Go Back" button should you get lost!

Go Back

If you try to print out these pages you are likely to discover a snag! In order for them to fit comfortably on computer screens I have formatted them to fit on the computer screen; however modern printer drivers will allow you to print *via Adobe Acrobat Reader* in two-by-two format, which comes out just about right.

Health warning

These notes cover *more* than the content of the lecture module. The examination will be designed on the basis of the material covered by the lectures and exercises only! However the lectures will make liberal use of topics from the first-year modules Probability A and B: (ST111/ST112). See Appendices A.1 and A.2 for sketch overviews of some of the more significant results from these modules.

Chapter 1

Stochastic processes

Contents

What is a *stochastic process*? The same as a *random process*: a random phenomenon evolving in time.

Definition 1.1 *A* stochastic process X *is a family* $\{X_t : t \in \mathcal{T}\}$ *of* random variables *indexed by a* time-set $\mathcal{T}$.

In this module $\mathcal{T}$ is always the set of non-negative integers; *discrete time*.

1.1. Illustrative examples

We begin with examples to show that the above abstract definition leads to many useful possibilities.

Example 1.2 ***Gambler's ruin.*** *We toss a sequence of fair coins. You pay me £1 for each head. I pay you £1 for each tail. I start with £10 and you start with £100. We* ***stop*** *when first one loses all.*

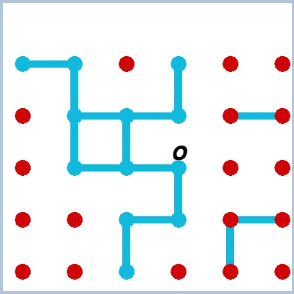
Here $X_n = x$ if you have £x after toss n, so $X_0 = 100$. Then $\{X_0, X_1, \ldots\}$ is a stochastic process. We analyse X later.

Question 1.3 *Questions which are important in applications:*

- *What is the chance you win?*
- *What is your expected gain?*
- *What is the expected number of coin tosses?*

See Malkiel (1973) for possible applications to finance.

Example 1.4 ***Percolation problems.*** *Consider a square lattice of pipes. Each pipe independently is* blocked *to the passage of liquid with probability* $1 - p$. *For site* n, *the random variable* X_n *equals zero if the liquid cannot flow from an* input site *at the origin* $\mathbf{o}$ *to* $\mathbf{n}$; *otherwise* X_n *equals* 1.

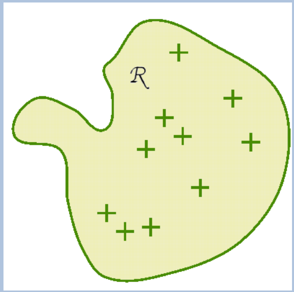


Question 1.5 *Questions which are important in applications:*

- *What is the probability that only finitely many* X*'s equal* 1?
- *How does this depend on the blocking probability* $1 - p$?

See `http://en.wikipedia.org/wiki/Aquifer` for a typical phenomenon from geology which motivates this kind of model.

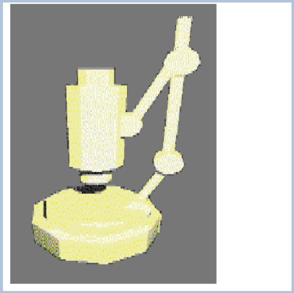
Example 1.6 ***Random point patterns.*** *Locations of major ore bodies in a given geographical region* $\mathfrak{R}$. *For each subset* $D \subseteq \mathfrak{R}$ *let* $X(D)$ *be the number of locations in* D.

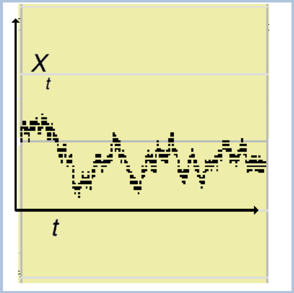


Question 1.7 *Questions which are important in applications:*

- *What stochastic models are appropriate for* X*?*
- *How to relate such a model to other geological knowledge?*

Example 1.8 ***Brownian motion.*** *A pollen particle, lit from the side, is observed through a microscope. Let X_t be the distance of the particle from a vertical reference plane at time t.*





Question 1.9 *Questions which are important in applications:*

- *What stochastic models are appropriate for X?*
- *What is the distribution of the random variable T, the first time the particle hits the reference plane?*

See http://www.sciences.demon.co.uk/wbbrowna.htm for some historical background.

1.2. Examples of state-spaces

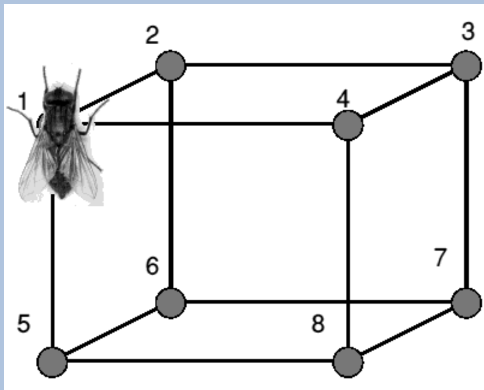
Definition 1.10 *The* state-space *of a random process* X *is the set of values taken by the various* X_t.

The variety of possible state-spaces is demonstrated from the examples above.

- The integers $\mathbb{Z}$ (gambler's ruin);
- the set $\{0, 1\}$ corresponding to "dry" and "wet" (percolation);
- $\{0, 1, 2, \ldots\}$ counting points in planar subsets (random point patterns);
- the real line $\mathbb{R}$ (Brownian motion).

WARWICK
Home Page
Title Page
Contents
Page 19 of 237
Go Back
Full Screen
Close
Quit

We will deal only with the case where the state-space is *discrete*: either the integers (as in gambler's ruin) or a finite subset. By re-labelling we can consider a finite subset as a subset of the integers. For example: a fly walking from vertex to vertex of a cube: label the cube vertices from 1 to 8.



WARWICK
Home Page
Title Page
Contents
Page 20 of 237
Go Back
Full Screen
Close
Quit

However even finite state-spaces can be *very* large.

Example 1.11 ***Changing words.*** *In a simple word-game, you are given a start word (say, "rat") and a finish word (say, "dog"), with the same number* $n = 3$ *of letters. You have to change from start to finish by changing one letter at a time, such that all intermediate steps are also words. So one solution is:*

`rat    rot   dot    dog`

Model the change process as a Markov chain: at each step choose a letter of the current word at random, and propose a change to one of the 25 other possible letters also at random. The step is blocked if the result is not a word.

The state space is large for $n > 2$*! (Less than* 26^n*, but maybe not much smaller ...) If the start and finish words are at all long then it is much easier to describe the chain as above rather than write down the state-space and the probabilities of making transitions from one state to another.*

We will focus on discrete-time, discrete-space random processes. The examples above show that there are all sorts of other variations, of great importance in applications: some are covered in ST333.

Chapter 2

Discrete Markov chains

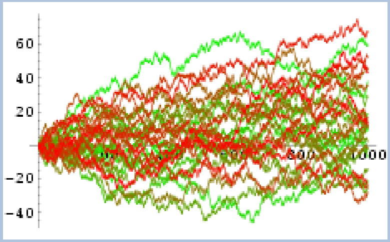
Contents

We will consider random processes using *discrete time* with *discrete time-space* satisfying the *Markov property*.

Example 2.1 ***Simple symmetric random walk.*** *State-space is* $\mathcal{S} = \mathbb{Z}$, *the integers* $\{0, \pm 1, \pm 2, \ldots\}$. *The process evolves by*

$$X_n \quad = \quad X_{n-1} \pm 1$$

with probabilities for jumps ± 1 *to be chosen below. We start* X *at* $X_0 = 0$ *for example.*
The sample space *is path-space, the space of all possible paths (infinite lists of integers, each differing from its predecessor by* ± 1*).*

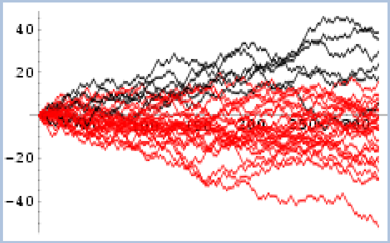


Thirty randomly chosen paths of length 1000.

WARWICK
Home Page
Title Page
Contents
Page 24 of 237
Go Back
Full Screen
Close
Quit

An event *is a subset of path-space, for example*

$$A \quad = \quad [\,\textit{paths staying below } 24 \textit{ up to time } n = 320\,]\ .$$



Thirty randomly chosen paths of length 320*. Of these,* 24 *do not exceed level* 24*.*

WARWICK
Home Page
Title Page
Contents
Page 25 of 237
Go Back
Full Screen
Close
Quit

How should one specify the probability $\mathbb{P}$ in the above?
One way, simple but inflexible: require that "all paths are equally likely". Then

1. $\mathbb{P}\left[\,\text{particular infinite path}\,\right] = 0$ so work with initial segments only: all initial segments of fixed initial length are equally likely.

2. Check consistency with the rule for evolution that resulting jumps are independent and equally likely to be $+1$ or -1 (a basic ST111 calculation!);

This approach is *special* to simple symmetric random walks (what for example if the $+1$ jump is more likely than the -1 jump?) so how can we do better? However we will come back to this approach in Section 4.3.

The main points to carry forward: From the "equally likely paths" condition we can calculate the conditional probability of a jump as follows:

$$\begin{aligned}\mathbb{P}\left[X_{n+1}=x_{n+1} \mid X_n=x_n, X_{n-1}=x_{n-1}, \ldots, X_0=x_0\right] &= \frac{1}{2} \\ \text{if } x_{n+1}=x_n \pm 1, & \\ &= 0 \\ \text{otherwise;} &\end{aligned}$$

and we notice

(a) The upcoming *jump* $X_{n+1}-X_n$ doesn't depend on the past (that is to say, the conditions that $X_n=x_n$, $X_{n-1}=x_{n-1}$, *etc*)

(b) $\mathbb{P}\left[\ldots \mid \ldots\right]$ doesn't depend on time n.

We shall carry these forward into a major generalization which is hugely useful: relax (a) slightly to allow $\mathbb{P}\left[\ldots \mid \ldots\right]$ to depend on the immediate present $X_n=x_n$ but nothing more.

Definition 2.2 ***Discrete Markov chain.*** *A random process* $X = \{X_n : n = 0, 1, \ldots\}$ *is said to be a* discrete state-space Markov chain *if*

(a) *the random variables* X_n *take values in a discrete state-space* $\mathcal{S}$*,*

(b) *the conditional probabilities of the future given the past depend only on the present:*

$$\begin{aligned} &\mathbb{P}\left[X_{n+1} = k_{n+1} | X_n = k_n, X_{n-1} = k_{n-1}, \ldots, X_0 = k_0\right] \\ &\quad = \quad \mathbb{P}\left[X_{n+1} = k_{n+1} | X_n = k_n\right] \end{aligned}$$

whenever both sides are well-defined, and for all n*, all* $k_0, \ldots,$ $k_{n+1} \in \mathcal{S}$.

Remark 2.3 *The conditional probabilities are not well-defined exactly when the conditioning event is impossible.*

We say that X has the *Markov property.* The correct intuition is:

$$\mathbb{P}\left[\text{ future } | \text{ present and past } \right] \quad = \quad \mathbb{P}\left[\text{ future } | \text{ present } \right]$$

Remark 2.4 *We can use the one-step transition probabilities to evaluate probabilities of events referring to path-space: see Theorem 2.11 below.*

Definition 2.5 ***Stationary transition probabilities.*** *The discrete state-space Markov chain is said to have* stationary transition probabilities *if the conditional probabilities*

$$p_{ij} \quad = \quad \mathbb{P}\left[X_{m+1} = j | X_m = i\right]$$

do not vary with time m whenever they are well-defined.

Notice that the stationarity of these one-step conditional probabilities can be shown to imply stationarity of more general conditional probabilities by the Markov property.

Example 2.6 ***A really simple meteorological example.*** *Consider an absurdly simplistic model: $X_n = 1$ if wet on day n and $X_n = 0$ if dry on day n; suppose X is a Markov chain with just two states $d = 0$ and $w = 1$, with transition probabilities*

- $\mathbb{P}\left[\,\textit{wet tomorrow} \mid \textit{wet today}\,\right] = 0.8$,
- $\mathbb{P}\left[\,\textit{dry tomorrow} \mid \textit{dry today}\,\right] = 0.7$.

We can deduce the other probabilities since, for example,

$$p_{dw} + p_{dd} \quad = \quad 1\,.$$

WARWICK
Home Page
Title Page
Contents
Page 30 of 237
Go Back
Full Screen
Close
Quit

Definition 2.7 ***Stationary transition probabilities and transition probability matrices.*** *If X is a discrete state-space Markov chain with stationary transition probabilities then*

(a) *the (one-step) transition probability (kernel) from j to k is denoted by*

$$p_{jk} \quad = \quad p(j,k) \quad = \quad \mathbb{P}\left[X_{m+1} = k | X_m = j\right]$$

(by stationarity this expression does not depend on time m);

(b) *the (one-step) transition probability matrix is denoted by*

$$\underline{\underline{P}} \quad = \quad [p(j,k)]_{j,k\in\mathcal{S}}\,;$$

(c) *similarly we speak of n-step transition probabilities and transition probability matrices, denoted by*

$$p^{(n)}_{jk} \quad = \quad p^{(n)}(j,k) \quad = \quad \mathbb{P}\left[X_{m+n} = k | X_m = j\right]$$

and

$$\underline{\underline{P}}^{(n)} \quad = \quad [p^{(n)}(j,k)]_{j,k\in\mathcal{S}}\,.$$

Remark 2.8 *The transition probability $p^{(0)}(j,k)$ is non-zero only when $j = k$ (when it equals 1). So $\underline{\underline{P}}^{(0)}$ is the identity matrix (0's off the diagonal, 1's on the diagonal).*

Remark 2.9 *We read $p(j,k)$ as "the probability that X goes from j to k in one step (conditional on the event that it begins at j)".*

Remark 2.10 *We often refer to transition probability matrices $\underline{\underline{P}}$ as* stochastic matrices. *They can be recognized as follows: the row sums are all equal to 1, and all entries are non-negative. For later use,* sub-stochastic matrices *are like stochastic matrices except that the row-sums can be less than 1. See Definition 3.1 below for a formal definition.*

Clearly, given this sort of information (the one-step transition probabilities, the initial state (or probabilities of being in various possible initial states) we can in principle compute any probability concerning the Markov chain. We say we know the *distribution* of the chain X if we know how to compute all such probabilities in principle.

Theorem 2.11 ***Distribution of discrete state-space Markov chain with stationary transition probabilities.*** *This is fully specified by (a) and (b) below.*

(a) its initial state X_0 (which could be random),

(b) its (one-step) transition probability matrix.

That is to say, given (a) and (b) we can calculate the probability of any event concerning the behaviour of X.

WARWICK
Home Page
Title Page
Contents
Page 33 of 237
Go Back
Full Screen
Close
Quit

Proof: Consider any *path* $\{k_0, k_1, \ldots, k_h\}$ over an initial time-segment 0, 1, …, h. Then

$$
\begin{aligned}
& \mathbb{P}\left[X \text{ follows } \{k_0, k_1, \ldots, k_h\}\right] \\
= \; & \mathbb{P}\left[X \text{ follows } \{k_0, k_1, \ldots, k_h\} | X \text{ follows } \{k_0, k_1, \ldots, k_{h-1}\}\right] \times \\
& \qquad \times \mathbb{P}\left[X \text{ follows } \{k_0, k_1, \ldots, k_{h-1}\}\right] \\
= \; & \mathbb{P}\left[X_h = k_h | X \text{ follows } \{k_0, k_1, \ldots, k_{h-1}\}\right] \times \\
& \qquad \times \mathbb{P}\left[X \text{ follows } \{k_0, k_1, \ldots, k_{h-1}\}\right] \\
= \; & \mathbb{P}\left[X_h = k_h | X_{h-1} = k_{h-1}\right] \times \mathbb{P}\left[X \text{ follows } \{k_0, k_1, \ldots, k_{h-1}\}\right] \\
= \; & \cdots = p(k_{h-1}, k_h) \times \\
& \qquad p(k_{h-2}, k_{h-1}) \times \ldots \times p(k_0, k_1) \times \mathbb{P}\left[X \text{ follows } \{k_0\}\right] \\
= \; & p(k_{h-1}, k_h) \times p(k_{h-2}, k_{h-1}) \times \ldots \times p(k_0, k_1) \times \mathbb{P}\left[X_0 = k_0\right]
\end{aligned}
$$

(where the colours indicate what is changing from line to line) which is what we require! $\square$

Example 2.12 ***A really simple meteorological example — continued.*** *Using simple arguments we find the one-step transition matrix:*

$$\underline{\underline{P}} \quad = \quad \begin{bmatrix} 0.7 & 0.3 \\ 0.2 & 0.8 \end{bmatrix}$$

Hence we can compute any probability we care to name. For example, if we start with a dry day on day 0 then the probability of a sequence $ddwd$ is given by $p_{dd}p_{dw}p_{wd} = 0.7 \times 0.3 \times 0.2 = 0.042$.

Remark 2.13 *The above definition and theorem have obvious modifications for the case of Markov chains with non-stationary transition probabilities. However the generalization of the following theory is then quite involved and not very revealing.*

It is important that matrix theory is able to give us clear formulae to do with computing probabilities for Markov chains. We shall see more of this later!

Theorem 2.14 ***Chapman-Kolmogorov equations.*** *Suppose X is a discrete state-space Markov chain. Then $\underline{\underline{P}}^{(n+m)} = \underline{\underline{P}}^{(n)}\underline{\underline{P}}^{(m)}$:*

$$p_{ij}^{(n+m)} = \sum_{k\in\mathcal{S}} p_{ik}^{(n)} p_{kj}^{(m)} . \tag{2.1}$$

Proof: Consider the conditional probability on the left-hand side:

$$\begin{aligned}
\mathbb{P}\left[X_{n+m}=j|X_0=i\right] &= \mathbb{P}\left[\left(\bigcup_{k\in\mathcal{S}}[X_{n+m}=j, X_n=k]\right)\middle| X_0=i\right] \\
&= \sum_{k\in\mathcal{S}} \mathbb{P}\left[X_{n+m}=j, X_n=k|X_0=i\right] \qquad \text{additive law} \\
&= \sum_{k\in\mathcal{S}} \mathbb{P}\left[X_{n+m}=j|X_n=k, X_0=i\right] \times \mathbb{P}\left[X_n=k|X_0=i\right] \\
&= \sum_{k\in\mathcal{S}} \mathbb{P}\left[X_{n+m}=j|X_n=k\right] \times \mathbb{P}\left[X_n=k|X_0=i\right] \qquad \text{Marko} \\
&= \sum_{k\in\mathcal{S}} p_{ik}^{(n)} p_{kj}^{(m)} \qquad \text{stationary transition probabilities}
\end{aligned}$$

as required. □

Comment for pedants. Note that the Markov property Definition 2.2 really only tells us

$$\mathbb{P}\left[X_{n+m}=k|X_n=i,X_0=j\right] \quad = \quad \mathbb{P}\left[X_{n+m}=k|X_n=i\right]$$

for the case $m=1$. But we can argue inductively: suppose it is true for $m-1$. Then

$$\begin{aligned}\mathbb{P}\left[X_{n+m}=k|X_n=i,X_0=j\right] \quad &= \\ \sum_{r\in\mathcal{S}}\mathbb{P}\left[X_{n+m}=k,X_{n+m-1}=r|X_n=i,X_0=j\right]&.\end{aligned}$$

Now

$$\begin{aligned} & \mathbb{P}\left[X_{n+m}=k,X_{n+m-1}=r|X_n=i,X_0=j\right] \\ = \quad & \mathbb{P}\left[X_{n+m}=k|X_{n+m-1}=r,X_n=i,X_0=j\right] \\ & \times\mathbb{P}\left[X_{n+m-1}=r|X_n=i,X_0=j\right] \\ = \quad & \mathbb{P}\left[X_{n+m}=k|X_{n+m-1}=r,X_n=i\right]\times\mathbb{P}\left[X_{n+m-1}=r|X_n=i\right] \\ = \quad & \mathbb{P}\left[X_{n+m}=k,X_{n+m-1}=r|X_n=i\right]\end{aligned}$$

where the last step but one uses the Markov property twice, once by way of the inductive hypothesis $\mathbb{P}\left[X_{n+m-1}=r|X_n=i,X_0=j\right] = \mathbb{P}\left[X_{n+m-1}=r|X_n=i\right]$. Hence

$$\begin{aligned} & \mathbb{P}\left[X_{n+m}=k|X_n=i,X_0=j\right] \\ = & \sum_{r\in\mathcal{S}} \mathbb{P}\left[X_{n+m}=k,X_{n+m-1}=r|X_n=i\right] \\ = & \mathbb{P}\left[X_{n+m}=k|X_n=i\right] . \end{aligned}$$

Theorem 2.15 ***Distribution of X_n.*** *Suppose the vector $a^{(n)}$ is given by*

$$a_i^{(n)} \quad = \quad \mathbb{P}\left[X_n = i\right] .$$

Then $a^{(n)} = a^{(0)} \underline{\underline{P}}^{(n)}$.

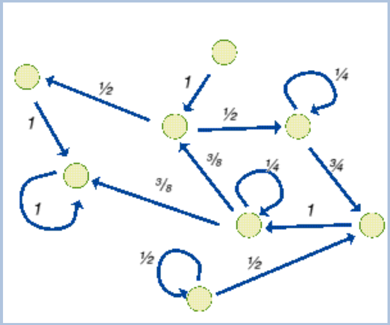
Proof:

$$\begin{aligned} a_i^{(n)} &= \mathbb{P}\left[X_n = i\right] = \sum_{k \in \mathcal{S}} \mathbb{P}\left[X_n = i | X_0 = k\right] \times \mathbb{P}\left[X_0 = k\right] \\ &= \sum_{k \in \mathcal{S}} p^{(n)}(k,i) a_k^{(0)} \end{aligned}$$

as required. $\square$

2.1. Examples

Later on we will need the convenient notion of a *state diagram* (directed graph, vertices the states, directed edges the possible one-step transitions). The following figure gives an example of this. A careful examination shows that the Markov chain given in pictorial form here has a state-space which somehow breaks down into separate regions, according to whether or not one can get from one state to another. In the following we examine this idea in more detail.



Example 2.16 ***Simple random walk.*** *Suppose* $\{Y_n : i \geq 1\}$ *is a sequence of independent identically distributed random variables with common distribution*

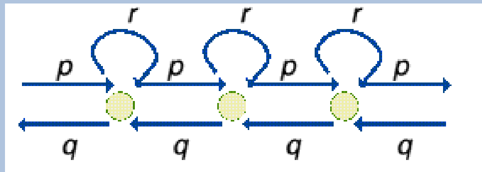
where $p + q + r = 1.$

$$\begin{aligned}
\mathbb{P}\left[Y = -1\right] &= q \\
\mathbb{P}\left[Y = 0\right] &= r \\
\mathbb{P}\left[Y = 1\right] &= p
\end{aligned}$$

Then the sequence of partial sums $\{X_n = X_{n-1} + Y_n : n \geq 0\}$ forms a Markov chain with tridiagonal transition probability matrix

$$\begin{bmatrix}
 & & \dots & & & & & \\
\dots & 0 & q & r & p & 0 & 0 & \dots \\
\dots & 0 & 0 & q & r & p & 0 & \dots \\
\dots & 0 & 0 & 0 & q & r & p & \dots \\
 & & \dots & & & & &
\end{bmatrix}$$

with the starting condition $a_i^{(0)} = \mathbb{P}\left[X_0 = i\right]$. The distribution of the Y_n is called the *jump distribution* of the random walk. The state diagram is as follows:



Question 2.17

- *with what frequency will X return to the origin 0?*
- *how long will X take to reach $\pm k$ if started at 0?*

Variations on random walks.

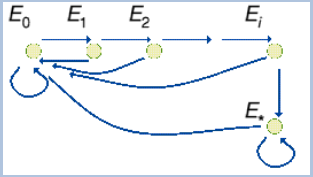
(a) absorbing or reflecting barriers,

(b) simple random walk on a lattice $\mathbb{Z}^2$ or $\mathbb{Z}^3$,

(c) replace jump distribution by arbitrary distribution on $\mathbb{Z}$.

Example 2.18 ***Sampling inspection or quality control.*** *Production line strategy: test every item until $i+1$ consecutive non-defective items are found. Then test each item at random with probability $1/r$ until a defective is found. Then test every item until ....*

Suppose defective probability is p and items are independent. The state-space can be taken to be



$$\{E_0, E_1, \ldots, E_i, E_*\}$$

where E_* is state of sampling 1 in every r while E_k is state of sampling every item with last k items being non-defective. Let X_m be the state at the time the m^{th} item is tested. Then $\{X_m : m \geq 0\}$ forms a Markov chain with transition matrix

$$\begin{bmatrix} p & q & 0 & 0 & \ldots & 0 \\ p & 0 & q & 0 & \ldots & 0 \\ p & 0 & 0 & q & \ldots & 0 \\ & & & \ldots & & 0 \\ p & 0 & 0 & 0 & \ldots & q \\ p/r & 0 & 0 & 0 & \ldots & 1-p/r \end{bmatrix}$$

Remark 2.19 *Why p/r? Consider*

$$(1/r) \times p \quad = \quad \mathbb{P}\left[\,\textit{item is chosen}\,\right] \times \mathbb{P}\left[\,\textit{item is defective}\,\right]$$

Remark 2.20 $q = 1 - p$ *in the above.*

Question 2.21

- *what is the long-run proportion of defectives among the items let through this procedure?*
- *what are the best choices for i and r given costs of inspection and penalties for letting faulty items through?*

Example 2.22 ***Brand switching.*** *Two supermarkets A and B with B inferior to A. If a customer shops at A having previously tried B then the customer never returns to B. The transition matrix (with appropriate coding of states $1 = A$ never having tried B; $2 = B$; $3 = A$ having tried B) might be:*

$$\begin{bmatrix} 0.8 & 0.2 & 0.0 \\ 0.0 & 0.7 & 0.3 \\ 0.0 & 0.0 & 1.0 \end{bmatrix}$$

Example 2.23 ***Queue in discrete time.*** *Customers arrive for service by single server, served in order of arrival. At each point of (discrete) time there is probability* p *of arrival, and arrivals at different time points are independent. So inter-arrival times* $\{A_n : n \geq 0\}$ *are independent identically distributed with Geometric distribution* $\mathbb{P}\left[A_n = r\right] = (1-p)^{r-1}p$*. The service times of customers* $\{S_n : n \geq 0\}$ *are independent identically distributed and independent of arrivals with common distribution*

$$\mathbb{P}\left[S_n = r\right] \quad = \quad b_r$$

where $b_1, b_2, \ldots$ *are fixed non-negative numbers adding up to* 1.

[NOTATION: this is an $M/G/1$ queue in D.G. Kendall's terminology.]

The number of people in the queue at time m is not in general a Markov chain! However let Q_n be the number in the queue at time of n^{th} departure. Then $\{Q_n : n \geq 0\}$ is a Markov chain (we call it an imbedded Markov chain) since $Q_{n+1} - Q_n$ is the number of people arrived during the service time of the $(n+1)^{st}$ customer minus 1 (the $(n+1)^{st}$ customer leaves at the end of service). For the difference

$Q_{n+1} - Q_n$ is independent of the past $Q_0, Q_1, \ldots, Q_{n-1}$ once Q_n is known.

Let N_n be the number of arrivals to the queue in the time-interval during which customer n is being served. Then

$$Q_{n+1} \quad = \quad \begin{cases} Q_n - 1 + N_{n+1} & \text{if } Q_n > 0 \\ N_{n+1} & \text{if } Q_n = 0 \end{cases}$$

and it can be shown that

$$\begin{aligned} &\mathbb{P}\left[N_n = j\right] = k_j = \\ &\quad = \sum_{r=1\ldots\infty} \mathbb{P}\left[j \text{ arrivals in service time} \mid \text{service time } = r\right] b_r \\ &\quad = \sum_{r=1\ldots\infty} ({}^r C_j) p^j (1-p)^{r-j} b_r \,. \end{aligned}$$

Hence a transition probability matrix for the Markov chain Q:

$$\begin{bmatrix} k_0 & k_1 & k_2 & k_3 & \dots \\ k_0 & k_1 & k_2 & k_3 & \dots \\ 0 & k_0 & k_1 & k_2 & \dots \\ 0 & 0 & k_0 & k_1 & \dots \\ & \dots & \dots & & \end{bmatrix}$$

Question 2.24

- *does the number in the queue exhibit equilibrium behaviour in a statistical sense?*
- *what is the distribution of the waiting time for a typical customer?*
- *what is the distribution of a busy period for the server?*

2.2. Classification of states - communicating classes

Suppose it is possible to travel from state a to state b and *vice versa*. Then states a and b will share many properties in common. Suppose $\underline{\underline{P}}$ is a stochastic matrix for a Markov chain with state space $\mathcal{S}$.

Definition 2.25 *Suppose $i, j \in \mathcal{S}$. Then*

(a) i leads to j *if there is some $n > 0$ with $p_{ij}^{(n)} > 0$. We write $i \to j$.*

(b) i, j intercommunicate *if $i \to j$ and $j \to i$. We write $i \leftrightarrow j$.*

(c) i, j *are in the same* communicating class of states *if either*

(i) $i \leftrightarrow j$ *or*

(ii) $i = j$.

and then we write $i \sim j$.

Theorem 2.26 *The relation $\sim$ is indeed an equivalence relation (it is reflexive, symmetric, and transitive).*

Proof: Most of this is rather obvious:

reflexive: $i \sim i$ since $i = i$;

symmetric: $i \sim j$ means $j \sim i$ direct from definition;

transitive: Suppose $i \sim j$ and $j \sim k$. Then $i \sim k$. If either of $i = j$ or $j = k$ then $i \sim k$ follows immediately. Otherwise: if $i \to j$ and $j \to k$ then we have $p_{ij}^{(n)} > 0$ and $p_{jk}^{(m)} > 0$. But it then follows from the Chapman-Kolmogorov equations (Theorem 2.14) that

$$p_{ik}^{(n+m)} \quad \geq \quad p_{ij}^{(n)} \times p_{jk}^{(m)}$$

and so $i \to k$ …now an obvious modification of the argument shows $k \sim i$.

$\square$

Thus in Example 1.11 we see that "rat" and "dog" are in the same communicating class. In the 4-letter variant, what about "pass" and "fail"? Decisions about intercommunicability are not always trivial! Division into communicating classes depends solely on the pattern of non-zero entries in the transition probability matrix $\underline{\underline{P}}$. To analyze the behaviour of the chain it more or less suffices to know how it behaves in each communicating class.

Definition 2.27 *(a) the state $i \in \mathcal{S}$ is* essential *if for all $j \in \mathcal{S}$ we have $i \to j$ implies $j \to i$. Otherwise i is* inessential.

(b) the stochastic matrix $\underline{\underline{P}}$ is irreducible *if all states intercommunicate, so that the state space $\mathcal{S}$ is one single communicating class.*

Examples:

- Consider brand-switching example: "A not B", B are inessential states. "A after B" is essential.

- Consider random walk example with *absorbing barriers* at 0, n (walk stops if ever it reaches 0 or n). Then the states $1, \ldots, n-1$ form one inessential communicating class. State 0 on its own forms an essential communicating class, as does n.

- Consider quality control, queue, random walk with no barriers. Except in degenerate cases (eg $p = 0$ for the random walk example) the whole state-space is a single (therefore essential) communicating class, and hence $\underline{\underline{P}}$ is irreducible.

WARWICK
Home Page
Title Page
Contents
Page 52 of 237
Go Back
Full Screen
Close
Quit

The proof of the following lemma is just a matter of thinking about the definitions!

Lemma 2.28 *Suppose $i \longrightarrow k$ and i is essential. Then k is essential. Furthermore, $i \sim k$.*

Proof: If $i \longrightarrow k$ and i is essential then, by Definition 2.27, $k \longrightarrow i$. Hence $i \sim k$. Moreover suppose $k \longrightarrow j$ for some other state j. Then $i \longrightarrow k$ and $k \longrightarrow j$ combine to show $i \longrightarrow j$. Therefore $j \longrightarrow i$, since i is essential. But now $j \longrightarrow i$ and $i \longrightarrow k$ combine to show $j \longrightarrow k$, and this conclusion shows that k too is essential, as required. $\square$

2.3. Classification of states - periodicity

Definition 2.29 ***Periodicity.*** *The* period *of a state is the greatest common divisor of times at which the chain might return to the state. Thus*

(a) the state i has period gcd$\{n > 0 : p_{ii}^{(n)} > 0\}$ if $i \to i$.

(b) if $i \to i$ does not hold then the period of i is not defined.

(c) if i has period 1 then it is said to be aperiodic.

Lemma 2.30 *Suppose that $i \to i$ and i has period d. Then there is $N > 0$ such that for all $k \geq N$ we have $p_{ii}^{(kd)} > 0$.*

The proof is in Appendix B.
In words, if a state i has period d then the chain has positive chance of returning to i at time kd for *all* sufficiently large integers k.

WARWICK
Home Page
Title Page
Contents
Page 54 of 237
Go Back
Full Screen
Close
Quit

Theorem 2.31 *Periodicity is a class property.*

Proof:
Notation: $a|b$ means positive integer a divides integer b. Suppose that $i \leftrightarrow j$ and i has period d.
Since $i \leftrightarrow j$ we can find n, m such that $p_{ij}^{(n)}$, $p_{ji}^{(m)}$ are both positive.
Since period$(i) = d$ we know $p_{ii}^{(k)} > 0$ means $d|k$.
If $p_{jj}^{(r)} > 0$ then $p_{ii}^{(n+r+m)} \geq p_{ij}^{(n)} p_{jj}^{(r)} p_{ji}^{(m)} > 0$ (follow the paths!) and so by definition of periodicity $d|n + r + m$. Since we also know $d|n + m$ (arguing from $p_{ii}^{(n+m)} \geq p_{ij}^{(n)} p_{ji}^{(m)} > 0$) it follows $d|r$ and hence state j has period d as required. $\square$

Definition 2.32 ***Cyclically moving subclasses.*** *Suppose C is a communicating class with period d then for all $i \in C$ and all integers $r > 0$*

$$G_r(i) \quad = \quad \{j \in C : p_{ij}^{(kd+r)} > 0 \textit{ for some } k\}$$

is a cyclically moving subclass *of C.*

Home Page

Title Page

Contents

Page 55 of 237

Go Back

Full Screen

Close

Quit

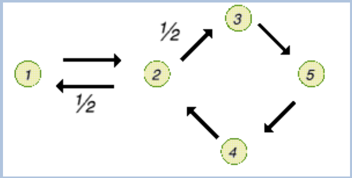
Remark 2.33 *$G_{d+r}(i) = G_r(i)$ and $G_0(i), \ldots, G_{d-1}(i)$ form a partition of C. Note also if C is essential and $j \in G_r(i)$ then*

$$\sum_{h \in G_{r+n}(i)} p_{jh}^{(n)} \quad = \quad 1$$

so the chain moves steadily through the cyclically moving subclasses. The transition probability matrix exhibits a corresponding block form. Moreover $\underline{\underline{Q}} = \underline{\underline{P}}^d$ decomposes into d irreducible aperiodic stochastic submatrices arranged down its diagonal. This means if we sample the chain at every d time-steps then it will appear to be moving only in one of the cyclically moving subclasses.

The notion of cyclically-moving subclasses can be used to prove the following result: if X is an irreducible Markov chain on a state-space $\mathcal{S}$ then consider the "double" (X, X') on $\mathcal{S} \times \mathcal{S}$, where X, X' are independent copies. The double is irreducible if and only if X has period 1 (is aperiodic), and otherwise has d communicating classes where d is the period of X.

Example 2.34 ***Periodicity.*** *Consider the Markov chain with the following state-diagram and transition matrix:*



$$\begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ \frac{1}{2} & 0 & \frac{1}{2} & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix}$$

and notice it is irreducible. Now write down all possible return times to 2 *(for example). These are:*

$$2\;4\;6\;8\;\ldots$$

so the period is 2*. The cyclically moving sub-classes are* $\{2,5\}$ *and* $\{1,3,4\}$*. Notice that by Theorem 2.31 we find the period is* 2 *whatever state we consider! For example return times to* 3 *are*

$$4\;6\;8\;10\;\ldots$$

and the greatest common divisor is again 2 (not 4, *which does not divide* 6 *for example).*

2.4. Classification of states - recurrence or transience?

Up to now our classification of states has taken no notice of the actual probabilities: the concepts of communicating class, essential states, and periodicity depend only on whether the transition probabilities are positive or negative, nothing more. This is about to change!

For a motivating example, consider the following hypothetical game. You bet on a sequence of independent coin tosses using a single unfair coin. You gain £1 for each head, lose £1 for each tail. Your initial capital is £10. When this is reduced to £0 the game is over (your unnamed opponent however has unlimited credit …). What is the probability that you are reduced to bankruptcy after 10 coin tosses? 11 coin tosses? 12? …

A simple Markov chain analysis reveals this to be a biased simple random walk with absorbing barrier at 0. If the probability of heads lies strictly between 0 and 1 then there are two communicating classes: $\{0\}$ is essential and aperiodic, $\{1, 2, \ldots\}$ is inessential and of period 2. Our question concerns the probability

$$\mathbb{P}\left[X \text{ first hits } 0 \text{ at time } n\right] .$$

Trivially (from periodicity) this is zero for odd n. What about the rest?

A full analysis can be obtained by special methods (see Section 4.3). However we can deal with this sort of problem systematically in a way which generalizes easily.

Definition 2.35 ***First-passage time distribution.*** *For all i, $j \in \mathcal{S}$ the distribution of T_{ij}, the first-passage time from i to j, is defined by $\mathbb{P}\left[T_{ij} = n | X_0 = i\right] = f_{ij}^{(n)}$, where*

$$f_{ij}^{(n)} = \mathbb{P}\left[X_n = j, X_m \neq j \text{ for } m = 1, \ldots, n-1 | X_0 = i\right] .$$

Also $\mathbb{P}\left[T_{ij} = \infty | X_0 = i\right] = 1 - f_{ij}^{(*)}$ where $f_{ij}^{(*)} = \sum_{n=1\ldots\infty} f_{ij}^{(n)}$.

This leads directly to a complementary pair of state properties which are often crucial in applications.

Definition 2.36 ***Recurrence and transience.*** *The state j is*

(a) recurrent (or persistent) if $f_{jj}^{(*)} = 1$,

(b) transient if $f_{jj}^{(*)} < 1$.

2.4.1. First-step decomposition

To find first-passage probabilities one solves linear equations:

$$\begin{aligned} f_{ij}^{(1)} &= p_{ij} \\ f_{ij}^{(n+1)} &= \sum_{\ell:\ell\neq j} p_{i\ell} f_{\ell j}^{(n)} = \sum_{\ell:\ell\to j,\ell\neq j} p_{i\ell} f_{\ell j}^{(n)} \,. \end{aligned}$$

(Note that $\sum_{\ell:\ell\to j,\ell\neq j} p_{i\ell} f_{\ell j}^{(n)}$ is the same as $\sum_{\ell:\ell\neq j} p_{i\ell} f_{\ell j}^{(n)}$, but omitting some zero terms!) To see this, divide the event in the conditional probability

$$\mathbb{P}\left[X \text{ first hits } j \text{ at time } n+1 | X_0 = i\right]$$

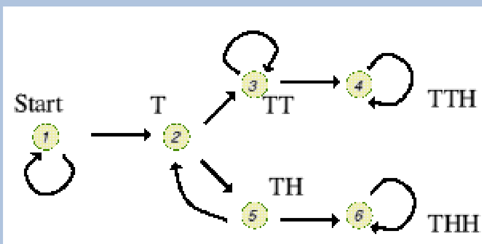
into a partition according to the first state visited by X after time 0). The most commonly used form first adds over n:

$$f_{ij}^{(*)} = p_{ij} + \sum_{\ell:\ell\to j,\ell\neq j} p_{i\ell} f_{\ell j}^{(*)} \,. \qquad (2.2)$$

This is often referred to as the *first-step decomposition*

Later we will find a more systematic matrix-based approach to this, but in fact for small Markov chains the above is a very effective way to find first-passage probabilities. Here is an example.

Example 2.37 ***Two players*** A ***and*** B ***choosing sequences.*** *Player* A *chooses a sequence of three symbols from* $\{H, T\}$*. Player* B *then chooses another. They observe a sequence of independent fair coin tosses. First triple to occur wins! Suppose for example that* A *chooses* THH *while* B *chooses* TTH*. We can code progress of the game as a Markov chain*



$$\begin{bmatrix} \frac{1}{2} & \frac{1}{2} & 0 & 0 & 0 & 0 \\ 0 & 0 & \frac{1}{2} & 0 & \frac{1}{2} & 0 \\ 0 & 0 & \frac{1}{2} & \frac{1}{2} & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & \frac{1}{2} & 0 & 0 & 0 & \frac{1}{2} \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

and we must study $f_{14}^{(*)}$ *and* $f_{16}^{(*)}$.

WARWICK
Home Page
Title Page
Contents
Page 62 of 237
Go Back
Full Screen
Close
Quit

Observe that the first-step decomposition gives

$$f_{14}^{(*)} = \frac{1}{2} f_{14}^{(*)} + \frac{1}{2} f_{24}^{(*)}$$

so $f_{14}^{(*)} = f_{24}^{(*)}$. Also

$$\begin{aligned} f_{24}^{(*)} &= \frac{1}{2} f_{34}^{(*)} + \frac{1}{2} f_{54}^{(*)} \\ f_{34}^{(*)} &= \frac{1}{2} + \frac{1}{2} f_{34}^{(*)} \\ f_{54}^{(*)} &= \frac{1}{2} f_{24}^{(*)} + \frac{1}{2} f_{64}^{(*)} \\ f_{64}^{(*)} &= 0 \end{aligned}$$

From here we can see $f_{54}^{(*)} = \frac{1}{2} f_{24}^{(*)}$ and so $f_{24}^{(*)} = \frac{1}{2} f_{34}^{(*)} + \frac{1}{4} f_{24}^{(*)}$. Thus $f_{24}^{(*)} = \frac{2}{3} f_{34}^{(*)}$. But also we can deduce from these equations that $f_{34}^{(*)} = 1$. Hence we see $f_{14}^{(*)} = \frac{2}{3}$. It is now easy to check that $f_{16}^{(*)} = \frac{1}{3}$

You should notice how these calculations are not much more than applied common sense, once one has taken on board the idea of the first-step decomposition. However in more complicated Markov chains one might wish for an even more systematic way to obtain the answer: later on in this section that is exactly what we will obtain! Moreover it is still not clear how we should deal with the *infinite* system of linear equations which will arise from the motivating coin-tossing example. This too will be dealt with later.

2.4.2. First-passage decomposition

There is a useful convolution sum: the *first-passage decomposition*

$$p_{ij}^{(n)} = \sum_{m=1\ldots n} f_{ij}^{(m)} p_{jj}^{(n-m)} \qquad \text{for } n > 0\,. \tag{2.3}$$

Prove this by dividing the event in the conditional probability

$$\mathbb{P}\left[X_n = j | X_0 = i\right]$$

into a partition according to the first time X visits state j: this is easy to see how to do after thinking about the Chapman-Kolmogorov equations and the first-step decomposition!
Notice that we may use $m = n - u$ to obtain

$$p_{ij}^{(n)} = \sum_{u=0\ldots n-1} f_{ij}^{(n-u)} p_{jj}^{(u)} \qquad \text{for } n > 0\,.$$

Converting into generating functions,

$$
\begin{aligned}
P_{ij}(s) &= \sum_{n=0}^{\infty} p_{ij}^{(n)} s^n \\
F_{ij}(s) &= \sum_{n=1}^{\infty} f_{ij}^{(n)} s^n \,,
\end{aligned}
$$

we obtain

$$
\begin{aligned}
P_{ij}(s) &= \sum_{n=0}^{\infty} p_{ij}^{(n)} s^n \\
&= p_{ij}^{(0)} + \sum_{n=1}^{\infty} p_{ij}^{(n)} s^n \\
&= p_{ij}^{(0)} + \sum_{n=1}^{\infty} \sum_{u=0}^{n-1} p_{jj}^{(u)} s^u f_{ij}^{(n-u)} s^{n-u}
\end{aligned}
$$

$$
\begin{aligned}
&= p_{ij}^{(0)} + \sum_{u=0}^{\infty} \sum_{n=u+1}^{\infty} p_{jj}^{(u)} s^u f_{ij}^{(n-u)} s^{n-u} \\
&= p_{ij}^{(0)} + \sum_{u=0}^{\infty} p_{jj}^{(u)} s^u \sum_{n=u+1}^{\infty} f_{ij}^{(n-u)} s^{n-u} \\
&= p_{ij}^{(0)} + \sum_{u=0}^{\infty} p_{jj}^{(u)} s^u \sum_{n=1}^{\infty} f_{ij}^{(n)} s^n \\
&= p_{ij}^{(0)} + F_{ij}(s) P_{jj}(s)
\end{aligned}
$$

and so

$$
P_{ij}(s) = p_{ij}^{(0)} + F_{ij}(s) P_{jj}(s) \tag{2.4}
$$

when $|s| < 1$.

Theorem 2.38 *The state j is*

(a) recurrent precisely when $\sum_n p_{jj}^{(n)} = \infty$,

(b) transient precisely when $\sum_n p_{jj}^{(n)} < \infty$, and then

$$\sum_n p_{jj}^{(n)} = \frac{1}{1 - f_{jj}^{(*)}}.$$

Notice that the criterion $\sum_n p_{jj}^{(n)} < \infty$ is the expected number of returns to j. So we see j is transient if and only the expected number of returns to j is finite!

Proof: We use a fact from analysis: if $\{a_n\}$ is a sequence of non-negative numbers and if $g(z) = \sum_n a_n z^n$ is convergent for $|z| < 1$ then

$$\lim_{z \uparrow 1} g(z) = \sum_n a_n$$

in the sense that one is finite if and only if the other is, and if both are finite then they are equal.

Now j is recurrent if and only if $f_{jj}^{(*)} = 1$ if and only if $\lim_{z\uparrow 1} F_{jj}(z) = 1$, where the last step uses the fact from analysis given above. Now we can use the generating function relationship: from

$$P_{jj}(s) = 1 + F_{jj}(s) P_{jj}(s)$$

we can deduce

$$F_{jj}(s) = \frac{P_{jj}(s) - 1}{P_{jj}(s)}$$

and so $\lim_{z\uparrow 1} F_{jj}(z) = 1$ if and only if $\lim_{z\uparrow 1} P_{jj}(z) = \infty$, (and now once again we use the fact from analysis) if and only if $\sum_n p_{jj}^{(n)} = \infty$. $\square$

There is an easy corollary with a very similar proof:

Corollary 2.39 *If j is transient then $p_{ij}^{(n)} \to 0$ as $n \to \infty$.*

Proof: If $i = j$ then this is easy: otherwise $\sum_n p_{jj}^{(n)}$ couldn't converge!
If $i \neq j$ then the generating function relationship becomes

$$P_{ij}(s) \quad = \quad F_{ij}(s) P_{jj}(s)$$

and we can use the fact from analysis twice to see

$$\begin{aligned} \sum_n p_{ij}^{(n)} &= P_{ij}(1) = \lim_{z \uparrow 1} P_{ij}(z) \\ &= \lim_{z \uparrow 1} F_{ij}(z) P_{jj}(z) = F_{ij}(1) P_{jj}(1) \end{aligned}$$

which is finite since j is transient (shows $P_{jj}(1) < \infty$) and F_{ij} is a generating function using probabilities summing to at most 1 (so $F_{ij}(1)$ must be no more than 1). $\square$

Theorem 2.40 *Recurrence and transience are class properties.*

Proof: Suppose i, j intercommunicate. So $p_{ij}^{(m)}$, $p_{ji}^{(n)}$ are both positive for some n and m. It follows,

$$p_{jj}^{(m+r+n)} \quad \geq \quad p_{ji}^{(m)} p_{ii}^{(r)} p_{ij}^{(n)}$$

and summing over r, substitution in $\sum_r p_{jj}^{(r)}$, allows us to deduce that if i is recurrent then so is j. Reversing the rôles of i, j completes the argument. $\square$

So we speak of transient and recurrent communicating classes and need to check only one state per class.
Can we ever move from a recurrent state into a different communicating class?

Theorem 2.41 *If i is recurrent and $i \to j$ then $i \sim j$ and $f_{ij}^{(*)} = f_{ji}^{(*)} = 1$. Thus recurrent implies essential and inessential implies transient.*

Proof: We know the following:

- i is recurrent, so $f_{ii}^{(*)} = 1$;
- $i \to j$, so $p_{ij}^{(n)} > 0$ for some n.

Use first-step decomposition:

$$f_{ki}^{(*)} = p_{ki} + \sum_{\ell \neq i} p_{k\ell} f_{\ell i}^{(*)}$$

which simplifies, as $f_{ii}^{(*)} = 1$,

$$f_{ki}^{(*)} \quad = \quad \sum_{\ell} p_{k\ell} f_{\ell i}^{(*)} \quad = \quad \sum_{\ell} p_{k\ell} \sum_{m} p_{\ell m} f_{mi}^{(*)}$$

where we resubstitute in the last step. Now exchange the order of this double summation of non-negative terms:

$$= \sum_{m} \left(\sum_{\ell} p_{k\ell} p_{\ell m} \right) f_{mi}^{(*)} = \sum_{m} p_{km}^{(2)} f_{mi}^{(*)} = \sum_{m} p_{km}^{(n)} f_{mi}^{(*)}$$

WARWICK
Home Page
Title Page
Contents
Page 72 of 237
Go Back
Full Screen
Close
Quit

Now this applies when $k = i$ and then we see

$$1 = f_{ii}^{(*)} \quad = \quad \sum_m p_{im}^{(n)} f_{mi}^{(*)}$$

and so

$$0 \quad = \quad \sum_m p_{im}^{(n)} \left(1 - f_{mi}^{(*)}\right)$$

since transition probabilities add up to one. But this is a sum of non-negative terms adding up to zero! The only way this can happen is if *every* term is zero! Since $p_{ij}^{(n)}$ is positive (for the special n discussed above) we see

$$0 \quad = \quad 1 - f_{ji}^{(*)}$$

and so

$$f_{ji}^{(*)} \quad = \quad 1$$

which forces $j \to i$. But this means i, j are in the same communicating class, since we are given $i \to j$, and so j must also be recurrent. The proof is completed by showing $f_{ij}^{(*)} = 1$; but this follows simply by switching the rôles of i, j. $\square$

Generally "essential" does not imply "recurrent". However:

Theorem 2.42 *If C is a finite essential class then it is recurrent.*

Proof: Suppose that $C = \{1, 2, \ldots, m\}$ is an essential class. Then we know $p_{ij}^{(n)} > 0$ for i in C means that j is in C. It follows that for all n

$$1 = \sum_r p_{ir}^{(n)} = \sum_{r \in C} p_{ir}^{(n)} .$$

Now if C were transient we would find every summand $p_{ir}^{(n)}$ in the above would converge to zero as n tends to infinity (use Corollary 2.39). Exchanging limits with (finite!) sum would show

$$1 = \lim \sum_{r \in C} p_{ir}^{(n)} = 0$$

a contradiction! So we conclude that C must be recurrent. $\square$

(But there *are* infinite transient essential classes!)
Here is a simple example to illustrate some of the above theory.

Example 2.43 *Consider a* 3×3 *transition probability matrix*

$$\underline{\underline{P}} = \begin{bmatrix} \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}.$$

In this simple case we can do all *calculations. For example:*

- $f_{11}^{(*)} = \mathbb{P}\left[X_1 = 1 | X_0 = 1\right] + \mathbb{P}\left[X_2 = 1 X_1 \neq 1 | X_0 = 1\right] + \ldots$ *giving* $f_{11}^{(*)} = 1/3 + 0 + 0 + \ldots = 1/3$ *(so* 1 *is transient), while* $p_{11}^{(n)}$ *is equal to*
    - 1 *if* $n = 0$
    - $1/3$ *if* $n = 1$
    - $1/3^2$ *if* $n = 2 \ldots$

  and therefore we can sum the geometric series to discover $\sum_{n=0}^{\infty} p_{11}^{(n)} = 3/2$*, which agrees with Theorem 2.38.*
- $f_{22}^{(*)} = 1$ *as* $\mathbb{P}\left[X_2 = 2 | X_0 = 2\right] = 1$ *(so* 2 *is recurrent), while* $p_{22}^{(n)}$ *is equal to* 1 *if* n *is even, to* 0 *if* n *is odd. Hence* $\sum_{n=0}^{\infty} p_{22}^{(n)} = \infty$*, as it should be since* 2 *is recurrent.*

So in both cases Theorem 2.38 is confirmed (as we knew it had to be – we proved it must be true!).

Chapter 3

The fundamental matrix $\underline{\underline{G}}$

Contents

Matrices can be used to answer questions about Markov chains.

3.1. Definition

Definition 3.1 ***Substochastic matrices.*** *The matrix* $\underline{\underline{Q}}$ *is* substochastic *if all its entries are non-negative and if all its rowsums are no greater than* 1*. Recall stochastic matrices are substochastic matrices with rowsums equal to* 1*.*

Probability can "leak away" in a substochastic matrix.

Definition 3.2 ***Fundamental matrix.*** *The fundamental matrix* $\underline{\underline{G}}$ *of a substochastic matrix* $\underline{\underline{Q}}$ *is the matrix given by*

$$\underline{\underline{G}} \quad = \quad \underline{\underline{\mathbb{I}}} + \underline{\underline{Q}} + \underline{\underline{Q}}^2 + \underline{\underline{Q}}^3 + \ldots$$

*(*ie*: add up all the matrix powers of* $\underline{\underline{Q}}$*).*

Remark 3.3 *Note the entries of* $\underline{\underline{G}}$ *can take values in the interval* $[0, \infty]$*.*

Remark 3.4 *If* $\underline{\underline{P}}$ *is a transition probability matrix then* $\underline{\underline{G}} = \sum_{n=0}^{\infty} \underline{\underline{P}}^{(n)}$*. Arguing as in* Exercise 4.7 *we see* G_{ij} *is the mean number of times* $X = j$ *if* X *is started at* i*.*

Example 3.5 ***Simple example.*** *Consider the* 2*-state chain with* $\underline{\underline{P}}$ *given by*

$$\begin{bmatrix} \frac{1}{2} & \frac{1}{2} \\ 0 & 1 \end{bmatrix}$$

Then we can compute the n*-step transition probability matrix* $\underline{\underline{P}}^{(n)}$ *as*

$$\begin{bmatrix} \left(\frac{1}{2}\right)^n & 1 - \left(\frac{1}{2}\right)^n \\ 0 & 1 \end{bmatrix}$$

and sum to obtain the fundamental matrix $\underline{\underline{G}}$

$$\begin{bmatrix} 2 & \infty \\ 0 & \infty \end{bmatrix}$$

Example 3.6 *Suppose N_{ij} is the random variable which is the number of times in state j given $X_0 = i$. Then (from above) $G_{ij} = \mathbb{E}\left[N_{ij}\right] = P_{ij}(1)$, the mean occupation time. How to find the fundamental matrix given the first-passage probabilities $f_{ij}^{(*)}$?*

Use the generating function relationship $P_{ij}(s) = p_{ij}^{(0)} + F_{ij}(s)P_{jj}(s)$. Hence

$$G_{jj} = \begin{cases} \infty & \text{if } j \text{ is recurrent} \\ \frac{1}{1-f_{jj}^{(*)}} & \text{if } j \text{ is transient} \end{cases} \tag{3.1}$$

and if i, j are different then

$$G_{ij} = \begin{cases} \infty & \text{if } j \text{ recurrent and } i \to j \\ 0 & \text{if it is not true that } i \to j \\ \frac{f_{ij}^{(*)}}{1-f_{jj}^{(*)}} & \text{if } j \text{ transient}\,. \end{cases} \tag{3.2}$$

In particular note that G_{ij} is always finite if j is transient.

Example 3.7 ***Simple example continued.*** *Since $G_{11} = 2$ we can deduce $f_{11}^{(*)} = \frac{1}{2}$.*

3.2. Properties

Useful case: $\mathcal{T}$ is a transient class or the union of all transient classes, $\underline{\underline{Q}}$ is the substochastic matrix which is the stochastic matrix $\underline{\underline{P}}$ restricted to $\mathcal{T}$. It is important to know that the fundamental matrix built from the full transition probability matrix $\underline{\underline{P}}$ and then restricted to (T) is the same as the fundamental matrix built from the substochastic matrix $\underline{\underline{Q}}$ obtained by restricting $\underline{\underline{P}}$ to $\mathcal{T}$. In a temporary and obvious notation, we want to show

$$\underline{\underline{G}}(\underline{\underline{P}})|_{\mathcal{T}} \quad = \quad \underline{\underline{G}}(\underline{\underline{P}}|_{\mathcal{T}}) \quad = \quad \underline{\underline{G}}(\underline{\underline{Q}})$$

(see Example 3.8 below for an explicit example). We can show from transience that $\underline{\underline{Q}}^{(n)}$ is the n-step transition probability matrix $\underline{\underline{P}}^{(n)}$ restricted to $\mathcal{T}$. The key idea is that if i, j are in $\mathcal{T}$ and k is not then it is not possible for both $i \to k$ and $k \to j$. See Exercise **??**. Consequently we have

$$\underline{\underline{G}} \quad = \quad \underline{\underline{\mathbb{I}}} + \underline{\underline{Q}} + \underline{\underline{Q}}^2 + \ldots$$

and so we can argue as for geometric series to show $\underline{\underline{G}} = \underline{\underline{\mathbb{I}}} + \underline{\underline{Q}}\,\underline{\underline{G}}$.

WARWICK
Home Page
Title Page
Contents
Page 81 of 237
Go Back
Full Screen
Close
Quit

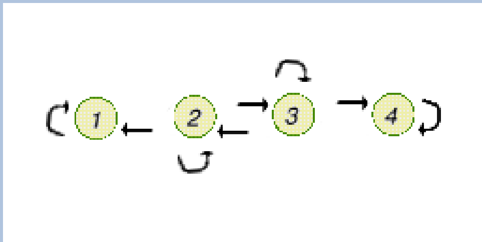
In general we can use Equation (3.1) and Equation (3.2), and the transience of all the states in $\mathcal{T}$, to show that all the entries of the fundamental matrix $\underline{\underline{G}}$ of $\underline{\underline{Q}}$ are finite; so we may subtract $\underline{\underline{Q}}\,\underline{\underline{G}}$ from both sides to obtain

$$\underline{\underline{\mathbb{I}}} \quad = \quad (\underline{\underline{\mathbb{I}}} - \underline{\underline{Q}})\underline{\underline{G}}\,.$$

If in addition $\mathcal{T}$ is finite then we can use matrix algebra, since we are now dealing with finite matrices with finite entries, to deduce

$$\underline{\underline{G}} \quad = \quad (\underline{\underline{\mathbb{I}}} - \underline{\underline{Q}})^{-1}\,.$$

Example 3.8 *Consider the transition probability matrix $\underline{\underline{P}}$ given by*



$$\begin{bmatrix} 1 & 0 & 0 & 0 \\ \frac{1}{3} & \color{red}{\frac{1}{3}} & \color{red}{\frac{1}{3}} & 0 \\ 0 & \color{red}{\frac{1}{3}} & \color{red}{\frac{1}{3}} & \frac{1}{3} \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

(entries corresponding to the set $\mathcal{T}$ of all transient states are in red) and notice we may immediately deduce part *of $\underline{\underline{G}}$:*

$$\begin{bmatrix} \infty & 0 & 0 & 0 \\ \infty & \color{red}{?} & \color{red}{?} & \infty \\ \infty & \color{red}{?} & \color{red}{?} & \infty \\ 0 & 0 & 0 & \infty \end{bmatrix}.$$

So now look at the substochastic matrix $\underline{\underline{Q}}$ corresponding to the non-border part of $\underline{\underline{P}}$ (which is to say, to the set $\mathcal{T}$ of transient states):

$$\begin{bmatrix} \color{red}{\frac{1}{3}} & \color{red}{\frac{1}{3}} \\ \color{red}{\frac{1}{3}} & \color{red}{\frac{1}{3}} \end{bmatrix}$$

and compute $\underline{\underline{\mathbb{I}}} - \underline{\underline{Q}}$:

$$\begin{bmatrix} \frac{2}{3} & -\frac{1}{3} \\ -\frac{1}{3} & \frac{2}{3} \end{bmatrix}$$

and its inverse

$$\begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$$

and so deduce the full fundamental matrix $\underline{\underline{G}}$:

$$\begin{bmatrix} \infty & 0 & 0 & 0 \\ \infty & 2 & 1 & \infty \\ \infty & 1 & 2 & \infty \\ 0 & 0 & 0 & \infty \end{bmatrix}$$

Here is a more extensive example.

Example 3.9 *Animal behaviour is often modelled by Markov chains. Consider the following idealized model for bat behaviour in a given day: states are*

1 *sleep at start of day*

2 *hang around*

3 *search for food*

4 *search for water*

5 *socialize*

6 *return to sleep*

Model the transitions between these states as a Markov chain with *transition matrix*

$$\underline{\underline{P}} = \begin{bmatrix} \frac{1}{4} & \frac{3}{4} & 0 & 0 & 0 & 0 \\ 0 & 0 & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \\ 0 & \frac{2}{3} & 0 & \frac{1}{3} & 0 & 0 \\ 0 & \frac{2}{3} & \frac{1}{3} & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}.$$

Here we have coloured blue those entries corresponding to the largest single transient class $\mathcal{T} = \{2, 3, 4, 5\}$. *(Notice however that* $\{1\}$ *is a further transient class which we are leaving out for now.) Correspondingly we obtain*

$$\underline{\underline{Q}} = \begin{bmatrix} 0 & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \\ \frac{2}{3} & 0 & \frac{1}{3} & 0 \\ \frac{2}{3} & \frac{1}{3} & 0 & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix}.$$

Now compute (in `Splus` or Mathematica *...)*

$$\left(\underline{\underline{\mathbb{I}}} - \underline{\underline{Q}}\right)^{-1} = \left(\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} - \begin{bmatrix} 0 & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \\ \frac{2}{3} & 0 & \frac{1}{3} & 0 \\ \frac{2}{3} & \frac{1}{3} & 0 & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix} \right)^{-1}$$

$$= \begin{bmatrix} 4 & 1.5 & 1.5 & 1 \\ 4 & 2.625 & 1.875 & 1 \\ 4 & 1.875 & 2.625 & 1 \\ 4 & 1.5 & 1.5 & 2 \end{bmatrix}$$

giving the full fundamental matrix

$$\begin{bmatrix} 1.333 & 4 & 1.5 & 1.5 & 1 & \infty \\ 0 & 4 & 1.5 & 1.5 & 1 & \infty \\ 0 & 4 & 2.625 & 1.875 & 1 & \infty \\ 0 & 4 & 1.875 & 2.625 & 1 & \infty \\ 0 & 4 & 1.5 & 1.5 & 2 & \infty \\ 0 & 0 & 0 & 0 & 0 & \infty \end{bmatrix}$$

The top-left corner is calculated either by applying the same process to the 1×1 *substochastic matrix* $[1/4]$ *(very easy) or by direct ar-*

guments using the geometric series. The rest of the top row follows from the remarks following Example 3.6.
Thus the mean time spent hanging around in any one day is 4 *units of time . . .*

3.3. Trouble with infinity

If $\mathcal{T}$ is infinite then there may be an *infinite* number of solutions: infinite matrices are tricky and can have more than one right- or left-inverse!

Example 3.10 *Cautionary example: Consider the infinite matrix:*

$$\begin{bmatrix} 0 & 1 & 0 & \ldots & \\ 0 & 0 & 1 & 0 & \ldots \\ 0 & 0 & 0 & 1 & \ldots \\ & & \ldots & & \end{bmatrix}$$

(the identity matrix shifted one column to the right) then any *of the following matrices serves as a right-inverse!*

$$\begin{bmatrix} x & 0 & 0 & \ldots \\ 1 & 0 & 0 & \ldots \\ 0 & 1 & 0 & \ldots \\ 0 & 0 & 1 & \ldots \\ & & \ldots & \end{bmatrix}$$

(the identity matrix shifted one row down, with an arbitrary entry x added at top-right corner).

However there is always a "smallest non-negative inverse" and this turns out to be the correct answer for $\underline{\underline{G}}$.

Theorem 3.11 *$\underline{\underline{G}}$ is the minimal non-negative solution $\underline{\underline{Z}}$ of*

$$\underline{\underline{Z}} = \underline{\underline{\mathbb{I}}} + \underline{\underline{Q}}\,\underline{\underline{Z}}.$$

(This is a re-write of $\underline{\underline{\mathbb{I}}} = (\underline{\underline{\mathbb{I}}} - \underline{\underline{Q}})\underline{\underline{Y}}$ to avoid minus signs, which interact badly with infinite entries: we don't want to consider $\infty - \infty$!)

Proof: Clearly $\underline{\underline{G}}$ is *one* possible such solution. So suppose that $\underline{\underline{Z}}$ is another: we have

$$\underline{\underline{Z}} = \underline{\underline{\mathbb{I}}} + \underline{\underline{Q}}\,\underline{\underline{Z}}$$

and also all entries of $\underline{\underline{Z}}$ are non-negative. Then

$$\begin{aligned}\underline{\underline{Z}} &= \underline{\underline{\mathbb{I}}} + \underline{\underline{Q}}\,\underline{\underline{Z}} = \underline{\underline{\mathbb{I}}} + \underline{\underline{Q}} + \underline{\underline{Q}}^2\underline{\underline{Z}} = \\ &= \ldots = \underline{\underline{\mathbb{I}}} + \underline{\underline{Q}} + \ldots + \underline{\underline{Q}}^n + \underline{\underline{Q}}^{n+1}\underline{\underline{Z}}\end{aligned}$$

WARWICK
Home Page
Title Page
Contents
Page 90 of 237
Go Back
Full Screen
Close
Quit

where the last term has non-negative entries, being made up of matrices with non-negative entries multiplied together. Hence entry by entry we find

$$\underline{\underline{Z}} \quad \geq \quad \underline{\underline{\mathbb{I}}} + \underline{\underline{Q}} + \ldots + \underline{\underline{Q}}^n$$

and taking the limit we find

$$\underline{\underline{Z}} \quad \geq \quad \underline{\underline{\mathbb{I}}} + \underline{\underline{Q}} + \ldots + \underline{\underline{Q}}^n + \ldots = \underline{\underline{G}}$$

as required. □

Example 3.12 ***Simple symmetric random walk with one absorbing barrier.*** *Suppose* 0 *is the absorbing barrier. For example it can be shown in this case that starting from any position* $j \geq 2$ *one makes on average* 2 *visits to position* 1 *before absorption takes place. (Calculate the* $\underline{\underline{Q}}$ *matrix by restriction from* $\underline{\underline{P}}$*, solve*

$$\underline{\underline{\mathbb{I}}} \quad = \quad (\underline{\underline{\mathbb{I}}} - \underline{\underline{Q}})\underline{\underline{Y}}\,,$$

and figure out the minimal non-negative solution $\underline{\underline{G}}$ *to find* $G_{i1} = 2$ *for* $i = 1, 2, \ldots$*)*

To see how this works, calculate (see Exercise 5.6 for details)

$$\begin{bmatrix} 1 & -\frac{1}{2} & 0 & \dots & \\ -\frac{1}{2} & 1 & -\frac{1}{2} & 0 & \dots \\ 0 & -\frac{1}{2} & 1 & -\frac{1}{2} & 0 \\ 0 & 0 & -\frac{1}{2} & 1 & -\frac{1}{2} \\ & \dots & & & \end{bmatrix} \times \begin{bmatrix} Y_{11} & \dots \\ Y_{21} & \dots \\ Y_{31} & \dots \\ Y_{41} & \dots \\ \dots & \end{bmatrix} = \begin{bmatrix} 1 & \dots \\ 0 & \dots \\ 0 & \dots \\ 0 & \dots \\ \dots & \end{bmatrix}$$

This gives

$$\begin{aligned} Y_{11} - \frac{1}{2}Y_{21} &= 1 \\ -\frac{1}{2}Y_{11} + Y_{21} - \frac{1}{2}Y_{31} &= 0 \\ -\frac{1}{2}Y_{21} + Y_{31} - \frac{1}{2}Y_{41} &= 0 \\ \dots & \end{aligned}$$

and hence

$$\begin{aligned} Y_{21} &= 2Y_{11} - 2 \\ Y_{31} &= 3Y_{11} - 4 \\ Y_{41} &= 4Y_{11} - 6 \end{aligned}$$

and in general

$$Y_{n1} = nY_{11} - (2n - 2)\,.$$

Now if Y_{11} is smaller than 2 we find eventually Y_{n1} becomes negative. On the other hand if $Y_{11} = 2$ then $Y_{n1} = 2$ for all n. So the smallest possible non-negative Y has $Y_{n1} = 2$ for all n. Hence $G_{n1} = 2$ for all n.
When is there more than one non-negative solution to $\underline{\underline{\mathbb{I}}} = (\underline{\underline{\mathbb{I}}} - \underline{\underline{Q}})\underline{\underline{Y}}$?
When there is a non-zero non-negative solution to $\underline{h} = \underline{\underline{Q}}\,\underline{h}$ for $\underline{h}$ a vector. This motivates the following result:

Theorem 3.13 *Let A be a subset of state-space $\mathcal{S}$, say $A = \{1, 2, \ldots\}$, and let $\underline{\underline{Q}}$ be the stochastic matrix restricted to A. Let*

$$a_i^{(n)} = \mathbb{P}\left[\,\text{the process stays in } A \text{ up to time } n | X_0 = i\right]$$

and let $a_i = \lim a_i^{(n)}$. Then the vector $\underline{a} = (a_1, a_2, \ldots)^T$ is the maximal solution of $\underline{h} = \underline{\underline{Q}}\,\underline{h}$ with $0 \leq h_i \leq 1$ for all i.

Proof:

(i) Show that the vector $\underline{a}$ is a solution. Clearly

$$1 \quad \geq \quad a_i^{(1)} \geq a_i^{(2)} \geq \ldots \geq 0$$

so (existence of limit of a decreasing sequence bounded below) we do indeed have

$$a_i^{(n)} \to a_i$$

and (strictly this uses basic results from ST213, but it is intuitively clear) we also have

$$a_i \quad = \quad \mathbb{P}\left[\,\text{the process stays in } A \text{ for ever } | X_0 = i\right] .$$

Now use first-step decomposition:

$$a_i^{(n+1)} \quad = \quad \sum_{j \in A} p_{ij} a_j^{(n)}$$

so in the limit (allowing exchange of sum and limit) we have

$$a_i \quad = \quad \lim_{n\to\infty} \sum_{j \in A} p_{ij} a_j^{(n)} = \sum_{j \in A} p_{ij} \lim_{n\to\infty} a_j^{(n)} .$$

The exchange of sum and limit is in fact justified by noticing the summand $p_{ij}a_j^{(n)}$ is dominated by p_{ij}, and

$$\sum_{j\in A} p_{ij} \quad \leq \quad 1\,.$$

So exchange is permitted by a result from the theory of infinite series. (In ST213 this result is generalized to integrals as the *Dominated Convergence Theorem*, namely Corollary 5.17.) We now deduce

$$a_i \quad = \quad \sum_{j\in A} p_{ij}a_j$$

and so $\underline{a} = \underline{\underline{Q}}\,\underline{a}$.

(ii) Show that $\underline{a}$ is maximal. This is similar to the proof of Theorem 3.11. Suppose we have $\underline{g} = \underline{\underline{Q}}\,\underline{g}$ where $0 \leq g_i \leq 1$ for all i in A. Let $\underline{e}$ be a vector of the same length as $\underline{g}$ but filled with 1's. Then $\underline{g} \leq \underline{e}$ with inequality holding entry by entry.

We notice $a_i^{(1)} = \sum_{j \in A} p_{ij}$ so that $\underline{a}^{(1)} = \underline{\underline{Q}}\,\underline{e}$. (We could say, $\underline{a}^{(0)} = \underline{e}$)

But applying the non-negative matrix $\underline{\underline{Q}}$ to both sides of $\underline{g} \leq \underline{e}$, we find

$$\underline{g} \quad = \quad \underline{\underline{Q}}\,\underline{g} \quad \leq \quad \underline{\underline{Q}}\,\underline{e} = \underline{a}^{(1)}\,.$$

Similarly we can apply $\underline{\underline{Q}}$ to both sides again to find

$$\underline{\underline{Q}}\,\underline{g} \quad = \quad \underline{g} \quad \leq \quad \underline{\underline{Q}}\underline{a}^{(1)} \quad = \quad \underline{a}^{(2)}\,,$$

and repeatedly to find $\underline{g} \leq \underline{a}^{(n)}$, and thus finally in the limit $\underline{g} \leq \underline{a}$ as required.

$\square$

We say $\underline{a}$ is a *harmonic measure* or a *right-invariant vector* for $\underline{\underline{Q}}$.

Corollary 3.14 *Either $\underline{a}$ is identically zero or* $\sup a_i = 1$.

Proof: Otherwise consider $\underline{a}^* = \underline{a}/k$, for $k = \sup_{i \in A} a_i$. $\square$

We can now summarize the situation.

Theorem 3.15 *If there are only a finite number of transient states then there is a unique solution to* $(\underline{\underline{\mathbb{I}}} - \underline{\underline{Q}})\underline{\underline{Y}} = \underline{\underline{\mathbb{I}}}$ *and the only bounded solution to* $\underline{h} = \underline{\underline{Q}}\,\underline{h}$ *(using* $A = \mathcal{T}$ *with* A *as in Theorem 3.13) is the zero solution. If there are an infinite number of transient states then either*

(a) $\sup_i h_i = 0$ *and the chain is sure to end up in one of the recurrent classes, or*

(b) $\sup_i h_i = 1$ *and there is a positive probability the chain stays forever among the transient states.*

The case (b) can be very complicated - there can be an uncountable number of essentially different ways in which the chain might stay among the transient states!

3.4. Fundamental matrix calculations

Example 3.16 ***Simple asymmetric random walk with one absorbing barrier.*** *Let p be the probability of jumping away from the barrier and q be the probability of jumping towards it. Suppose $p+q=1$. If $q \geq p$ then the above theory shows the walk is certain to hit the barrier. If $q < p$ and the walk starts i units away from the barrier then there is a positive chance $1-(q/p)^i$ of never hitting the barrier (but in fact of drifting out to infinity).*

To see this notice

1. there is one absorbing class 0 (if the barrier is located at state 0) and one transient class $\{1, 2, 3, \ldots\}$;
2. the $\underline{\underline{Q}}$ matrix for $1, 2, 3, \ldots$ is given by

$$\begin{bmatrix} 0 & p & 0 & 0 & \ldots \\ q & 0 & p & 0 & \ldots \\ 0 & q & 0 & p & \ldots \\ & \ldots & & & \end{bmatrix}$$

WARWICK
Home Page
Title Page
Contents
Page 98 of 237
Go Back
Full Screen
Close
Quit

3. solve $\underline{h} = \underline{\underline{Q}}\,\underline{h}$ with $0 \leq h_i \leq 1$:

$$\begin{aligned} h_1 &= ph_2 \\ h_2 &= qh_1 + ph_3 \text{ so that } h_3 - h_2 = (q/p)^2 h_1 \\ &\dots \\ h_n &= qh_{n-1} + ph_{n+1} \text{ so that } h_{n+1} - h_n = (q/p)^n h_1 \end{aligned}$$

and so we deduce (summing a finite geometric series)

$$\begin{aligned} h_n &= h_1(1-(q/p)^n)/(1-q/p) \text{ if } q, p \text{ are not equal} \\ & \quad nh_1 \text{ if } q = p\,. \end{aligned}$$

So if $q < p$ we can use $h_i = 1-(q/p)^i$; otherwise there is no suitable $\underline{h}$.

Example 3.17 ***Renewal theory as in Exercise 2.7.*** *The state is the age of a machine in weeks at the start of week n, given that a machine n weeks old is replaced at the start of week i with probability g_n independently of all other past events! We obtain a Markov chain which is irreducible (if all g_i are nonzero!).* However *if we alter $g_1 = 1$ then we obtain $\mathcal{T} = \{2, 3, \ldots\}$ and can apply* Corollary 3.14 *to $A = \mathcal{T}$ to obtain:*

$$h_n = (1 - g_n) h_{n+1}$$

and hence

$$h_{n+1} = h_2 / ((1 - g_2)(1 - g_3) \ldots (1 - g_n)) \,.$$

We can now argue we can construct $\underline{h}$ with $0 \leq h_i \leq 1$ for all i exactly when the infinite product $(1 - g_2)(1 - g_3) \ldots$ is non-zero! In that case we can put

$$h_{n+1} = (1 - g_{n+1})(1 - g_{n+2}) \ldots$$

and we find from previous theory

$$\mathbb{P}\left[X \in A \textit{ for ever } | X_0 = n\right] = (1 - g_n)(1 - g_{n+1}) \ldots$$

Finally we deduce for the original chain that the single irreducible class $\{0, 1, 2, \ldots\}$ must be transient if $(1-g_2)(1-g_3)\ldots$ is non-zero, since there is a positive chance of never returning to 0.

More generally we may consider absorption at a set.

Definition 3.18 ***Absorption probabilities.*** *If $A \subseteq \mathcal{S}$ then*

$$f^{(*)}_{iA} \quad = \quad \mathbb{P}\left[X_n \in A \text{ for some } n > 0 | X_0 = i\right] \,.$$

Theorem 3.19 *Absorption probabilities satisfy an equation derived from a first-step decomposition: they are the minimal solution to*

$$f^{(*)}_{iA} \quad = \quad \sum_{k \in A} p_{ik} + \sum_{k \notin A} p_{ik} f^{(*)}_{kA} \,.$$

Proof: (Sketch) Use first-step decomposition on $\mathbb{P}\left[X \text{ visits } A | X_0 = i\right]$ to show equation holds for $f^{(*)}_{iA}$. Argue as in Theorem 3.11 to show the $f^{(*)}_{iA}$ form the minimal solution. □

A useful particular case is when $A = C$ is a recurrent class and we consider absorption probabilities from the union of all transient

WARWICK
Home Page
Title Page
Contents
◀◀ ▶▶
◀ ▶
Page 101 of 237
Go Back
Full Screen
Close
Quit

classes $\mathcal{T}$. Let $\underline{F}_C^{(*)}$ be the vector of such absorption probabilities $f_{iC}^{(*)}$. Then we can use Theorem 3.19 to deduce a matrix formula for the column vector $\underline{F}_C^{(*)}$ using $\underline{\underline{G}}$ and other easily computable matrices. For consider

$$f_{iC}^{(*)} - \sum_{k \notin C} p_{ik} f_{kC}^{(*)} = \sum_{k \in C} p_{ik} \,.$$

and (since $\mathcal{T}$ is the union of all transient states and $A = C$ is a recurrent class) the left-hand side is

$$f_{iC}^{(*)} - \sum_{k \in \mathcal{T}} p_{ik} f_{kC}^{(*)} \quad = \quad i^{th} \text{ entry of } (\underline{\underline{\mathbb{I}}} - \underline{\underline{Q}})\underline{F}_C^{(*)}$$

while the right-hand side is

$$\sum_{k \in C} p_{ik} 1 \quad = \quad i^{th} \text{ entry of } \underline{\underline{R}}\,\underline{e}$$

where $\underline{\underline{R}}$ is the matrix of one-step transition probabilities

$$[p_{ij} \text{ for } i \in \mathcal{T}, j \in C] \,,$$

and $\underline{e}$ is a column vector full of 1's, as many entries as states in C. Putting these together, we have obtained

$$(\underline{\underline{\mathbb{I}}} - \underline{\underline{Q}})\underline{F}_C^{(*)} \quad = \quad \underline{\underline{R}}\,\underline{e}$$

Now $\underline{\underline{G}}$ is an inverse of $\underline{\underline{\mathbb{I}}} - \underline{\underline{Q}}$ so

Corollary 3.20

$$\underline{F}_C^{(*)} \quad = \quad \underline{\underline{G}}\,\underline{\underline{R}}\,\underline{e} \tag{3.3}$$

where $\underline{\underline{G}}$ is the fundamental matrix for $\mathcal{T}$.

We will use this in Section 4.1.4. [1]
An alternative derivation of this (which works for infinite state space case as well) decomposes $F_{iC}^{(*)}$ by both time and location of the first jump from $\mathcal{T}$ to C.

[1] Suppose t is the number of transient states, and c is the number of states in the recurrent class C. We already know $\underline{\underline{G}}$ is a matrix of shape $t \times t$ rows by columns. The matrix $\underline{\underline{R}}$ is of shape $t \times c$ rows by columns, while $\underline{e}$ is a column vector with c entries. You can check for yourself that these matrix shapes are exactly what is required for the matrix multiplication $\underline{\underline{G}}\,\underline{\underline{R}}\,\underline{e}$ to make sense and equal the column vector $\underline{F}_C^{(*)}$!

Theorem 3.21 *Suppose $\mathcal{T}$ is the union of all transient classes and C is any union of recurrent classes such that $f^{(*)}_{iC} = 1$ for all $i \in \mathcal{T}$. Let T_{iC} be the time from i till first hitting C. Then*

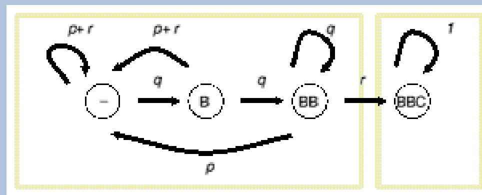
$$[\mathbb{E}\,[T_{iC}] : i \in \mathcal{T}] \quad = \quad \underline{\underline{G}}\,\underline{e}\,.$$

Here $\underline{e}$ is a column vector full of 1's, as many entries as there are states in $\mathcal{T}$.

Proof: Because $\mathcal{T}$ is the union of all transient classes and C is any union of recurrent classes such that $f^{(*)}_{iC} = 1$ for all $i \in \mathcal{T}$, we can argue that if $X_0 = i$ is in $\mathcal{T}$ then the number of times in sites in $\mathcal{T}$ until X leaves $\mathcal{T}$ will be exactly the number of times in sites in $\mathcal{T}$ until X first hits C. This is because recurrent states are essential (use Theorem 2.41). Thus once one visits a recurrent state j then all further states visited have to be in the same communicating class as j (and, in particular, recurrent). But now $\underline{\underline{G}}_{ij}$ is exactly the mean number of times in j in C until X leaves $\mathcal{T}$. It follows that the column vector $\underline{\underline{G}}\,\underline{e}$ has as i^{th} element the entry $\sum_{j\in\mathcal{T}} \underline{\underline{G}}_{ij}$, which is the mean sum of numbers of times in transient states as required. □

Example 3.22 *For each of a sequence of independent trials the outcomes are A, B, or C with probabilities p, q or r. How long to wait for the pattern BBC?*

We construct a state-diagram and compute the stochastic matrix $\underline{\underline{Q}}$ for the transient class $\{-, B, BB\}$:



$$\begin{bmatrix} p+r & q & 0 & 0 \\ p+r & 0 & q & 0 \\ p & 0 & q & r \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

and hence ($\underline{\underline{Q}}$ is in red) we get $\underline{\underline{G}} = (\underline{\underline{1}} - \underline{\underline{Q}})^{-1}$: we find $\underline{\underline{G}}$ is given by

$$\frac{1}{qr}\begin{bmatrix} \frac{1-q}{q} & 1-q & q \\ \frac{1-q-rq}{q} & 1-q & q \\ \frac{p}{q} & p & q \end{bmatrix}.$$

WARWICK
Home Page
Title Page
Contents
Page 105 of 237
Go Back
Full Screen
Close
Quit

Hence $\underline{\underline{G}}\,\underline{e}$ is the column vector

$$\begin{bmatrix} 1/(q^2 r) \\ (1-rq)/(q^2 r) \\ (1-r-rq)/(q^2 r) \end{bmatrix} .$$

Hence answer as required using Theorem 3.21: expected waiting time is $(q^2 r)^{-1}$ (the full distribution can be found using generating functions).

An example of use of the formula $\underline{F}_C^{(*)} = \underline{\underline{G}}\,\underline{\underline{R}}\,\underline{e}$ is to be found in the treatment of inbreeding in the following section. Here is a simple example.

Example 3.23 *Consider a simple random walk on* $\{0, 1, 2, 3, 4\}$ *with probability* p *of a* $+1$ *jump and probability* q *of a* -1 *jump, except that the transition* $0 \longrightarrow 1$ *has probability* 1 *and so does the transition* $4 \longrightarrow 3$*. What is the probability of visiting state* 0 *before state* 4 *if you start at state* 1*?*

To answer this, first *change the chain* to make 0 and 4 into absorbing states. Then we have recurrent classes $\{0\}$, $\{4\}$, and a single transient class $\{1, 2, 3\}$, with transition probability matrix

$$\begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ q & 0 & p & 0 & 0 \\ 0 & q & 0 & p & 0 \\ 0 & 0 & q & 0 & p \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

with the substochastic matrix for the transient states marked out in red.

The fundamental matrix for the transient class is given by

$$\underline{\underline{G}} = \left(\underline{\underline{1}} - \begin{bmatrix} 0 & p & 0 \\ q & 0 & p \\ 0 & q & 0 \end{bmatrix} \right)^{-1} = \frac{1}{1-2pq} \begin{bmatrix} 1-pq & p & p^2 \\ q & 1 & p \\ q^2 & q & 1-pq \end{bmatrix}.$$

The $\underline{\underline{R}}$ matrix for the recurrent class $\{0\}$ is given by

$$\underline{\underline{R}} = \begin{bmatrix} q \\ 0 \\ 0 \end{bmatrix}$$

and the corresponding $\underline{e}$ vector by

$$\underline{e} = \begin{bmatrix} 1 \end{bmatrix}.$$

So the probability of ending up in state 0 (which is what we require!) is given by

$$\underline{\underline{G}}\,\underline{\underline{R}}\,\underline{e} = \frac{1}{1-2pq}\begin{bmatrix} 1-pq & p & p^2 \\ q & 1 & p \\ q^2 & q & 1-pq \end{bmatrix} \times \begin{bmatrix} q \\ 0 \\ 0 \end{bmatrix} \times \begin{bmatrix} 1 \end{bmatrix}$$

$$= \frac{1}{1-2pq}\begin{bmatrix} (1-pq)q \\ q^2 \\ q^3 \end{bmatrix}.$$

So result is $(1-pq)q/(1-2pq)$.
Notice that for this question it is just as fast to proceed by using first-step decomposition: let

$$g_i = \mathbb{P}\left[\,\text{visit } 0 \text{ before } 4 \,|X_0 = i\right]$$

and use first-step decomposition to note that

$$\begin{aligned} g_0 &= 1 \\ g_1 &= qg_0 + pg_2 \\ g_2 &= qg_1 + pg_3 \\ g_3 &= qg_2 + pg_4 \\ g_4 &= 0 \end{aligned}$$

which leads to

$$\begin{aligned} g_3 &= qg_2 \\ g_2 &= qg_1 + pqg_2 \\ g_2 &= \frac{q}{1-pq} g_1 \\ g_1 &= qg_0 + \frac{pq}{1-pq} g_1 \end{aligned}$$

and (using $g_0 = 1$) we recover the result obtained above using matrices!

In general the matrix technique has the advantage of being systematic when the chain is large or lacks structure: the first-step decomposition (and then solution by simultaneous linear equations) can be *much* faster for small (2×2!) matrices or structured set-ups (random walks!).

Chapter 4

Applications

Contents

In this chapter we discuss three different applications of Markov chains: Genetics, Branching processes, Random walks.

Genetics is now a hugely important area of application for probability: it always was important because of animal/plant breeding, and genetic counselling; now it stands fair to be the most significant application area of all together with mathematical finance, with the developments of DNA testing, genetic manipulation, the Human Genome Project, . . .

Branching processes are also important in biology, because they model extinction and survival of populations. Historically they arose from work on nuclear chain reactions.

Random walks appear in all sorts of applied probability problems:

- approximating Brownian motion of small particles in fluids;
- movement of foraging animal along a river bank;
- price of a share;

  . . .

They also crop up in the course of solution of many other problems; often the easiest way to see whether a given Markov chain is recurrent or transient is to compare it to a random walk!

4.1. Genetics

4.1.1. Terminology

- *Genes:* small areas on *chromosomes*. The position of a gene on the chromosome is its *locus*. Each gene is in one of several different forms or *alleles*.
  - Humans: 23×2 chromosomes.
  - *Drosophila* fly: 4

  Humans are *diploid*: chromosomes are usually paired in *homologous pairs* (except sex chromosomes, of which more later).

  Other possibilities:
  - *monoploid*: 1
  - *triploid*: 3
  - *tetraploid*: 4 example: *tomatoes*.

  (are humans less advanced than tomatoes?)

- *Genotype*: genetic composition of an individual
- *Phenotype*: physical appearance of a genotype.

Notice: genotype controls phenotype, but some aspects of genotype may not be manifested *at all* in the phenotype!

Example 4.1 *At a specific* locus *in a* diploid *individual there are two* alleles A, a, *and* A *is* dominant *over* a. *So* a *is* recessive.

Genotype:	AA	Aa	aa
Phenotype:	A	A	a

Example 4.2 *Simplified genotype/phenotype structure for human blood system.*

Genotype:	AA	AO	AB	BB	BO	OO
Phenotype:	A	A	AB	B	B	O

See for example

BBC clip of "Hancock's Half-hour: the Blood Donor"

Diploid reproduction Each offspring takes 1 chromosome of each type (= *homologous pair*) from each parent: $Aa \times Aa$ leads to

	AA	Aa	aa
Probabilities	$1/4$	$1/2$	$1/4$

Theorem 4.3 ***Hardy-Weinberg law – Case of*** 2 ***alleles in diploid population*** *Whatever the initial distribution of genotypes, random mating yields a* stationary distribution *after just one generation, with ratios*

$$\begin{array}{ccc} AA & Aa & aa \\ \phi^2 & 2\phi\rho & \rho^2 \end{array}$$

for some ϕ, ρ, *where* $\phi + \rho = 1$.

Proof: Suppose the initial proportions are

$$\begin{array}{ccc} AA & Aa & aa \\ u & 2v & w \end{array}$$

then tabulation of the various possibilities for random mating, and addition of the corresponding probabilities for offspring, leads to

$$\begin{array}{ccc} AA & Aa & aa \\ (u+v)^2 & 2(u+v)(w+v) & (w+v)^2 \end{array}$$

as required. A typical calculation is built up out of smaller calculations such as,

$$
\begin{aligned}
& \mathbb{P}\left[Aa \text{ and parents were } Aa \text{ and } Aa\right] \\
= \; & \mathbb{P}\left[Aa | \text{ parents were } Aa \text{ and } Aa\right] \times \mathbb{P}\left[\text{ parents were } Aa \text{ and } Aa\right] \\
= \; & \mathbb{P}\left[Aa | \text{ parents were } Aa \text{ and } Aa\right] \times 4v^2 \\
= \; & (1/2) \times 4v^2 \quad = \quad 2v^2 .
\end{aligned}
$$

(Details of calculations in Appendix D)

After a further generation we would obtain (same calculations!)

$$
\begin{array}{ll}
AA : & ((u+v)^2 + (u+v)(w+v))^2 \\
AB : & 2((u+v)^2 + (u+v)(w+v))((w+v)^2 + (u+v)(w+v)) \\
aa : & ((w+v)^2 + (u+v)(w+v))^2
\end{array}
$$

which in turn leads to

$$
\begin{array}{ccc}
AA & Aa & aa \\
(u+v)^2 & 2(u+v)(w+v) & (w+v)^2
\end{array}
$$

if we simplify using $u + 2v + w = 1$.

$\square$

4.1.2. Sex-linked genes

Consider sex-determination by chromosomes labelled X, Y. For most mammals,[1]

- XX *homogametic* female
- XY *heterogametic* male

(so humans have 22 pairs of homologous *autosomes*, 1 pair of sex chromosomes). The sex chromosomes X, Y are of different lengths and hence some characteristics are X*-linked* and some are Y*-linked*. (X-linked: controlled by gene on X-chromosome without homologue on Y-chromosome; ...).

[1] But not all: according to a recent letter to Nature, the duck-billed platypus has **ten** sex-linked chromosomes which work like this: $XXXXX : XXXXX$ for female, $XXXXX : YYYYY$ for male.

WARWICK
Home Page
Title Page
Contents
◀◀ ▶▶
◀ ▶
Page 119 of 237
Go Back
Full Screen
Close
Quit

Example 4.4 *haemophilia, some kinds of colour-blindness are controlled by recessive X-linked genes:*

$$\begin{array}{ccc} \textit{normal male} & & \textit{carrier female} \\ A- & \times & Aa \end{array}$$

yields

1	$A-$	$a-$	AA	Aa
$Prob$	$1/4$	$1/4$	$1/4$	$1/4$
	(male)	*(male)*	*(female)*	*(female)*
		(haem)		*(carrier)*

4.1.3. Random mating for X-linked gene

(Unreasonable to assume random mating for haemophilia! rather more reasonable for colour-blindness.)
We commence with the following probabilities for (respectively) female genotypes and male genotypes:

	Female			Male	
Initial	AA	Aa	aa	$A-$	$a-$
	u_0	$2v_0$	w_0	r_0	s_0

Set

$$p_0 = \mathbb{P}[\text{ random female transmits } A] = u_0 + v_0 \,,$$
$$q_0 = \mathbb{P}[\text{ random female transmits } a] = v_0 + w_0 \,.$$

Next generation has the following genotype frequencies:

$$p_0 r_0 \quad p_0 s_0 + q_0 r_0 \quad q_0 s_0 \qquad\qquad r_1 = p_0 \quad s_1 = q_0$$

See this for example by noting that a female of the next generation has probability $p_0 s_0$ of obtaining A from her mother, a from her father; and probability $q_0 r_0$ of obtaining a from her mother, A from her

Home Page
Title Page
Contents
Page 121 of 237
Go Back
Full Screen
Close
Quit

father. So she has probability $p_0 s_0 + q_0 r_0$ of being Aa. On the other hand a male *must* obtain the Y chromosome from his father (or otherwise he wouldn't be a male!) and has probability p_0 of obtaining A from his mother. So he has probability p_0 of being $A-$.
Set

$$\begin{aligned} p_1 &= \mathbb{P}\left[\text{ random female transmits } A\right] = \frac{1}{2}(p_0 + r_0)\,, \\ q_1 &= \mathbb{P}\left[\text{ random female transmits } a\right] = \frac{1}{2}(q_0 + s_0)\,. \end{aligned}$$

and repeating to n^{th} generation we deduce

$$\begin{aligned} p_{n+1} &= \frac{1}{2}(p_n + r_n) \\ q_{n+1} &= \frac{1}{2}(q_n + s_n) \\ r_{n+1} &= p_n \\ s_{n+1} &= q_n \end{aligned}$$

and we can put these together to see that for $n \geq 2$ we have

$$p_n \quad = \quad \frac{1}{2}(p_{n-1} + p_{n-2})$$

and so

$$(p_n - p_{n-1}) = -\frac{1}{2}(p_{n-1} - p_{n-2})\,.$$

This can be used to show (exercise!) that as n tends to infinity so $p_n - p_0$ converges to

$$\frac{2}{3}(p_1 - p_0)$$

and hence

$$p_n \quad \to \quad \alpha = (p_0 + 2p_1)/3 = (2p_0 + r_0)/3\,.$$

We deduce (for example, by computing $r_{n+1} = p_n \to (2p_0 + r_0)/3$) that the limiting frequency of genotypes in the population is given by

AA	Aa	aa	$A-$	$a-$
α^2	$2\alpha(1-\alpha)$	$(1-\alpha)^2$ (for females)	α	$1-\alpha$ (for males)

At the phenotype level this means the recessive X-linked gene has effect in $1-\alpha$ of the males but only $(1-\alpha)^2$ of the females.

4.1.4. Inbreeding and Markov chains

Now we use all the theory of section 2.

Example 4.5 ***Sib-mating*** *Consider the case of an X-linked gene for a diploid population with 2 alleles A, a. The mating process is as follows: first mate two individuals to start things off. Then at each succeeding stage mate a randomly chosen male offspring with randomly chosen female offspring. (Presupposes there will always be males, females to choose from!)*

This produces a Markov chain in which the state at each generation is the male genotype and the female genotype from which one mates:

$$
\begin{array}{lcccccc}
 & 1 & 2 & 3 & 4 & 5 & 6 \\
1 : A - \times AA & 1 & 0 & 0 & 0 & 0 & 0 \\
2 : A - \times Aa & \frac{1}{4} & \frac{1}{4} & 0 & \frac{1}{4} & \frac{1}{4} & 0 \\
3 : A - \times aa & 0 & 0 & 0 & 0 & 1 & 0 \\
4 : a - \times AA & 0 & 1 & 0 & 0 & 0 & 0 \\
5 : a - \times Aa & 0 & \frac{1}{4} & \frac{1}{4} & 0 & \frac{1}{4} & \frac{1}{4} \\
6 : a - \times aa & 0 & 0 & 0 & 0 & 0 & 1
\end{array}
$$

From the state diagram one quickly works out communicating classes: recurrent classes $\{1\}$, $\{6\}$ formed by absorbing states, $\{2, 3, 4, 5\}$ forming a transient class. The transient substochastic matrix $\underline{\underline{Q}}$ is given in red in the above table. The fundamental matrix for the single transient class is given by

$$3(\underline{\underline{\mathbb{I}}} - \underline{\underline{Q}})^{-1} \quad = \quad \begin{bmatrix} 8 & 1 & 2 & 4 \\ 4 & 5 & 1 & 8 \\ 8 & 1 & 5 & 4 \\ 4 & 2 & 1 & 8 \end{bmatrix}$$

and so we may directly apply Equation (3.3) at the end of Section 2: the vector of $f_{i1}^{(*)}$ for i running through 2, 3, 4, 5 is obtained as

$$(\underline{\underline{\mathbb{I}}} - \underline{\underline{Q}})^{-1} \underline{\underline{R}}\,\underline{e} \quad = \quad \begin{bmatrix} 2/3 \\ 1/3 \\ 2/3 \\ 1/3 \end{bmatrix}$$

where $\underline{\underline{R}} = [\frac{1}{4}, 0, 0, 0]^T$ and $\underline{e} = [1]$. In particular, if we begin by mating $a-$ with AA then the chance we end up eventually with $A-$ and AA is $2/3$.

Similarly if $C = \{1, 6\}$ then the vector of expectations $[\mathbb{E}\,[T_{iC}]]$ is given by

$$(\underline{\underline{\mathbb{I}}} - \underline{\underline{Q}})^{-1}\underline{e} \quad = \quad \begin{bmatrix} 5 \\ 6 \\ 6 \\ 5 \end{bmatrix}$$

where now

$$\underline{e} \quad = \quad \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}$$

and so from state 5 ($a - \times Aa$) we reduce to just one allele in the entire population in an average of five matings.

Other examples can be generated with different breeding strategies (for example "Parent-Child", treated in some past examinations). Again a Markov chain analysis gives a quick and efficient answer.

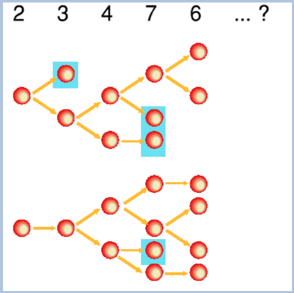
Example 4.6 ***Parent-Child-mating***
Consider the case of a gene for a diploid population with 2 *alleles* A, a. *The mating process is as follows: first mate two individuals to start things off, and nominate one of these as individual* J_0. *Then at each succeeding stage* n *mate a randomly chosen offspring (of appropriate gender)* J_n *with* J_{n-1}. *Use fundamental matrix theory to compute the probability that the breeding strategy results in a pair of individuals which are both purely dominant, if the chain starts in one of* $2, 4, 5, 6, 8$, *where states are numbered by (parent×child):*

$$\begin{array}{llllll} 1: & AA\times AA & 2: & AA\times Aa & 3: & AA\times aa \\ 4: & Aa\times AA & 5: & Aa\times Aa & 6: & Aa\times aa \\ 7: & aa\times AA & 8: & aa\times Aa & 9: & aa\times aa \end{array} .$$

Answer in Exercise 6.6.

4.2. Branching processes

Staying with the genetics theme, one can ask the following question: will a mutant gene establish itself in a population? will it die out? or might it spread widely throughout the population? Clearly this has something to do with the "fitness" of the gene mutation: if the mutation is too damaging then the organism in question will die early and leave no progeny, so the gene mutation will die out. Similarly, if the mutation is very favourable then it will spread very fast. But what if things are more evenly balanced?



Or, how can we model the early stages of an influenza epidemic? will it spread through the population or will it peter out? Or, in a nuclear reactor, how can we model the population of neutrons? as each neutron encounters a U_{235} atom it can cause the atom to undergo fission and send out two more neutrons. What will a chain reaction look like? How can we control it?

Or, how can we model the effect of radiation therapy on the growth of a tumour?

One general feature is shared by all these cases:

Each entity in a given generation yields a random number of off-spring of known distribution, independently of other entities in that generation and of the past.

So statistics of the number X_n in the n^{th} generation are defined by

(a) the number X_0 in the initial generation;

(b) the distribution of size of a typical family Z.

Notice that X so defined will be a Markov chain, since all that is required in order to determine X_{n+1} is X_n the size of the previous generation. If Z_1, Z_2, ...are independent identically distributed family-sizes then the transition probability is given by

$$\begin{aligned} p_{ij} &= \mathbb{P}\left[X_{n+1} = j | X_n = i\right] \\ &= \mathbb{P}\left[Z_1 + \ldots + Z_i = j | X_0 = i\right] \end{aligned}$$

State classification: $\{0\}$ is certainly an essential communicating class on its own, hence 0 is an absorbing state; moreover if $\mathbb{P}\left[Z = 0\right] > 0$ then it is possible to reach $\{0\}$ from anywhere else and $\{0\}$ is the only recurrent class the population process is unable to grow; unless $\mathbb{P}\left[Z > 1\right] > 0$; if $\mathbb{P}\left[Z = 0\right] > 0$, $\mathbb{P}\left[Z > 1\right] > 0$, and $\mathbb{P}\left[Z = 1\right] > 0$ then it can be shown (exercise!) that $\{1, 2, 3, \ldots\}$ form another transient communicating class, and the question is whether the chain is certain to leave the transient class or whether it can stay there forever with positive probability (in which case it must go off to infinity!).

4.2.1. Generating functions

If Z is a random variable with non-negative integer values then we define the generating function $G(z)$ for $|z| \leq 1$ as follows:

$$G(z) \quad = \quad \sum_{m=0}^{\infty} z^m \mathbb{P}\left[Z = m\right] \quad = \quad \mathbb{E}\left[z^Z\right]$$

We shall have need of $G(z)$ the generating function for the typical family-size Z, also of $F_n(z)$ the generating function of the number X_n in the n^{th} generation.
Notice that we can use standard arguments for applying probability generating functions to random sums of random variables, as in Exercise 7.2:

$$\begin{aligned} F_{n+1}(z) \quad &= \quad \mathbb{E}\left[z^{X_{n+1}}\right] \\ &= \quad \sum_{j=0}^{\infty} \mathbb{E}\left[z^{X_{n+1}} | X_n = j\right] \mathbb{P}\left[X_n = j\right] \end{aligned}$$

$$= \sum_{j=0}^{\infty} \mathbb{E}\left[z^{Z_1+\ldots+Z_j} | X_n = j\right] \mathbb{P}\left[X_n = j\right]$$

$$= \sum_{j=0}^{\infty} \mathbb{E}\left[z^{Z_1+\ldots+Z_j}\right] \mathbb{P}\left[X_n = j\right]$$

$$= \sum_{j=0}^{\infty} \mathbb{E}\left[z^{Z_1}\right] \ldots \mathbb{E}\left[z^{Z_j}\right] \mathbb{P}\left[X_n = j\right]$$

$$= \sum_{j=0}^{\infty} \mathbb{E}\left[z^{Z_1}\right] \ldots \mathbb{E}\left[z^{Z_1}\right] \mathbb{P}\left[X_n = j\right]$$

(taking the product j times)

$$= \sum_{j=0}^{\infty} (\mathbb{E}\left[z^{Z}\right])^j \, \mathbb{P}\left[X_n = j\right]$$

$$= \sum_{j=0}^{\infty} G(z)^j \, \mathbb{P}\left[X_n = j\right]$$

$$= F_n(G(z)) \, .$$

In the special case of $X_0 = 1$ we find $F_0(z) = \mathbb{E}\left[z^{X_0}\right] = z$ and so can deduce

$$\begin{aligned} F_1(z) &= F_0(G(z)) \;=\; G(z) \\ F_2(z) &= F_1(G(z)) = G(G(z)) \\ &\ldots \\ F_n(z) &= F_{n-1}(G(z)) = G(\ldots(G(z))\ldots) = G(F_{n-1}(z))\,. \end{aligned}$$

(This can also be seen by conditioning on X_1 instead of X_n in the previous argument.)

4.2.2. Extinction probability

The crucial question in each of our applications is as follows: will the process die out with probability one? Or, we need to know how to calculate the *extinction probability*

$$f_{10}^{(*)} = \mathbb{P}\left[X_n = 0 \text{ for all sufficiently large } n \mid X_0 = 1\right]\,.$$

WARWICK
Home Page
Title Page
Contents
Page 133 of 237
Go Back
Full Screen
Close
Quit

We argue

$$\begin{aligned} f_{10}^{(*)} &= \mathbb{P}\left[X_n = 0 \text{ for all sufficiently large } n \mid X_0 = 1\right] \\ &= \lim_{n\to\infty} \mathbb{P}\left[X_n = 0, X_{n+1} = 0, \ldots \mid X_0 = 1\right] \\ &= \lim_{n\to\infty} \mathbb{P}\left[X_n = 0 \mid X_0 = 1\right] \end{aligned}$$

(the step to the limit really requires arguments from ST213, but we assume it here as it is quite reasonable.)
Notice we want $f_{i0}^{(*)}$ as defined in Definition 2.35. Using independence of different families, we can compute $f_{i0}^{(*)} = \left(f_{10}^{(*)}\right)^i$.
Let $x_n = \mathbb{P}\left[X_n = 0 | X_0 = 1\right]$, the probability of extinction by time n, so $x_n = \sum_{r=0}^n f_{10}^{(r)}$. Notice that

$$x_n = F_n(0)$$

from definition of F_n. Suppose that $X_0 = 1$. We find

$$x_{n+1} = F_{n+1}(0) = G(F_n(0)) = G(x_n)$$

so the sequence

$$G(0) = \mathbb{P}\left[X_1 = 0 | X_0 = 1\right] = x_1,\ x_2,\ \ldots$$

is given by iterating G.
There are two obvious cases:

- $G(0) = 0$ (no family is ever empty!) means $x_n = 0$ for ever so no extinction!
- $G(0) = 1$ (all families are empty!) means $x_n = 1$ for all n so we get extinction immediately.

Theorem 4.7 ***Formula for the extinction probability*** *The extinction probability α is the smallest non-negative root of the equation*

$$\alpha = G(\alpha)\,.$$

Proof: The two cases $G(0) = 0$ or 1 are already covered. If $0 < G(0) = \mathbb{P}\left[Z = 0\right] < 1$ then we argue as follows:
G is an increasing function and its derivative G' is positive (consider

$$G'(z) = \sum_{j=0}^{\infty} j z^{j-1}\,\mathbb{P}\left[Z = j\right]$$

and not all terms are zero if we know $\mathbb{P}\left[Z = 0\right]$ is smaller than 1)

WARWICK
Home Page
Title Page
Contents
Page 135 of 237
Go Back
Full Screen
Close
Quit

This means, $0 \leq x < y \leq 1$ implies $G(x) < G(y)$
Now $G(0) = \mathbb{P}[Z = 0] = x_1 > x_0 = 0$.
Hence $x_2 = G(x_1) > x_1 = G(x_0) \ldots$ hence $x_{n+1} = G(x_n) > x_n = G(x_{n-1})$ so the x sequence is increasing. As it is bounded above by 1 it must converge. Its limit α is the extinction probability, and must by continuity solve $\alpha = G(\alpha)$.
The next part is similar to the argument of Theorem 3.11. Suppose β is another non-negative solution of $\beta = G(\beta)$. Then we must have

$$\beta \geq 0 = x_0 \,.$$

Hence, applying G to both sides,

$$\begin{aligned} G(\beta) &= \beta \geq x_1 = G(x_0) \\ \ldots & \\ G(\beta) &= \beta \geq x_{n+1} = G(x_n) \end{aligned}$$

so $\beta \geq \alpha$ as required. $\square$

The above can be awkward to apply: exact calculations are feasible only for a limited range of family-size probability generating functions, for example the *fractional-linear probability generating functions*

$$G(z) = \frac{a + bz}{1 - cz} .$$

(We need conditions on coefficients a, b, c to ensure we get a probability generating function here: they must be non-negative and $a + b + c = 1$. See Exercise 7.3.)

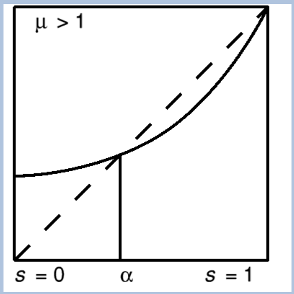
Theorem 4.8 ***Criterion based on mean family-size*** *Suppose* $G(0) > 0$. *Consider* $\mu = G'(1)$ *the mean family-size. Then*

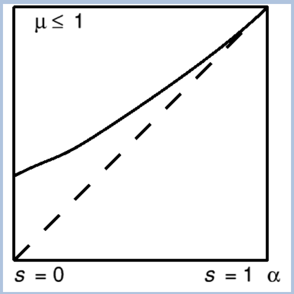
- $\mu \leq 1$ *implies certain extinction*
- $\mu > 1$ *implies extinction is* not *certain.*

We describe $\mu > 1$ as the *supercritical case*, $\mu < 1$ as the *subcritical* case, and $\mu = 1$ as the *critical case*.

Proof: Consider the graph of $G(z)$ for z in the range $[0, 1]$. It is continuous and monotonic nondecreasing (since G' is non-negative).

Moreover the same is true for the derivative G'. These facts mean the graph *must* be one of the two possibilities below.





We build up these pictures in stages:

- Mark in $G(0) > 0$ (our theorem statement assumes this!) and $G(1) = 1$;
- Then consider the case $G'(1) > 1$ and note a crossing argument shows there must be a solution $\alpha = G(\alpha)$ with $0 < \alpha < 1$. For otherwise the continuous curve $y = G(x) - x$ cannot cross from above 0 at $x = 0$, to below 0 at $x = 1-$. ($G(x) - x < 0$ is required for $x = 1-$ if $G'(1) - 1 > 0$.)

- Then consider the case $G'(1) \leq 1$. Now the continuous curve $y = G(x) - x$ has a strictly negative slope everywhere in $(0, 1)$, since $G'(x)$ is strictly increasing to 1 over this range (examine the power series

$$G'(x) = p_1 + 2p_2 x + 3p_3 x^2 + \ldots$$

which has non-negative coefficients not all zero). So $y = G(x) - x$ must strictly decrease from $G(0)$ to 0. This means $G(x) - x$ must not equal 0 in $(0, 1)$.

$\square$

Remark 4.9 *If $X_0 = n$ then $\mathbb{P}\left[\text{ eventual extinction }\right] = \alpha^n$ by independence.*

Example 4.10 *Suppose the number of offspring is 0 (prob p), 1 (prob q), or 2 (prob r), where $p + q + r = 1$ and neither of p, r are zero. Then*

$$\mathbb{E}\left[Z\right] = q + 2r$$
$$G(z) = p + qz + rz^2$$

Solve $\alpha = G(\alpha)$, *which is to say*

$$p + (q-1)\alpha + r\alpha^2 \quad = \quad 0$$

(and notice $q - 1 = -(p + r)$*) to get two roots*

$$\frac{(r+p) \pm \sqrt{((r-p)^2)}}{2r} \quad = \quad 1 \text{ or } p/r$$

If $\mathbb{E}[Z] \leq 1$ *then* $q + 2r \leq 1$ *and so* $r \leq p$*. Then* $p/r \geq 1$ *and so extinction must be* sure.
If $\mathbb{E}[Z] > 1$ *then* $q + 2r > 1$ *and so* $r > p$*. Then* $p/r < 1$ *and so extinction probability is* $p/r < 1$*: Theorem 4.7* and *Theorem 4.8 both check out!*

4.2.3. Application to cell populations

Consider the effect of X-*radiation therapy* on a small tumour.

Example 4.11 *Cells in the tumour divide according to a branching process as in the example immediately preceding, but with* $q = 0$ *and* *with* $r = 1 - p = \frac{1}{2}(1 + \epsilon)$. *Suppose* $\epsilon > 0$, *so that the branching process is supercritical by Theorem 4.8 and calculation of the mean family-size* $1 + \epsilon$. *By Theorem 4.7 the extinction probability is given by* $p/r = (1 - \epsilon)/(1 + \epsilon)$.

The effect of radiation is to alter the probability of a tumour cell successfully dividing: the new probability of a cell having two immediate descendants changes to $r' = 1 - p' = \frac{1}{2}(1 + \epsilon\beta)$ *for some positive* $\beta < 1$. *Now* $1+\epsilon\beta$ *is still greater than* 1, *so extinction is still not sure, but its probability increases to* $p'/r' = (1 - \epsilon\beta)/(1 + \epsilon\beta)$.

Typical figures: $\epsilon = 1/9$, $\beta = 1/11$, *leading to*
$\mathbb{P}\left[\,\text{tumour growth}\,\right] = 20\%$ *without therapy,*
$\mathbb{P}\left[\,\text{tumour growth}\,\right] = 2\%$ *with therapy.*

Example 4.12 *We can also study the total number* K *of tumour cells produced from a single original tumour cell. Let*

$$R(s) \quad = \quad \mathbb{P}\left[K = 0\right] + s\,\mathbb{P}\left[K = 1\right] + s^2\,\mathbb{P}\left[K = 2\right] + \ldots$$

be the probability generating function of this number. Notice that if the tumour branching process is supercritical then there is a positive *probability of this number being infinite and so*

$$R(1) \quad = \quad \mathbb{P}[K=0] + \mathbb{P}[K=1] + \mathbb{P}[K=2] + \ldots \quad < \quad 1\,.$$

Let the number of immediate descendants of the original cell be Z. Then

$$K \quad = \quad 1 + K_1 + \ldots + K_Z$$

where $K_1, \ldots, K_Z$ are the independent random total numbers of cells produced from each of the immediate descendants. Using the same argument as we used to produce $F_{n+1}(s) = F_n(G(s))$, we deduce $R(s) = sG(R(s))$.

In the particular case $G(s) = p + rs^2$ as above, this leads to a quadratic equation for $R(s)$ which we can solve:

$$R(s) \quad = \quad \frac{1 \pm \sqrt{(1 - 4prs^2)}}{2rs}$$

Which root to take? If we take the positive root then R tends to infinity at $s = 0$, which is silly, so we must take the negative root:

$$R(s) \quad = \quad \frac{1 - \sqrt{(1 - 4prs^2)}}{2rs}$$

with a value at $s = 0$ of zero (evaluate by l'Hôpital's rule: differentiate top and bottom, let s tend to zero).
The mean total number can be found by calculating $R'(1)$. Actually it also can be found without solving $R(s) = sG(R(s))$. Obviously it will be infinite if the process is supercritical, since there will be a positive chance of an infinite total number! If the process is not supercritical then K is a proper random variable and hence $\mathbb{E}[K] = R'(1)$.
But

$$R'(s) = G(R(s)) + sG'(R(s))R'(s)$$

hence

$$R'(1) = G(R(1)) + G'(R(1))R'(1) .$$

Now use $R(1) = 1$ (vital here that K is a proper finite random variable), $G(1) = 1$, $G'(1) = \mathbb{E}[Z]$ and so deduce

$$\mathbb{E}[K] = 1 + \mathbb{E}[Z]\,\mathbb{E}[K] .$$

We deduce:

- in the critical case ($\mathbb{E}[Z] = 1$) we have

$$\mathbb{E}[K] = \infty$$

- in the subcritical case ($\mathbb{E}[Z] < 1$) we have

$$\mathbb{E}[K] = 1/(1 - \mathbb{E}[Z])$$

In fact we can use the explicit form of the probability generating function $R(s)$ to compute all the probabilities $\mathbb{P}[K = n]$! For we know from the definition of a probability generating function (Equation C.1) that

$$R(s) \quad = \quad \sum_{n=0}^{\infty} \mathbb{P}[K = n]\, s^n \,.$$

WARWICK
Home Page
Title Page
Contents
Page 144 of 237
Go Back
Full Screen
Close
Quit

On the other hand, we can expand $R(s)$ using the generalized binomial theorem:

$$
\begin{aligned}
R(s) &= \frac{1-\sqrt{(1-4prs^2)}}{2rs} \\
&= \frac{1}{2rs} - \Big(\frac{1}{2rs} - \frac{1}{2} \times \frac{4prs^2}{2rs} - \frac{1}{3!} \times \frac{1}{2} \times \frac{-1}{2} \times \frac{(4prs^2)^2}{2rs} \\
&\qquad -\frac{1}{4!} \times \frac{1}{2} \times \frac{-1}{2} \times \frac{3}{2} \times \frac{(4prs^2)^3}{2rs} \\
&\qquad -\frac{1}{5!} \times \frac{1}{2} \times \frac{-1}{2} \times \frac{3}{2} \times \frac{-5}{2} \times \frac{(4prs^2)^4}{2rs} - \ldots\Big) \\
&= ps + \frac{4!}{2!2!} p^2 r s^3 + \frac{6!}{3!3!} p^3 r^2 s^5 + \ldots + \frac{(2n)!}{n!n!} p^n r^{n-1} s^{2n-1} + \ldots .
\end{aligned}
$$

Equating terms in the two convergent power series, we deduce

$$
\mathbb{P}\left[K = 2n-1\right] = \frac{(2n)!}{n!n!} p^n r^{n-1} \tag{4.1}
$$

for $n = 1, 2, \ldots$.

The coefficients $\frac{(2n)!}{n!n!}$ are called the *Catalan numbers* and arise in all kinds of mathematical contexts.

There are also more general applications to social behaviour! Look up Malcolm Gladwell's *New Yorker* article on "The Tipping Point" at

http://www.gladwell.com/1996/1996_06_03_a_tipping.htm

which describes super/sub-criticality in human behaviour — without once mentioning the crucial underlying mathematics!

Example 4.13 *Here is an epidemiological example taken from Gladwell's article, with added mathematics. Suppose a large city has* 1000 *tourists from a specific area in a summer, and suppose each of these carries an infection of an untreatable strain of twenty-four hour 'flu. Suppose anyone who has this 'flu will pass it on to an average of two-per-cent of the people they meet over the twenty-four hours of infection. Suppose further that tourist and city-dweller alike meet an average of fifty people a day.*
Under mild assumptions, we can model the total number infected in a given summer by the total number of individuals in a branching

WARWICK
Home Page
Title Page
Contents
◀◀ ▶▶
◀ ▶
Page 146 of 237
Go Back
Full Screen
Close
Quit

process with family-size distribution Binomial$(50, 0.02)$. Each individual infects on average 1 person in the twenty-four hours of their infectious period. From Theorem 4.8 we know there is probability 1 that only a finite ($=$ small in reality!) number will be infected that summer.

If on the other hand tourist and city-dweller alike meet a few more people, say $50(1+\varepsilon)$ for $\varepsilon > 0$ (for whatever reason — perhaps a public transport slowdown makes things more crowded) then the average family-size will increase beyond 1, and there arises a positive chance that an infinite ($=$ large in reality![2]) number will be infected. How likely is this?

Exact calculation would have to be numerical. But we can make approximations. Approximate Binomial$(50(1+\varepsilon), 0.02)$ by Poisson$(1+\varepsilon)$, with family-size probability generating function

$$G(s) \quad = \quad \exp\left(-(1+\varepsilon)(1-s)\right) .$$

[2]The city population is obviously not infinite, so an infinite number will not be infected. But in our mathematical model "infinity" approximates a large fraction of the large city population.

By Theorem 4.7 we know the probability of extinction must be the least non-negative root $\alpha = 1 - \delta$ of $\alpha = G(\alpha)$. Approximating $\exp(u) \approx 1 + u + u^2/2$ for small u, we obtain

$$1 - \delta \quad \approx \quad 1 - (1+\varepsilon)\delta + \frac{1}{2}(1+\varepsilon)^2\delta^2$$

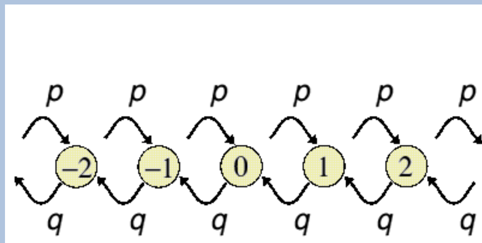
and so $\delta \approx 2\varepsilon$ corresponds to the approximate probability of *non*-extinction for infections resulting from a single infective.
However we are dealing with 1000 infectives. The probability that at least one of them produces a super-critical branching process is therefore approximately (for small ε)

$$1 - (1 - 2\varepsilon)^{1000} \quad \approx \quad 2000\varepsilon$$

(expanding the power using the binomial theorem, then ignoring higher-order terms).
Putting this altogether, it shows that the average number of people met in a day increases from 50 to just $50(1 + 1/4000) = 50.0125\ldots$ then there is a 50% chance of a major epidemic!
This extreme sensitivity demonstrates the importance of branching process theory.

4.3. Random walks

Consider simple random walk on the integers, with $p + q = 1$ and such that neither p nor q are zero:



This is *symmetric* if $p = q = \frac{1}{2}$, otherwise *asymmetric*. The state-space is irreducible, being one essential class (it may be transient, it may be recurrent!). Note that the period is 2. (Sometimes we allow the walk to stay where it is with probability r, and $p + q + r = 1$: if $r > 0$ then the period will be 1.)

Applications

(a) A prototype for random processes arising in many areas including biology, operational research, stock markets.

(b) Simple case of random walks arising in sequential statistical methods.

(c) A multidimensional version is used as a model for polymer molecules.

Paths of random walks

Recall the discussion about path-space at the start of Section 2. A *path* for a discrete-time random process is simply a list of states $\{x_a, x_{a+1}, \ldots, x_m\}$ which might be visited in immediate sequence by the random process at times a, $a + 1$, $\ldots$, m. Thus we can speak of the probability that X follows a given path. For a simple random walk we must have $x_n = x_{n-1} \pm 1$.

4.3.1. Transition probabilities

We begin with an easy distributional fact.

Theorem 4.14 ***Distribution of a random walk*** *Consider a simple random walk X begun at 0 at time 0 (for convenience' sake). The distribution of the random variable X_n is given for even $n + k$ by*

$$\mathbb{P}\left[X_n = k | X_0 = 0\right] \quad = \quad \binom{n}{\frac{n+k}{2}} p^{(n+k)/2} q^{(n-k)/2} \,.$$

Here we use the usual convention that the combinatorial coefficient

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

is zero if r exceeds n or is negative.

Notice the effect of periodicity:

$$\mathbb{P}\left[X_n = k | X_0 = 0\right] \quad = \quad 0 \text{ if } n + k \text{ is odd!}$$

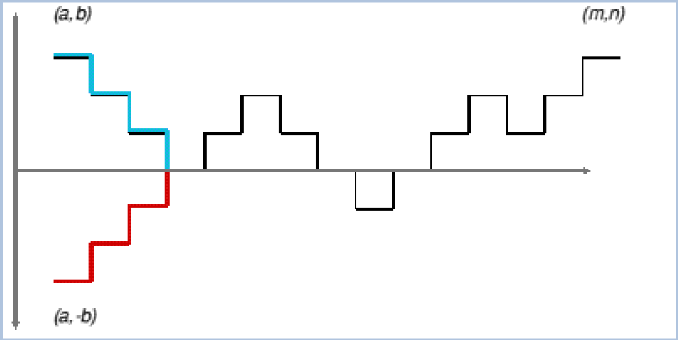
Proof: To go from 0 to k in n time steps requires k more up-jumps u than down-jumps d: hence we have $u + d = n$ and $u - d = k$. Therefore $u = \frac{1}{2}(n+k)$ and $d = \frac{1}{2}(n-k)$. The result follows by the independence of different jumps of the random walk. $\square$

WARWICK
Home Page
Title Page
Contents
Page 152 of 237
Go Back
Full Screen
Close
Quit

4.3.2. Reflection principle

Random walks are very symmetric, and this allows us to use a special technique.

Theorem 4.15 ***Reflection Principle*** *Consider a simple random walk. Suppose that $a < m$ and b, n are of the same sign. There are as many paths from b at time a to n at time m hitting the x-axis as there are paths from $-b$ at time a to n at time m.*



Proof: For every such path from (a, b) to (m, n) we obtain a unique path from $(a, -b)$ to (m, n) by reflecting (in the x-axis) the initial segment up to the first time that the path hits the x-axis. This sets up a one-to-one correspondence, so there are as many paths of one kind as the other. $\square$

Note that there is a reversed statement as well (paths from (a, b) to $m, \pm n$).

4.3.3. Application to first-passage times

Consider the problem of finding

$$f_{0k}^{(n)} = \mathbb{P}\left[T_{0k} = n\right] .$$

We can use the reflection principle to find this exactly! Suppose $k \geq 1$ for convenience.

Theorem 4.16 ***Distribution of first passage time for random walk***
Suppose $k \geq 1$. When $n + k$ is even,

$$\mathbb{P}\left[T_{0k} = n\right] \quad = \quad \frac{k}{n}\binom{n}{\frac{1}{2}(n+k)} p^{\frac{1}{2}(n+k)} q^{\frac{1}{2}(n-k)} .$$

If $n + k$ is odd, the probability is zero.

Clearly T_{0k} can only be one of k, $k+2$, $k+4$, ...if it is finite at all. If $T_{0k} = n = k+2r$ then there must be $k+r$ up-jumps and r down-jumps. Hence we can compute (for $n = k+2r$; notice $r = \frac{1}{2}(n-k)$):

$$
\begin{aligned}
f_{0k}^{(n)} \quad &= \\
&= \sum \mathbb{P}\left[\, \text{path from } 0 \text{ at } 0 \text{ to } k \text{ at } n, \text{ below } k \text{ before then} \,\right] \\
&= \sum p^{k+r} q^{r} \qquad \text{(summing over all such paths)} \\
&= p^{k+r} q^{r} \#\{\text{ paths } (0,0) \text{ to } (n,k), \text{ otherwise below } k \,\} \\
&= p^{k+r} q^{r} \#\{\text{paths } (0,0) \text{ to } (n-1,k-1), \text{ always below } k \,\} \\
&= p^{k+r} q^{r} \Big(\#\{\text{ paths } (0,0) \text{ to } (n-1,k-1) \,\} \\
&\qquad - \#\{\text{ paths } (0,0) \text{ to } (n-1,k-1) \text{ hitting } k \,\} \Big)
\end{aligned}
$$

(where $\#A$ denotes the number of elements in the set A) and we can apply the reflection principle of Theorem 4.15 to compute

$$\begin{aligned} &\#\{ \text{ paths } (0,0) \text{ to } (n-1,k-1) \text{ hitting } k \} \\ = \quad &\#\{ \text{ paths } (0,2k) \text{ to } (n-1,k-1) \} \end{aligned}$$

and hence (using $n = k + 2r$):

$$\begin{aligned} f_{0k}^{(n)} &= p^{k+r} q^r \Big(\#\{ \text{ paths } (0,0) \text{ to } (n-1,k-1) \} - \\ &\qquad \#\{ \text{ paths } (0,2k) \text{ to } (n-1,k-1) \} \Big) \\ &= p^{k+r} q^r \left(\binom{k+2r-1}{k+r-1} - \binom{k+2r-1}{k+r} \right) \end{aligned}$$

We can now use the identity

$$\binom{n-1}{n-r-1} - \binom{n-1}{n-r} = \left(\frac{n-r}{n} - \frac{r}{n} \right) \binom{n}{n-r} = \frac{n-2r}{n} \binom{n}{n-r}$$

to deduce the result.

WARWICK
Home Page
Title Page
Contents
Page 156 of 237
Go Back
Full Screen
Close
Quit

4.3.4. Generating function approach

An alternative approach uses generating functions. Using the notation of Section 2.4 and first-step decomposition we find

$$\begin{aligned} F_{0,1}(s) &= \mathbb{E}\left[s^{T_{0,1}}\right] \\ &= s\,\mathbb{P}\left[X_1 = 1\right] + \mathbb{E}\left[s^{1+T_{-1,1}} | X_1 = -1\right]\mathbb{P}\left[X_1 = -1\right] \\ &= ps + qsF_{-1,1}(s) \end{aligned}$$

(either the first step is up - get right there! - or down - need to travel from -1). Now a symmetry and independence argument tells us to go from -1 to 1 we need to go first from -1 to 0, then from 0 to 1, and so

$$F_{-1,1}(s) \quad = \quad (F_{0,1}(s))^2 \,.$$

This allows us to deduce a quadratic equation for $F_{0,1}(s)$ as follows: $F_{0,1}(s) = ps + qs(F_{0,1}(s))^2$ and so

$$F_{0,1}(s) \quad = \quad \frac{1 \pm \sqrt{(1-4pqs^2)}}{2qs}$$

(use $p + q = 1$ here!).

WARWICK
Home Page
Title Page
Contents
◀◀ ▶▶
◀ ▶
Page 157 of 237
Go Back
Full Screen
Close
Quit

Which branch of root? We know $F_{0,1}(s)$ is continuous (so it must be the same sign for all s) and bounded above by 1. But if we take the positive sign then the expression exceeds 1 if $s < 1$, and even goes to infinity as s tends to 0. Therefore the correct answer must be as given below.

Theorem 4.17

$$F_{0,1}(s) \quad = \quad \frac{1-\sqrt{1-4pqs^2}}{2qs},$$

and easily

$$F_{0,k}(s) \quad = \quad \left(\frac{1-\sqrt{1-4pqs^2}}{2qs}\right)^k.$$

Now repeated differentiation will give $f_{0k}^{(n)}$ by

$$\left[\frac{\mathrm{d}^n}{\mathrm{d}\,s^n}F_{0,k}(s)\right]_{s=0} \quad = \quad n!f_{0k}^{(n)}$$

(use theory of probability generating functions).
Also

$$
\begin{aligned}
f_{0k}^{(*)} &= \lim_{s\to 1} F_{0,k}(s) = \left[\frac{1-|1-2p|}{2q}\right]^k \\
&= (p/q)^k \qquad \text{if } p < \tfrac{1}{2}, \\
&= 1 \qquad \text{if } p \geq \tfrac{1}{2}.
\end{aligned}
$$

(This agrees with earlier calculations!)
The following result follows easily.

Theorem 4.18 ***Transience/recurrence for the one-dimensional simple random walk*** *The simple random walk X is recurrent if symmetric, and transient if asymmetric.*

What about random walks in higher dimensions? Theorem 2.38 and Theorem 4.14 give the answer, together with *Stirling's approximation*:

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

in the sense that LHS divided by RHS tends to a positive limit as n tends to infinity.

4.3.5. Case of two-dimensional symmetric simple random walk

A two-dimensional symmetric random walk (we won't deal with asymmetric case, which rather easily is shown to be transient using the one-dimensional results!) moves on the integer lattice $\{(m,n) : m,n \text{ integers }\}$ in the plane, at each step moving with equal probability to each one of its four neighbours. By rotating the lattice through 45^{deg} we see that at each step it is equally likely to move

- up and left;
- up and right;
- down and left;
- down and right;

thus it can viewed as produced by two one-dimensional symmetric simple random walks each driving a coordinate, and so we can borrow from Theorem 4.14 to deduce

$$p^{(2n)}_{(0,0),(0,0)} \quad = \quad \left(p^{(2n)}_{00}\right)^2 = \left(4^{-n}\binom{2n}{n}\right)^2$$

Now use Stirling's formula to show

$$\binom{2n}{n} \quad = \quad \frac{(2n)!}{n!\;n!} \quad \sim \quad \frac{\sqrt{2\pi}(2n)^{2n+\frac{1}{2}}}{2\pi(n^{n+\frac{1}{2}})^2} \quad = \quad \frac{2^{2n+\frac{1}{2}}}{\sqrt{2\pi n}}\,.$$

Applying Theorem 2.38 we deduce that the random walk is transient only when $\sum_n p^{(2n)}_{(0,0),(0,0)}$ converges. However

$$\sum_n p^{(2n)}_{(0,0),(0,0)} \quad = \quad \sum_n \left(4^{-n}\binom{2n}{n}\right)^2 \quad \sim \quad \sum_n \frac{1}{\pi n} = \infty\,.$$

Thus we have the following result.

Theorem 4.19 ***Recurrence of two-dimensional symmetric simple random walk*** *The symmetric simple random walk in two dimensions is recurrent.*

4.3.6. Case of three-dimensional symmetric simple random walk

A three-dimensional symmetric random walk (we won't deal with asymmetric case, which again is easily shown to be transient) moves on the integer lattice

$$\{(m,n,p):\ m,n,p \text{ integers }\}$$

in 3-space, at each step moving with equal probability to each one of its eight neighbours. By rotating the lattice we again see that it can be viewed as three one-dimensional symmetric simple random walks driving each of the coordinates, and so we can borrow from Theorem 4.14 to deduce

$$p^{(2n)}_{(0,0,0),(0,0,0)} \quad = \quad \left(p^{(2n)}_{00}\right)^3 = \left(4^{-n}\binom{2n}{n}\right)^3 .$$

Now use Stirling's formula again: applying Theorem 2.38 we deduce that the random walk is transient only if $\sum_n p^{(2n)}_{(0,0,0),(0,0,0)}$ converges.

However

$$\sum_n p^{(2n)}_{(0,0,0),(0,0,0)} \quad = \quad \sum_n \left(4^{-n}\binom{2n}{n}\right)^3 \leq \sum_n \frac{1}{(\pi n)^{3/2}} < \infty\,.$$

Thus we obtain the following result.

Theorem 4.20 ***Transience of high-dimensional random walks*** *The symmetric simple random walk in three (or more!) dimensions is transient.*

Example 4.21 ***Lost in space*** *The* Enterprise *spaceship is lost in space, with all navigational instruments destroyed in a clash with the Romulans. It can make hyperspace jumps which move it at random on an integer lattice in* 3*-space, according to a symmetric simple random walk. The above calculations show that with positive probability it will never return to the starbase at* $(0,0,0)$.

Example 4.22 ***Lost in space – continued*** *Spock managed to make a single star observation before the final destruction of the navigation instruments, which allows Captain Kirk to exert some control over* *the hyperspace configuration so as to fix the symmetric simple random walk of hyperspace jumps to lie in a* planar *integer lattice containing both the* Enterprise *and the starbase at* $(0, 0, 0)$*. The above calculations now show that there is probability* 1 *of returning to the starbase. (However further calculations would show that the time of return could be very far in the future: the mean return time is infinite!)*

Chapter 5

Renewal Theory

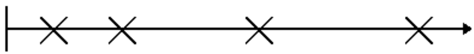
Contents

Example 5.1 *Consider a piece of electronic equipment. Suppose we replace a crucial component whenever it fails. What can be said about the mean number of failures occurred by time t? This is a fundamental question from* reliability theory.

5.1. Definition of renewal process

Suppose component lifetimes are T_1, T_2, …; independent, identically distributed, positive, integer-valued, may be *defective*: which is to say, ($\mathbb{P}\left[T_i = \infty\right] > 0$). Then component *renewals* occur at times

$$\begin{aligned} S_1 &= T_1\,, \\ S_2 &= S_1 + T_2 = T_1 + T_2\,, \\ \ldots & \\ S_{n+1} &= S_n + T_{n+1} = T_1 + T_2 + \ldots + T_{n+1}\,. \end{aligned}$$



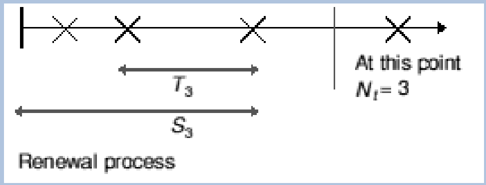
Failure times of electronic component

We also define N_t as the number of renewals to have happened by time t: so that

$$N_t = \#\{m > 0 : S_m \leq t\}$$

and we have a basic identity connecting the *counting* and *renewal epoch* points of view:

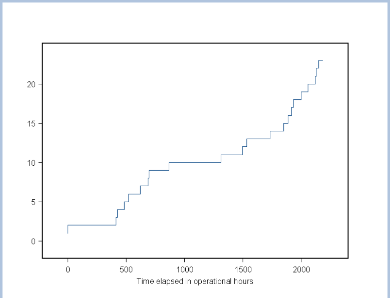
$$[N_t \geq n] = [S_n \leq t] \,.$$



Example 5.2 ***Airconditioning equipment in Boeing 727 aircraft*** *You notice when one of these fails! they have to be replaced immediately. Here is some data taken from Cox and Snell (1981, Example T), concerning intervals between failures (operating hours) of airconditioning equipment in a Boeing 720 aircraft.*

413	14	58	37	100	65	9	169	447	184	36
201	118	34	31	18	67	57	62	7	22	34

Here is a graph of numbers of units used against time elapsed in operational hours.



Example 5.3 *Consider a Markov chain X with $X_0 = m$. Let S_n be the (discrete) time at which X returns to state m for the n^{th} time. By the Markov property it follows that the random variables $T_n = S_n - S_{n-1}$ are identically distributed and independent: this is exactly what we need for the return times $\{S_n : n > 0\}$ to form a renewal process.*

Indifferently we call both the counting process $\{N_t : t \geq 0\}$ and the epoch process $\{S_n : n > 0\}$ by the name *renewal process*. We shall here deal only with renewal processes where the T_n (equivalently the S_n) are whole-number valued: that is to say, *discrete-time renewal* *processes*. Continuous-time renewal processes are broadly similar except that one can end up with infinitely many renewals in bounded time intervals - a complication we do not want to consider!

Notation

$$
\begin{aligned}
p_k &= \mathbb{P}\left[T_1 = k\right] \\
u_n &= \mathbb{P}\left[\text{ renewal at time } n\ \right] \qquad \text{for } n \geq 1 \\
&= \mathbb{P}\left[S_m = n \text{ for some } m\right] \qquad \text{for } n \geq 1.
\end{aligned}
$$

To make later formulae simpler we set $u_0 = 1$. We call $\{u_n : n \geq 0\}$ the *renewal sequence*.
In the non-defective case (*i.e.* $\mathbb{P}\left[T_1 < \infty\right] = 1$) we write

$$
\begin{aligned}
\mu &= \mathbb{E}\left[T_1\right] \\
\lambda &= \frac{1}{\mu}
\end{aligned}
$$

with the convention $1/\infty = 0$.

We say the renewal sequence has *period* $\gcd\{k : p_k > 0\}$ (just like for the Markov chain case!). If the renewal sequence has period 1 we say it is *aperiodic*. The following lemma is essentially the same as Lemma 2.30 but applied to renewal processes rather than periodic Markov chains; we re-state it here for convenience.

Lemma 5.4 *from number theory: if the period is* 1 *then for some sufficiently large integer* M *we have* $u_n > 0$ *for all* $n > M$.

For proof see Appendix B.

5.2. Delayed renewal processes

It is convenient to introduce the idea of a renewal process where the first component has a different lifetime from the rest (perhaps because it is not completely new ...). Consider T_0' independent of the T_1, T_2, ..., non-negative, integer-valued, of a *different* distribution: suppose

$$\mathbb{P}\left[T_0' = k\right] = a_k \text{ for } k \geq 0.$$

The *delayed renewal process* corresponding to delay distribution given by T_0' is given by

$$\begin{aligned}
S_0' &= T'_0\,, \\
S'_1 &= T'_0 + T_1\,, \\
S'_2 &= S_1' + T_2 = T'_0 + T_1 + T_2\,, \\
&\ldots \\
S'_{n+1} &= S_n' + T_{n+1} = T'_0 + T_1 + T_2 + \ldots + T_{n+1}\,.
\end{aligned}$$

We write

$$\begin{aligned} v_n &= \mathbb{P}\left[\,\text{(delayed) renewal at time } n\,\right] \\ &= \mathbb{P}\left[S'_m = n \text{ for some } m \geq 0\,\right]\,. \end{aligned}$$

We call $\{v_n : n \geq 0\}$ the *delayed renewal sequence*, using $\{a_k\}$ for the delay distribution.
The original case is also called the *zero-delayed renewal sequence*.

5.3. Renewal equations

We can use a *first-passage decomposition* just as with Markov chains (Section 2.4.2) to decompose the event $[\text{ renewal at time } n\ ]$ according to $[T_1 = k]$ for $k = 1, 2, \ldots, n$ to get

$$u_n = \sum_{k=1}^{n} p_k u_{n-k} = \sum_{k=0}^{n-1} p_{n-k} u_k\,. \tag{5.1}$$

(Recall $u_0 = 1$ for the zero-delayed renewal sequence.)
In the delayed renewal case, decompose by $[T_0' = k]$ for $k = 0, 2, \ldots, n$ to get

$$v_n = \sum_{k=0}^{n} a_k u_{n-k} = \sum_{k=0}^{n} a_{n-k} u_k\,. \tag{5.2}$$

Just as in the arguments in Section 2.4.2 concerning first-passage decomposition, we now define generating functions. In the zero-delayed case we set

$$F(s) = \sum_{k=1}^{\infty} p_k s^k \quad \text{defined for } |s| \leq 1$$

and

$$U(s) = \sum_{k=0}^{\infty} u_k s^k \quad \text{defined for } |s| < 1.$$

This makes sense if we consider the Markov chain example of renewals being returns to a given state: the p_k here is the probability that first return happens at time k (namely $f^{(k)}_{mm}$ while u_k is the probability that the Markov chain finds itself back at the original state at time k (namely $p^{(k)}_{mm}$. Just as for the Markov chain case we can argue

$$U(s) = 1 + U(s)F(s)$$

(uses the renewal equation) and, by letting $s \uparrow 1$, using the techniques of Theorem 2.38, we obtain the following result.

Theorem 5.5

$$\sum_{k=1}^{\infty} p_k < 1$$

if and only if

$$\sum_{k=0}^{\infty} u_k < \infty$$

and then

$$\sum_{k=0}^{\infty} u_k \quad = \quad \frac{1}{1-\sum_{k=1}^{\infty} p_k}$$

Example: Suppose a component is equally likely to fail on any one of the first six days of its operation. Then we have:

k	1	2	3	4	5	6
p_k	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$

and hence

$$F(s) \quad = \quad \frac{s+s^2+s^3+s^4+s^5+s^6}{6}$$

whence

$$U(s) = \frac{1}{1-F(s)} \quad = \quad \frac{6(1-s)}{6-7s+s^7}\,.$$

By repeated differentiation we can now discover various renewal probabilities. For example

$$u_5 = \frac{1}{5!}\left[\frac{\mathrm{d}^5\, U(s)}{\mathrm{d}\, s^5}\right]_{s=0} \quad = \quad \frac{2401}{7776} = 0.3088\,.$$

5.4. Stationary renewal processes

We *assume* the renewal process is nondefective, aperiodic, and $\mu < \infty$. Here is an important special case of a delayed renewal process: $a_k = c_k = \lambda \sum_{i=k+1}^{\infty} p_i$.

Example 5.6 *Check this example is honest, in the sense that* $\sum c_k = 1$.

The clever point about this particular choice of $a_k = c_k$ is that the resulting delayed renewal process is *stationary*.

Theorem 5.7 *In the above case, the so-called* stationary renewal process, *we have*

$$v_n \quad = \quad \mathbb{P}\left[\, \textit{delayed renewal at } n \,\right] \quad = \quad \lambda$$

for all n.

Proof: we use induction and the renewal equations.
Induction: Observe that $v_0 = c_0 = \lambda \sum_{i=1}^{\infty} p_i = \lambda$ as required, so the result certainly holds for $n = 0$. Now suppose that it holds for

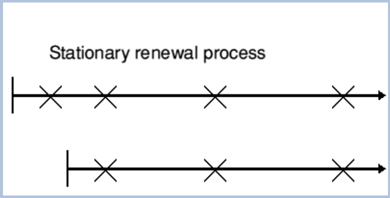
WARWICK
Home Page
Title Page
Contents
Page 177 of 237
Go Back
Full Screen
Close
Quit

$n-1$ so that $v_{n-1}=\lambda$. Use the delayed renewal equation on v_n to see:

$$
\begin{aligned}
v_n &= \sum_{k=0}^{n} c_{n-k}u_k = \lambda\left(\sum_{k=0}^{n}\sum_{i=n-k+1}^{\infty} p_i u_k\right) \qquad \text{using formula for } c_{n-k}\\
&= \lambda\left(\sum_{k=0}^{n-1}\sum_{i=n-k+1}^{\infty} p_i u_k\right) + \lambda\left(u_n\sum_{i=1}^{\infty} p_i\right)\\
&= \lambda\left(\sum_{k=0}^{n-1} u_k \sum_{i=n-k}^{\infty} p_i\right) - \lambda\left(\sum_{k=0}^{n-1} u_k p_{n-k}\right) + \lambda\left(u_n\sum_{i=1}^{\infty} p_i\right)\\
&= v_{n-1} + \lambda\left(u_n - \sum_{k=0}^{n-1} u_k p_{n-k}\right)\\
&\qquad \text{using renewal equation for first summand}\\
&\qquad \text{and } \textstyle\sum_{i=1}^{\infty} p_i = 1 \text{ for third summand}\\
&= \lambda
\end{aligned}
$$

where the last part vanishes by using the zero-delayed renewal equation. $\square$

Corollary 5.8 *If one cuts off the initial segment of a stationary renewal process then it is still stationary. Of course the cut-off point must be non-random, or at least independent of the renewal process itself.*



Proof: immediate from Theorem 5.7 for the following reasons:

- clearly a delayed renewal process;
- from Theorem 5.7 we know $v_n = \lambda$;
- use v_n and (delayed) renewal equation (5.2) to deduce that delay distribution $\{a_k : k = 0, 1, \ldots\}$ must be as given in discussion preceding statement of Theorem 5.7.

$\square$

5.5. Coupling and the renewal theorem

We continue to *assume* the renewal process is nondefective, aperiodic, and that $\mu < \infty$.
We want to prove that at large time *all* such renewal processes behave approximately as if they were stationary. This is interesting from the point of view of reliability theory (it allows us to answer the basic question, what is the average rate of renewals) and also vital for an application to Markov chains which we give in the next chapter.

The coupling construction from here to start of Theorem 5.9 is not examinable

Consider two independent stationary renewal processes $N^{(1)}$ and $N^{(2)}$ with the same lifetime distribution $\{p_i\}$, and consider times of *simultaneous renewals*: set

$$N_k^* \quad = \quad \#\left\{r \leq k : N_r^{(1)} - N_{r-1}^{(1)} = N_r^{(2)} - N_{r-1}^{(2)} = 1\right\}$$

to be the number of simultaneous renewals by time k. Then

- This is also a renewal process!
- by independence we see

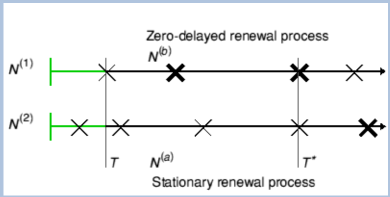
$$v_n^* = \mathbb{P}\left[N^* \text{ renewal occurs at time } n\right] = (v_n)^2 = \lambda^2 > 0$$

So, by Theorem 5.5, N^* is nondefective. Also by Theorem 5.7 we see that the mean value $1/\lambda^*$ of the time to the first simultaneous renewal must be finite.

Let T be the first renewal time of $N^{(1)}$ and notice the following:

- $N_k^{(b)} = N_{k+T}^{(1)} - N_T^{(1)}$ (consider $N^{(1)}$ onwards from time of first incident for $N^{(1)}$) defines a zero-delayed renewal process;
- $N_k^{(a)} = N_{k+T}^{(2)} - N_T^{(2)}$ (consider $N^{(2)}$ onwards from time of first incident for $N^{(1)}$) defines a stationary renewal process (use Corollary 5.8 and independence of $N^{(1)}$ and $N^{(2)}$);
- There is T^* at which these two processes experience a simultaneous renewal for the first time! This happens at either the *first* or the *second* simultaneous renewal for the original processes, and therefore has finite mean.

Now we build a *new* renewal process M by *coupling* these two. Start with $N^{(b)}$ up until time T^* and then switch to use the renewals of $N^{(a)}$ using formulae,



$$
\begin{aligned}
M_k &= N_k^{(b)} && \text{while } k \leq T^* ;\\
M_k &= N_k^{(a)} - N_{T^*}^{(a)} + N_{T^*}^{(b)} && \text{once } k > T^* .
\end{aligned}
$$

The resulting process M is indeed a renewal process (times between renewals are still independent with the right distribution), but starts as a zero-delayed renewal process and eventually (after time T^*) goes on to be a stationary renewal process!

Example:

Lightbulbs are renewed at random intervals of time. In one universe (the (a) universe) light bulbs are renewed according to a stationary renewal process. In another universe (the (b) universe) a zero-delayed renewal process is used. Clearly there is no difference between the possibilities

- the (b) universe is used throughout;
- the (b) universe is used initially, but we switch over to the (a) universe at the first time when lightbulbs are renewed simultaneously in the two universes.

For after all, the renewals in either scenario are characterized as taking place at independent intervals of time with the same distribution! Now the fundamental question for us is the average number of renewals occurring in the zero-delayed renewal process at large time. This is the same as considering the following for large n:

$$
\begin{aligned}
&\mathbb{E}\left[\#\left\{M \text{ renewals by time } n\right\}\right]/n \\
&= \Big(\mathbb{E}\left[\#\left\{N^{(b)} \text{ renewals by } T^*\right\}\right] \\
&+\mathbb{E}\left[\#\left\{N^{(a)} \text{ renewals by } n \text{ since } T^*\right\}\right]\Big)/n \\
&= \mathbb{E}\left[\#\left\{N^{(a)} \text{ renewals by } n\right\}\right]/n \\
&\qquad +\Big(\mathbb{E}\left[\#\left\{N^{(b)} \text{ renewals by } T^*\right\}\right] \\
&\qquad\qquad -\mathbb{E}\left[\#\left\{N^{(a)} \text{ renewals by } T^*\right\}\right]\Big)/n
\end{aligned}
$$

But the last term is bounded above by the mean of T^* (which we have seen above to be finite!) divided by n, and so tends to 0 as n tends to ∞, while the first term is

$$\mathbb{E}\left[\#\left\{N^{(a)} \text{ renewals by } n\right\}\right]/n \quad = \quad \left(\sum_{k=0}^{n} v_k\right)/n = (n\lambda)/n = \lambda\,.$$

End of non-examinable material

This gives us the fundamental results about renewals:

Theorem 5.9 ***Elementary renewal theorem*** *If the renewal process is* aperiodic, nondefective, *and* $\sum_{n=1}^{\infty} np_n = \mu < \infty$

$$\frac{1}{n+1}\sum_{k=0}^{n} u_k \quad \to \quad \lambda$$

where $\lambda = 1/\mu$, *as* n *tends to* ∞. *(In fact it holds for* $\sum_{n=1}^{\infty} np_n = \mu = \infty$ *also: take* $\lambda = 0$.*)*

By the same basic arguments we easily find

Theorem 5.10 ***Renewal theorem*** *If the renewal process is aperi-*odic, nondefective, *and* $\sum_{n=1}^{\infty} np_n = \mu < \infty$, *then for fixed* m *we have*

$$\sum_{k=n+1}^{n+m} u_k \quad \to \quad m\lambda$$

where $\lambda = 1/\mu$, *as* n *tends to* ∞, *and in particular*

$$u_n \quad \to \quad \lambda \qquad \text{as } n \text{ tends to } \infty\,.$$

The periodic case is now dealt with easily, simply by sampling at periodic times:

Corollary 5.11 *If the renewal process is* nondefective, *and* $\sum_{n=1}^{\infty} np_n = \mu < \infty$, *and* $\gcd\{n : p_n > 0\} = d > 1$ *then*

$$u_{nd} \quad \to \quad \lambda_d \qquad \text{as } n \text{ tends to } \infty$$

where λ_d *is given by*

$$\lambda_d \quad = \quad \begin{cases} 1/(\sum_k kp_{kd}) & \text{if } \sum_k kp_{kd} < \infty \\ 0 & \text{otherwise.} \end{cases}$$

Moreover

$$\sum_k kp_{kd} \quad = \quad \frac{1}{d}\sum_k kp_k\,.$$

Example 5.12 *Suppose* $\sum p_k = 1$ *and* $\mu = \sum kp_k < \infty$*. Let* F_r *be the residual lifetime of the component in use at time* r*, so*

$$F_r \quad = (\text{ time of next renewal after } r\ ) - r\,.$$

Then

$$\mathbb{P}\left[F_r = k\right] \quad \to \quad (\mathbb{P}\left[\text{ lifetime } \geq k\right])/\mu \qquad \text{as } r \text{ tends to } \infty\,.$$

To see this we will use the fact

$$u_j \quad = \quad \mathbb{P}\left[\text{ renewal at time } j\ \right] \quad = \quad \sum_{n=1}^{\infty}\mathbb{P}\left[S_{n-1} = j\right]$$

(setting $S_0 = 0$). First consider the LHS

$$\begin{aligned}
\mathbb{P}\left[F_r = k\right] &= \sum_{n=1}^{\infty} \mathbb{P}\left[S_n = k + r, S_{n-1} \leq r\right] \\
&= \sum_{n=1}^{\infty} \sum_{j=0}^{r} \mathbb{P}\left[S_{n-1} = j, T_n = k + r - j\right] \\
&= \sum_{n=1}^{\infty} \sum_{j=0}^{r} \mathbb{P}\left[S_{n-1} = j\right] \mathbb{P}\left[T_n = k + r - j\right] \\
&= \sum_{j=0}^{r} \sum_{n=1}^{\infty} \mathbb{P}\left[S_{n-1} = j\right] p_{k+r-j} \\
&= \sum_{j=0}^{r} u_j p_{k+r-j} = \sum_{v=0}^{r} u_{r-v} p_{k+v}
\end{aligned}$$

and the last sum converges to the RHS as a consequence of $u_{r-v} \rightarrow \lambda$, $\sum_k p_k = 1$, and a theorem from the theory of infinite sums (equivalently, the dominated convergence theorem given as Corollary 5.17 in ST213).

The significant point here is that the (equilibrium) residual lifetime is *not* like a typical lifetime! In fact the limiting distribution is that of the initial lifetime for a *stationary* renewal process.

Example 5.13 *Suppose X is a Markov chain, irreducible, aperiodic, recurrent, begun at state 1. Suppose also that the mean time to return is finite ($\mathbb{E}[T_{ii}] < \infty$). Consider the renewal process of returns to the initial state 1.*
Renewal probabilities are $u_n = p_{11}^{(n)}$ and indeed the first-passage decomposition is exactly the renewal equation.
The resulting renewal process is nondefective, aperiodic, and $\mu < \infty$ by positive recurrence. So (Theorem 5.10)

$$u_n \quad = \quad p_{11}^{(n)} \quad \to \quad 1/\mu = 1/\,\mathbb{E}[T_{11}]\ .$$

A similar result involving delayed renewal processes applies if the process is started somewhere other than state 1.
So the long-term behaviour of Markov chains follows from renewal theory: more on this in the next section.

Example 5.14 *Given a renewal process, suppose that L_r is the total life of the item in use at time r. Then*

$$\begin{aligned}
\mathbb{P}\left[L_r = k\right] &= \sum_{n=1}^{\infty}\sum_{j=0}^{k-1} \mathbb{P}\left[S_{n-1} = r-j, S_n = k+r-j\right] \\
&= \sum_{j=0}^{k-1} u_{r-j} p_k \qquad \text{if } r > k \\
&\to (k p_k)/\mu
\end{aligned}$$

Here the last sum $\sum_{j=0}^{k-1} u_{r-j}p_k$ converges to $(p_k + p_k + \ldots + p_k)\lim_{r\to\infty} u_r$ (summing k terms), using the Renewal theorem 5.10 to deduce

$$\lim_{r\to\infty} u_r = 1/\mu\,.$$

In particular we find the mean total lifetime of the item in use at a particular (large) time r is

$$\left(\sum_{j=0}^{\infty} k^2 p_k\right)/\mu = \mathbb{E}\left[T^2\right]/\mu \geq \mu$$

WARWICK
Home Page
Title Page
Contents
◀◀ ▶▶
◀ ▶
Page 189 of 237
Go Back
Full Screen
Close
Quit

with equality only when the typical time between renewals is non-random*!*

This is an example of *length-biased sampling*: the item in use at a typical time is more likely to be long-lived, as short-lived items are less likely to be observed!
Here is another case where differing renewal theoretic assumptions make a real difference to what is the best decision.

Example 5.15 *An eminent statistician, visiting the Department of Mathematics of the National University of Singapore, leaves his office one morning in search of a lavatory.*[1] *He turns to the* right *and walks along the corridor for a while till he finds the object of his search. However a notice says it is closed for renovations. What to do? He has to choose between:*

Strategy A continuing in the same direction;

Strategy B reversing direction.

[1]Being American, he thinks of this as going in search of a rest-room . . .

The best decision depends on whether the statistician believes

Case A lavatories are scattered along the corridors independently at random, with probability p of a given room being a lavatory;

Case B lavatories are dispersed at regular intervals, separated from each other at intervals of ℓ rooms.

In Case A simple computations show that if he continues to the right then he will encounter X_r rooms which are not lavatories before finding the next lavatory, where

$$\mathbb{P}\left[X_r = k\right] \quad = \quad (1-p)^r p\,.$$

(So X_r is *Geometrically distributed* with success probability p.) On the other hand if he reverses direction then he must pass x rooms before he returns to his office, then encounter X_l further rooms which are not lavatories before finding the next lavatory, where X_l is distributed as X_r. Hence Strategy A means he has to travel X_r rooms to the nearest lavatory, and total distance travelled is $2(x + X_r)$. Strategy B means he has to travel $x+X_l$ rooms to the nearest lavatory, and total distance travelled is $2(x + X_l)$. So, on the grounds of getting quickly to the lavatory, the best plan in Case A is to adopt Strategy A.

But in Case B even simpler computations show, if he adopts Strategy A then he has to travel ℓ rooms to the nearest lavatory, and total distance travelled is $2(x + \ell)$. Strategy B means he has to travel $x + (\ell - x) = \ell$ rooms to the nearest lavatory, and total distance travelled is 2ℓ. So, on the grounds of getting back to the office as soon as possible, the best plan in Case B is to adopt Strategy B.

Clearly in general the correct strategy depends on

1. the statistician's prior probabilities for which of Cases A or B might hold;
2. the statistician's comparative utilities for quickly finding a lavatory and quickly returning to his office.

Crucially important generalizations of renewal theory apply to observations of biological sections under a microscope, where it really matters that we do not overestimate (*eg*) the typical diameter of a population of cells. This constitutes the subject of *stereology*.

Other examples:

- intervals between bus arrivals;
- lengths between flaws in cotton yarn;
- statistical analysis of times-to-failure;
- Geiger counters with *dead-time* periods.

Chapter 6

Limiting behaviour

Contents

So far in this module we have described how to classify a Markov chain state-space, in particular learning how to divide it into transient and recurrent parts. We have seen how questions such as "where will the chain exit from the transient region?" and "how long will it stay in the transient region?" can be answered using the fundamental matrix $\underline{\underline{G}}$. These are the basic questions relating to the transient part of a chain. We must now complete the picture by seeing how to treat the recurrent part. Here the main question is "in which part of a recurrent class will a chain be after a long period of time?", and it is answered by finding expressions which approximate the n-step transition probabilities after a long time n.

From an applied point of view, in applied probability, economically important questions often concern long-run behaviour of Markov chains (example: for a queue, what is expected waiting time to service once queue has attained statistical equilibrium?). Further, recent ideas of *Markov chain Monte Carlo* in statistics use the equilibrium behaviour of Markov chains to provide an effective and versatile way of solving realistic statistical problems. In this final section we investigate such questions.

6.1. Limits for n-step transition probabilities

Start with the first-passage decomposition described in Section 2.4.2:

$$p_{ii}^{(n)} \quad = \quad \sum_{r=1}^{n} f_{ii}^{(r)} p_{ii}^{(n-r)}$$

This is the same as the renewal equation $u_n = \sum_{r=1}^{n} p_r u_{n-r}$; in fact the Markov property allows us to apply renewal theory as described in the previous section (see an example there for further description of this connection!). Immediately we find (use Theorem 5.5, Corollary 5.11, argument in Exercise 5.12; alternatively the Corollary 2.39):

- i transient:

$$p_{ii}^{(n)} \quad \to \quad 0 \text{ as } n \to \infty$$

- i recurrent, period d: then $f_{ii}^{(*)} = \sum_{r=1}^{\infty} f_{ii}^{(r)} = 1$ and we find

$$p_{ii}^{(nd)} \quad \to \quad \begin{cases} \frac{d}{\mathbb{E}[T_{ii}]} & \text{if } \mathbb{E}\left[T_{ii}\right] < \infty\,, \\ 0 & \text{if } \mathbb{E}\left[T_{ii}\right] = \infty\,. \end{cases}$$

Definition 6.1 *Suppose i is recurrent. Then*

- i *is* positive-recurrent *if* $\mathbb{E}[T_{ii}] < \infty$,
- i *is* null-recurrent *if* $\mathbb{E}[T_{ii}] = \infty$.

Now consider the behaviour of $p_{ij}^{(n)}$ when $i \neq j$. Using delayed renewal process theory ($v_n = \sum_{r=0}^{n} a_{n-r} u_r$) we find

- j transient:

$$p_{ij}^{(n)} \quad \to \quad 0 \text{ as } n \to \infty\,;$$

- i does not lead to j:

$$p_{ij}^{(n)} \quad = \quad 0 \text{ anyway.}$$

Consider the case when we have $i \to j$; j recurrent, period 1:

$$p_{ij}^{(n)} = \sum_{r=1}^{n} f_{ij}^{(r)} p_{jj}^{(n-r)} \quad \to \quad \sum_{r=1}^{\infty} f_{ij}^{(r)} \lim_{n\to\infty} p_{jj}^{(n-r)}$$

now use the theory of infinite sums, alternatively the dominated convergence theorem given as Corollary 5.17 in ST213

$$= \begin{cases} \sum_{r=1}^{\infty} \frac{f_{ij}^{(r)}}{\mathbb{E}[T_{jj}]} & \text{if } \mathbb{E}\left[T_{jj}\right] < \infty\,, \\ \sum_{r=1^{\infty}} f_{ij}^{(r)} \times 0 & \text{if } \mathbb{E}\left[T_{jj}\right] = \infty\,. \end{cases}$$

Hence we have

- $i \to j$; j recurrent, period 1:

$$p_{ij}^{(n)} \quad \to \quad \begin{cases} \frac{f_{ij}^{(*)}}{\mathbb{E}[T_{jj}]} & \text{if } \mathbb{E}\left[T_{jj}\right] < \infty\,, \\ 0 & \text{if } \mathbb{E}\left[T_{jj}\right] = \infty\,. \end{cases}$$

- $i \to j$; j recurrent, period d: obvious but more complicated modification of above statement!

For an example of the modifications required, consider the case of simple random walk on $\{\ldots, -1, 0, 1\}$ with a reflecting barrier at 1, so that $p_{10} = 1$, and $p_{i,i+1} = p > 1/2$ if $i \leq 0$, so that the walk is positive-recurrent. This has period 2. Consequently p_{00}^{2n+1} is always zero. On the other hand, if we study the random walk at even times only, and if it begins at 0, then it behaves as an *aperiodic* positive-recurrent Markov chain with state-space $\{\ldots, -4, -2, 0\}$, and we may deduce

$$p_{00}^{2n} \quad \to \quad \frac{1}{\mathbb{E}\left[T_{00}/2\right]} \quad = \quad \frac{2}{\mathbb{E}\left[T_{00}\right]}\,.$$

Notice also, we may compute $\mathbb{E}\left[T_{00}\right] = 2p + (1-p)(1 + \mathbb{E}\left[T_{-1,0}\right]$ (use first-step decomposition). The generating function $F(s)$ for $T_{-1,0}$ has been computed in Theorem 4.17, and $\mathbb{E}\left[T_{-1,0}\right] = F'(1) = 1/(2p-1)$, so we can deduce the value of $\mathbb{E}\left[T_{00}\right] = 2p + (1-p)/(2p-1)$ and hence the limiting value $(2p-1)/(1-3p+4p^2)$ of p_{00}^{2n}. Notice we need $p > 1/2$ if these values are to be positive and finite!

Theorem 6.2 *Positive/null-recurrence are class properties.*

Proof: We already know recurrence is a class property. So it is enough to show, if $i \leftrightarrow j$ and j is null-recurrent then i must be null-recurrent. So suppose for some r, s we have $p_{ji}^{(r)}$, $p_{ij}^{(s)}$ are both positive. Then we know $p_{jj}^{(r+n+s)} \geq p_{ji}^{(r)} p_{ii}^{(n)} p_{ij}^{(s)}$ and the left-hand side converges to zero. (Null-recurrence of j, use above results.) This means the right-hand side must converge to zero also; since $p_{ji}^{(r)}$, $p_{ij}^{(s)}$ are both positive we find $p_{ii}^{(n)} \to 0$ and consequently i is null-recurrent as required (use above results). $\square$

Theorem 6.3 *A finite essential class C is positive-recurrent.*

Proof: Suppose that $C = \{1, \ldots, m\}$. Then we have for each fixed i in C

$$\sum_{j=1}^{m} p_{ij}^{(n)} \quad = \quad 1\,.$$

If C were null-recurrent then $p_{ij}^{(n)} \to 0$ for i, j in C by the above results, and we could argue

$$\lim_{n\to\infty} \sum_{j=1}^{m} p_{ij}^{(n)} \quad = \quad \sum_{j=1}^{m} \lim_{n\to\infty} p_{ij}^{(n)} = 0$$

which would be a contradiction! So C must be positive-recurrent. □

Notice it is vital we have a *finite* class C here, since otherwise the exchange of limit and infinite sum

$$\lim_{n\to\infty} \sum_{j=1}^{m} \ldots \quad = \quad \sum_{j=1}^{m} \lim_{n\to\infty} \ldots$$

is not necessarily true!

Example 6.4 *Consider a symmetric simple random walk X begun at 0. We know from Theorem 4.14 and the proof of Theorem 4.19 that*

$$p_{00}^{(2n)} \quad = \quad \binom{2n}{n} 2^{-2n} \quad \approx \quad \sqrt{2\pi} n^{-\frac{1}{2}}$$

(using Stirling's formula as in Theorem 4.19), and this converges to zero as n tends to ∞. So here we have an infinite essential class (recall the random walk is irreducible!) which is recurrent (Theorem 4.19) but not positive-recurrent!

6.2. Stationary distributions

We now know how transition probabilities behave in the limit for a Markov chain. But in the limit do they tie together to form a distribution? We *can't* assume this is so, exactly because of the point noted above:

$$\lim_{n\to\infty} \sum_j p_{ij}^{(n)}$$

and

$$\sum_j \lim_{n\to\infty} p_{ij}^{(n)}$$

aren't necessarily equal!

Definition 6.5 *A* stationary *or* invariant *or* equilibrium *distribution for the stochastic matrix* $\underline{\underline{P}}$ *is a row vector* $\underline{u}$ *such that*

$$\underline{u} = \underline{u}\,\underline{\underline{P}}$$

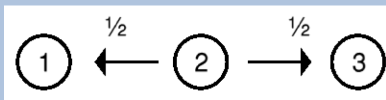
with non-negative entries adding up to 1.

Remark 6.6 *The following simple examples show the limits of what we might be able to say:*

- *If* $\mathbb{P}\left[X_0 = i\right] = u_i$ *for all* i *and if* $\underline{u}$ *is an invariant distribution then*

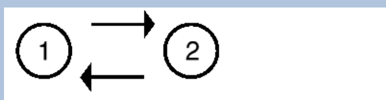
$$\mathbb{P}\left[X_n = i\right] = u_i \text{ for all } i, \text{ all } n.$$

- *There may be more than one stationary distribution! Consider the simple* 3*-state Markov chain:*



and notice there are at least two invariant distributions: $[1, 0, 0]$ *and* $[0, 0, 1]$.

- *Even if the stationary distribution is unique, when the chain is periodic it may not always be the limiting distribution. Consider the simple* 2*-state Markov chain:*



and notice that successive distributions for X_n *could be* $[1, 0]$, $[0, 1]$, $[1, 0]$, $\ldots$.

- *Finally consider the* 3*-state Markov chain in which transitions* $1 \to 2$, $2 \to 3$, $3 \to 1$ *all have probability* $1/3$ *while tran-*

sitions $3 \to 2$, $2 \to 1$, $1 \to 3$ *all have probability* $2/3$. *By symmetry the distribution* $[1/3, 1/3, 1/3]$ *is invariant, and it turns out to be the only one.*

We are interested in cases where there is a unique stationary distribution which is also the limiting distribution. The main result for the important finite-state-space case is:

Theorem 6.7 Equilibrium Theorem. *Suppose that* X *is a finite-state-space irreducible aperiodic Markov chain with stochastic matrix* $\underline{\underline{P}}$. *Then* $\mathbb{P}\left[X_n = i\right] \to u_i$, whatever the starting distribution of X_0, *where* $\underline{u} = (u_i)$ *is the unique row vector such that*

$$\underline{u} \quad = \quad \underline{u}\,\underline{\underline{P}}\,,$$

the entries in $\underline{u}$ *are non-negative, and* $\sum_i u_i = 1$.

Proof: By Theorem 6.3 the state-space must be positive recurrent, so our Markov chain is ergodic. Applying renewal theory as in Section 6.1 we see

$$p_{ij}^{(n)} \quad \to \quad \frac{1}{\mathbb{E}\left[T_{jj}\right]}$$

as n tends to infinity.
So set $u_i = 1/\mathbb{E}[T_{ii}]$ and take limits in $\underline{\underline{P}}^{(n)}\underline{\underline{P}} = \underline{\underline{P}}^{(n+1)}$. This is permitted because this is really a finite sum:

$$\sum_k p_{ik}^{(n)} p_{kj} \quad = \quad p_{ij}^{(n+1)} .$$

We get

$$\sum_k u_k p_{kj} = u_j .$$

Similarly we can take limits in $\sum_k p_{ik}^{(n)} = 1$ to see $\sum_k u_k = 1$. (Again it is crucial that the sum is finite.)
So $\underline{u}$ is indeed stationary. If $\underline{v}$ is another stationary distribution then just take limits:

$$\begin{aligned} \underline{v}\,\underline{\underline{P}}^n &= \underline{v} \\ v_j &= \left(\sum_k v_k \, \mathbb{P}[X_n = j | X_0 = k]\right) \quad \to \quad \left(\sum_k v_k u_j\right) = u_j \end{aligned}$$

so $\underline{v} = \underline{u}$.

Finally notice that the starting position i is irrelevant to the limiting value, as required. $\square$

Remark 6.8 *Irreducible aperiodic positive-recurrent Markov chains are called* ergodic.

Remark 6.9 *At every step we use finiteness of the state-space in order to exchange the limiting operation with the (finite) sum.*

Example 6.10 *Consider the Markov chain given by the stochastic matrix $\underline{\underline{P}}$ as follows:*

$$\underline{\underline{P}} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & \frac{1}{4} & 0 & \frac{3}{4} & 0 \\ \frac{1}{3} & 0 & \frac{1}{3} & 0 & \frac{1}{3} \\ 0 & \frac{1}{2} & 0 & \frac{1}{2} & 0 \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & 0 & 0 \end{bmatrix}$$

We find there is one transient class $\{3, 5\}$ and two positive recurrent aperiodic classes $\{1\}$, $\{2, 4\}$. Hence we can use Section 6.1 *to find* *the limiting values of much of $\underline{\underline{P}}^n$:*

$$\underline{\underline{P}}^n \quad \rightarrow \quad \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & ? & 0 & ? & 0 \\ ? & ? & 0 & ? & 0 \\ 0 & ? & 0 & ? & 0 \\ ? & ? & 0 & ? & 0 \end{bmatrix}$$

On $\{2,4\}$ we can use Theorem 6.7: solve

$$[u_2, u_4]\,\underline{\underline{Q}} \quad = \quad [u_2, u_4]$$

and $u_2 + u_4 = 1$ where $\underline{\underline{Q}}$ is the relevant substochastic matrix for $\{2,4\}$ (actually stochastic since the class is essential!):

$$\underline{\underline{Q}} \quad = \quad \begin{bmatrix} \frac{1}{4} & \frac{3}{4} \\ \frac{1}{2} & \frac{1}{2} \end{bmatrix}.$$

We deduce

$$\begin{aligned} \frac{1}{4}u_2 + \frac{1}{2}u_4 &= u_2 \\ \frac{3}{4}u_2 + \frac{1}{2}u_4 &= u_4 \end{aligned}$$

WARWICK
Home Page
Title Page
Contents
◀◀ ▶▶
◀ ▶
Page 208 of 237
Go Back
Full Screen
Close
Quit

and hence

$$\begin{aligned} u_2 + 2u_4 &= 4u_2 \\ 3u_2 + 2u_4 &= 4u_4 \end{aligned}$$

which yields (twice over!)

$$2u_4 = 3u_2 \,.$$

How to get the actual values of u_2, u_4*? We* always *have to use the fact the the stationary distribution must add up to one: using* $u_2 + u_4 = 1$ *we immediately find* $u_2 = 2/5$ *and* $u_4 = 3/5$.

Starting from $\{3, 5\}$ *it suffices to know* $f_{3i}^{(*)}$ *and* $f_{5i}^{(*)}$*. (See discussion in Section 6.1 again!). Either by working with the fundamental matrix as in Section 3, or by solving the first-step decomposition equations as in Section 2.4.1, we find* $f_{51}^{(*)} = 3/5$, $f_{52}^{(*)} = f_{54}^{(*)} = 2/5$, $f_{31}^{(*)} = 4/5$, $f_{32}^{(*)} = f_{34}^{(*)} = 1/5$*, and so ultimately the full limiting* *form of* $\underline{\underline{P}}^n$ *:*

$$\underline{\underline{P}}^n \quad \to \quad \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & \frac{2}{5} & 0 & \frac{3}{5} & 0 \\ \frac{4}{5} & \frac{2}{25} & 0 & \frac{3}{25} & 0 \\ 0 & \frac{2}{5} & 0 & \frac{3}{5} & 0 \\ \frac{3}{5} & \frac{4}{25} & 0 & \frac{6}{25} & 0 \end{bmatrix}$$

Example 6.11 *Here we solve a particular case of the quality control example 2.18. Suppose the strategy is to test every item until* $i = 4$ *consecutive non-defective items are found. Then test each item at random with probability* $1/r$ *until a defective is found. Then test every item until . . . . We can compute the long-run limiting probability that the system is in* E_**, the non-faulty mode.*
First compute the transition probability matrix:

$$\underline{\underline{P}} \quad = \quad \begin{bmatrix} p & q & 0 & 0 & 0 \\ p & 0 & q & 0 & 0 \\ p & 0 & 0 & q & 0 \\ p & 0 & 0 & 0 & q \\ p/r & 0 & 0 & 0 & 1 - p/r \end{bmatrix} .$$

Now solve $\underline{\pi}\,\underline{\underline{P}} = \underline{\pi}$:

$$\begin{aligned}
p(\pi_0 + \pi_1 + \pi_2 + \pi_3) + \frac{p}{r}\pi_* &= \pi_0 \\
q\pi_0 &= \pi_1 \\
q\pi_1 &= \pi_2 \\
q\pi_2 &= \pi_3 \\
q\pi_3 + (1 - \frac{p}{r})\pi_* &= \pi_*
\end{aligned}$$

leads to

$$\begin{aligned}
\pi_1 &= q\pi_0 \\
\pi_2 &= q^2\pi_0 \\
\pi_3 &= q^3\pi_0 \\
\pi_* &= \frac{rq^4}{p}\pi_0
\end{aligned}$$

We need one more fact to get a solution. Use

$$\pi_0 + \pi_1 + \pi_2 + \pi_3 + \pi_* = 1$$

and simplify (using $p = 1 - q$ *and the finite geometric series* $1 + q + q^2 + q^3 = (1 - q^4)/(1 - q)$*) to deduce*

$$\pi_* \quad = \quad \frac{rq^4}{1 - (1 - r)q^4} \,.$$

Example 6.12 ***Application to the Gibbs sampler*** *The above theory has a surprising and important application to Bayesian statistics. Here is a very simplified example.*
Suppose we need to know $p(x, y) = \mathbb{P}\left[X = x, Y = y\right]$ *for two discrete random variables* X, Y *taking values in* $\{1, \ldots, n\}$*, and such that* $p(x, y) > 0$ *for all possible pairs* (x, y)*. Typically we find it much easier to state the conditional probabilities*

- $a(x, y) = \mathbb{P}\left[X = x | Y = y\right]$,
- $b(y, x) = \mathbb{P}\left[Y = y | X = x\right]$

(they often arise naturally from a Bayesian analysis if X, Y *are parameters conditioned on observation of data).*

For many purposes it is enough if we can simulate approximately from the joint distribution of X, Y: for example we can then calculate sample means to approximate theoretical means, et cetera.
So we construct an ergodic Markov chain Z on $\{1, \ldots, n\} \times \{1, \ldots, n\}$ as follows: given a first try $Z_0 = (X_0, Y_0) = (x_0, y_0)$ we pick $Y_1 = y_1$ according to the conditional distribution $b(\cdot, x_0)$ and then pick $X_1 = x_1$ according to the conditional distribution $a(\cdot, y_1)$. Hence we get $Z_1 = (X_1, Y_1) = (x_1, y_1)$. We continue this procedure (it is extremely easy to program on the computer!).
It is easy to check that the conditions of Theorem 6.7 apply (aperiodic, irreducible, finite state-space) and so there is a unique invariant distribution $u(x, y)$ and

$$\mathbb{P}\left[X_n = x, Y_n = y\right] \quad \rightarrow \quad u(x, y) \,.$$

But also it is easy to compute that the joint probability mass function $p(x, y)$ is another invariant distribution (simple conditional probability!). It follows that

$$\mathbb{P}\left[X_n = x, Y_n = y\right] \quad \rightarrow \quad p(x, y) \,.$$

So for sufficiently large n we know $\mathbb{P}[X_n = x, Y_n = y]$ approximates the probability mass function we want.

Remark 6.13 *How large should n be? For a number of years this was a difficult issue. Recently two US computer scientists, Propp and Wilson, have produced a remarkably simple coupling argument to show how one can often modify the simulation so that*

- *it runs for a random length of time*
- *it produces a result (X, Y) which is exactly of the correct distribution!*

What about infinite state-space? Here are two results without proof which lay out the basic position.

Theorem 6.14 *Suppose that C is an aperiodic essential class, with associated stochastic matrix $\underline{\underline{P}}$. We do* not *suppose C to be finite. The set of solution vectors*

$$\underline{u} = \underline{u}\,\underline{\underline{P}}$$

such that $\sum_k |u_k| < \infty$ *is given by* $\{a\underline{\rho} : a$ *is a real number* $\}$ *where*

$$\rho_i = \begin{cases} \frac{1}{\mathbb{E}[T_{ii}]} & \text{if } i \text{ is positive recurrent;} \\ 0 & \text{if } i \text{ is null-recurrent or transient.} \end{cases}$$

Corollary 6.15 *If* C *is an aperiodic positive-recurrent class, with associated stochastic matrix* $\underline{\underline{P}}$*, then*

$$\underline{u} = \underline{u}\,\underline{\underline{P}},$$

$$\sum_k u_k = 1,$$

has a unique solution

$$u_i = \frac{1}{\mathbb{E}\left[T_{ii}\right]}$$

for all i *and*

$$p_{ij}^{(n)} \rightarrow u_j$$

as n *tends to infinity.*

This is a stronger version of Theorem 6.7, as it applies to the important case of infinite state-space and tells us then that aperiodicity and positive-recurrence suffice!

References

This is a rich hypertext bibliography. In the PDF version of this document, publishers are linked to their homepages, and links are given to the online Warwick University library catalogue `WEBCAT`. Remember to check that `WEBCAT` is searching all collections!

Cox, D. R. and E. J. Snell (1981). *Applied Statistics: Principles and Examples*. London: Chapman and Hall, [WEBCAT].

Grimmett, G. R. and D. R. Stirzaker (1982). *Probability and Random Processes*. The Clarendon Press Oxford University Press, [WEBCAT].

Häggström, O. (2002). *Finite Markov chains and algorithmic applications*, Volume 52 of *London Mathematical Society Student Texts*. Cambridge: Cambridge University Press, [WEBCAT].

Jones, P. W. and P. Smith (2001). *Stochastic Processes*. London: Arnold, [WEBCAT].

Malkiel, B. G. (1973). *A random walk down Wall Street*. New York: Norton, [WEBCAT].

Norris, J. (1997). *Markov Chains*. Cambridge University Press, [WEBCAT].

Ross, S. M. (1997). *Introduction to Probability Models*. Academic Press, [WEBCAT].

Appendix A

First year probability

These notes are concluded by appendices providing supplementary material. For your convenience, we begin with an appendix giving a (very) rapid summary of some of the material covered in first-year probability. You are expected to be well-versed in *all* of the material covered in the first year, not just the material below!

A.1. Reminder about probability

In this section we review some topics from the first-year courses Probability A and B (ST111/ST112) which we will need throughout this term.

Here are the basic elements of probability. See the first year probability course ST111 for more details. Check the first section of the online quizzes if you need to confirm your understanding: you should be able to do all the questions correctly with the help of paper and pencil.

i. The *Sample Space* Ω; the set of all possible outcomes. Example: set of all possible sequences of heads and tails arising from ten independent coin tosses.

ii. An *Event* $A \subseteq \Omega$; subset of sample space. Example: set of outcomes for which the cumulative number of heads is never strictly less than the cumulative number of tails.

Technicality: actually A must be a *measurable subset*. At this level all you need to know is

iii. The *Probability* $\mathbb{P}[A]$ *of an event* A. It is always the case that

$$0 \quad \leq \quad \mathbb{P}[A] \quad \leq \quad 1\,. \tag{A.1}$$

iv. The *Additive Law of probabilities*:

$$\mathbb{P}[A \cup B] \quad = \quad \mathbb{P}[A] + \mathbb{P}[B] - \mathbb{P}[A \cap B]\,. \tag{A.2}$$

v. The σ*-Additive Law of probabilities*. Indeed, if A_1, $A_2 \ldots$ form a disjoint countable sequence of events then

$$\mathbb{P}\left[\bigcup_i A_i\right] \quad = \quad \sum_i \mathbb{P}[A_i]\,. \tag{A.3}$$

(a) to be measurable, A must be "constructible" from very elementary outcomes, in some technical sense;

(b) this technical sense is so very general that if you can write down how to tell whether or not an outcome is in A then it is almost inevitable that A is measurable! The second-term course ST213 *Mathematics of Random Events* takes these issues further.

vi. Moreover if A_1, A_2 ... are as in (v) above and also $\bigcup_i A_i = \Omega$ (the sequence is a *partition* of the sample space) then for any event B we have

$$\mathbb{P}\left[B\right] \quad = \quad \sum_i \mathbb{P}\left[B \cap A_i\right] . \qquad \text{(A.4)}$$

vii. *Conditional Probability*: suppose A, B are events and $\mathbb{P}\left[B\right] > 0$. Then we define the conditional probability of the event A conditional on the event B by

$$\mathbb{P}\left[A|B\right] \quad = \quad \frac{\mathbb{P}\left[A \cap B\right]}{\mathbb{P}\left[B\right]} . \qquad \text{(A.5)}$$

If $\mathbb{P}\left[B\right] = 0$ then $\mathbb{P}\left[A|B\right]$ is undefined.

viii. We say events A and B are *independent* if $\mathbb{P}\left[A|B\right] = \mathbb{P}\left[A\right]$. More generally (and more symmetrically) we say events A and B are independent if $\mathbb{P}\left[A \cap B\right] = \mathbb{P}\left[A\right] \times \mathbb{P}\left[B\right]$.

ix. The events A_1, A_2 ... form an *independent sequence of events* if all possible "multiplication rules" hold good: that is to say, if for any finite subsequence $A_{i_1}, \ldots, A_{i_n}$ we have

$$\mathbb{P}\left[\prod_{k=1}^{n} A_{i_k}\right] = \mathbb{P}\left[A_{i_1}\right] \times \ldots \times \mathbb{P}\left[A_{i_n}\right] . \qquad \text{(A.6)}$$

A.2. Reminder about random variables

To speak fluently about applied probability we also need the idea of a random variable. See the first year probability course ST111 for more details. Check the second section of the online quizzes if you need to confirm your understanding: you should be able to do all the questions correctly with the help of paper and pencil.

i. A *random variable* is simply a function $X : \Omega \to \mathbb{R}$ on sample space, taking an outcome for argument, yielding (usually) a real number. Example: the total numbers of heads scored after ten coin tosses. Or the first coin toss after which the cumulative number of heads is strictly less than the cumulative number of tails.

Technicality: actually it must be a *measurable function*. At this level all you need to know is

(a) to be measurable, a function must be "constructible" from very elementary functions, in some technical sense;

(b) this technical sense is so very general that if you can *write down* explicitly how to tell whether or not a random variable exceeds any specified level

ii. The *distribution function* of a random variable X:

$$F(x) \quad = \quad F_X(x) \quad = \quad \mathbb{P}\left[X \leq x\right] . \qquad \text{(A.7)}$$

The distribution function defines the statistical behaviour of the random variable X considered in isolation – what we call its *distribution*.

iii. There is an important distinction between *discrete* and *(absolutely) continuous random variables*. A discrete random variable takes values in a countable range (often the integers). Its distribution can conveniently be considered via its probability mass function

$$p(x) \quad = \quad p_X(x) \quad = \quad \mathbb{P}\left[X = x\right] . \qquad \text{(A.8)}$$

A random variable which is (absolutely) continuous ranges over a continuum and furthermore has a probability density

---

then it is almost inevitable that it is a measurable function! The second-term course ST213 *Mathematics of Random Events* takes these issues further.

function: there is a non-negative function f_X which integrates to 1 such that

$$\mathbb{P}\left[X \leq x\right] \quad = \quad F_X(x) \quad = \quad \int_{-\infty}^{x} f_X(u)\,\mathrm{d}\,u \qquad \text{(A.9)}$$

for all x.

iv. The *expectation of a random variable* X:

$$\mathbb{E}\left[X\right] \quad = \quad \sum_x x p_X(x) \qquad \text{(A.10)}$$

for the discrete case,

$$\mathbb{E}\left[X\right] \quad = \quad \int_{-\infty}^{\infty} x f_X(x)\,\mathrm{d}\,x \qquad \text{(A.11)}$$

for the absolutely continuous case.

This distinction is important for technical reasons, but at a more sophisticated level of the theory (as described in ST213) it becomes evident that the theories and results for discrete and (absolutely) continuous random variables are essentially identical — given systematic exchanges of integrals and sums. Note that some random variables are in some sense mixed between discrete and (absolutely) continuous, while others (arising for example in the theory of fractals) are neither!

v. Random variables X and Y are *independent* if

$$\mathbb{P}\left[X \leq x, Y \leq y\right] \quad = \quad \mathbb{P}\left[X \leq x\right] \times \mathbb{P}\left[Y \leq y\right] \qquad (A.12)$$

for all x and y.

Appendix B

Periodicity proof

In lectures we will use the following result:

Suppose that $i \to i$ *and* i *has period* d*. Then there is* $N > 0$ *such that for all* $k \geq N$ *we have* $p_{ii}^{(kd)} > 0$.

In words, if a state i has period d then the chain has positive chance of returning to i at time kd for *all* sufficiently large integers k.

The proof uses Euclid's algorithm in an elementary way.

Proof: *(this proof is not examinable!)* Without loss of generality we may suppose $d = 1$ (otherwise simply look only at the times $d, 2d, 3d, \ldots$).

Suppose possible return times are $t_1, t_2, \ldots$. If

$$\gcd\{t_1, t_2, \ldots\} \quad = \quad 1$$

then we have to show there is M, such that for all $m \geq M$ we may express m as a sum $\lambda_1 t_1 + \ldots \lambda_k t_k$ for some k (perhaps depending on m and *non-negative* $\lambda_1, \ldots \lambda_k$.

Actually, because the gcd is non-increasing as we add more t_k, and is integer-valued, we need consider only

$$\gcd\{t_1, \ldots, t_k\} \quad = \quad 1$$

for some fixed k.

Clearly all numbers $t_1, 2t_1, 3t_1, \ldots$ can be obtained as sums in this way. This deals with all numbers which are reduced to zero modulo t_1.

Now we may use the Euclidean algorithm to express each of the numbers 1, 2, …, $t_1 - 1$, t_1 as a sum $\alpha_1 t_1 + \ldots + \alpha_k t_k$ where the

α_i are (possibly negative) integers (note, the expression for t_1 itself is trivial!). Set

$$M = \max\{|\alpha_1|t_1 + \ldots + |\alpha_k|t_k\}$$

where the $\max$ ranges over the t_1 different expressions for these numbers.
Then each of the numbers $M+1$, $M+2$, $\ldots$, $M+t_1$ can be expressed as sums with non-negative coefficients. But now we may also obtain all subsequent numbers simply by adding multiples of t_1. $\square$

Appendix C

Crash course in generating functions

Here are the bare facts about generating functions. For more details see (for example) the first year Probability B course ST112, or Grimmett and Stirzaker (1982, Chapter 5).

Given a bounded sequence of numbers a_0, a_1, ... we can form the *generating function*

$$G(s) \quad = \quad \sum_{n=0}^{\infty} a_n s^n \, .$$

Here s is a dummy variable, ranging over (for example) $[0,1)$. Think of this as a way of "encoding" the sequence in terms of a formula: using for example the theory of sums and series we can figure out a definite formula for $G(s)$. For example,

- if $a_0 = 1/2$, $a_1 = 1/2$, all other a_i are zero, then we deduce $G(s) = (1+s)/2$;
- if $a_n = 2^{-n-1}$ for all n then (sum the geometric series $\sum 2^{-n-1} s^n$) we deduce $G(s) = 1/(2-s)$.

We most often use generating functions to encode the sequence of probabilities of a non-negative integer-valued random variable taking on its range of values;

$$G_X(s) \quad = \quad \sum_{n=0}^{\infty} s^n \, \mathbb{P}\left[X = n\right] \, . \tag{C.1}$$

However this is not the only case: consider the definition of $P_{ij}(s)$ in section 2.4.2 concerning the first-passage decomposition.
Given the formula for the generating function $G(s)$ we can always recover the sequence $a_0, a_1, \ldots$ by differentiation. For example

$$
\begin{aligned}
a_0 &= G(0) = \left[\sum_n a_n s^n\right]_{s=0} \\
a_1 &= G'(0) \\
a_2 &= \frac{1}{2}G''(0) \\
\ldots \\
a_n &= \frac{1}{n!}G^{(n)}(0) \qquad \text{(C.2)}
\end{aligned}
$$

where G' is the first derivative, and $G^{(n)}$ the n^{th} derivative, of G.
Actually in practice one typically works backwards: given a generating function G one guesses what sequence $a_0, a_1, \ldots$ leads to G by (for example) looking up tables of generating functions.
In case G_X is the generating function for a random variable X, we

can deduce

$$\begin{aligned} 1 &= G(1) = \left[\sum_n a_n s^n\right]_{s=1} \\ \mathbb{E}[X] &= G'(1) \end{aligned}$$

and in fact it is possible to obtain the variance of X indirectly by using the second derivative. Of course if X actually has infinite expectation then we find $G'(1)$ is also infinite ....
In case X and Y are independent non-negative integer-valued random variables, we find

$$G_{X+Y}(s) = G_X(s)G_Y(s)\,. \tag{C.3}$$

This is the real reason why generating functions are useful: at the price of dealing with the indirect concept of a generating function we can work with products of things. If we were to insist on working directly with the probabilities we would have to deal with a *convolution sum*:

$$\mathbb{P}[X+Y=n] = \sum_{r=0}^{n} \mathbb{P}[X=r] \times \mathbb{P}[Y=n-r]\,. \tag{C.4}$$

Another example of this is to be found in the exploitation of the first-passage decomposition using Equation (2.4).

Finally, we should note that a variation on this last arises when we consider the generating function of a random sum of independent identically distributed random variables: if N is the random number of summands, and if X_r is a typical random summand, and if S is the value of the sum, so

$$S = \sum_{r=1}^{N} X_r \,,$$

then a simple argument using conditional expectation shows

$$G_S(s) = G_N(G_{X_1}(s)) \,. \tag{C.5}$$

Appendix D

Genetics calculations

Here are the details of genetics calculations for Theorem 4.3 (the Hardy-Weinberg law).

Recall

$$\begin{aligned} & \mathbb{P}\left[Aa \text{ and parents were } Aa \text{ and } Aa\right] \\ = \; & \mathbb{P}\left[Aa \mid \text{parents were } Aa \text{ and } Aa\right] \times \mathbb{P}\left[\text{ parents were } Aa \text{ and } Aa\right] \\ = \; & \mathbb{P}\left[Aa \mid \text{parents were } Aa \text{ and } Aa\right] \times 4v^2 \\ = \; & (1/2) \times 4v^2 \quad = \quad 2v^2 \,. \end{aligned}$$

We can add further calculations, best laid out in a table:

Parents	Prob:	AA	Aa	aa
AA $\times$ AA	u^2	1	0	0
AA $\times$ Aa	$4uv$	$1/2$	$1/2$	0
AA $\times$ aa	$2uw$	0	1	0
Aa $\times$ Aa	$4v^2$	$1/4$	$1/2$	$1/4$
Aa $\times$ aa	$4vw$	0	$1/2$	$1/2$
aa $\times$ aa	w^2	0	0	1

Totals come out as $u^2+2uv+v^2 = (u+v)^2$ for AA, $2uv+2uw+2v^2+2vw = 2(u+v)(w+v)$ for Aa, $v^2+2vw+w^2 = (v+w)^2$ for aa. This gives the proportions for the first generation of a population obeying the Hardy-Weinberg law. For the second and further generations, consider

$$\begin{aligned}
\tilde{u} &= (u+v)^2 \\
2\tilde{v} &= 2(u+v)(w+v) \\
\tilde{w} &= (v+w)^2
\end{aligned}$$

and simply apply the previous formulae over again! the proportions for generation 2 must be

$$\begin{aligned}
\tilde{\tilde{u}} &= (\tilde{u}+\tilde{v})^2 = ((u+v)^2 + (u+v)(w+v))^2 \\
&= ((u+2v+w)(u+v))^2 = (u+v)^2 \\
2\tilde{\tilde{v}} &= 2(\tilde{u}+\tilde{v})(\tilde{w}+\tilde{v}) \\
&= 2((u+v)^2 + (u+v)(w+v))((v+w)^2 + (u+v)(w+v)) \\
&= 2(u+2v+w)(u+v)(u+2v+w)(w+v) = (u+v)(w+v) \\
\tilde{\tilde{w}} &= (\tilde{v}+\tilde{w})^2 \\
&= ((u+v)(w+v) + (v+w)^2)^2 \\
&= ((u+2v+w)(w+v))^2 = (w+v)^2
\end{aligned}$$

where we factorize out and use $u + 2v + w = 1$ repeatedly. Since we obtain

$$\tilde{\tilde{u}} = \tilde{u}\,, \qquad 2\tilde{\tilde{v}} = 2\tilde{v}\,, \qquad \tilde{\tilde{w}} = \tilde{w}$$

the Hardy-Weinberg law must follow.