ST333 Applied Stochastic Processes: Outline notes Spring Term 2001-2002

Author: Wilfrid S. Kendall,
Statistics, University of Warwick

IMPORTANT NOTICE: These notes refer to the lectures as given in 2001-2002. They are likely to be helpful for subsequent years: however students should bear in mind that the coverage of topics is likely to change.

Therefore: rely on the notes you make while attending lectures, rather than these online notes!

This is `st333.tex` (`Windows 2k`) version 4.3.
Last edited: 14:55:31, 31/05/2005.

1. Introduction

"There was things which he stretched, but mainly he told the truth."
In: *The Adventures of Huckleberry Finn*, Chapter 1, by Mark Twain (S. L. Clemens) 1835–1910.

1.1. The simplest possible example

The first part of the course deals with how one can construct and analyze and apply Markov chains for which time varies continuously. Clearly this can be important in applications where it is unnatural to suppose that time is discretized.

We begin our discussion of continuous-time Markov chains with a simple illustration.

Example 1.1 ***(The "Switcher" Markov chain):*** *Consider the following model for the weather. It is either dry or wet. If dry, then after an exponential time of mean $1/\mu$ it turns wet. If wet, then after an exponential time of mean $1/\nu$ it turns dry. Because of the forgetful property of the exponential distribution this process has a Markov property: given the information of whether it is currently wet or dry, no further information about the past is of any help in predicting the future! We could work out the transition probability that the weather is wet at time t after it is dry (it is a sum of exponentials), and we can see that in some sense the weather changes from dry to wet at rate μ from wet to dry at rate ν. Our calculated probability could be used to determine the long-run equilibrium probability distribution of whether it is wet or dry at large time t.*

All the Markov chains in this course turn out to be rather like this, except that they may have n states not just 2 or (actually more common) a countably-infinite number of states. The example illustrates terms used in continuous-time Markov chain theory (in our case, always with countable state-space).

1.2. Basic Definitions

Definition 1.2 *A random process* $\{X_t : t \geq 0\}$ *has the Markov property if for all times* $s, t \geq 0$ *it is the case that*

$$\mathbb{P}\left[\, X_{t+s} = j | X_s = i,\ \textit{sundry other facts about past at } s \,\right] \quad = \quad \mathbb{P}\left[\, X_{t+s} = j | X_s = i \,\right]$$

does not depend on the "sundry other facts". It has stationary transition probabilities if this conditional probability does not depend on s *either.*

If X has a Markov property then it makes sense to talk about the conditional probability of it moving from i to j over a time period t. Just as in the discrete-time case we can thus introduce transition probabilities.

Definition 1.3 *The transition probability function of a Markov chain (with stationary transition probabilities) is the function*

$$p_t(i,j) \quad = \quad \mathbb{P}\left[\, X_{s+t} = j | X_s = i \,\right] .$$

If X has the Markov property then we call it a *Markov chain*. In fact all the Markov chains in this course have the following additional properties:

Standing Assumption 1 [1] *All Markov chains in this course are continuous time, countable discrete state space, with stationary transition probability functions (unless specified otherwise). Also their (stationary) transition probability functions $p_t(x, y)$ are assumed to be continuous at time $t = 0$:*

$$\lim_{t\to 0} p_t(x, y) \quad = \quad p_0(x, y) \quad = \quad \begin{cases} 1 & \text{if } x = y \\ 0 & \text{otherwise.} \end{cases}$$

It should be clear why the continuity assumption is sensible!

1.3. More Examples

Here are some more examples of the kinds of Markov chains which we will consider.

- *Poisson counting process*. Here N_t measures the number of "incidents" occurring in the time interval $[0, t]$. The process N is random, but moves only from n to $n + 1$ at constant rate μ, independent of the value of n.
- *(Pure) Birth process*. Here X_t measures the number of particles in a population, where each particle gives birth randomly at constant rate, independently of the others.

[1] abbreviated to SA I

- *Birth-death process*. As above, but now the particles also die randomly at a constant (different) rate.
- *$M/M/1$ queue*. Customers arrive randomly at a constant rate (so cumulative number forms a Poisson process!) and are served in turn with exponentially distributed service times.
- *S-I-R epidemic*. An epidemic in a finite population of Susceptibles, moving through Infective stage to Removal stage.

Later in the course we look at processes which are not Markov: for example queues with non-exponential service times, and processes which are actually random patterns of points distributed in space (where the Markov property is not applicable).

1.4. Fundamental theorems

Just as in discrete time, we can now establish a relationship between different values of the transition probability function according to the truism, to get there from here you have go somewhere in between first!

Theorem 1.4 *The transition probability function satisfies the* Chapman-Kolmogorov equations*:*

$$p_{s+t}(i,j) = \sum_k p_s(i,k)p_t(k,j)$$

where the sum ranges over the (countable) state-space.

Proof: This is just as is given in ST202. □

From this we can draw strong conclusions. For example it follows from this (together with SA I) that the transition probabilities $p_t(i,j)$ are uniformly continuous in the time variable t for each fixed i.

But now we come upon a problem. We can't pick a smallest positive time t, so we can't settle on a transition probability matrix which multiplies up to produce all transition probabilities. The obvious answer is to differentiate instead (if we can!):

Definition 1.5 *Define the (instantaneous)* transition rate *from x to y:*

$$q(x,y) \quad = \quad \lim_{t\to 0} \frac{p_t(x,y) - p_0(x,y)}{t}$$

Remark 1.6 *It is convenient to use the $o(t)$ notation here. We say in the above case*

$$p_t(x,y) - p_0(x,y) \quad = \quad q(x,y)t + o(t)$$

where by $o(t)$ we mean some unspecified quantity depending on t and such that

$$\frac{o(t)}{t} \quad \to \quad 0 \qquad \text{as } t \to 0\,.$$

In fact $p_0(x,y)$ is one or zero, depending on whether x, y are equal or not. So we set

$$q(x,y) \quad = \quad \lim_{t\to 0} \frac{p_t(x,y)}{t} \text{ if } x \neq y$$

$$q(x) \quad = \quad \lim_{t \to 0} \frac{1 - p_t(x,x)}{t} = -q(x,x)$$

We arrange all the q functions in the *Q-matrix*:

$$\underline{\underline{Q}} \quad = \quad \{q(x,y)\}_{x,y}$$

Example 1.7 ***(The "Switcher" example 1.1, continued):*** $\underline{\underline{Q}}$ *-matrix:*

$$\begin{bmatrix} -\mu & \mu \\ \nu & -\nu \end{bmatrix}$$

For, arguing informally (we do it rigorously later),

$$p_t(1,1) \approx 1 - \mathbb{P}\,[\ \textit{at least one transition by time } t\ ] = \exp(-\mu t)$$

from which it follows $q(1,1) = -\mu$. *The rest follows similarly.*

But do these derivatives exist in general? In fact we can use SA I and the Chapman-Kolmogorov equations to show *either* q(x) exists *or* it equals ∞. Infinite $q(x)$ corresponds to a Markov chain leaving x at infinite rate and none of ours will do this. So we impose the following general assumption throughout this course:

Standing Assumption 2 [2] *All Markov chains in this course have finite* $q(x)$ *for all* x.

[2] abbreviated to SA II

From this it follows all the derivatives $q(x, y)$ exist and are finite.

We shall not prove any of this (though much of it is first-year Analysis). But notice that in the special case of finite-state-space (*eg:* the "Switcher" example 1.1) SA II is just a consequence of SA I. SA II can only be violated when there is bad behaviour at infinity.

For finite-state-space chains we can differentiate under the summation sign in

$$p_t(x, x) = 1 - \sum_{y:y\neq x} p_t(x, y)$$

to show

$$q(x) = -q(x, x) \quad = \quad \sum_{y:y\neq x} q(x, y)\,.$$

This motivates a further basic assumption.

Standing Assumption 3 [3] *All Markov chains in this course have*

$$q(x) = -q(x, x) \quad = \quad \sum_{y:y\neq x} q(x, y)\,.$$

So the Q-matrix has finite non-negative off-diagonal elements, finite non-positive diagonal elements, and zero row-sums (compare the stochastic matrix in discrete-time Markov chain theory!).

We summarize this as a Theorem.

[3] abbreviated to SA III

Theorem 1.8 ***(Differentiability at 0):***. *If X is a Markov chain as above (satisfying SA I, SA II, SA III) then its transition probability function is differentiable at 0 with derivatives encoded by the $\underline{\underline{Q}}$-matrix*

$$\begin{bmatrix} q_{11} & q_{12} & q_{13} & \cdots \\ q_{21} & q_{22} & q_{23} & \cdots \\ q_{31} & q_{32} & q_{33} & \cdots \\ \cdots & \cdots & \cdots & \cdots \end{bmatrix} = \begin{bmatrix} -q_1 & q_{12} & q_{13} & \cdots \\ q_{21} & -q_2 & q_{23} & \cdots \\ q_{31} & q_{32} & -q_3 & \cdots \\ \cdots & \cdots & \cdots & \cdots \end{bmatrix}$$

If we set $\underline{\underline{P}}(t)$ to be the matrix $\{p_t(x,y)\}_{x,y}$ of transition probabilities then $\underline{\underline{Q}}$ is formally the derivative of $\underline{\underline{P}}(t)$.

Can we use this to get information on all of $\underline{\underline{P}}(t)$?

The answer is *yes*: with a little more work one can justify formally differentiating the Chapman-Kolmogorov equations with respect to s to get the Kolmogorov backward differential equations:

Theorem 1.9 *Under SA I, SA II, SA III, for all time t the $\underline{\underline{P}}(t)$ matrices satisfy the* Kolmogorov backward differential equations*:*

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}t} p_t(i,j) &= \sum_k q(i,k) p_t(k,j) \\ \frac{\mathrm{d}}{\mathrm{d}t} \underline{\underline{P}}(t) &= \underline{\underline{Q}}\ \underline{\underline{P}}(t) \end{aligned}$$

with initial conditions

$$\begin{aligned} p_0(i,j) &= \begin{cases} 1 & \text{if } i = j \\ 0 & \text{otherwise;} \end{cases} \\ \underline{\underline{P}}(0) &= \text{identity matrix.} \end{aligned}$$

Proof: Just differentiate the Chapman-Kolmogorov equations with respect to the s time variable (the "backwards" time variable). Subject to sorting out the formalities[4] of exchanging differentiation and infinite summation, this does the trick. □

Does this set of differential equations have a unique solution? If so then we can specify continuous-time Markov chains via their $\underline{\underline{Q}}$-matrices, just as we did so effectively in the discrete-time case.

Unfortunately uniqueness can fail. We have uniqueness for finite-state-space, and in general there is always a minimal solution, but uniqueness can fail if the chain may "explode to infinity" in finite time.

On the other hand we could differentiate with respect to t, the "forwards" time variable. This yields the following result.

Theorem 1.10 *Subject to additional regularity conditions as well as SA I, SA II, SA III, for all time t the $\underline{\underline{P}}(t)$ matrices satisfy the* Kolmogorov forward differential

[4] This is why this is a *formal* derivation of the Kolmogorov backward differential equations

equations*:*

$$\frac{\mathrm{d}}{\mathrm{d}t}p_t(i,j) = \sum_k p_t(i,k)q(k,j)$$

$$\frac{\mathrm{d}}{\mathrm{d}t}\underline{\underline{P}}(t) = \underline{\underline{P}}(t)\,\underline{\underline{Q}}$$

with initial conditions

$$p_0(i,j) = \begin{cases} 1 & \text{if } i=j \\ 0 & \text{otherwise;} \end{cases}$$

$$\underline{\underline{P}}(0) = \text{identity matrix.}$$

Proof: Just differentiate the Chapman-Kolmogorov equations with respect to the t time variable (the "forwards" time variable). Subject to sorting out the formalities [5] of exchanging differentiation and infinite summation, this does the trick. However to make this work we need additional regularity conditions beyond SA I, SA II, SA III . We do not go into this. □

We will see that we can get more information from the forward differential equations than from the backward differential equations . . . however in some cases they may not hold (but we will not worry about this). Again, if we can solve the forward differential equations then the rates specify the Markov chain uniquely.

In most of the cases we consider, the forward and backward differential equations have unique solutions and so the rates specify the Markov chain uniquely.

[5] Again, this is why this is a *formal* derivation of the Kolmogorov forward differential equations

1.5. Detailed Examples

Exercise 1.11 ***(The "Switcher" example 1.1, continued):*** *To make things even simpler, suppose* $\mu = \nu = 1$. *Observe that the matrix square* $\underline{\underline{Q}}^2$ *is*

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

which equals $-2\underline{\underline{Q}}$. *Hence by induction* $\underline{\underline{Q}}^n = (-2)^{n-1}\underline{\underline{Q}}$ *for* $n \geq 1$. *Now compute the matrix exponential*

$$\exp(t\underline{\underline{Q}}) = \underline{\underline{I}} + \sum_{n\geq 1}(-2)^{n-1}t^n\underline{\underline{Q}}/n! = \underline{\underline{I}} + \frac{1}{2}(1-e^{-2t})\underline{\underline{Q}}$$

as a good guess for the solution of the Kolmogorov forward or backward differential equations (as $(\mathrm{d}/\mathrm{d}t)\exp(t\underline{\underline{Q}})$ *should equal* $\underline{\underline{Q}}\exp(t\underline{\underline{Q}}) = \exp(t\underline{\underline{Q}})\underline{\underline{Q}}$*). Hence we obtain a closed form expression for* $\underline{\underline{P}}(t)$*:*

$$\begin{bmatrix} e^{-t}\cosh t & e^{-t}\sinh t \\ e^{-t}\sinh t & e^{-t}\cosh t \end{bmatrix}$$

and we can read off the long-run equilibrium distribution behaviour: all entries converge to $\frac{1}{2}$.

(I have given a brief summary of the sinh, cosh *notation in Appendix A.4.)*

An alternative approach (and good exercise!) is to write out the Kolmogorov forward differential equations for $p_t(1,1)$, $p_t(1,2)$ and solve them. The next example demonstrates this approach in a different case and gives practice in dealing with the forward differential equations.

Exercise 1.12 *Consider the Markov chain with $\underline{\underline{Q}}$-matrix as given:*

$$\begin{Bmatrix} -\lambda & \lambda/2 & \lambda/2 \\ \lambda/2 & -\lambda & \lambda/2 \\ \lambda/2 & \lambda/2 & -\lambda \end{Bmatrix}$$

Thus it has 3 states, moves from the current state at rate λ, moves to any other state at rate $\lambda/2$. The forward differential equations for $p_t(1,1)$, $p_t(1,2)$, $p_t(1,3)$ are given below:

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}t}p_t(1,1) &= -\lambda p_t(1,1) + (\frac{1}{2}\lambda)(p_t(1,2)+p_t(1,3)) \\ \frac{\mathrm{d}}{\mathrm{d}t}p_t(1,2) &= -\lambda p_t(1,2) + (\frac{1}{2}\lambda)(p_t(1,1)+p_t(1,3)) \\ \frac{\mathrm{d}}{\mathrm{d}t}p_t(1,3) &= -\lambda p_t(1,3) + (\frac{1}{2}\lambda)(p_t(1,1)+p_t(1,2)) \end{aligned}$$

The smart way to solve such equations is to use integrating factors, as described in Appendix A.1, together with whatever other clues there may be: by probability (!) we know

$$p_t(1,1)+p_t(1,2)+p_t(1,3) = 1$$

and so

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}p_t(1,1) &= -\lambda p_t(1,1) + (\frac{1}{2}\lambda)(1 - p_t(1,1)) \\ &= -(\frac{3}{2}\lambda)p_t(1,1) + \frac{1}{2}\lambda\end{aligned}$$

Now notice that

$$\frac{\mathrm{d}}{\mathrm{d}t}p_t(1,1) + \alpha p_t(1,1) = e^{-\alpha t}\frac{\mathrm{d}}{\mathrm{d}t}(e^{\alpha t}p_t(1,1))$$

so that if we set $u_t(i) = e^{+3\lambda t/2}p_t(1,i)$ *then*

$$\frac{\mathrm{d}}{\mathrm{d}t}u_t(1) = (\frac{1}{2}\lambda)e^{+3\lambda t/2}.$$

Hence (using $u_0(1) = 1$*) we find*

$$u_t(1) = \frac{e^{+3\lambda t/2} + 2}{3}$$

and thus

$$p_t(1,1) = \frac{1 + 2e^{-3\lambda t/2}}{3}.$$

Using the symmetry between states 2 *and* 3*, and probability,*

$$p_t(1,2) = p_t(1,3) \quad = \quad \frac{1 - p_t(1,1)}{2} \quad = \quad \frac{1 - e^{-3\lambda t/2}}{3}$$

Consequently we have achieved a complete solution of the forward differential equations. Notice that as t tends to infinity so the distribution converges to an equilibrium distribution which is independent of starting place.

There are not many finite-state-space Markov chains which are both easy enough to solve exactly *and* useful in practice. Here is one example.

Example 1.13 ***(Cross-over in genetics):*** *As you know from ST202 Section* **??**, *in diploid sexual reproduction a pair of chromosomes for the child is obtained by selecting one chromosome from the corresponding pair of the father, and one from the mother. However this picture is slightly over-simplified when one starts to consider two or more genes at different loci on the chromosome. Briefly, the process of combination of the chromosomes is such that* crossover *can occur: the chromosome selected from the father's pair is unreeled from the pair progressively and from time to time the unreeling process switches to the homologous chromosome (terminology as explained in ST202 Section* **??***). This happens at random, and leads to a Markov chain $\{X_t : t \geq 0\}$, where $X_t = 0, 1, 2$ depending on considering the chromosome pair for the child at location t along the pair; if this is derived from the same two chromosomes as at location 0 then $X_t = 2$, if there is just one chromosome in common then $X_t = 1$, if there are none in common then $X_t = 0$. Under reasonable assumptions (about crossover happening at random at*

rate λ, in fact according to a Poisson process as described in Definition 2.1) the $\underline{\underline{Q}}$-matrix for X is given by

$$\underline{\underline{Q}} = \begin{bmatrix} -2\lambda & 2\lambda & 0 \\ \lambda & -2\lambda & \lambda \\ 0 & 2\lambda & -2\lambda \end{bmatrix}.$$

We can then derive forward differential equations and obtain in particular

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}t} p_t(1,1) &= 2\lambda p_t(1,0) - 2\lambda p_t(1,1) + 2\lambda p_t(1,2) \\ &= 2\lambda \left(1 - p_t(1,1)\right) - 2\lambda p_t(1,1) \\ &= 2\lambda - 4\lambda p_t(1,1)\,, \\ p_0(1,1) &= 1\,. \end{aligned}$$

This allows us to employ the integrating factor method:

$$\begin{aligned} e^{-4\lambda t} \frac{\mathrm{d}}{\mathrm{d}t} \left(e^{4\lambda t} p_t(1,1)\right) &= 2\lambda \\ e^{4\lambda t} p_t(1,1) &= \frac{e^{4\lambda t}}{2} - \frac{1}{2} \\ p_t(1,1) &= \frac{1}{2}\left(1 + e^{-4\lambda t}\right) \\ p_t(1,0) = p_t(1,2) &= \frac{1}{4}\left(1 - e^{-4\lambda t}\right)\,. \end{aligned}$$

In general most finite-state-space Markov chains are not so easy to "solve" as this, though one could use *Matlab*, *Maple* or *Mathematica*. (Though where there is symmetry one can often proceed as above.) And infinite-state-space Markov chains are harder still ... unfortunate, because they are typically more useful in practice! Instead, as we will see in Chapter 2, we have to work out how to use the $\underline{\underline{Q}}$-matrix indirectly.

1.6. Summary

We can use this theory to specify a Markov chain, by

- requiring Standing Assumptions SA I, SA II, SA III to be in force,
- specifying $q(x)$ the rate of leaving state x,
- specifying $q(x, y)$ the rate of transitting from state x to state y (subject to SA III).

The differentiability results above tell us

- $p_t(x, x) = 1 - q(x) \times t + o(t)$ as $t \to 0$,[6]
- $p_t(x, y) = q(x, y) \times t + o(t)$ as $t \to 0$, $x \neq y$,
- if the backward differential equations can be solved without explosion then we know

[6] Here $o(t)$ is a correction term with $o(t)/t \to 0$ as $t \to 0$

  - $p_t(x, y)$,
  - if the forward differential equations can be solved then $p_t(x, y)$ is known in regular situations.

We will discuss instances where it is possible to solve backward or forward differential equations. However in general it is more important to know

- how to simulate the process X given $\underline{\underline{Q}}$,
- how to use $\underline{\underline{Q}}$ to draw comparisons between X and better understood processes.
- how to use $\underline{\underline{Q}}$ to study large-time behaviour.

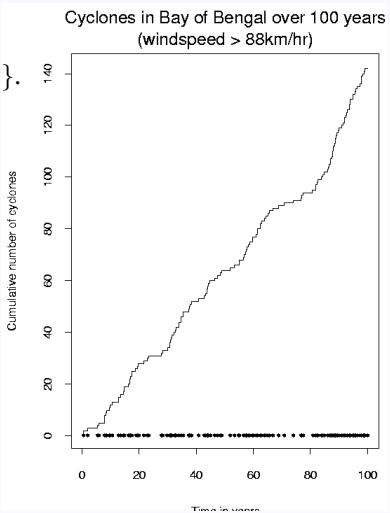
2. Linear Birth-Death

Most of the useful simple continuous-time Markov chains have countably-infinite state-space made up of the non-negative integers $\{0, 1, 2, \ldots\}$. They are prescribed by giving the transition rates $q(x, y)$ for x, y non-negative integers: and the simplest are such that

(a) the only non-zero transition rates are those for which $x = y \pm 1$ (or of course $x = y$);

(b) $q(x, x+1)$, $q(x, x-1)$ depend linearly on x.

These are the *linear birth-death processes*.



2.1. The Poisson process

In the simplest case only $q(x, x+1) = q(x) = -q(x, x)$ are non-zero, and they are constant (do not depend on x at all). This is an important model for $\{N_t : t \geq 0\}$

where N_t counts the number of incidents (of a specified kind) occurring in the time interval $[0, 1]$.

There are many many applications: modelling times of child-births, industrial accidents, imperfections in cotton yarns, alpha-particle decay, extreme weather incidents, ... But also the Poisson process is highly significant as a building block for all kinds of other random processes, as we will see later in this course.

Here is an actual dataset which could be (and has been!) modelled by a Poisson process: the times of 142 cyclones (storms with wind-speeds exceeding 88km/hr) in the Bay of Bengal, recorded over a 100-year period.

Definition 2.1 *The random process N is a* Poisson (counting) process *of rate λ if it:*

(i) is a Markov chain satisfying SA I, SA II, SA III ;

(ii) is a counting process, so $q(x, y) = 0$ if $y < x$;

(iii) sees no jumps of size greater than 1, so $q(x, y) = 0$ if $|x - y| > 1$;

(iv) sees jumps at a constant rate, so $q(x, x + 1) = q(x) = -q(x, x) = \lambda$ for $x = 0, 1, 2, \ldots$.

Now we can build the backward differential equations:

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}t} p_t(n, n) &= -\lambda p_t(n, n) \\ \frac{\mathrm{d}}{\mathrm{d}t} p_t(n, n+k) &= -\lambda p_t(n, n+k) + \lambda p_t(n+1, n+k) \end{aligned}$$

with boundary conditions

$$\begin{aligned} p_0(n,n) &= 1 \\ p_0(n,n+k) &= 0 \text{ for } k = 1,2,\ldots \end{aligned}$$

These backward differential equations can be solved! (This is unusual ....) Use the integrating factor approach: solve the $\frac{\mathrm{d}}{\mathrm{d}t}p_t(n,n)$ differential equation by considering $\frac{\mathrm{d}}{\mathrm{d}t}(e^{\lambda t}p_t(n,n))$, then work up using induction. Answer is

$$p_t(n,n+k) = (\lambda t)^k e^{-\lambda t}/k!$$

for $k = 0,1,2,\ldots$.

Example 2.2 *Suppose N_t is the number of children born at Coventry Walsgrave Hospital over $[0,t]$. Suppose N is a Poisson process of rate λ as above. Then the number N_t born by time t has a Poisson distribution of mean λt.*

How do we construct a Poisson process (for example if we want to simulate it)?

Theorem 2.3 ***(Construction of Poisson process):*** *Take $T_1, T_2, \ldots,$ independent random variables of common exponential distribution, mean $1/\lambda$. Set*

$$\begin{aligned} N_t &= n \text{ if } T_1 + \ldots + T_n \le t < T_1 + \ldots + T_{n+1} \\ N_t &= 0 \text{ if } t < T_1 \end{aligned}$$

then N is a Poisson counting process of rate λ.

Proof: This depends on the memoryless property of exponential random variables:

$$\mathbb{P}\left[T > t+s | T > s\right] \quad = \quad \mathbb{P}\left[T > t\right] \text{ for all } s,t > 0$$

if and only if T is exponential. (Use definition of conditional probability, take logs of both sides, to see that

$$f(t) \quad = \quad \log \mathbb{P}\left[T > t\right]$$

is additive in t. Since f is continuous, $f(t) = a \times t$ for some constant a.) Because of the memoryless property it follows that N is a Markov process with stationary transition probabilities. For suppose that $u_n = t_1 + \ldots + t_n \leq t$ and consider the conditional probability

$$\begin{aligned}
&\mathbb{P}\left[N_{t+s} \geq n+k | N_t = n, T_1 = t_1, \ldots, T_n = t_n\right] \\
&= \mathbb{P}\left[T_1 + \ldots + T_{n+k} \leq t+s | N_t = n, T_1 = t_1, \ldots, T_n = t_n\right] \\
&\quad = \mathbb{P}\big[T_1 + \ldots + T_{n+k} \leq t+s \\
&\qquad\qquad | T_{n+1} > t - u_n, T_1 = t_1, \ldots, T_n = t_n\big] \\
&\quad = \mathbb{P}\big[T_{n+1} + \ldots + T_{n+k} \leq t+s-u_n \\
&\qquad\qquad | T_{n+1} > t - u_n, T_1 = t_1, \ldots, T_n = t_n\big] \\
&= \mathbb{P}\left[T_{n+1} + \ldots + T_{n+k} \leq t+s-u_n | T_{n+1} > t-u_n\right] \\
&\qquad \text{by independence} \\
&= \mathbb{P}\left[T_{n+1} + \ldots + T_{n+k} \leq s\right] \text{ (so doesn't depend on past)}
\end{aligned}$$

by the memoryless property, since it implies that $T_{n+1} - (t - u_n), T_{n+2} \ldots, T_{n+k}$ under the condition $T_{n+1} > t - u_n$ have the same joint distribution as $T_{n+1}, T_{n+2}, \ldots, T_{n+k}$.

Now all we have to do is check the transition rates. (We assume SA I, SA II, SA III hold). We first observe

$$\begin{aligned} \sum_{k>n+1} q_{n,k} &= \lim_{t\to 0} \mathbb{P}\left[\, N_t > n+1 | N_0 = n \,\right]/t \\ &= \lim_{t\to 0} \mathbb{P}\left[\, T_1 + T_2 < t \,\right]/t \\ &= \lim_{t\to 0} \int_0^t \lambda^2 u e^{-\lambda u} du/t = 0 \end{aligned}$$

by the theory of differentiation of integrals.

Clearly $q_{n,k} = 0$ if $k < n$, so it remains (by SA III) to compute *either* $q_{n,n}$ *or* $q_{n,n+1}$. We consider

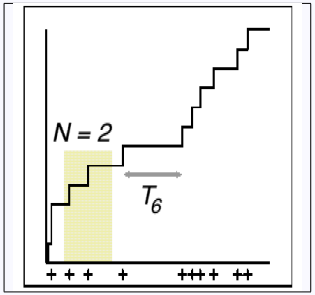
$$\begin{aligned} q_{n,n} &= \lim_{t\to 0}(\mathbb{P}\left[\, N_t = n | N_0 = n \,\right] - 1)/t \\ &= \lim_{t\to 0}(\mathbb{P}\left[\, T_1 > t \,\right] - 1)/t = \lim_{t\to 0}(e^{-\lambda t} - 1)/t \\ &= -\lambda \end{aligned}$$

This completes the proof. □

The above theorem tells us how to write a program to simulate a Poisson process (simulation of exponential random variables is easy!). It is also a valuable prototype for several arguments that follow: first prove a process is Markov

(often we will merely glance over this step) and then compute the rates. Furthermore often a simple argument (as here) will show that most of the rates are zero.

A consequence of our solution of the backward differential equations, together with the Markov property, is the following result which often allows us to compute probabilities of events concerned with Poisson processes.



Theorem 2.4 Independent increments of the Poisson (counting) process *The increment* $N_{t+u} - N_t$ *of a Poisson process of rate* λ *is Poisson of mean* λu*, and is independent of the past (incidents over* $[0, t]$*).*

Proof: Consider that by the Markov property

$$\begin{aligned} &\mathbb{P}\left[\, N_{t+u} - N_t = k | N_t = n, T_1 = t_1, \ldots, T_n = t_n \,\right] \\ &= \quad \mathbb{P}\left[\, N_{t+u} - N_t = k \,\right] \end{aligned}$$

and the right-hand side is the appropriate Poisson probability by our calculations above. □

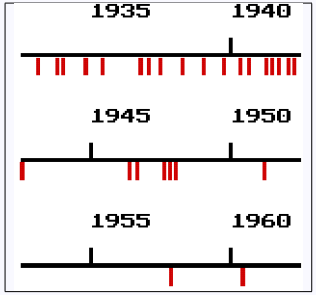
Theorems 2.3, 2.4 suggest two complementary views of a Poisson (counting) process: one may focus on either the inter-incident times T_i or on the counts over

specified time intervals $N(t, t+u) = N_{t+u} - N_t$. You should note the fundamental relation:

$$\mathbb{P}\left[T_1 + \ldots + T_k > t\right] = \mathbb{P}\left[N_t < k\right]$$

which has been used implicitly several times above.

There are a huge number of applications, of which we will discuss just one in lectures (see also the exercises).



Example 2.5 *Coal mining disasters in the UK. (Jarrett [1]) Consider the first* 10 *years of the plot of coal mining disasters* [7] *given below (note to the historically minded: why might this decade be considered different from subsequent decades?). Do disasters cluster in time?*

There are 18 disasters in the first 10 years (1933-1942) which suggests we estimate λ by $1.8 = 18/10$. Let us define a clustered incident as a disaster which is followed less than $\delta = 0.25$ years later by another (this is rather ad hoc, but the example is merely illustrative). Then we *expect* about $\lambda t(1 - e^{-\lambda\delta}) = 6.52$ clustered incidents in this period (see below) on the basis of Poisson variation. Since we observe 6 clustered incidents, there does not appear

[7] A disaster means 10 or more dead.

to be any clustering. (A similar calculation for the next 20 years shows we expect 0.95 clustered incidents but only see 3: obviously a more sophisticated analysis is required!)

To explain the calculation of the expectation we have to anticipate the *Strong Markov Property* described in Section 4. If N is a Poisson (counting) process of rate λ then if S_n is the time of the n^{th} incident it is reasonable (and in fact true by this strong Markov property) that $\{N(S_n + s) - N(S_n) : s \geq 0\}$ forms a Poisson process of rate λ (this is actually obvious for a Poisson process constructed according to Theorem 2.3). Then the probability that the n^{th} incident is a clustered incident is equal to $(1 - e^{-\lambda\delta})$. It follows that the expected number of clustered incidents before time t is given by

$$\begin{aligned}
&\sum_n \mathbb{P}\left[\, S_n \leq t \text{ and another incident follows within } \delta \,\right] \\
&= \sum_n (1 - e^{-\lambda\delta})\, \mathbb{P}\left[\, S_n \leq t \,\right] = (1 - e^{-\lambda\delta})\, \mathbb{E}\left[\, N_t \,\right] \\
&= \lambda t (1 - e^{-\lambda\delta})
\end{aligned}$$

as required. (See Appendix A.3 for remarks on why $\mathbb{E}\left[\, N_t \,\right] = \sum_n \mathbb{P}\left[\, S_n \leq t \,\right]$.)

A slightly more sophisticated analysis can be based on recognizing that a Poisson process of rate λ over $[0, h]$ (say) can be built by

(a) collecting Poisson(λh) points,

(b) placing them independently and uniformly on $[0, h]$.

(Checking this is a nice exercise!) Then the probability of there being r clustered incidents can be computed exactly.

Note that here we are using the Poisson counting process as a kind of benchmark for purely random distribution of incidents. This is one very typical use: indeed the Poisson counting process is sometimes described as "the" completely random point process. We shall return to this in Section 6.

Another crucial rôle for the Poisson counting process is as a building block for other processes. Before we see this we first need a few more theorems on what happens when we change the process in various ways.

Suppose you are waiting at a bus-stop, and buses headed for Warwick University campus arrive according to a Poisson process of rate λ, while buses headed for Coventry Airport arrive according to a Poisson process of rate μ, independently of the Warwick University buses. Consider the process which simply counts the number of buses arrived, without regard for where they are headed. This too is Poisson ....

Theorem 2.6 ***(Superposition of independent Poisson processes):*** *If N and M are independent Poisson counting processes of rates λ, μ then their sum $N + M$ is a Poisson process of rate $\lambda + \mu$.*

Proof: The main part of the work lies in showing the Markov property. This follows if $N_{t+s}+M_{t+s}-(N_t+M_t)$ is independent of the past of both N and M at time t. It is sufficient to show that $C = [N_{t+s}-N_t = i]$ and $D = [M_{t+s}-M_t = j]$ are independent of the composite event $A \cap B$, where A, B are events in the past of N, M respectively at time t.

But by definition of conditional probability

$$\mathbb{P}\left[C \cap D | A \cap B\right] \quad = \quad \frac{\mathbb{P}\left[C \cap D \cap A \cap B\right]}{\mathbb{P}\left[A \cap B\right]}$$

and by independence

$$\mathbb{P}\left[C \cap D \cap A \cap B\right] \quad = \quad \mathbb{P}\left[C \cap A\right] \times \mathbb{P}\left[D \cap B\right] ,$$

while

$$\mathbb{P}\left[A \cap B\right] = \mathbb{P}\left[A\right] \times \mathbb{P}\left[B\right] .$$

Hence

$$\mathbb{P}\left[C \cap D | A \cap B\right] \quad = \quad \mathbb{P}\left[C|A\right] \times \mathbb{P}\left[D|B\right] = \mathbb{P}\left[C\right] \times \mathbb{P}\left[D\right] ,$$

where the last step uses the Markov property (actually Theorem 2.4 for Poisson processes).

It follows that

$$\begin{aligned} &\mathbb{P}\left[N_{t+s} + M_{t+s} - (N_t + M_t) = k | A \cap B\right] \\ &\quad = \quad \mathbb{P}\left[N_{t+s} + M_{t+s} - (N_t + M_t) = k\right] , \end{aligned}$$

since this can be broken up into events as above. It remains to check that the rates are correct. But this is easy: for example

$$\begin{aligned} q_{n,n+1} \quad &= \quad \lim_{t \to 0} \mathbb{P}\left[N_t + M_t = n + 1 | N_0 + M_0 = n\right] / t \\ &= \quad \lim_{t \to 0} (\lambda + \mu) t e^{-(\lambda+\mu)t} / t = \lambda + \mu \end{aligned}$$

using the result that the sum of two independent Poisson random variables is itself Poisson. □

Suppose you are waiting at this bus-stop after reluctantly waking up for the morning after the night before, which you spent partying away. You are only intermittently conscious, and do not count any buses arriving during your flashes of micro-sleep. What can be said about the number of buses you actually succeed in counting?

Theorem 2.7 ***(Poisson counts):*** *If N is a Poisson (counting) process of rate λ, and A is any (measurable) subset of the time axis $[0, \infty)$ of length measure a, then the number $N(A)$ of incidents counted by N which lie in A is Poisson of mean λa.*

Proof: We shall prove this only for the case when the set A is the union of a sequence of *disjoint* intervals

$$A \quad = \quad (s_1, t_1) \cup (s_2, t_2) \cup \ldots \cup (s_n, t_n)\,.$$

(More general cases are not hard: one uses techniques derived from ST213) Then

$$N(A) \quad = \quad N(s_1, t_1) + N(s_2, t_2) + \ldots + N(s_n, t_n)$$

and we can apply Theorem 2.4 sequentially to see this is a sum of independent Poisson random variables of means adding up to

$$\lambda \times (t_1 - s_1 + t_2 - s_2 + \ldots + t_n - s_n) \quad = \quad \lambda \times a\,.$$

□

Theorem 2.8 ***(Independent thinning of Poisson processes):*** *If N is a Poisson (counting) process of rate λ and it is thinned by independent removal of incidents with probability $1-p$, then the thinned process $\overline{N}$ (counting the remaining incidents) is a Poisson (counting) process of rate $p\lambda$.*

Proof: First check $\overline{N}$ is Markov. This is almost obvious, because the thinning is independent, and so we shall say no more about it. Now check the rates. Since we are reducing the numbers of incidents the only crucial matter is to compute

$$\begin{aligned} q_{n,n+1} &= \lim_{t\to 0} \mathbb{P}\left[\,\overline{N}_t = 1 | \overline{N}_0 = 0\,\right]/t \\ &= \lim_{t\to 0} \mathbb{P}\left[\text{ no thinning by } t\,\right] \mathbb{P}\left[\,N_t = 1 | N_0 = 0\,\right]/t \\ &\quad + \lim_{t\to 0} \mathbb{P}\left[\,N_t > 1, \text{ only 1 not thinned by } t | N_0 = 0\,\right]/t \\ &= \lim_{t\to 0} p \times \lambda t e^{-\lambda t}/t = p\lambda\,, \end{aligned}$$

since the second limit must be zero. □

Minor generalizations of the Poisson (counting) process include the following:

- the time-inhomogeneous Poisson (counting) process (the rate λ is allowed to vary deterministically with time);
- the mixed Poisson (counting) process (the rate is constant in time but is itself a random variable);

- the Cox point process or doubly stochastic Poisson process (the rate λ is itself a random process ... or perhaps even more generally a *random measure*).
- even more generally than this: allowing the intensity to depend on the point process realization itself (*self-exciting point processes*).

2.2. Birth processes

Now suppose that N_t counts a population of individuals, each of which breeds independently at constant rate λ.

Example 2.9 ***(1992 Assessed Exercise):*** *Cosmic rays arrive at atmosphere top, at height y miles, with vertical velocity $-v$ miles $\cdot$ second^{-1}. As they fall they collide occasionally with atmosphere molecules and give birth to other particles at rate α per second, which in turn given birth to others and so forth. All particles travel with constant vertical velocity $-v$ miles per second. Assume that X_0 particles arrive simultaneously at atmosphere top at 0 seconds and set the number of particles at t seconds to be X_t. Then we might consider X to be a Markov chain with rates given by $X \to X + 1$ at rate αX.*

Definition 2.10 ***(Pure Birth process):*** *The process N as above is a pure birth process, birth rate λ, if:*

1. *it is a Markov chain (satisfying SA I, SA II, SA III);*
2. *it is a counting process ($q_{x,y} = 0$ if $x > y$);*
3. *no jumps are greater than 1 ($q_{x,y} = 0$ if $|y - x| > 1$);*
4. *new-born individuals arrive at a rate proportional to the current population ($q_x = -q_{x,x} = \lambda x$ for $x = 1, 2, \ldots$). (Note: if $x = 0$ process never changes!)*

The backward differential equations are:

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}t} p_t(n,n) &= -\lambda n p_t(n,n) \\ \frac{\mathrm{d}}{\mathrm{d}t} p_t(n,n+k) &= -\lambda n p_t(n,n+k) + \lambda n p_t(n+1,n+k) \end{aligned}$$

(for $k = 1, 2, \ldots$) with

$$\begin{aligned} p_0(n,n) &= 1\,; \\ p_0(n,n+k) &= 0 \quad \text{for } k > 0\,. \end{aligned}$$

The integrating factor method[8] produces the following representation for their solution: for $k = 1, 2, \ldots$

$$p_t(n, n+k) \quad = \quad e^{-n\lambda t} n\lambda \int_0^t e^{n\lambda s} p_s(n+1, n+k)\, \mathrm{d}s$$

while for $k = 0$ we see immediately

$$p_t(n, n) = e^{-n\lambda t}\,.$$

Notice the non-trivial state-space classification of this process! (Communicating classes are $\{0\}, \{1\}, \{2\} \ldots$.)

The forward differential equations are easier to solve:

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}t} p_t(1,1) &= -\lambda p_t(1,1) \\ \frac{\mathrm{d}}{\mathrm{d}t} p_t(1,k) &= -\lambda k p_t(1,k) + \lambda(k-1) p_t(1,k-1) \end{aligned}$$

(for $k = 2, 3, \ldots$) with

$$\begin{aligned} p_0(1,1) &= 1\,; \\ p_0(1,k) &= 0 \text{ for } k > 1\,. \end{aligned}$$

[8] Differentiate $e^{-n\lambda s} p_s(n, n+k)$.

Their solution[9] (spot the geometric distribution!) is

$$\begin{aligned} p_t(1,k) &= e^{-\lambda t}(1-e^{-\lambda t})^{k-1} \\ &= e^{-\lambda k t}(e^{\lambda t}-1)^{k-1}\,. \end{aligned}$$

Exercise 2.11 *(Not hard, but messy.) Show that the backwards and forwards differential equations have the same solution!*
Hint: *From the forwards equation formula* $p_t(1,1+k) = (1-\exp(-\lambda t)p_t(1,k)$.
Hint: *the transition probability* $p_t(n,n+k)$ *is the probability that* n *independent pure birth processes starting with just* 1 *individual each give rise to a total of* $n+k$ *individuals at time* t.

An alternative approach calculates a partial differential equation for the generating function

$$G_t(z) = \sum_k z^k p_t(1,k)\,.$$

See Exercise 3.5 for a slightly more general version.

Exercise 2.12 *The link with the Geometric distribution leads to a curious property for the pure birth process:*

$$\mathbb{P}\left[\, N_t = 1+k_1+k_2 | N_t \geq 1+k_1\,, N_0 = 1\,\right] = \mathbb{P}\left[\, N_t = 1+k_2 | N_0 = 1\,\right]\,.$$

[9] Try a geometric type of answer: $p_t(1,k) = a_k(t)(1-a_k(t))^{k-1}$ with $a_k(0) = 1$.

To see this, compute the conditional probability ratio using transition probability functions, then sum the geometric series: the left-hand side then equals

$$\begin{aligned} & p_t(1, 1+k_1+k_2) / \sum_{r>0} p_t(1, 1+k_1+r) \\ &= a(1-a)^{k_1+k_2} / \sum_{r>0} a(1-a)^{k_1+r} \\ &= a(1-a)^{k_1+k_2} / (1-a)^{k_1} = p_t(1, 1+k_2) \end{aligned}$$

Theorem 2.13 ***(Construction of pure birth process):*** *Suppose that $T_1, T_2, \ldots$ are independent exponential random variables, such that the mean of T_n is $(n\lambda)^{-1}$. Set*

$$\begin{aligned} N_t &= n+1 \text{ if } T_1 + \ldots + T_n \le t < T_1 + \ldots + T_{n+1} \\ N_t &= 1 \text{ if } t < T_1 \, . \end{aligned}$$

Then N is a pure birth process of birth rate λ, with initial value $N_0 = 1$. (Start at 1: different from Theorem 2.3.)

Proof: As with Theorem 2.3 we must prove the Markov property (proof is very similar so we omit it), assume SA I, SA II, SA III, and check rates. Here is a sketch of the rates computations. It is crucial to check that

$$\mathbb{P}\left[\text{ more than one birth before } t | N_0 = n \right] = o(t)$$

which works as in Theorem 2.3. Then show

$$\begin{aligned} q_{n,n} &= \lim_{t\to 0}(\mathbb{P}[N_t = n | N_0 = n] - 1)/t \\ &= \lim_{t\to 0}(\mathbb{P}[T_n > t] - 1)/t \\ &= \lim_{t\to 0}(e^{-n\lambda t} - 1)/t = -n\lambda\,. \end{aligned}$$

This completes the proof. □

Exercise 2.14 *Find differential equations for the moments* $\mu_k(t) = \mathbb{E}[N_t^k]$ *of* N*, hence find the mean and variance as functions of time. Do this by computing the first derivative of* $\mu_k(t) = \sum_n n^k p_t(1,n)$ *term by term, and seeing from the forward differential equations that*

$$(d/dt)\mu_k(t) = -\lambda\mu_{k+1}(t) + \lambda\sum_n n(n+1)^k p_t(1,n)$$

so (setting $k = 1, 2$*) finding solutions (if* $N_0 = 1$*)*

$$\begin{aligned} \mu_1(t) &= e^{\lambda t} \\ \mu_2(t) &= 2e^{2\lambda t} - e^{\lambda t} \\ \textbf{\textit{Var}}(N_t) &= e^{\lambda t}(e^{\lambda t} - 1) \end{aligned}$$

Hence $\textbf{\textit{Var}}(N_t)/(\mathbb{E}[N_t])^2$ *converges to* 1 *as time increases.* In fact $N_t/\mathbb{E}[N_t]$ *converges in distribution to an exponential distribution of mean* 1.

Exercise 2.15 *The generalized birth process has $q_n = \lambda_n$ and $q_{n,n+1} = \lambda_n$ for a general sequence $\{\lambda_1, \lambda_2, \ldots\}$. Construct this using independent exponential random variables. Show that it "explodes" in finite time if $\sum_n 1/\lambda_n$ converges. (This is an example of a Markov chain for which the backwards differential equations can have several different solutions ....)*

Example 2.16 ***(Coalescent death process):*** *This is a process which arises in the study of coalescence and of genetics: it is a Markov chain on $\{0, 1, 2, \ldots\}$ satisfying the standing assumptions SA I, SA II, SA III, with rates*

$$\begin{aligned} q(x, x-1) &= \frac{1}{2}\lambda(x-1)x\,, \\ q(x, x) &= -\frac{1}{2}\lambda(x-1)x\,. \end{aligned}$$

In coalescence theory we start with X_0 particles, which combine together in a pairwise fashion, at a rate proportional to the number of pairs: interest focuses on the distribution of the time at which all particles have finally combined into one single conglomerate. In genetics this arises when we study the family tree of a sample from a population. At time 0 there are X_0 subjects. Running time backwards, at time $-t$ we let X_t denote the number of female ancestors of these subjects. Under reasonable assumptions including random mating we may suppose that $\{X_t : t \geq 0\}$ behaves as above.

Using analyses of the subjects' DNA one can deduce facts concerning the basic arrangement of the resulting family tree, and then use Markov chain Monte Carlo

simulation methods to derive a posterior distribution for the time T of the birthday of the latest common ancestor of the sample! These ideas are being vigorously developed: see for example the Oxford University Mathematical Genetics home page.

Exercise 2.17 *For a pure birth process N we have*

$$\mathbb{P}\left[\, N_t \leq n \,\right] = \mathbb{P}\left[\, T_1 + \ldots + T_n > t \,\right] .$$

Now we can say something about the density of a sum of independent exponential random variables of different means: T_1, …, T_n with means $1/\lambda_1$, …, $1/\lambda_n$, no two of which are equal. Consider the moment generating function of the sum (with argument p):

$$\text{mgf}(T_1 + \ldots + T_n) \quad = \quad \prod_k \lambda_k/(\lambda_k - p)$$

Because the λ_k are all different we can solve for A_k in

$$\prod_k \lambda_k/(\lambda_k - p) \quad = \quad \sum_k A_k \lambda_k/(\lambda_k - p)$$

and find

$$A_m \quad = \quad \prod_{k:k\neq m} \lambda_k/(\lambda_k - \lambda_m)$$

(this is pretty! use partial fractions, or polynomials:

$$\prod_k \lambda_k \quad = \quad \sum_m A_m \lambda_m \times \left(\prod_{k:k\neq m} (\lambda_k - p) \right)$$

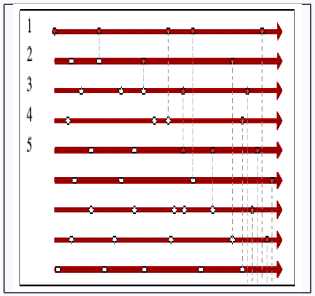
then set $p = \lambda_m$ and solve for A_m.) Deduce from uniqueness of moment-generating functions that the sum has density

$$\sum_k A_k \lambda_k \exp(-\lambda_k t)\,.$$

Hence

$$\mathbb{P}\,[\,N_t \leq n\,] \quad = \quad \sum_k A_k \exp(-\lambda_k t)\,.$$

The pure birth process can be usefully considered as arising from a construction using a "microscopic model". This describes the behaviour of individuals $1, 2, 3, \ldots$. First construct a countable set of independent Poisson processes. Each represents an individual. Initially no individuals are alive except for the first n of them (in the figure $n = 1$).



An alive individual gives birth whenever an incident occurs on her Poisson process track. Birth

is implemented by making alive the unborn individual with the lowest number.

Exercise 2.18 *Check the total number of individuals is a pure birth process. First prove the Markov property: this follows easily from Theorem 2.4. Assume SA I, SA II, SA III. Show that*

$$\mathbb{P}\,[\ \textit{more than one birth before } t | N_0 = n\,] \quad = \quad o(t)$$

then compute

$$\begin{aligned} q_{n,n} &= \lim_{t\to 0} \mathbb{P}\,[\,N_t = n | N_0 = n\,] - 1)/t \\ &= \lim_{t\to 0} (e^{-n\lambda t} - 1)/t = -n\lambda \end{aligned}$$

since the superposition of n independent Poisson(λ) processes is a Poisson($n\lambda$) process (Theorem 2.6). Hence the rates are correct!

Notice how the Poisson counting process is used as a building block. This will be a continuing theme of these lectures.

2.3. Birth-death processes

Suppose N_t counts individuals which reproduce at rate λ but also die at rate μ.
Examples: a culture of bacteria; an epidemic in a large population; the cosmic ray example if collisions either give rise to a pair of cosmic rays or nothing; ...

Definition 2.19 ***(Linear birth-death process):*** *N is a birth-death process, birth rate λ, death rate μ, if:*

(i) it is a Markov chain (satisfying SA I, SA II, SA III);

(ii) it is never negative ($q_{x,y} = 0$ if $y < 0$);

(iii) no jumps are greater than 1 ($q_{x,y} = 0$ if $y > x + 1$);

(iv) individuals arrive and leave at rates proportional to the current population ($q_{x,x+1} = \lambda x$, $q_{x,x-1} = \mu x$, $q_x = -q_{x,x} = (\lambda + \mu)x$, for $x = 0, 1, 2, \ldots$).

The backward differential equations are:

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}p_t(n,m) \quad &= \\ -(\lambda+\mu)np_t(n,m) + \lambda np_t(n+1,m) &+ \\ +\mu np_t(n-1,m) &\end{aligned} \tag{1}$$

for $m = 0, 1, 2, \ldots$ with $p_0(n,n) = 1$; $\quad p_0(n,m) = 0$ for $m \neq n$. (Here we use the useful convention, $p_t(n,m) = 0$ if either n or m are negative.)

Similarly the forward differential equations are:

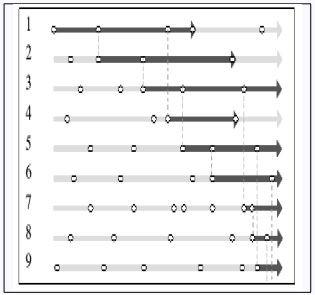
$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}p_t(n,m) \quad &= \\ -(\lambda+\mu)mp_t(n,m) + \lambda(m-1)p_t(n,m-1) &+ \\ +\mu(m+1)p_t(n,m+1) &\end{aligned} \tag{2}$$

for $m = 0, 1, 2, \ldots$) with

$$p_0(n, n) = 1; \qquad p_0(n, m) = 0 \text{ for } m \neq n.$$

To solve this we would have to work with probability generating functions and hence partial differential equations (see Exercise 3.5). This does not seem illuminating. Instead we follow up the "microscopic model" work of above.

The birth-death process can also be derived from a "microscopic model" describing the behaviour of individuals 1, 2, 3, ... First construct a countable set of independent Poisson processes, one per possible individual. Whenever an incident occurs on an alive individual's Poisson track it gives birth with probability p, dies with probability $1 - p$. Birth is implemented by making alive the individual with the lowest number who has not been alive before.



Exercise 2.20 *Check the total number of alive individuals at time t is a birth-death process. What are the birth and death rates λ, μ? (Follow the now-established procedure: Markov property via Theorem 2.4; assume SA I, SA II, SA III ; check the crucial*

$$\mathbb{P}\,[\ \textit{more than one birth or death before } t | N_0 = n\ ] = o(t)$$

which now needs a little more thought: you must allow for a birth leading to a further birth or death; then compute rates using Theorem 2.6.)

Exercise 2.21 *Study the differential equations for mean and variance of linear birth-death process, modifying the procedure carried out for the pure birth process.*

Note that the birth-death process has just two communicating classes: $\{0\}$, $\{1, 2, 3, \ldots\}$ (if $\lambda, \mu > 0$).

Exercise 2.22 *Hidden inside the birth-death process is a random walk (**Hint:** N_t changes by ± 1 every time there is a birth or death). Find the probability that the birth-death process becomes extinct!*

Answer: See ST202 Section **??**. Let X_n be the value of N each time there is a birth or death. Then $X_{n+1} = X_n \pm 1$, using $+1$ with probability p, -1 with probability $1 - p$ if $X \neq 0$. If $X_0 = 1$ then the probability of ever hitting 0 is $\min\{1, (1-p)/p\}$.

Exercise 2.23 *Hidden inside the birth-death process is a branching process (**Hint:** consider the immediate offspring of the initial individuals in the microscopic model above). Find the family-size probability generating function and hence again the extinction probability. Also find the probability generating function of the total population, in case extinction is certain.*

Answer: See ST202 Section **??**. Family-size generating function is $G(s) = (1-p)/(1-ps)$. Extinction probability is smallest nonnegative root η of $s = G(s)$, hence $\eta = \min\{1, (1-p)/p\}$ again. If extinction is certain then probability

generating function of total population size is $R(s)$ solving $R(s) = sG(R(s))$. Solution is

$$R(s) = \frac{1-\sqrt{1-4p(1-p)s}}{2p}.$$

If $p < 1/2$ then mean total population size is

$$R'(1) = \frac{1-p}{1-2p}.$$

Definition 2.24 ***(Birth-death-immigration process):*** *The process N is a birth-death-immigration process, birth rate λ, death rate μ, immigration rate α, if:*

(i) it is a Markov chain (satisfying SA I, SA II, SA III);

(ii) it is never negative ($q_{x,y} = 0$ if $y < 0$);

(iii) no jumps are greater than 1 ($q_{x,y} = 0$ if $|y - x| > 1$);

(iv) individuals arrive and leave at rates as follows $q_{x,x+1} = (\alpha+\lambda x)$, $q_{x,x-1} = \mu x$, and consequently $q_x = -q_{x,x} = \alpha + (\lambda + \mu)x$, for $x = 0, 1, 2, \ldots$.

The backward differential equations are:

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}p_t(n,m) &= \\ -(\alpha+(\lambda+\mu)n)p_t(n,m) &+ (\alpha+\lambda n)p_t(n+1,m) + \\ &+\mu n p_t(n-1,m)\end{aligned} \tag{3}$$

(for m = 0,1,2, . . .) with

$$p_0(n,n) = 1; \qquad p_0(n,m) = 0 \text{ for } m \neq n\,.$$

(Again we use the useful convention, $p_t(n,m) = 0$ if either n or m are negative.)

Similarly the forward differential equations are:

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}p_t(n,m) \quad &= \\ -(\alpha + (\lambda+\mu)m)p_t(n,m) &+ (\alpha + \lambda(m-1))p_t(n,m-1) + \\ +\mu(m+1)p_t(n,m+1) & \end{aligned} \tag{4}$$

(for m = 0,1,2, . . .) with

$$p_0(n,n) = 1; \qquad p_0(n,m) = 0 \text{ for } m \neq n\,.$$

We won't try to solve either of these . . . .

Note the state-space is just one communicating class if $\alpha, \lambda, \mu > 0$.

Exercise 2.25 *Make a construction of the birth-death- immigration process using a "microscopic model".*

In this case the process will never die away. Can it settle down to a statistical equilibrium?

Exercise 2.26 ***(Birth-death-immigration process equilibrium distribution):*** *Arguing informally, we expect $p_t(n,m) \to \pi_m$ as time increases if there is an equilibrium, and so we should get the π_m if we set $\frac{\mathrm{d}}{\mathrm{d}t}p_t(n,m)$ to zero. The backwards*

differential equations don't help (by SA III we get $0 = 0$!) but the forwards differential equations give

$$\begin{aligned} -(\alpha + (\lambda+\mu)m)\pi_m + (\alpha + \lambda(m-1))\pi_{m-1} + \mu(m+1)\pi_{m+1} \\ = \quad 0\,. \end{aligned}$$

Setting $m = 0$, and using a convention that π_m is zero for negative m, we find $\alpha\pi_0 = \mu\pi_1$. Proceeding,

$$\mu(m+1)\pi_{m+1} \quad = \quad (\alpha + (\lambda+\mu)m)\pi_m - (\alpha + \lambda(m-1))\pi_{m-1}$$

and so

$$\pi_m \quad = \quad \frac{\alpha(\lambda+\alpha)\ldots((m-1)\lambda+\alpha)\pi_0}{\mu^m m!}\,.$$

Obviously we must have $\sum_m \pi_m = 1$, and this sum can converge only if $\lambda < \mu$. Hence we get a formula for π_m: for example

$$\begin{aligned} \pi_0 \quad &= \quad \left[\sum_m \alpha(\lambda+\alpha)\ldots((m-1)\lambda+\alpha)/(\mu^m m!)\right]^{-1} \\ &= \quad (1-\lambda/\mu)^{\alpha/\lambda} \end{aligned}$$

(the last step follows by expanding using the generalized binomial theorem).

In matrix terms we see what the above is doing is taking the evolution equation for the vector of equilibrium probabilities

$$\pi \quad = \quad \pi \underline{\underline{P}}(t)$$

and differentiating to get

$$0 \quad = \quad \pi \underline{\underline{Q}} \, .$$

This applies in general: when we want to find the equilibrium distribution $\{\pi_n\}_n$ of a Markov chain with $\underline{\underline{Q}}$-matrix $\underline{\underline{Q}} = \{q(n,m)\}_{n,m}$ then we must solve the system of linear equations derived from the above matrix equation:

$$0 \quad = \quad \sum_n \pi_n q(n,m) \, . \tag{5}$$

This system does not give all the information we need in order to obtain equilibrium: as long as the state-space is irreducible it defines everything in terms of (say) π_0, and then we must fix π_0 using

$$\sum_n \pi_n \quad = \quad 1 \, .$$

(Of course sometimes we find that the sum on the left-hand side will not converge unless π_0, and thus all the π_n, vanish: in this case equilibrium cannot exist: compare the case $\lambda \geq \mu$ above.)

Exercise 2.27 *Neutrons enter a nuclear reactor at rate α, and then collide with Uranium 235 atoms at rate β per particle. A collision has probability p of giving rise to two neutrons, $1-p$ of giving rise to none.*

This can be formulated as a birth-death-immigration process, birth rate $p\beta$, death rate $(1-p)\beta$, immigration rate α. We can compute the equilibrium distribution of the number of neutrons in the reactor as above. Furthermore, if π_0 is the

equilibrium distribution probability of there being no neutrons in the reactor, then consider the mean long-run proportion of time that there are no neutrons present. This is

$$\mathbb{E}\left[\frac{1}{T}\int_0^T \mathbb{I}_{[X_t=0]}\,\mathrm{d}t | X_0 = x_0\right] \quad = \quad \frac{1}{T}\int_0^T p_t(x_0,0)\,\mathrm{d}t\,,$$

which converges to π_0.

Exercise 2.28 *Describe a birth-death-immigration process with zero birth, death rates, positive immigration rate.*

Exercise 2.29 *Show that the superposition of independent birth-death processes with the same birth and death rates is again a birth-death process with these rates. What is the corresponding result for a birth-death-immigration process?*

There is a related (strictly speaking nonlinear) process which is almost as simple, combining

- **birth** $X \to X+1$ at rate λX;
- **death** $X \to X-1$ at rate μX;
- **immigration** $X \to X+1$ at rate α;
- **emigration** $X \to X-1$ at rate β if $X > 0$.

Note however that the emigration rate does not depend linearly on X, which leads to complications. An immigration-emigration process will be important in the next section when we study queues. It is also worth noting that *these* processes are themselves a sub-class of a further kind of process.

Definition 2.30 ***(Generalized birth-death processes):*** *These are Markov chains obeying SA I, SA II, SA III, for which the state-space is* $\{0, 1, 2, 3, \ldots\}$*, and the only transitions allowed are* ± 1*, but no further restrictions are placed on the rates.*

Note that a particular example of this kind of process is given by the generalized birth process defined in Exercise 2.15.

3. Queueing theory

This section introduces the *vast* area of queueing theory. A queue is made up of several components:

- arrival process (*eg*: Poisson process),
- service time distribution (*eg*: exponential distribution),
- number of servers (*eg*: 1),
- waiting capacity (*eg*: infinite),
- queue discipline (*eg*: first-in-first-out, abbreviated to FIFO).

A common notation encodes arrival, service, and server characteristics by *eg* $M/M/1$ for Markov input (Poisson process), Markov service time (exponential distribution), and 1 server.

We shall deal with three representative examples:

- $M/M/1$ (the easiest case),
- $M/M/c_1/c_2$ (c_1 servers, capacity c_2),
- $M/G/1$ (general service time distribution).

We shall omit discussion of the modern research area of queueing networks. Here are the basic questions which we will address:

- Equilibrium behaviour:
  - is there a limiting equilibrium?
  - what is it?
- Measures of effectiveness:
  - what is the mean queue length?
  - what is the probability of an empty queue?
  - what is the mean waiting time?

  We shall be able to deal with all of the above.
- Transient behaviour:
  - on start-up,
  - more generally, when conditions change.

  These are hard; we shall leave them alone!

3.1. The Markov single-server queue ($M/M/1$)

The simplest queue is $M/M/1$, defined as follows:

Definition 3.1 ***($M/M/1$ queue):*** *People arrive according to a Poisson(λ) process, are served in order and service times are independent Exponential(μ) random variables.*

The following theorem follows the pattern established in Section 2, and its proof is left as an exercise.

Theorem 3.2 *The number N_t in the system of a $M/M/1$ queue is a Markov chain, in fact an immigration-emigration process.*

Proof: The proof follows the usual pattern: establish Markov property using memoryless property of exponential distribution and Theorem 2.4; assume SA I, SA II, SA III ; show that the probability of more than one arrival-service incident before time t is $o(t)$; compute non-zero rates. Here

$$\left.\begin{array}{rcl} -q_{n,n} &=& \lambda+\mu \\ -q_{0,0} &=& \lambda \\ q_{n,n+1} &=& \lambda \\ q_{n+1,n} &=& \mu \end{array}\right\} \text{ if } n>0\,,$$

□

The forward differential equations are

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}t}p_t(n,m) &= -(\lambda+\mu)p_t(n,m) \\ &\quad +\lambda p_t(n,m-1)+\mu p_t(n,m+1) \end{aligned}$$

(for $m=1,2,\ldots$)

$$\frac{\mathrm{d}}{\mathrm{d}t}p_t(n,0) = -\lambda p_t(n,0)+\mu p_t(n,1)$$

with

$$p_0(n,n) = 1; \qquad p_0(n,m) = 0 \text{ for } m \neq n\,.$$

Note there is just one communicating class (if λ, $\mu > 0$).

We can solve the equilibrium equations in the same way as for the birth-death-immigration process. The equilibrium equations are:

$$\begin{aligned} 0 &= -(\lambda+\mu)\pi_m + \lambda\pi_{m-1} + \mu\pi_{m+1} \qquad (\text{for } m > 0) \\ 0 &= -\lambda\pi_0 + \mu\pi_1 \end{aligned}$$

and they are solved (together with $\pi_0 + \pi_1 + \pi_2 + \ldots = 1$) when $\lambda < \mu$ by

$$\pi_m = (1-\rho)\rho^m$$

where $\rho = \lambda/\mu$ is the *utilization factor*.

Exercise 3.3 *When $\lambda \geq \mu$ (arrival rate exceeds service rate) then show that there can be no equilibrium. Spot an imbedded random walk with reflecting barrier, and hence use ST202 theory to show*

(a) if $\lambda = \mu$ then the queue-length will keep revisiting 0;

(b) if $\lambda > \mu$ then the queue-length will eventually visit zero for the last time and then never revisit it again.

Solving the equilibrium equations is not a totally pleasant experience. In this and many other Markov chains it is often easier to solve a simpler set of equations.

Theorem 3.4 ***(The equations of detailed balance):*** *Consider the equations balancing the "probability flux" across edges connecting each pair of states:*

$$\pi_m q_{m,n} \quad = \quad \pi_n q_{n,m}$$

the equations of detailed balance. If we can find a solution

$$\{\pi_n : n \text{ runs through all states } \}$$

to these equations for which $\sum \pi_n = 1$ *then this also solves the equilibrium equations*

$$0 \quad = \quad \sum_m \pi_m q_{m,n} \, .$$

Proof: Suppose the π_n solve the equations of detailed balance. Then

$$\sum_m \pi_m q_{m,n} \quad = \quad \sum_m \pi_n q_{n,m} \quad = \quad \pi_n \sum_m q_{n,m} = 0$$

by SA III. □

This means it is always worth trying to solve the equations of detailed balance (which are quite easy). If you get a solution then it solves the equilibrium equations and if $\pi_0 + \pi_1 + \pi_2 + \ldots = 1$ then it is an equilibrium distribution. If not, little is lost!

Exercise 3.5 *In fact the equations of detailed balance (together with* $\pi_0 + \pi_1 + \pi_2 + \ldots = 1$*) can be solved exactly when the underlying Markov chain* X*, started*

with the equilibrium distribution $\{\pi_n\}$*, is* reversible*: that is,* $\{X_t : t \in [0,T]\}$ *is statistically the same as its reversal* $\{X_{T-t} : t \in [0,T]\}$*. One way round this is easy:*

$$\begin{aligned} \pi_m q_{m,n} &= ((d/dt)\,\mathbb{P}\,[\,X_0 = m, X_T = n\,])_{t=0} \\ &= ((d/dt)\,\mathbb{P}\,[\,X_0 = n, X_T = m\,])_{t=0} = \pi_n q_{n,m}\,. \end{aligned}$$

The other way requires more detailed analysis (also ideas from Section 4) so we leave it.

Exercise 3.6 *See for yourself how easy it is to solve the equations of detailed balance for the* $M/M/1$ *queue.*

Exercise 3.7 ***(Tandem queues and reversibility):*** *Consider one* $M/M/1$ *queue feeding immediately into another queue which also has exponential service times. Reversibility makes it particularly easy to analyze this situation. For in equilibrium the first queue looks the same when reversed in time. This means the* departures *occur at instants of a Poisson process (since they are* arrivals *in reversed time!). This is known as* Burke's theorem*. Now departures from the first queue are arrivals at the second queue, so we can therefore use* $M/M/1$ *theory to analyze this second queue too.*

Exercise 3.8 *Consider the Markov chain on* 3 *states with* $\underline{\underline{Q}}$*-matrix defined by* $q_{1,2} = 1$*,* $q_{2,3} = 1$*,* $q_{3,1} = 1$*, all other non-diagonal entries zero. Show that the equations of detailed balance have no solution in this case. (Assume that SA I, SA II, SA III hold!)*

Exercise 3.9 *Consider the Markov chain on 3 states with $\underline{\underline{Q}}$-matrix defined by all non-diagonal entries being set equal to 1. Show that the equations of detailed balance have a solution in this case. (Assume that SA I, SA II, SA III hold!)*

Remark 3.10 *Notice that the property of reversibility carries over easily to Markov chains which are restrictions of an original Markov chain. Suppose X is a reversible Markov chain on state space $\{0, \pm 1, \pm 2, \ldots\}$ (all the integers), and suppose for example that we construct a new Markov chain Y on a restricted state space $\{0, 1, 2, \ldots\}$ (for example) simply by forbidding transitions which take the chain out of the new state space. To be specific, we alter the old Q-matrix by deleting all rows and columns corresponding to states not in the new state space, and then alter all the diagonal entries to ensure row sums remain at zero in accordance with SA III.*

If $\underline{\pi}$ is the equilibrium distribution for X then the restriction of $\underline{\pi}$ to the non-negative integers will satisfy the equations of detailed balance for the new chain Y. So we can deduce that the equilibrium probabilities for Y must be proportional to the entries of $\underline{\pi}$ restricted to the new state space.

An example of the above in action can be detected in Erlang's formula, Theorem 3.15.

We now consider measures of effectiveness for the $M/M/1$ queue. Suppose it is in equilibrium, so $\rho = \lambda/\mu$ is smaller than 1. Let N be the number in the system at equilibrium, while N_q is the number in the queue (excludes the one being served). We find directly:

- Mean number in system: $\mathbb{E}[N] = \rho/(1-\rho)$;
- Probability of empty queue: $\mathbb{P}[N = 0] = 1 - \rho$;
- Mean number in queue: $\mathbb{E}[N_q] = \rho^2/(1-\rho)$.

What about the waiting time distribution? (It is important here that we have FIFO queue discipline.) Let W be the time to wait suffered by a new arrival at equilibrium; let W_q be the time this individual spends in the queue. (This person is described as a virtual arrival: someone arriving at a queue in equilibrium. In some sense their experience is a typical one.)

Then

$$\mathbb{P}[W_q = 0] = \mathbb{P}[N = 0] (= \pi_0) = 1 - \rho\,,$$

while conditional on N with $N > 0$ we have that W_q is the sum of $N = N_q + 1$ independent identically distributed service times, each distributed as Exponential(μ). We can argue using moment generating functions to prove that (since $N_q = N - 1$ conditional on $N > 0$ will be geometrically distributed as for N above, by a discrete memoryless property for geometric distributions) conditional on $N > 0$ we have that W_q is distributed as Exponential($\mu(1-\rho)$).

Hence the waiting time distribution:

$$\mathbb{P}[W_q \leq t] \quad = \quad 1 - \rho \exp(-\mu(1-\rho)t)$$

for $t > 0$ so we get the mean waiting time in the queue:

$$\mathbb{E}[W_q] = \frac{\rho}{\mu(1-\rho)}\,.$$

Similarly we get the mean waiting time in the system:

$$\mathbb{E}[W] = \frac{1}{\mu} + \frac{\rho}{\mu(1-\rho)} = \frac{1}{\mu(1-\rho)}.$$

Here are two important relationships which generalize (for example) to the $M/G/1$ queue. Check them!

Exercise 3.11 ***(Little's formula):***

$$\mathbb{E}[N] \quad = \quad \lambda\,\mathbb{E}[W].$$

Exercise 3.12 ***(Pollaczek-Khintchine formula):***

$$\mathbb{E}[W_q] \quad = \quad \lambda(\textbf{\textit{Var}}[S] + \mathbb{E}[S]^2)/(2(1-\rho))$$

where S is a typical service time.

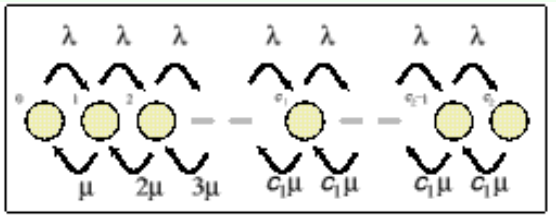
3.2. Multiserver queues ($M/M/c_1/c_2$)

Now suppose there are c_1 servers but room for only c_2 customers in the system: customers who arrive when there is no room are turned away and never return. Interest focusses on the probability of a customer being turned away when the queue is in statistical equilibrium.

Example 3.13 ***(Model of a telephone exchange):*** *Calls arrive according to a Poisson process of rate λ, there are c servers, and calls are lost if they arrive when all servers are occupied. Services, arrivals are independent.*

Definition 3.14 ***($M/M/c_1/c_2$ queue):*** *Customers arrive according to a Poisson process of rate λ, there are c_1 servers each serving at rate μ, (services, arrivals are independent) the system is restricted to a capacity at most c_2, and customers are lost to the system if they arrive when the system is full.*

Following arguments which are now familiar, in an $M/M/c_1/c_2$ queue the number of customers in the system is a Markov chain, with $\underline{\underline{Q}}$-matrix as summarized in the diagram.



Now we apply the equations of detailed balance. We see

$$\begin{aligned} \pi_i &= \lambda/(i\mu)\pi_{i-1} \quad \text{for } i = 1, 2, \ldots c_1 \\ \pi_i &= \lambda/(c_1\mu)\pi_{i-1} \quad \text{for } i = c_1 + 1, \ldots c_2 \end{aligned}$$

and we can solve these together with $\sum \pi_i = 1$. In particular for $c_1 = c_2 = c$ we obtain the following result.

Theorem 3.15 ***(Erlang's formula):*** *For an $M/M/c/c$ queue, service rate μ, arrival rate λ, we find the equilibrium probability of being turned away is*

$$\pi_c = ((\lambda/\mu)^c/c!) \times \pi_0$$

$$= \frac{(\lambda/\mu)^c/c!}{1+\lambda/\mu+\ldots+(\lambda/\mu)^c/c!}$$

Remarkably, this generalizes to $M/G/c/c$ for which service times are independent and identically distributed but no longer exponential. See Appendix A.2 for a sketch of how this works.

Exercise 3.16 *On the TODAY programme (Radio 4, 7/2/95) there was a discussion of provision of intensive care beds by the NHS. Suppose a ward has c beds, the arrival rate is λ, time in bed is Exponential(μ), and the usual independence assumptions obtain. We have an $M/M/c/c$ queue, and in equilibrium (always attained) if $\rho = \lambda/\mu$ then*

$$\pi_r = \pi_0\rho^r/r!$$

hence we deduce the probability of turning a patient away is π_c. Notice we can also calculate an efficiency measure: expected number of beds in use is

$$\begin{aligned}\sum_{r=0}^{c} r\pi_r &= \sum_{r=0}^{c} r\pi_0\rho^r/r! = \sum_{r=1}^{c} \pi_0\rho^r/(r-1)! \\ &= \sum_{r=0}^{c-1} \rho\pi_r = \rho(1-\pi_c)\end{aligned}$$

We can also discuss whether one big ward ($2c$ beds, input rate 2λ) is better than two small wards (c beds each, input rate λ each). (Note implicit use of the Poisson process superposition theorem, Theorem 2.6!).

Finally, we can discuss the effect of two degrees of care: intensive care, and a less expensive continuous care for less critical cases. This leads to a Markov chain (X, Y) where X counts intensive beds occupied, Y counts less intensive beds occupied. Detailed balance:

$$\begin{aligned} p\lambda\pi_{m,n} &= (m+1)\mu\pi_{m+1,n} \\ (1-p)\lambda\pi_{m,n} &= (n+1)\mu\pi_{m,n+1} \end{aligned}$$

assuming intensive cases arrive at rate $p\lambda$, less intensive cases at rate $(1-p)\lambda$, both served at rate μ. If there are c_1 intensive care, c_2 continuous care beds, then we can find the two probabilities of being turned away. Detailed balance breaks down if less intensive cases can occupy available intensive care beds, but is restored by using a different state-space ....
It should be clear from this how useful detailed balance can be.

Here is another exercise to help you see how detailed balance helps one to do crucial calculations rather easily.

Exercise 3.17 ***(Showers at Stanford University):*** *In Stanford University there is a sports centre containing six showers. Each of these showers independently has probability $\frac{1}{2}$ of being operational. Knowing this, someone enters the shower area to discover that five out of the six showers are occupied. What should be their posterior probability of the unoccupied shower being operational?*

Answer: As is common in applied probability, how one answers this depends on what further assumptions one is prepared to make. Suppose we assume that people

are served by the k operational showers according to an $M/M/k/k$ queue, arrival rate λ, service rate μ, and also assume that equilibrium has been attained. (The $M/M/k/k$ assumption implicitly carries the additional assumption that people arriving to find no available showers do not wait for one to become available, but go elsewhere.) Clearly k must equal either 5 or 6.

Without loss of generality we suppose the occupied showers to be numbers 1, 2,..., 5, and the unoccupied shower to be number 6.

We can use Erlang's formula (Theorem 3.15) to deduce that

$$\begin{aligned} &\text{if } k = 6 \text{ then} \\ \mathbb{P}[\text{ showers 1, 2,..., 5 occupied }] &= \frac{\rho^5/5!}{1 + \ldots + \rho^6/6!} \times \frac{1}{6}, \\ &\text{if } k = 5 \text{ then} \\ \mathbb{P}[\text{ showers 1, 2,..., 5 occupied }] &= \frac{\rho^5/5!}{1 + \ldots + \rho^5/5!}, \end{aligned}$$

using $\rho = \lambda/\mu$ to denote the utilization factor. Accordingly the posterior probability that in fact $k = 6$ is given by

$$\frac{2^{-6} \times \frac{\rho^5/5!}{1+\ldots+\rho^6/6!} \times \frac{1}{6}}{2^{-6} \times \frac{\rho^5/5!}{1+\ldots+\rho^6/6!} \times \frac{1}{6} + 2^{-6} \times \frac{\rho^5/5!}{1+\ldots+\rho^5/5!}} = \frac{1}{1 + 6 \times \frac{1+\ldots+\rho^6/6!}{1+\ldots+\rho^5/5!}}.$$

So we see that the posterior probability can range from a maximum of $1/7$ in case of zero utilization factor $\rho = 0$, down to 0 in case of high utilization factor $\rho = \infty$.

The case of high utilization factor accords with intuition (if the showers are heavily used then it is most unlikely that an operational shower would be left vacant), while the case of zero utilization factor agrees with a simpler analysis assuming only that there were five people before you, who have chosen showers at random from operational showers. □

3.3. Queues with general service-times ($M/G/1$)

The number of customers in the system in an $M/G/1$ queue is no longer Markov! (Think of an extreme case where each service lasts one unit of time. Then you can predict the time of the next departure exactly if you know not only that someone is being served now but also when the last departure and arrival took place.)

Definition 3.18 ***($M/G/1$ queue):*** *Customers arrive according to a Poisson process of rate λ, there is a single server whose service time has a specified distribution (not exponential!), and services, arrivals are independent.*

The cure is to observe the embedded Markov chain.

Definition 3.19 ***(The embedded Markov chain):*** *Let X_n be the number of people in the system just after the departure of customer n. Then $\{X_1, X_2, \ldots\}$ is a discrete-time Markov chain.*

Note the Markov property for X follows rapidly from Theorem 2.4. Note also that

$$X_{n+1} \quad = \quad X_n - 1 + A_{n+1} \text{ if } X_n \geq 1$$

$$X_{n+1} = X_n + A_{n+1} = A_{n+1} \text{ if } X_n = 0$$

where A_n is the number of customers to arrive during the service of customer n . Let

$$\begin{aligned} k_n &= \mathbb{P}\left[\, n \text{ customers arriving during } S \,\right] \\ &= \mathbb{E}\left[\, \mathbb{P}\left[\, n \text{ customers arriving during } S | S \,\right]\,\right] \end{aligned}$$

(see Appendix A.5 for details of how to compute with such conditional probabilities and expectations) so that

$$\sum_n n k_n = \lambda\, \mathbb{E}\left[\, S \,\right].$$

Then the stochastic matrix $\mathbb{P}$ of transition probabilities for the embedded chain is as follows:

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

If this transition matrix is aperiodic and irreducible (no periodic cycles such as $1 \to 2 \to 1 \to 2 \dots$, can get anywhere from anywhere else) then it is sensible to look to see if it has an equilibrium distribution. Irreducibility and aperiodicity follow if the mean number of arrivals in a typical service, $K'(1)$, is positive.

Then an equilibrium distribution, if it exists, will solve $\underline{\pi} = \underline{\pi}\,\underline{\underline{P}}$, leading to

$$\pi_i \quad = \quad \pi_0 k_i + \sum_{j=1}^{i+1} \pi_j k_{i-j+1}$$

which is more conveniently expressed using generating functions:

Exercise 3.20

$$\Pi(z) \quad = \quad \Pi(0)(1-z)K(z)/(K(z)-z)$$

where

$$\Pi(z) \quad = \quad \sum_j \pi_j z^j\,, \qquad K(z) \quad = \quad \sum_j k_j z^j$$

Notice that by differentiation we can now find the π_i in terms of π_0, once we know $K(z)$.

If the embedded Markov chain is positive-recurrent then it turns out that the $M/G/1$ queue has a stationary distribution and satisfies Little's formula and the Pollaczek-Khinchine formula. We will show why this is true in the remainder of this section.

Theorem 3.21 ***(Little's formula):*** *Let N, W have the distribution of the number in the system and the waiting time for completion of service, for someone who arrives when the queue is in statistical equilibrium (the so-called virtual arrival). Then*

$$\mathbb{E}\,[\,N\,] \quad = \quad \lambda\,\mathbb{E}\,[\,W\,]\,.$$

We will only sketch the proof of this, and we do so under the assumption (which can of course be proved!) that N, W for the virtual arrival are the same in distribution as for a typical arrival in the Poisson(λ) stream of incoming customers. (We will make the same assumption below while proving the *Pollaczek-Khintchine formula*.)

Proof: Consider the point at which someone leaves the queue. They leave behind N people, mean $\mathbb{E}[N]$. On the other hand they had to wait time W before their service was complete. During that time, conditional on the value of W, there arrived Poisson(λW) people (since arrivals, services are independent, and arrivals before and after this person are independent by Theorem 2.4). Hence the number of people who arrive while they wait has mean $\mathbb{E}[Poisson(\lambda W)] = \lambda \mathbb{E}[W]$. (See Appendix A.5 for details.) □

(Actually Little's formula can be applied to situations which are much more general than this.)

Exercise 3.22 *Recall that*

$$\Pi(z) \quad = \quad \Pi(0)(1-z)K(z)/(K(z)-z).$$

Show that $\sum_j \pi_j = \Pi(1)$ *can equal* 1 *only if*

$$K'(1) \quad = \quad \sum_j jk_j \quad = \quad \lambda \mathbb{E}[S] < 1\,.$$

(Notice numerator and denominator above both vanish when $z = 1$*. So use l'Hôpital's rule to compute* $\Pi(1)$ *as a ratio of derivatives evaluated at* $z = 1$*.)*

Notice that $\Pi(1) = 1$ *implies*

$$1 - \lambda \mathbb{E}[S] \quad = \quad 1 - K'(1) \quad = \quad \pi_0$$

which is a useful bonus!

Theorem 3.23 ***(Pollaczek-Khintchine formula):*** *Consider an* $M/G/1$ *queue, with arrival rate* λ *and typical service time* S*. Let* $\rho = \lambda \mathbb{E}[S]$ *be the utilization factor (note this agrees with the definition in Section 3.1 for an* $M/M/1$ *queue).Then the expected waiting time in the queue is given by*

$$\mathbb{E}[W_q] \quad = \quad \frac{\lambda(\mathbf{Var}[S] + \mathbb{E}[S]^2)}{2(1-\rho)}.$$

Proof: As for Little's formula (3.21), we prove this for the embedded Markov chain and assume it carries over to the $M/G/1$ queue itself. The proof proceeds via a number of exercises.

Exercise 3.24 *Differentiate the expression for* $\Pi(z)$ *twice and set* $z = 1$ *to obtain an expression for* $\Pi'(1)$ *the mean number* $\mathbb{E}[N]$ *in the system at equilibrium. You will find it easier to deal with*

$$\Pi(z)(K(z) - z) \quad = \quad \Pi(0)(1-z)K(z)$$

and you should get

$$\mathbb{E}[N] \quad = \quad \frac{2\Pi(0)K'(1) + K''(1)}{2(1 - K'(1))}.$$

Exercise 3.25 *Use Little's formula,* $\mathbb{E}[N] = \lambda \mathbb{E}[W]$*, also the facts*

$$\begin{aligned} K'(1) &= \mathbb{E}[Poisson(\lambda S)] = \lambda \mathbb{E}[S], \\ \textbf{\textit{Var}}[Poisson(\lambda S)] &= K''(1) + K'(1) - K'(1)^2, \\ \textbf{\textit{Var}}[Poisson(\lambda S)] &= \mathbb{E}[\textbf{\textit{Var}}[Poisson(\lambda S)|S]] \\ &\quad + \textbf{\textit{Var}}[\mathbb{E}[Poisson(\lambda S)|S]] \\ &= \lambda \mathbb{E}[S] + \lambda^2 \textbf{\textit{Var}}[S] \\ 1 - \lambda \mathbb{E}[S] &= 1 - K'(1) = \pi_0 = \Pi(0) \end{aligned}$$

(the last being as established above in Example 3.22) to establish the Pollaczek-Khintchine formula.

Remember, we only have the Pollaczek-Khintchine formula for the embedded Markov chain. Transferring this to the $M/G/1$ queue requires an argument using the pure maths techniques of *ergodic theory*, which allow us to use time-averages (for which the embedded Markov chain gives information) to give information about statistical averages or expectations (which concern the $M/G/1$ queue). However we do not enter on this here. □

Exercise 3.26 *Jobs arrive at a computer central processing unit (*cpu*) according to a Poisson process and at rate* 0.5 *per unit of time. The* cpu *serves at an average of* 1 *unit of time per job.*

If service times are exponential (thus if we have an $M/M/1$ *queue) then we*

can use the work of Section 3.1 to compute

$$\mathbb{E}\,[\,W_q\,] \quad = \quad 1.$$

Suppose the service time S is not exponential, and has variance σ_S^2. Then the Pollaczek-Khintchine formula tells us

$$\mathbb{E}\,[\,W_q\,] \quad = \quad \frac{1}{2}(1+\sigma_S^2).$$

Thus we can exactly quantify the benefits to be gained from modifying the system to gain greater regularity in the service time. For $\sigma_S^2 = 1$ if the service time is exponential, so if we have benefits attached to lower expected service time and costs attached to arranging greater regularity then this shows how to decide.

4. More Markov properties

In this section we consider the Strong Markov Property, and simulation of continuous time Markov chains.

4.1. The strong Markov property

Recall that the Markov property says, given the present state of the Markov chain then the past is irrelevant to predictions of the chain's future behaviour. But consider the present at a random time, for example the random time at which the chain first leaves its initial state. Does a Markov property still hold? In general the answer is NO:

Example 4.1 *Consider the 2-state switcher Markov chain X, and consider the random time T which is the first time after 1 such that $X(T) = 1$, and X holds at state 1 between T and $T+1$ if $X(T-1) = 1$, or makes a switch to 0 before $T+1$ if $X(T-1) = 0$.*

Example 4.2 *Consider a toaster which takes an exponential(1) length of time S to toast a piece of bread. Then X_t, defined to be 0 if toast not yet toasted by time t, 1 otherwise, is a Markov chain from the memoryless property of the exponential distribution. Consider the random time T defined to be the greatest of $S-1$ and 0.*

The answer is YES as long as the random time does not try to look ahead into the future. In particular, the answer is YES if the random time is a stopping

time as given below. But we need an extra standing assumption to exclude some pathological examples.

Standing Assumption 4 [10] *The sample paths of the Markov chain are right-continuous with left limits, jumping finitely often in each bounded time interval.*

This stops us considering silly examples where the Markov chain twitches around. Now we can define "fair" random times, which don't peek into the future (unlike the examples above).

Definition 4.3 ***(Stopping time):*** *A random time T is a stopping time for a Markov chain $\{X_t : t \geq 0\}$ if for all $t \geq 0$ the event $[T \leq t]$ can be determined from knowledge of the path $\{X_s : s \leq t\}$.*

This allows us to state (but not prove!) the Markov property for this class of random times.

Theorem 4.4 ***(Strong Markov property):*** *Suppose that $\{X_t : t \geq 0\}$ is a Markov chain satisfying SA I, SA II, SA III, IV. Then it possesses the Strong Markov property: the present state at a stopping time T is sufficient to predict the future:*

$$\mathbb{P}\left[\, X_{T+t} = n | X_T = m, \text{ the past } X_s \text{ for } s \leq T \,\right] \quad = \quad p_t(m, n)$$

for any stopping time T.

[10] abbreviated to SA IV

Example 4.5 *Consider the* stopping time T *which is the time of the first jump of* X*. Clearly* $[T \leq t]$ *is determined from the path of* X *up to* t*.*

Conditional on $[T > s]$ the process $\{X_{t+s} : t \geq 0\}$ has the same Q-matrix as the original chain and starts in the same place, so has the same statistical behaviour. So $T - s$ given $[T > s]$ has the same distribution as T. This gives us an important result:

Theorem 4.6 ***(Exponential holding times):*** *Suppose that* $\{X_t : t \geq 0\}$ *is a Markov chain satisfying SA I-IV, with* $X_0 = i$*. Consider the stopping time* T *which is the time of the first jump of* X*. Then* T *has an exponential distribution of rate* q_i*.*

Proof: It is easy to see that T is exponential from the ordinary Markov property:

$$\mathbb{P}\left[T > t + s | T > s\right] \quad = \quad \mathbb{P}\left[T > t\right]$$

hence T is exponential(μ).

Calculating the rate takes more thought. We have

$$\begin{aligned} &\lim_{t\to 0}(1 - \mathbb{P}\left[X_t = i | X_0 = i\right])/t \\ &= \lim_{t\to 0} \mathbb{P}\left[X_t \neq i | X_0 = i\right]/t \\ &\leq \quad \lim_{t\to 0} \mathbb{P}\left[T \leq t\right]/t \end{aligned}$$

so $q_i \geq \mu$. On the other hand it can be shown (though we will not show it!) that

$$\mathbb{P}\left[\, X \text{ jumps more than once before } t | X_0 = i \,\right] \quad = \quad o(t)$$

as a consequence of SA I, SA II, SA III, SA IV. Hence

$$\begin{aligned}
&(\mathbb{P}\left[\, T \leq t \,\right] - \mathbb{P}\left[\, X_t \neq i | X_0 = i \,\right])/t \\
&= \mathbb{P}\left[\, X \text{ jumps but returns to } i \text{ at } t | X_0 = i \,\right]/t \\
&\leq \mathbb{P}\left[\, X \text{ jumps more than once before } t | X_0 = i \,\right]/t \\
&= o(t)/t \quad \rightarrow \quad 0
\end{aligned}$$

as $t \rightarrow 0$. This establishes the result. □

4.2. Simulation

A very practical application of this is to show how to simulate any Markov chain obeying SA I, SA II, SA III, SA IV. Computers produce pseudo-random variates U which can be taken to be independent and uniformly distributed over $[0, 1]$. Basic theory of simulation shows how to produce exponential random variables of mean 1 via $-\log(U)$ (and thus of general mean via multiplication by a constant). Moreover if we want a discrete-time Markov chain (say) which jumps to $i = 1$, 2, ... with probabilities p_1, p_2, ... then we can arrange this by drawing a uniform random variable U and deciding

$$\text{to jump to 1 if } U \quad < \quad p_1 \,,$$

$$\text{to jump to 2 if } p_1 \le U \quad < \quad p_1 + p_2 \,,$$
$$\dots$$

Then the following theorem tells us how to simulate a continuous-time Markov chain (compare constructions of Theorem 2.3 and Theorem 2.13).

Theorem 4.7 ***(Simulation of a Markov chain):*** *Suppose that* $\{X_t : t \ge 0\}$ *is a Markov chain satisfying SA I, SA II, SA III, SA IV, so in particular it does not explode in finite time (SA IV). Then it may be constructed as follows. First build a discrete time Markov chain* $\{Y_n : n = 0, 1, 2, \dots\}$ *with stochastic matrix given by* $p_{ij} = q_{ij}/q_i$ *if* $i \ne j$, $p_{ii} = 0$. *The paths of* X *are given by following the paths of* Y, *but holding at* $Y_n = i$ *for a time of duration* T_n/q_i, *where* $T_1, T_2, \dots$, *are independent and identically distributed exponential random variables of mean* 1, *independent of the jump process* $\{Y_n : n = 0, 1, 2, \dots\}$.

Actually even if the chain can explode in finite time it is still possible to use this construction to build a "minimal" version which can be used to construct examples which solve the Kolmogorov backwards differential equations.

5. Epidemics

Definition 5.1 ***(The cycle of infection):*** *The cycle of events for an individual in an epidemic can be broken down as follows.*

- *Start off as Susceptible,*
- *Exposed to infection, infected, become Latent,*
- *Become Infective,*
- *Exhibit symptoms,*
- *Removal (could be death, could be recovery to be immune).*

The interval between exposure to infection and becoming infectious is the *latent period*.

The interval between exposure to infection and exhibiting symptoms is the *incubation period*.

The interval between becoming infectious and removal is the *infectious period*.

We shall further assume that infection arises from contact with other infectives, and that the whole population *mixes homogeneously*.

Complications (which we won't consider) — carriers, vectors, non-homogeneous mixing, multiple varieties of disease . . .

5.1. Deterministic infection processes

Suppose that the infectious period extends indefinitely (no removals), and that the overall population is large with just one initial infective. We consider an approximate deterministic model.

Let x be the number of susceptibles, y be the number of infectives, so that the total population is $n = x + y$. Then (neglecting stochastic effects)

$$\Delta x = -\alpha x y \Delta t$$

under homogeneous mixing. For this we use the heuristic

$$\Delta x \quad = \quad \mathbb{E}\left[\, \Delta x | \text{Past at time } t \,\right] .$$

We change the time-scale so that $\alpha = 1$. In the limit $\Delta t \to 0$ we get

$$\frac{\mathrm{d}x}{\mathrm{d}t} \quad = \quad -x(n-x)$$

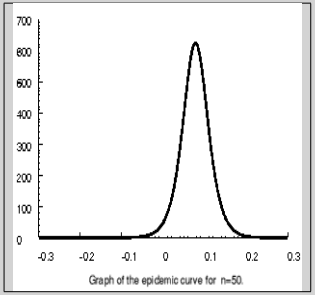
with solution

$$\begin{aligned} x &= \frac{n(n-1)}{n-1+e^{nt}} \\ y &= \frac{n}{1+(n-1)e^{-nt}} \quad = \quad n-x \end{aligned}$$

It is more informative to look at the *epidemic curve*

$$-\frac{\mathrm{d}x}{\mathrm{d}t} \quad = \quad xy \quad = \quad \frac{n^2(n-1)e^{nt}}{(n-1+e^{nt})^2}$$

graphed here for the case of $n = 50$. Note the characteristic peak of the epidemic, at time $t \approx 0.08$, when the rate of infections is at its highest level.



5.2. Epidemics without removals

Obviously a stochastic approach is better, and we can provide one! Let X_t, Y_t be the numbers of susceptibles and infectives at time t. Then $X_t + Y_t = n$ and $\{Y_t : t \geq 0\}$ is a Markov chain with rates

$$Y \to Y + 1 \qquad \text{at rate } \alpha XY = \alpha Y(n - Y)$$

while $Y < n$. Again a time change permits us to put $\alpha = 1$. This is called the *simple epidemic* or *epidemic without removals*.

How long will it be before everyone is infected? (Note that this question is unanswerable from the deterministic point of view.) The answer is a sum of independent but not identically distributed exponential times (use Theorem 4.7) with

expectation approximately $2\log n/(\alpha n)$ (for large n with α fixed). It is also possible to compute the limiting distribution of $\alpha n T - 2\log n$ using moment generating functions.

The approximate expectation follows by reasoning thus. The time T_y for $y \to y+1$ is exponential of rate $\alpha y(n-y)$ so has mean $(\alpha y(n-y))^{-1}$. Hence the expected total time is

$$\begin{aligned}\mathbb{E}\left[T_1 + \ldots + T_{n-1}\right] &= \sum_y (\alpha y(n-y))^{-1} \\ &= (\alpha n)^{-1} \sum_y (y^{-1} + (n-y)^{-1}) \\ &\approx 2\log n/(\alpha n)\end{aligned}$$

using $1 + 1/2 + \ldots + 1/n \approx \log(n+1)$ for large n.

We can even derive information about the exact distribution, using the discussion about sums of exponential random variables of different means to be found in Section 2.2 (the pure birth process as defined in Definition 2.10). Suppose that m is the greatest integer no larger than $n/2$: we write $m = [n/2]$. Then the exponential random variables $T_1, \ldots, T_m$ all have different means. Hence by the work in Section 2.2.

$$\mathbb{P}\left[T_1 + \ldots + T_m > t\right] = \sum_y \left(\prod_{j \neq y} \frac{\lambda_j}{\lambda_j - \lambda_y} \right) \exp(-\lambda_y t)$$

We can think of $T_1 + \ldots + T_m$ as the "half-life" of the epidemic.

5.3. Epidemics with removals

Now we allow for removals: infectives can be removed from the population (either by dying, or by recovering to become immune). There is a deterministic version, but we go straight to the stochastic version. This is called the *general epidemic* or *epidemic with removals*[11]. We take a Markov chain model. Let (S, I, R) be a Markov chain with state-space $\{0, 1, 2, \ldots\}^3$. Assuming SA I, SA II, SA III, SA IV, the chain is defined by the rates

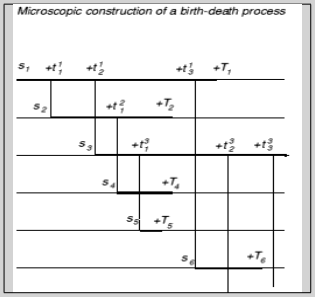
$$\begin{aligned} S \to S-1,\ I \to I+1 \qquad & \text{at rate } \alpha SI\,, \\ I \to I-1,\ R \to R+1 \qquad & \text{at rate } \beta I \end{aligned}$$

Notice that $S+I+R$ remains fixed: we set $S+I+R=n$ where n is the total population size. Notice too that the rate of infection, αSI, implies homogeneous mixing: the rate of removal, βI, more or less implies an exponential infectious period. Homogeneous mixing is a drastic assumption which should be challenged (eg by comparing theoretical results with simulation studies): infectious period assumptions are typically less serious, at least for issues of whether an initially small epidemic may become widespread. Finally notice that I behaves rather like a birth-death process, but is not Markov on its own. However (I, S) is Markov in its own right, since we can then compute the number of removals as $R = n-I-S$. The question we shall address in the remainder of this section is, how to relate I to a birth-death process in a useful way? Particularly, what can we say about the

[11] Or sometimes the $S-I-R$ *epidemic*.

probability of the epidemic becoming widespread? We answer this by drawing a comparison with birth-death processes: we make a simultaneous "microscopic" construction of a birth-death process and an epidemic with removals in such a way that we can get bounds on the above probability.

Here is the first part of the construction. We build a birth-death process, in a way which is a slight variation on that in Section 2.3.



Suppose we have an infinite population, each individual i represented by a Poisson(α) process $\{t_1^i, t_2^i, t_3^i, \ldots\}$ which controls the infections they make and an exponential(β) random time T_i which indicates their removal. The figure illustrates the construction.

Initially the infected set of individuals are $\{1, 2, \ldots, i_0\}$ and the remaining individuals are $\{i_0, i_0 + 1, \ldots\}$. Infections and removals follow the rules:

If i is infected at time s_i then

- remove i at time $s_i + T_i$;
- at times $s_i + t_j^i$ previous to removal time use i to infect the individual with lowest number who is neither infected nor removed by time $s_i + t_j^i$.

All Poisson processes and exponential random variables are independent of each other. Set X_t to be the number of infectives at time t, and Y_t to be the number of removals at time t.

Theorem 5.2 ***(Identifying the birth-death process):*** *The random process* $\{X_t : t \geq 0\}$ *constructed above is a birth-death process of birth rate* α *and death rate* β.

Proof: We have already considered a very similar construction in Section 2.2, except that here the Poisson processes begin at the infection times not at time 0.

The Markov property follows as usual, using Theorem 2.4 and the memoryless property of the exponential distribution. Assuming SA I, SA II, SA III, SA IV, it remains for us to check the rates. First we show we can ignore the chance of two or more birth or death incidents (that is, infections or removals) in a small time interval $[0, t]$.

$$\begin{aligned}
&\mathbb{P}[\text{ More than one birth before } t | X_0 = x\,] \\
&\leq \mathbb{P}[\,2 \text{ plus incidents in } x+1 \text{ Poisson}(\alpha) \text{ processes in } [0,t]\,] \\
&= 1 - \exp(-\alpha(x+1)t) - \alpha(x+1)t\exp(-\alpha(x+1)t) \\
&= o(t) \qquad \textit{eg}\text{: by differentiation or by power series}
\end{aligned}$$

$$\begin{aligned}
&\mathbb{P}[\,1 \text{ birth and some deaths before } t | X_0 = x\,] \\
&\leq \mathbb{P}[\,1 \text{ incident in } x \text{ Poisson}(\alpha) \text{ processes in } [0,t]\,]
\end{aligned}$$

$$\begin{aligned} &\times \mathbb{P}\,[\,1 \text{ plus } x+1 \text{ Exponential}(\beta) \text{ times in } [0,t]\,] \\ &= \alpha x t \exp(-\alpha x t)(1-\exp(-(x+1)\beta t)) \\ &= o(t) \qquad \textit{eg}\text{: by differentiation or by power series} \end{aligned}$$

$$\begin{aligned} &\mathbb{P}\,[\text{ No births, more than 1 death before } t | X_0 = x\,] \\ &\leq \mathbb{P}\,[\,2 \text{ plus } x \text{ Exponential}(\beta) \text{ times in } [0,t]\,] \\ &= 1-\exp(-x\beta t) - x(1-\exp(-\beta t))\exp(-(x-1)\beta t) \\ &\qquad \text{(by Binomial distribution!)} \\ &= o(t) \qquad \textit{eg}\text{: by differentiation or by power series} \end{aligned}$$

We deduce

$$\mathbb{P}\,[\text{ more than one birth or death incident } | X_0 = x\,] = o(t)\,.$$

Now the rates can be computed rather easily:

$$q_{x,y} = 0 \quad \text{if } |x-y| > 1\,;$$

$$\begin{aligned} q_{x,x} &= \lim_{t\to 0}(\mathbb{P}\,[\,X_t = x | X_0 = x\,] - 1)/t \\ &= \lim_{t\to 0}(\mathbb{P}\,[\text{ no births or deaths by } [0,t] | X_0 = x\,] - 1)/t \\ &\qquad \text{(using above to simplify the event!)} \\ &= \lim_{t\to 0}(\exp(-x\alpha t)\exp(-x\beta t) - 1)/t \;=\; -(\alpha+\beta)x \end{aligned}$$

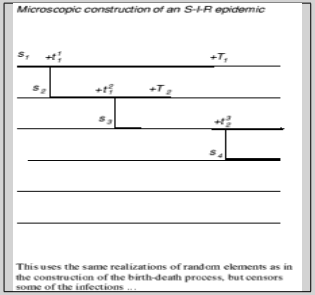
$$
\begin{aligned}
q_{x,x+1} &= \lim_{t\to 0} \mathbb{P}\left[\, X_t = x+1 | X_0 = x \,\right]/t \\
&= \lim_{t\to 0} \mathbb{P}\left[\, \text{ one birth, no deaths by } [0,t] | X_0 = x \,\right]/t \\
&\qquad \text{(using above to simplify the event!)} \\
&= \lim_{t\to 0} (x\alpha t \exp(-x\alpha t)\exp(-x\beta t))/t \;=\; \alpha x
\end{aligned}
$$

$$
\begin{aligned}
q_{x,x-1} &= \lim_{t\to 0} \mathbb{P}\left[\, X_t = x-1 | X_0 = x \,\right]/t \\
&= \lim_{t\to 0} \mathbb{P}\left[\, \text{ no births, one death by } [0,t] | X_0 = x \,\right]/t \\
&\qquad \text{(using above to simplify the event!)} \\
&= \lim_{t\to 0} (\exp(-x\alpha t) x\beta t \exp(-x\beta t))/t \;=\; \beta x
\end{aligned}
$$

This establishes the result. □

The second part of the construction needs to reduce the infection rate to allow for there being fewer people to infect as the epidemic progresses. We do this by attaching to each infection incident t_j^i a "censoring" random variable U_j^i uniformly distributed over $[0,1]$, independent of each other and of the Poisson processes and Exponential random variables. The new recipe is as follows:

Initially the infected set of individuals are $\{1, 2, \ldots, i_0\}$ and the remaining individuals are



$\{i_0+1,\dots n\}$ (there will never be an infection beyond that of individual n). Infections and removals follow the rules:

If i is infected at time s_i then

- remove i at time $s_i + T_i$;
- at times $s_i + t^i_j$ previous to removal time use i to infect the individual with lowest number who is neither infected nor removed by time $s_i+t^i_j$. *However* this infection is allowed to proceed only if we have $(n-i_0)U^i_j < S(s_i + t^i_j)$, which is an event of conditional probability $S(s_i + t^i_j)/(n-i_0)$ when conditioned on past of S, I, R at t.

Consequently the infection times s_i will be different from those in the previous construction. We set

- S_t to be the number of individuals in $\{1,\dots,n\}$ who are neither infected nor removed by time t;
- I_t to be the number of individuals in $\{1,\dots,n\}$ who are infected but not yet removed by time t;
- R_t to be the number of individuals in $\{1,\dots,n\}$ who are removed by time t.

Clearly $S + I + R = n$ for all time. Note also that once S reaches 0 then all infections *stop*. The construction is illustrated in the figure, which is based on the Poisson processes and Exponential distributions used in the birth-death process construction above.

Theorem 5.3 ***(Identifying the epidemic with removals):*** *This (S, I, R) is an epidemic with removals, with rates*

$$\begin{aligned} S \to S-1,\ I \to I+1 \qquad & \text{at rate } \alpha SI/(n-i_0)\,, \\ I \to I-1,\ R \to R+1 \qquad & \text{at rate } \beta I\,. \end{aligned}$$

Proof: The Markov property is essentially as before, except that we also use the independence of the "censoring" random variables U_j^i . Similarly the proof of

$$\mathbb{P}\,[\text{ More than 1 infection/removal incident by } t | S_0, I_0, R_0\,] = o(t)$$

is essentially the same (after all, we are *reducing* the number of such incidents by the censoring mechanism!).

Transition rate computations are as follows. First consider removal: ($I \to I-1, R \to R+1$). The rate is:

$$\begin{aligned} &\lim_{t\to 0} \mathbb{P}\,[\,I_{t_0+t} = i-1, R_{t_0+t} = r+1 | S_{t_0} = s, I_{t_0} = i, R_{t_0} = r\,]\,/t \\ &= \lim_{t\to 0} \mathbb{P}\big[\text{ no infections (potential or otherwise),} \\ &\qquad \text{one removal in } [t_0, t_0+t] | S_{t_0} = s, I_{t_0} = i, R_{t_0} = r\big]/t \end{aligned}$$

$$= \lim_{t\to 0}\left[\begin{array}{c}\exp(-\alpha i t)\\ \text{(no infections)}\end{array}\right]\times$$
$$\times\left[\begin{array}{c} i(1-\exp(-\beta t))\exp(-(i-1)\beta t)\\ \text{(one removal — use Binomial probability)}\end{array}\right]/t$$
$$= \quad i\beta\,.$$

The rate of infection ($S \to S-1$, $I \to I+1$) is:

$$\lim_{t\to 0}\mathbb{P}\left[\,S_{t_0+t}=s-1, I_{t_0+t}=i+1 | S_{t_0}=s, I_{t_0}=i, R_{t_0}=r\,\right]/t$$
$$= \lim_{t\to 0}\mathbb{P}\Big[\text{ one infection (which is actual),}$$
$$\text{and no removals in } [t_0,t_0+t] | S_{t_0}=s, I_{t_0}=i, R_{t_0}=r\Big]/t$$

$$= \lim_{t\to 0}\left[\begin{array}{c}\alpha i t\exp(-(i-1)\alpha t)\\ \text{(one Poisson infection)}\end{array}\right]\times$$
$$\times\left[\begin{array}{c}(s/s_0)\\ \text{(which is actual)}\end{array}\right]\times$$
$$\times\left[\begin{array}{c}\exp(-\beta i t)\\ \text{(no infections)}\end{array}\right]/t$$
$$= \alpha i s/s_0 \quad = \quad \alpha i s/(n-i_0)\,.$$

All other rates are zero save for the rate of standing still ($S\to S, I\to I, R\to R$) which is given by SA I, SA II. □

We have seen how to build both an epidemic with removals and a birth-death process using essentially the same sources of randomness (this is called coupling), and it is reasonably intuitive that the total number ever infected by the epidemic $(I_t + R_t)$ should be less than the total number ever infected by the birth-death process $(X_t + Y_t)$. But we need to prove this.

Theorem 5.4 ***(Comparison between the two processes):*** *Suppose that i_0 individuals are alive at time 0 in the birth-death process, and the same i_0 individuals are infectious at time 0 in the epidemic. Suppose that (X, Y), (S, I, R) are constructed as in Theorem 5.3 and Theorem 5.4. For all time we have*

$$I_t + R_t \quad \leq \quad X_t + Y_t \, .$$

Proof: The idea is to look at the times at which individuals get infected. Let

$$\begin{aligned} H_i^{epi} &= \text{time of infection of } i \text{ in epidemic} \, , \\ H_i^{bd} &= \text{time of infection of } i \text{ in birth-death process.} \end{aligned}$$

The result follows if $H_i^{epi} \geq H_i^{bd}$ for all i . This is certainly true for $i = 1, \ldots, i_0$. We establish it for all i by induction.

Suppose it is true for $i = 1, \ldots, j-1$. Set

$$\begin{aligned} M_{j-1}^{epi}(t) \quad = \quad & \text{number directly infected by } 1, \ldots, j-1 \\ & \text{in epidemic by time } t \\ & \text{(including original infectives)} \, , \end{aligned}$$

$$M_{j-1}^{bd}(t) \quad = \quad \text{number directly infected by } 1,\ldots,j-1 \text{ in birth-death process by time } t \text{ (including the original infectives)}.$$

Then by the inductive hypothesis we know

$$M_{j-1}^{epi}(t) \quad \leq \quad M_{j-1}^{bd}(t) \qquad \text{for all } t\,,$$

since by induction $H_i^{epi} \geq H_i^{bd}$ for all $i < j$ and subsequent birth-death infections by such i must include subsequent epidemic infections for the same i. So at any time t the first $1, \ldots, j-1$ individuals will have infected more people in the birth-death process than in the epidemic. But

$$\begin{aligned} H_j^{epi} &= \inf\{t : M_{j-1}^{epi}(t) = j\}\,, \\ H_j^{bd} &= \inf\{t : M_{j-1}^{bd}(t) = j\}\,, \end{aligned}$$

and so

$$H_j^{epi} \quad \geq \quad H_j^{bd}$$

and the result follows by induction. □

Example 5.5 *Suppose we say there is an* epidemic *if the total number of infected exceeds a predetermined threshold γn, for $\gamma \in (0,1)$. Consider that before the threshold is exceeded we have $I + R \leq \gamma n$. From Theorem 5.4 we deduce:*

if $X + Y$ never exceeds γn
then $I + R$ never exceeds γn

and so in the limit we have

$$\lim_{n\to\infty} \mathbb{P}[\ \textit{no epidemic}\ ] \geq \mathbb{P}[\ \textit{birth-death process dies out}\ ] = \min\left\{\left(\frac{\beta}{\alpha}\right)^{i_0}, 1\right\}$$

(computing eg *from embedded random walk or embedded branching process in birth-death process).*

Actually we can do better than this and get a lower bound on the probability of an epidemic as well! Before the threshold we have a bound on the infection rate *via* $S \geq (1-\gamma)n$:

$$\alpha IS/(n - i_0) \geq (1-\gamma)\alpha n I/(n - i_0)$$

and by increasing the censoring we may compare $I + R$ to $X^* + Y^*$ derived from a new birth-death process of birth-rate $(1-\gamma)\alpha n/(n - i_0)$ and death-rate β, in a manner similar to that of Theorem 5.4. Hence we may argue as in the Example 5.5 to obtain the following.

Theorem 5.6 ***(Whittle's threshold theorem):*** *Suppose that S, I, R together form an epidemic with removals such that*

$$\begin{aligned} S \to S-1,\ I \to I+1 \qquad & \text{at rate } \alpha SI/(n-i_0)\,, \\ I \to I-1,\ R \to R+1 \qquad & \text{at rate } \beta I\,. \end{aligned}$$

(Here $i_0 = I_0$ is the initial number of infectives.) Say there is an epidemic if the total proportion infected exceeds a predetermined threshold γn, for $\gamma \in (0,1)$. Then

$$\begin{aligned} \min\left\{\left(\frac{\beta}{\alpha}\right)^{i_0}, 1\right\} &\le \lim_{n\to\infty} \mathbb{P}[\ \textit{no epidemic}\ ] \\ &\le \min\left\{\left(\frac{\beta}{\alpha(1-\gamma)}\right)^{i_0}, 1\right\}. \end{aligned}$$

Remark 5.7 *This method of proof generalizes (with more work) to the case when infectious periods are of general distribution (but still independent): comparisons can then be drawn with branching processes instead of birth-death processes.*

Remark 5.8 *See my web page on the Eyam plague for some motivational background ...*

Exercise 5.9 ***(Influenza in a University of size n):*** *Suppose that the influenza outbreak proceeds as an epidemic with removals. Suppose that the infectious period averages 3 days (so $\beta = 1/3$) and suppose that the initial rate of infection is*

1. *Then Theorems 5.2, 5.3, 5.4 combine to show we can compare to a birth-death process with birth-rate $\alpha = 1$ and death-rate $\beta = 1/3$, thus if there is one infective initially then as $n \to \infty$ so*

$$\mathbb{P}[\ \textit{epidemic}\ ] \quad \leq \quad 1 - \beta/\alpha \quad = \quad 2/3\,.$$

Note that the full version of Whittle's theorem also gives an lower bound on the probability of an epidemic, which depends on the defining proportion γ.

6. Spatial processes

Many applications force us to consider random processes in space:

- Locations of cases of childhood leukaemia (random point patterns);
- Craters on Mars (random point patterns);
- Trees in forests (random point patterns);
- Dust particles on a collector (random disk patterns);
- Fibres in paper (random line patterns).

Here the Markov property does not have the same unifying power (there is no sense of a point of "time" dividing time-space into "past" and "future", because "time" is multi-dimensional here).

So even our simple Definition 2.1 of such a basic thing as a Poisson process does not work. But Theorem 2.7 makes no mention of Markov properties, past or future, so this suggests a more general definition, for a Poisson process as a random pattern of incidents or points, scattered about whether on the line or in the plane or whatever.

6.1. The spatial Poisson process

Definition 6.1 ***(The spatial Poisson point process):*** *Let D be a region (in $\mathbb{R}^2$, $\mathbb{R}^3$, ... or an even more general region, such as a sphere). A* stationary Poisson

point process *in D of* intensity λ *is defined as follows: for any (disjoint) $U_1, \ldots, U_h \subseteq D$, with $m(U_1), \ldots, m(U_h) < \infty$* [12]

(a) the number $N(U)$ of points in U is Poisson($\lambda m(U)$);

(b) $N(U_1), \ldots, N(U_h)$ are independent.

Theorems 2.4, 2.7 say that the random pattern of incidents given by the jumps of a Poisson counting process do indeed form a stationary Poisson point process on $\mathbb{R}$. Do such processes exist in general, for example on $\mathbb{R}^2$? And how can we construct them?

Theorem 6.2 ***(Relationship to uniformly random points):*** *Consider the random pattern of points in D (for $m(D) < \infty$) formed by generating a random number of points (of Poisson($\lambda m(D)$) distribution) distributed independently and uniformly in D. This is a stationary Poisson point process of intensity λ.*

Proof: There are many approaches — for example by probability generating functions. We outline an indirect proof.

If the recipe above didn't give a Poisson point process then we would have to have a contradiction

$$\mathbb{P}\left[\, N(U_1) = k_1, N(U_2) = k_2, \ldots N(U_m) = k_m | N(D) = n \,\right]$$

[12] Of course m is length, area, or volume measure as appropriate!

as given by the recipe and as given by the definition, for at least some partition $D_1, \ldots, D_m$ of D. But in each case we can compute the conditional probability as equalling

$$\frac{n!}{k_1! \ldots k_m!} p_1^{k_1} \ldots p_m^{k_m}$$

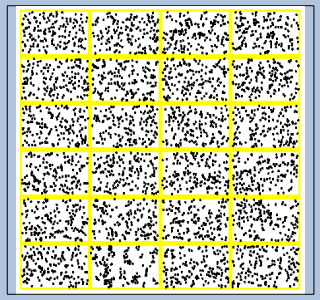
where $p_i = m(U_i)/m(D)$. (This is the well-known fact that conditioning a collection of independent Poisson counts results in a multinomial probability.) □

Remark 6.3 *If $m(D)$ is infinite the required process can be produced by taking the union of random point patterns produced on disjoint $D_1, D_2, \ldots$ of finite measure with union D.*[13]

Remark 6.4 *Problems with Poisson processes often result in calculations of $m(U)$ for suitable U, in order to apply (a) of Definition 6.1.*

Example 6.5 *Here is a figure giving simulations of 12 point patterns. Of these, 10 patterns use the recipe given above in Theorem 6.2, and so are Poisson point patterns. Of the other two, one is produced in a way which makes the arrangement of points slightly more orderly, and one in a way which makes the arrangement more clustered.*

[13] Strictly speaking, D must be σ-finite, which is to say, the countable union of pieces of finite measure. But this holds in all practical situations which I can think of right now.



Can you tell which is which? The answer, with further information on how the non-Poisson patterns were produced, is to be found in Appendix A.6 – but don't turn to that before you have made your best guess!

The human eye, being good at recognizing patterns, tends to over-interpret random point patterns, and thus to see clustering where in fact none exists.

Example 6.6 *Heather and pine seedlings. The illustration maps locations of pine seedlings in Central Sweden; shading indicates regions of heather coverage. The scientific question: is there an association between the heather coverage and the pine seedlings? If not then we might expect the point pattern of pine seedlings to behave like a stationary Poisson point process over all of D : so if H is the region covered by heather then*

$$N_{pine}(H)/m(H) \quad \approx \quad N_{pine}(H^c)/m(H^c)\,.$$

In fact calculations do not suggest any difference: consider for example the diagram's top half. There are 64 seedlings of which 43 lie in H. The area of H is $60m^2$ approx, the total area of the top half is $100m^2$. Now

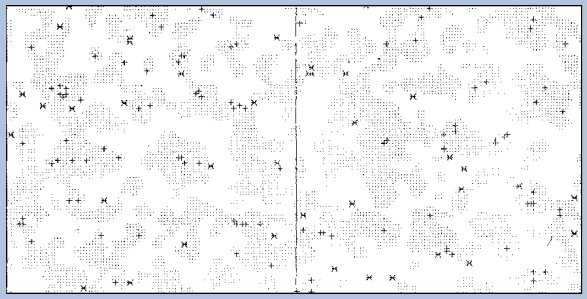
$$\textit{Poisson}(\lambda m(H))/m(H) - \textit{Poisson}(\lambda m(H^c))/m(H^c)$$

is approximately normally distributed of mean 0 *and variance* $\lambda(1/m(H) + 1/m(H^c))$ *but we find*

$$N(H)/m(H) - N(H^c)/m(H^c) \quad = \quad 0.192$$

is plausible under these assumptions: variance is 0.027 *if we estimate* λ *by its unbiased sufficient statistic*

$$N(D)/m(D) \quad = \quad 64/100\,.$$



Of course we may question the Poisson point process hypothesis. In H^c do the seedlings tend to lie closer to the heather boundary? This leads us to consider $N_{pine}(H_r \setminus H)$, which is distributed as Poisson($\lambda m(H_r \setminus H)$) for $H_r = \{x : dist(x, H) \leq r\}$. Now we can compute

$$\begin{aligned} &\mathbb{P}\,[\text{ no seedlings both within } r \text{ of } H \text{ but outside of } H\,] \\ &= \mathbb{P}\,[\,N_{pine}(H_r \setminus H) = 0\,] \quad = \quad \exp(-\lambda m(H_r \setminus H))\,. \end{aligned}$$

This suggests a simple statistical test, computing the nearest distance to H of seedlings in $H_r \setminus H$, then comparing to the distribution implicit above. We shall not perform this, because it would require computation of $m(H_r \setminus H)$ which is not convenient till I scan the picture in and use appropriate computer software!

Of course discretization errors can mess this up: so for example one might try plotting $N_{pine}(H_r \setminus H)$ against $\lambda m(H_r \setminus H)$ (estimating λ as above) and seeing if the plot looks linear (since $\mathbb{E}\,[\,N_{pine}(H_r \setminus H)\,] = \lambda m(H_r \setminus H)$).

But what is a significant deviation from linearity? This leads to the next topic.The same issue arises from finding out whether seedlings cluster or repel, by plotting and examining the number of pairs of seedlings within r of each other as a function of r.

6.2. Some statistics for Poisson processes

We take up the last point in the previous sub-section: studying clustering in a point pattern by counting close pairs of points. The procedure is as follows: first we must study the mean number $\lambda K(r)$ of points of the pattern within a ball of radius r and centred on a "typical" point which is itself not counted. We need the following theorem.

Theorem 6.7 ***(Slivnyak's theorem):*** *If $\{N(U) : U \subseteq D\}$ is a Poisson(λ) process then the process $\{N(U) : U \subseteq D\}$ conditional on a point being placed at $\mathbf{o} \in D$ is the point pattern produced by adding the point $\mathbf{o}$ to a Poisson(λ) process.*

Proof: (sketch) The main issue is to determine what we mean by a Poisson process conditioned to have a point at $\mathbf{o}$. After all, this is conditioning on an event of probability zero! Various routes lead to the answer, a sensible interpretation is

$$\begin{aligned}&\mathbb{P}\left[\, k \text{ points in } U \setminus \{\mathbf{o}\} | 1 \text{ point at } \mathbf{o} \,\right] \\ &= \lim_{\epsilon \to 0} \mathbb{P}\left[\, k \text{ points in } U \setminus \text{ ball}(\mathbf{o}, \epsilon) | 1 \text{ point in } \text{ ball}(\mathbf{o}, \epsilon) \,\right] .\end{aligned}$$

So consider

$$\begin{aligned}&\mathbb{P}\left[\, k \text{ points in } U \setminus \text{ ball}(\mathbf{o}, \epsilon) | 1 \text{ point in } \text{ ball}(\mathbf{o}, \epsilon) \,\right] \\ &= \mathbb{P}\left[\, \text{Poisson}(\lambda m(U \setminus \text{ ball}(\mathbf{o}, \epsilon))) = k \,\right] \\ &\to \mathbb{P}\left[\, \text{Poisson}(\lambda m(U \setminus \{\mathbf{o}\})) = k \,\right] \\ &= \mathbb{P}\left[\, \text{Poisson}(\lambda m(U)) = k \,\right]\end{aligned}$$

using the properties of a Poisson point process. □

Notice that Slivnyak's theorem 6.7 even has an application in the one-dimensional case. Consider for example buses arriving according to a Poisson(λ) process. By Slivnyak's theorem, after a typical bus arrival the time until another bus arrives has an Exponential(λ) distribution (see also Theorem 2.3, which gives an alternative approach in the one-dimensional case). However we can use Theorem 2.4, alternatively Definition 6.1(a), to deduce that the interval in which one might arrive to wait for a bus is always larger: in fact it has distribution the sum of two independent Exponential(λ) random variables, hence density

$$f(t) \;=\; \int_0^t \left[\lambda \exp(-\lambda s)\right] \times \left[\lambda \exp(-\lambda(t-s))\right] \, \mathrm{d}s \;=\; \lambda^2 t \exp(-\lambda t)$$

(for $t > 0$). This is an example of *length-biasing*: the interval in which one arrives is in some sense selected in a way biased towards longer intervals.

It follows, if we neglect the boundary effect of ∂D, that

$$\begin{aligned} \lambda K(r) &= \mathbb{E}\,[\,N(\,\text{ball}(\mathbf{o}, r)) - 1|\ \text{point at } \mathbf{o}\,] \\ &= \mathbb{E}\,[\,N(\,\text{ball}(\mathbf{o}, r))\,] \quad = \quad \lambda\pi r^2 \end{aligned}$$

(if dimension two! otherwise obvious changes). Anticipating some definitions,

Definition 6.8 ***(Ripley's K-function):*** *We define Ripley's K-function for a general point process N to be $K(r)$ given by*

$$\lambda K(r) \quad = \quad \mathbb{E}\,[\,N(\,\text{ball}(\mathbf{o}, r)) - 1|\ \textit{point at } \mathbf{o}\,]$$

where λ is the intensity (the "average number of points per unit area", defined below in Definition 6.19 for a general point process). Notice it is necessary to do some technical work to interpret the conditioning on an event of measure zero.

In the planar case consider

$$L(r) \quad = \quad \sqrt{\frac{K(r)}{\pi}}$$

which in the Poisson case is $L(r) = r$. This suggests an empirical approach to studying clustering relative to the Poisson process: plot the following empirical

analogue against r:

$$\sqrt{\frac{\hat{\chi}(r)}{\pi\hat{\lambda}^2}}$$

where

$$\hat{\chi}(r) \quad = \quad (\text{ number of ordered pairs closer than } r)/m(D)$$

and

$$\hat{\lambda} \quad = \quad (\text{ number of points in } D)/m(D)$$

based on the formula

$$\mathbb{E}\,[\,\hat{\chi}(r)\,] \quad = \quad \int_D \lambda K(r)(\lambda\,\mathrm{d}m)/(m(D))$$

(neglecting boundary effects!).

(There is a bonus effect here: taking the square-root stabilizes the Poisson variance. Of course this wouldn't work out so neatly in 3 dimensions, where a cube-root would be required in order to get a linear plot.)

Here are some remarks on the practical use of this idea before we turn to an example.

First, notice that in practice one should not neglect boundary effects. There are various correction techniques, all based on replacing $\hat{\chi}(r)$ by $\hat{\chi}_1(r)$ where

$$\hat{\chi}_1(r) \quad = \quad \sum k(x,y)$$

where the sum is over all points x, y of the pattern within r of each other, and $k(x, y)$ is a "fudge factor" which is larger if x or y is closer to the boundary, to compensate for bias due to the censoring effect of missing potential points lying outside the boundary. Careful choice of $k(x, y)$ makes $\hat{\chi}_1(r)$ unbiased for $\lambda^2 K(r)$.

Second, this method works under a general assumption that the point pattern is statistically "isotropic" and "homogeneous" (no preferred directions, no trends). We will discuss this further below in Definition 6.17.

Third, once we have made a plot we have to worry about the extend to which it deviates from linearity. It always will, because of random variation, so how much deviation can we allow before we should become suspicious of clustering or repulsion effects between points of the pattern? Simulation techniques come to the rescue here.

Definition 6.9 ***(Simulation envelopes):*** *Condition on the total number of observed points in D to eliminate the nuisance parameter λ. Under this conditioning on $N(D)$ the point pattern of the Poisson(λ) process is formed by independent uniform scattering of $N(D) = n$ points in D (see Theorem 6.2). (Sometimes called a binomial process in D with n points.) Simulate this $R = 199$ times (say) using the same D , the same number n of points, and for each simulation compute*

$$\sqrt{\frac{\hat{\chi}(r)}{\pi \hat{\lambda}^2}}$$

as a function of r . Compute the upper and lower simulation envelopes which are the maximum and minimum values for each r of these 199 simulated functions.

These give an indication of the amount of variation to be expected from a Poisson process with the same general features (D, total number of points) as the point pattern under consideration.

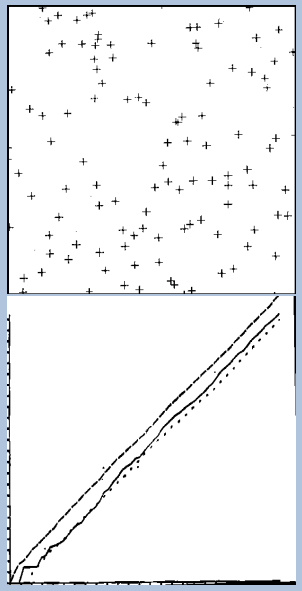
Indeed there is even a formal significance test based on the probability (computed by symmetry arguments)

$$\mathbb{P}\Big[\text{ actual } \sqrt{\frac{\hat{\chi}(r)}{\pi\hat{\lambda}^2}} \text{ computed from data lies outwith simulation envelopes at fixed } r\Big] \quad = \quad 0.01$$

(but in practice the general rule-of-thumb is more useful).

A refinement is to use *5% simulation envelopes*, where one takes not the maximum and minimum but the 97.5% and 2.5% quantiles (to smooth out random deviations).

Example 6.10 *(taken from [3]) Centres of α-particle traces on a detector are measured in a 60×60 square (1 unit is $2\mu m$). The L function is plotted together with upper and lower 5% simulation envelopes. There is some evidence suggesting repulsive interaction over the range $[1,3]$, since the L function touches the lower envelope at points in that range.*



Of course a further analysis would need to consider possibilities to explain such a repulsive interaction (equals deficit of point pairs separated in the range of 1 to 3μm).

6.3. Stochastic Geometry

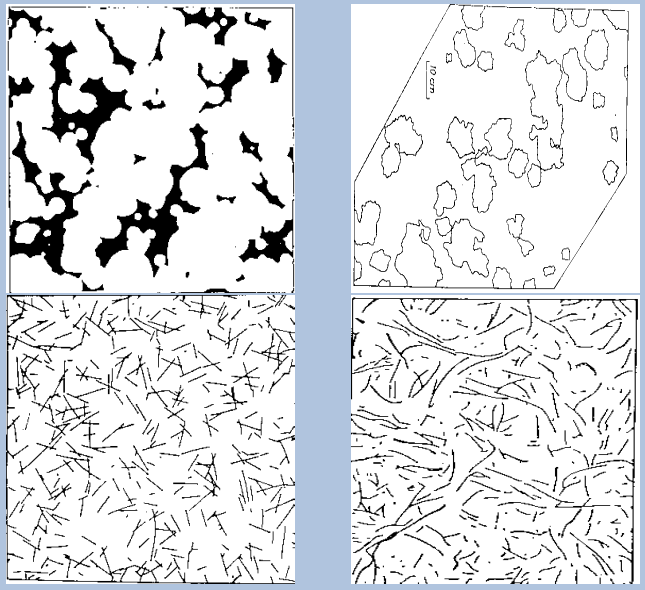
We have introduced a number of concepts which really need further explanation (homogeneity, isotropy, general point process) and these will be defined at the end of this final section. But first let us consider briefly a thriving spin-off of these ideas: stochastic geometry. Basically this concerns questions which arise when we mix concepts such as Poisson process with simple geometric constructions. For example:

Definition 6.11 ***(The Boolean model):*** *This is constructed on an unobserved point pattern produced by a Poisson point process. On each point (called a germ) one places a geometric set (called a grain) such as a disk, or a square, possibly with some degree of randomness (the disk radius could be random; the square might be randomly rotated). The union of all the grains is called the Boolean model. We define the intensity of the Boolean model as the intensity of the underlying germ process, and the volume fraction of the Boolean model as*

$$p = \mathbb{P}\left[\, \boldsymbol{o} \in \text{ Boolean model } \right]$$

(some thought shows this remains numerically unchanged if we replace $\boldsymbol{o}$ *by any other point — if we neglect edge-effects).*

This is important because it is (a) flexible, (b) easy to calculate with. For example consider the following figures, which illustrate different images which can be modelled by Boolean models with varying grains:



Here is the most basic calculation which we can carry out with the Boolean model:

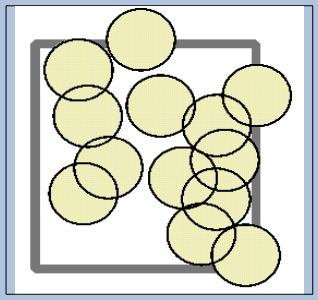
Theorem 6.12 ***(The volume fraction of a Boolean model):*** *Suppose for simplicity that the grains are disks of fixed radius r . Then*

$$p \quad = \quad 1 - \exp(-\lambda m(\text{ ball}(\boldsymbol{o}, r)))\,.$$

Proof: Consider

$$\begin{aligned} p &= \mathbb{P}[\,\mathbf{o} \in \text{ Boolean model }] \\ &= \mathbb{P}[\,N(\text{ ball}(\mathbf{o}, r)) \neq 0\,] \\ &= 1 - \exp(-\lambda m(\text{ ball}(\mathbf{o}, r)))\,. \end{aligned}$$

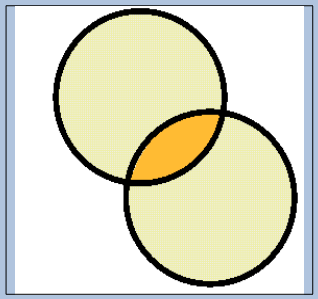
□



This proof illustrates a typical principle when calculating with Boolean models. First formulate the question, then turn it into a question about the underlying Poisson process of germs. Finally use the simple properties of a Poisson point process (Definition 6.1) to find the answer, which usually just involves a computation of an area.

Example 6.13 *Dust particles on a sensor region. Model this by a Boolean model Σ of disks of fixed radius r, with germ process of intensity λ. Neglecting edge-effects, determine the mean measure of that part of the region D which is not covered by the Boolean model Σ.*

We have to compute $\mathbb{E}\,[\,m(D \setminus \Sigma)\,]$. But this is given by



$$\begin{aligned}\mathbb{E}\,[\,m(D \setminus \Sigma)\,] &= \mathbb{E}\left[\int_D \mathbb{I}_{[x \notin \Sigma]}\,\mathrm{d}x\right] \\ &= \int_D \mathbb{P}\,[\,x \notin \Sigma\,]\,\mathrm{d}x \\ &= m(D)(1-p) \\ &= m(D)\exp(-\lambda\pi r^2)\,.\end{aligned}$$

It is a good exercise to compute the variance of $m(D \setminus \Sigma)$, *via*

$$\mathbf{Var}[m(D\setminus\Sigma)] \quad = \quad \mathbb{E}\,[\,m(D \setminus \Sigma)^2\,] - (\mathbb{E}\,[\,m(D \setminus \Sigma)\,])^2$$

and

$$\begin{aligned}\mathbb{E}\,[\,m(D \setminus \Sigma)^2\,] &= \mathbb{E}\left[\int_D \mathbb{I}_{[x \notin \Sigma]}\,\mathrm{d}x \int_D \mathbb{I}_{[y \notin \Sigma]}\,\mathrm{d}y\right] \\ &= \int_D \int_D \mathbb{P}\,[\,x \notin \Sigma, y \notin \Sigma\,]\,\mathrm{d}x\,\mathrm{d}y\end{aligned}$$

and

$$\begin{aligned}&\mathbb{P}\,[\,x \notin \Sigma, y \notin \Sigma\,] \\ &= \mathbb{P}\,[\,\text{Poisson}(\lambda m(\,\text{ball}(\mathbf{o}, r) \cup \,\text{ball}(x-y, r))) = 0\,] \\ &= \exp(-\lambda m(\,\text{ball}(\mathbf{o}, r) \cup \,\text{ball}(x-y, r)))\end{aligned}$$

so that all reduces to the geometric computation, what is the area of the intersection of two disks, radius r, separated by distance $|x-y|$ (see figure). This computation is left as an exercise!

6.4. General stationary point processes

We now very briefly cover some of the concepts mentioned in the first part of this section. First, the definition of a general (ie not Poisson) point process.

Definition 6.14 ***(Point process in D):*** *This a random locally finite point pattern with no coincident points. Hence*

- $N(U)$ *counts the number of points in* U*;*
- $N(U)$ *is finite for all bounded* U*;*
- $N(\text{ ball}(\boldsymbol{o}, \epsilon))$ *converges to* 0 *or* 1 *as* $\epsilon \to 0$*.*

The point process is the random process $\{N(U) : U \subseteq D\}$ *(strictly speaking, defined only for measurable* $U \subseteq D$ *).*

When are two point processes statistically the same? When they have the same *distribution*:

Definition 6.15 ***(Distribution of a point process in D):*** *This is the family of all joint distributions of point counts* $N(U)$ *(sometimes called the family of all finite-dimensional distributions):* $\{\mathcal{L}(N(U_1), N(U_2), \ldots, N(U_n)) : U_1, U_2, \ldots, U_n \subseteq$

$D, n = 1, 2, \ldots\}$ (strictly speaking, U_1, U_2, …, U_n must be measurable). Here $\mathcal{L}(X)$ is the distribution of the (possibly multivariate) random variable X.

It would be hard work to check all joint distributions of point counts in order to identify a point process distribution. (Though we did do this to prove Theorem 6.2.) There is an easier way. Because of the following remarkable result it suffices to check only the void probabilities (the probabilities of there being no points in sets U).

Theorem 6.16 ***(A consequence of Choquet's theorem on capacities, from advanced measure theory):*** *The distribution of a point process is determined by the void probability system* $\{\mathbb{P}[N(U) = 0] : U \subseteq D\}$.

(No proof!)

The statistical symmetries of a point process are of great importance.

Definition 6.17 ***(Stationarity, isotropy, motion-invariance):*** *A point process is stationary (also called homogeneous) if its distribution is the same when it is shifted. That is, if*

$$\begin{aligned}&\mathcal{L}(N(U_1), N(U_2), \ldots, N(U_n))\\ &= \mathcal{L}(N(U_1 + x), N(U_2 + x), \ldots, N(U_n + x))\end{aligned}$$

whenever

$$\begin{aligned}&U_1, U_2, \ldots, U_n, U_1 + x, U_2 + x, \ldots, U_n + x \subseteq D\,,\\ &\quad x \in D\,, \qquad n = 1, 2, \ldots\,.\end{aligned}$$

It is isotropic *if its distribution is the same when it is rotated. That is, if*

$$\mathcal{L}(N(U_1), N(U_2), \ldots, N(U_n)) = \mathcal{L}(N(\mathcal{R}U_1), N(\mathcal{R}U_2), \ldots, N(\mathcal{R}U_n))$$

whenever

$$U_1, U_2, \ldots, U_n, \mathcal{R}U_1, \mathcal{R}U_2, \ldots, \mathcal{R}U_n \subseteq D\,,$$
$$\mathcal{R} \text{ is a rotation}\,, \qquad n = 1, 2, \ldots\,.$$

It is motion-invariant *if it is both stationary and isotropic.*

Note that Theorem 6.16 says we really only have to check $\mathbb{P}\,[\,N(U_+x) = 0\,]$, respectively $\mathbb{P}\,[\,N(\mathcal{R}U) = 0\,]$, doesn't depend on x, respectively $\mathcal{R}$, to see if the process is stationary, respectively isotropic.

Theorem 6.18 ***(The stationary Poisson point process is motion-invariant):***

Proof: This is actually obvious from Definition 6.1! □

On the other hand the inhomogeneous Poisson point process (replace m in Definition 6.1 by a general measure which is not translation-invariant) is not stationary, and the Poisson doublet process (produced from the Boolean model using a two-point grain $\{(0,0),(0,1)\}$) is not isotropic.

Now we can consider basic characteristics of a point process. To start with, the mean number of points per unit area:

Definition 6.19 ***(Intensity measure of a point process):*** *This is defined by*

$$\mu(U) \quad = \quad \mathbb{E}\,[\,N(U)\,] \qquad \text{for } U \subseteq D\,.$$

If the point process is stationary then the intensity measure is translation-invariant, and so must be a multiple of area measure m. Then we define the *intensity λ of the point process* by $\mu(U) = \lambda m(U)$.

The intensity measure is analogous to the mean of a random variable. Similarly we can define measures which are analogous to higher-order moments:

Definition 6.20 ***(The k^{th}-order moment measure of a point process):*** *This is defined as follows: for $k = 2$ we have*

$$\mu_2(U \times V) \quad = \quad \mathbb{E}\,[\,N(U) \times N(V)\,]$$

while for general k we have

$$\mu_k(U_1 \times \ldots \times U_k) \quad = \quad \mathbb{E}\,[\,N(U_1) \times \ldots \times N(U_k)\,]\,.$$

The second-order moment measure $\mu_2(U \times V)$ can be viewed as a non-centred covariance. It is related (in the motion-invariant case) to Ripley's K-function (see Definition 6.8).

6.5. Examples of point processes

In this final section we run briefly through some alternative point process models. Obviously these are important if we want to go further than the rather rudimentary

statistical methods described in Section 6.1 and Section 6.2, since we then require point processes which might exhibit features which we believe we have detected in our dataset.

Definition 6.21 ***(Inhomogeneous Poisson point processes):*** *We have mentioned this in the previous subsection. Replace m by a general measure ν in Definition 6.1. We say the resulting process is driven by the intensity measure ν.*

Example 6.22 *If ν can be obtained from m by using a density function, so that $\nu(U) = \int_U f(x)m(\mathrm{d}x)$, and if f is bounded above by λ, then we may obtain the inhomogeneous Poisson point process driven by ν by using non-stationary thinning: start with a stationary Poisson point process of intensity λ, and retain each point x independently of the rest with retention probability $f(x)/\lambda$.*

Definition 6.23 ***(Cox point processes):*** *(Also called doubly-stochastic point processes.) These are defined as in Definition 6.21, but the driving measure is random!*

Definition 6.24 ***(Neyman-Scott cluster point processes):*** *Start with an unobserved stationary Poisson point process of cluster centres. Centered on each cluster centre x scatter a random number M_x of points y such that the scattering vectors $x-y$ are independent and identically distributed with prescribed scattering distribution F. The random variables M_x are independent of each other and of the scattering vectors and share the same cluster distribution C.*

Example 6.25 *The distribution of galaxies, which are thought to cluster around (unobserved) cluster centres.*

Notice that a Neyman-Scott cluster point process is automatically stationary. If the scatter vector distribution is rotationally symmetric then the point process is also motion-invariant. In any case the intensity of the point process is given by $\lambda \times \mathbb{E}[C]$, where λ is the intensity of the underlying Poisson point process. Its second-order moment measure is different from that for a Poisson process: it will be "larger at closer ranges", depending on F. A similar difference will be observed in the K-function.

Example 6.26 *The Neyman-Scott cluster point process can be viewed as a Boolean model in which the grain is formed by a typical cluster (number of points has distribution C, locations are distributed according to F).*

Example 6.27 *The doublet process described in the previous subsection can be viewed as a Neyman-Scott cluster point process. What is the cluster distribution? What is the scatter distribution F? (Hint: neither are very random!)*

Our final example is of a class of point processes for which the points repel instead of clustering.

Definition 6.28 ***(hard-core point processes):*** *Consider a Poisson point process in a bounded region D, conditioned so that no two points are closer together than $2r$ for some prescribed r. This is a hard-core point process with hard-core radius r.*

Variations on this: replace the "disk-like" hard-core by some other shape (square? triangle?); use a limiting argument to define the process over all $\mathbb{R}^2$; replace the

hard-core by a soft-core; further variations allow more complicated interactions between points to produce the famous Gibbs point processes. These provide a very flexible model for point patterns where repulsion between points is observed. Unfortunately they are not good at providing models for attraction (clustering). Until recently it was thought one would have to use the entirely different Neyman-Scott cluster point processes to model attraction. However Baddeley and Möller, also Van Lieshout, have developed [14] a generalization known as nearest-neighbour Markov point processes which can deal both with clustering and repulsion, and which embrace both Gibbs point processes and (suitably regular) Neyman-Scott cluster point processes. This is at the cutting edge of some very exciting research in this area . . . watch this space!

[14] And I have made a small contribution too.

References

[1] R.G. Jarrett. A note on the intervals between coal-mining disasters. *Biometrika*, 66:191–193, 1979. [JSTOR].

[2] F.P. Kelly. *Reversibility and Stochastic Networks*. John Wiley and Sons, Chichester, 1979. [OPAC].

[3] D. Stoyan, W.S. Kendall, and J. Mecke. *Stochastic Geometry and its Applications*. John Wiley and Sons, Chichester, 1987. [OPAC].

A. Appendices

These notes are concluded by appendices providing supplementary material. This supplementary material is generally provided for purposes of quick reference.

A.1. Crash course in integrating factors

Here is a sketch of how to use the integrating factor method to solve simple differential equations.

First a really simple example. Consider the differential equation

$$\frac{\mathrm{d}y}{\mathrm{d}t} \quad = \quad y + 1\,.$$

First re-arrange so the y-terms are all on the left-hand side:

$$\frac{\mathrm{d}y}{\mathrm{d}t} - y \quad = \quad 1\,.$$

Now look for a way to convert the left-hand side into the form $(1/F(t))\,\mathrm{d}(F(t)y)/\,\mathrm{d}t$ for some function $F(t)$. Here $F(t) = \exp(-t)$ will do:

$$\exp(t)\frac{\mathrm{d}}{\mathrm{d}t}\left(\exp(-t)y\right) \quad = \quad \frac{\mathrm{d}y}{\mathrm{d}t} - y \quad = \quad 1\,.$$

(Check this out by use of the formula for differentiating a product on the expression on the extreme left-hand side!) We can now re-arrange into the form of “derivative

of something" equals "function of t alone":

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(\exp(-t)y\right) \quad = \quad \exp(-t)\,.$$

Integrate both sides, and don't forget the arbitrary constant arising from the integration process:

$$\exp(-t)y \quad = \quad A + \int_0^t \exp(-s)\,\mathrm{d}s \quad = \quad A - \exp(-t)\,.$$

We now obtain the general solution

$$y \quad = \quad A\exp(t) - 1\,.$$

In general the constant A is determined by boundary conditions. In our problems these usually arise from requiring that the transition probability function $p_t(x, y)$ should take on appropriate values at $t = 0$:

$$p_0(x, y) \quad = \quad \begin{cases} 1 & \text{if } x = y \\ 0 & \text{if } x \neq y \end{cases}$$

Here is a slightly more involved example which does not integrate out in closed form. Consider the differential equation

$$\frac{\mathrm{d}y}{\mathrm{d}t} \quad = \quad ty + 1\,.$$

First re-arrange so the y-terms are all on the left-hand side:

$$\frac{\mathrm{d}y}{\mathrm{d}t} - ty \quad = \quad 1\,.$$

Now look for a way to convert the left-hand side into the form $(1/F(t))\,\mathrm{d}(F(t)y)/\,\mathrm{d}t$ for some function $F(t)$. Here $F(t) = \exp(-t^2/2)$ will do:

$$\exp(t^2/2)\frac{\mathrm{d}}{\mathrm{d}t}\left(\exp(-t^2/2)y\right) \quad = \quad \frac{\mathrm{d}y}{\mathrm{d}t} - ty \quad = \quad 1\,.$$

(Check this out by use of the formula for differentiating a product on the expression on the left-hand side!) We can now re-arrange into the form of "derivative of something" equals "function of t alone":

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(\exp(-t^2/2)y\right) \quad = \quad \exp(-t^2/2)\,.$$

Integrate both sides, and don't forget the arbitrary constant arising from the integration process:

$$\exp(-t^2/2)y \quad = \quad A + \int_{-\infty}^{t} \exp(-s^2/2)\,\mathrm{d}s\,.$$

(It doesn't really matter what you put in the lower limit of the integral: it all gets absorbed in the arbitrary constant A.)

It is well-known that we can't hope to evaluate the error function integral

$$\int_{-\infty}^{t} \exp(-s^2/2)\,\mathrm{d}s$$

in closed form. But it is usually enough for our purposes to have a well-defined and tabulated integral, and anyway we could write computer code to integrate it numerically if we had to!

We now obtain the general solution

$$y \quad = \quad A\exp(t^2/2) - \exp(t^2/2)\int_{-\infty}^{t} \exp(-s^2/2)\,\mathrm{d}s\,.$$

The crucial trick is to spot the *integrating factor* $F(t)$ (respectively $\exp(-t)$ and $\exp(-t^2/2)$ in the above examples). In complete generality, given

$$\frac{\mathrm{d}y}{\mathrm{d}t} \quad = \quad B(t)y + C(t)\,,$$

we re-arrange,

$$\frac{\mathrm{d}y}{\mathrm{d}t} - B(t)y \quad = \quad C(t)\,,$$

find $F(t)$ such that $\mathrm{d}F(t)/\,\mathrm{d}t = -B(t)F(t)$, hence deduce

$$\frac{1}{F(t)}\frac{\mathrm{d}}{\mathrm{d}t}\left(F(t)y\right) \quad = \quad C(t)\,,$$

and arrive at the task of integrating out

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(F(t)y\right) \quad = \quad C(t)F(t)$$

as

$$F(t)y \quad = \quad A + \int_{\ell}^{t} C(s)F(s)\,\mathrm{d}s\,,$$

for A an arbitrary constant (and convenient choice of lower limit ℓ in the integral).

A.2. Erlang's formula for M/G/c/c queues

Here is a sketch of how to derive Erlang's formula for $M/G/c/c$ queues.

The argument hangs on three facts.

- Firstly, we can represent an $M/G/c/c$ queue as a continuous-time Markov chain on a *continuous state space*. Since the service times are no longer exponential it no longer suffices simply to record the number of customers in the system. (See the discussion about this point at the start of Section 3.3.) However if we record for each of the n customers in the system
  - the time they have spent in the system so far;
  - the total time they require to be served;

  then the resulting state $(n; u_1, s_1, \ldots, u_n, s_n)$ *does* determine a Markov chain Y. Notice that the state space is not simple! it can be viewed as

the union of orthants of Euclidean spaces of dimensions $2, \ldots 2c$, together with an added extra state $*$ corresponding to the case of empty queue.

- Secondly, when we let c tend to ∞ the resulting $M/G/\infty$ queue X becomes *relatively* easy to analyze using the techniques of Section 6. For an $M/G/\infty$ queue, all customers are served as soon as they arrive. If we "mark" each Poisson arrival time a with a real number s indicating the length of service required, then the resulting random pattern of two-dimensional points is an inhomogeneous Poisson point process, as given in Definition 6.21. Some simple manipulations now show that the transformation of points sending (a, s) to $(a + s, s)$ does not alter the statistical behaviour of the random point pattern (it remains an inhomogeneous Poisson point process with the same intensity measure). This can be interpreted as saying that the $M/G/\infty$ queue is a reversible random process, since the time $a + s$ is the departure time. It is also a standard spatial Poisson process calculation to show that the number of points (a, s) such that $a \leq t < a_s$ has a Poisson distribution of mean $\lambda \mathbb{E}[S]$, where λ is the intensity of the arrival process and S is a typical service time. (Here we suppose that the $M/G/\infty$ queue has been in operation since time minus infinity, or equivalently that it is in equilibrium.) Accordingly

$$\begin{aligned} &\mathbb{P}[\text{ there are } n \text{ customers in an } M/G/\infty \text{ queue in equilibrium }] \\ &= \frac{\rho^n}{n!} \exp(-\rho)\,, \end{aligned}$$

where $\rho = \lambda \mathbb{E}[S]$ is the utilization factor analogous to the definition for an $M/G/1$ queue.

- Thirdly, the $M/G/c/c$ queue Y is simply a restriction of the $M/G/\infty$ queue X which allows only c customers to be in service at any one time. Accordingly we can use Remark 3.10 (or, more accurately, its generalization to continuous state space) to show that the equilibrium probabilities for Y are proportional to those of X.

From all this we can deduce that Erlang's formula (Theorem 3.15) extends:

$$\mathbb{P}[\text{ there are } n \text{ customers in an } M/G/c/c \text{ queue in equilibrium }] = \frac{\frac{\rho^n}{n!}}{1+\rho+\frac{\rho^2}{2}+\ldots+\frac{\rho^c}{c!}}, \quad \text{for } n = 0, 1, \ldots, c.$$

Of course to fully sort this out one needs to work out the details of the theory of continuous state space Markov chains, which is beyond the scope of these notes. However it is possible to get very close by using a kind of discretization: we can approximate a general service time S by a mixture of sums of exponential random variables, and view each customer as passing through several different *phases* of service. By this means we can get back to a countable state space Markov chain, in which the states carry information on the phases of service for each of the customers in the system. (See [2] for more details.) So this forms a possible setting for an interesting fourth-year MMathStat or MMORSE essay: work out these details, and then collect data from a setting where it might be plausible to approximate by

an $M/G/c/c$ queue (*eg:* the ATMs of one of the campus banks located between the Union and Rootes: not quite $M/G/c/c$) but perhaps $M/G/c/(c+2)$?!) and investigate how good the approximation might be.

A.3. Counting and indicator random variables

The following argument is typical of many occurring throughout applied probability and statistics.

Suppose N_t counts the number of incidents occurring before time t. Suppose further that incident n actually occurs at time S_n. Then it is an almost tautologous remark that

$$N_t \quad = \quad \sum_n \mathbb{I}_{[S_n \leq t]} \, .$$

(The number of incidents occurring before time t is the same as the number obtained by adding 1 for each incident which occurs before time t!)

Take expectations and exchange expectation and sum on right-hand side (exchange permitted because the summands are non-negative: see discussion of Fubini's theorem, Theorem **??** in ST213). We deduce

$$\begin{aligned} \mathbb{E}\left[N_t \right] &= \sum_n \mathbb{E}\left[\mathbb{I}_{[S_n \leq t]} \right] \\ &= \sum_n \mathbb{P}\left[S_n \leq t \right] . \end{aligned}$$

Here for the record is another argument of a similar nature: suppose M is a random variable taking non-negative integer values. Then

$$M = \sum_{n=1}^{\infty} \mathbb{I}_{[M \geq n]}$$

(check this out: true whether $M = 0, 1, 2, \ldots$). Taking expectations and exchanging with sum as before,

$$\begin{aligned} \mathbb{E}[M] &= \sum_{n=1}^{\infty} \mathbb{E}\left[\mathbb{I}_{[M \geq n]}\right] \\ &= \sum_{n=1}^{\infty} \mathbb{P}[M \geq n] . \end{aligned}$$

A.4. Hyperbolic functions in a nutshell

Here is a brief summary of the sinh, cosh notation.

Think of this notation as a short-hand for commonly occurring combinations of exponential functions:

$$\begin{aligned} \cosh(x) &= \frac{e^x + e^{-x}}{2} , \\ \sinh(x) &= \frac{e^x - e^{-x}}{2} . \end{aligned}$$

It is informative to sketch their graphs: $\cosh$ is symmetric about 0, equals 1 there, and grows like $\exp(x)/2$ at infinity; $\sinh$ is odd about 0, equals 0 there, and also grows like $\exp(x)/2$ at infinity.

The reason for the notation is that they behave in a manner which is *very* similar to that of trigonometric functions but with occasional minus signs going missing:

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}x}\cosh(x) &= \sinh(x)\,, \\ \frac{\mathrm{d}}{\mathrm{d}x}\sinh(x) &= \cosh(x)\,. \end{aligned}$$

Also

$$\cosh^2(x) - \sinh^2(x) = 1\,.$$

This is no coincidence: one of the basic facts in complex analysis is that it makes sense to write

$$\begin{aligned} \cos(x) &= \cosh(\sqrt{-1}x)\,, \\ \sin(x) &= \frac{\sinh(\sqrt{-1}x)}{\sqrt{-1}}\,; \end{aligned}$$

and in fact all the properties of hyperbolic functions can be deduced from those of trigonometric functions by using these relations.

Why the name? Because the hyperbolic functions provide a good parametrization for the hyperbola conic curve, and for hyperboloid surfaces. But discussion of this would take us too far afield.

A.5. Computing probabilities for arrivals in an $M/G/1$ service period

Here is a summary of how first-year probability is used to compute such things as k_n, the probability of n customers arriving in a service period of an $M/G/1$ queue.

We use the techniques of *indicator functions* and of *iterated expectations*. Consider

$$k_n \quad = \quad \mathbb{P}\left[\, n \text{ customers arrive during service period } \; S \,\right] .$$

First use indicator functions:

$$k_n \quad = \quad \mathbb{E}\left[\, \mathbb{I}_{[n \text{ customers arrive during service period } \; S]} \,\right] .$$

Now use iterated expectations, conditioning on the value of S:

$$k_n \quad = \quad \mathbb{E}\left[\, \mathbb{E}\left[\, \mathbb{I}_{[n \text{ customers arrive during service period } \; S]} | S \,\right] \,\right] .$$

Now unravel the indicator function:

$$k_n \quad = \quad \mathbb{E}\left[\, \mathbb{P}\left[\, n \text{ customers arrive during service period } \; S | S \,\right] \,\right] .$$

You will be able to compute the conditional probability using known formulae about the Poisson process of arrivals, which is independent of the service time S (see Theorem 2.4). Hence

$$k_n \quad = \quad \mathbb{E}\left[\frac{(\lambda S)^n \exp(-\lambda S)}{n!} \right]$$

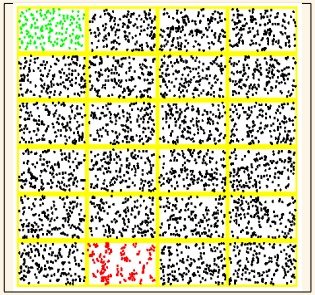
where λ is the arrival rate of the Poisson process, and so calculation of k_n reduces to knowing how to calculate some expectations to do with S.

A similar but simpler exercise shows why $\mathbb{E}[Poisson(\lambda W)] = \lambda \mathbb{E}[W]$. For

$$\begin{aligned} \mathbb{E}[Poisson(\lambda W)] &= \mathbb{E}[\mathbb{E}[Poisson(\lambda W)|W]] \\ &= \mathbb{E}[\lambda W] \end{aligned}$$

by computing the inside, conditional, expectation, where the conditioning allows us to treat W as constant.

A.6. Poisson and non-Poisson realizations



Here is the figure of Example 6.5 in which the non-Poisson realizations are picked out in different colours. The clustered or "attractive" realization is picked out in red, and should have been easily identifiable as more clustered than the norm. It was produced by placing about 55 invisible parent points according to a Poisson process, and then for each parent placing three visible daughters, each independently displaced in x and y directions using normally distributed displacements of standard deviation about $1/21$ of the length of the y-axis.

The orderly or "repulsive" realization is much harder to spot: here it is picked out in green. It was produced by similarly displacing about 160 points laid out randomly on a square lattice. A trained observer should be able to spot this pattern as unusual because of the relative absence of large empty spaces – but this effect is much harder to notice than clustering!

End of outline notes