MAST30001: Stochastic Modelling

Assoc. Prof. Mark Holmes
Room 209, Peter Hall Building
holmes.m@unimelb.edu.au
x41309

- Lectures and tutorials
- 2 Assignments, each worth 10%
- final exam, worth 80%

Course Outline

- Probability Review
- Discrete time Markov chains
- Poisson Process
- Continuous-time Markov chains
- Queueing theory
- Renewal theory

But first...... motivation!

Motivation

We want to understand how a complex system works. Real-world experimentation can be

- too slow,
- too expensive,
- possibly too dangerous/immoral,
- may not deliver insight.

The alternative is to build a physical, mathematical or computational model that captures the essence of the system that we are interested in (think: NASA).

Modelling

We develop an *imitation* of the system. It could be, for example,

- physical: a small replica of a marina development,
- mathematical: a set of equations describing the relations between stock prices,
- computational: a computer simulation that reproduces a complex system (think: the paths of planets in the solar system).

We may use a model

- to understand the evolution of a system,
- to understand how 'outputs' relate to 'inputs', and
- to decide how to influence a system.

Why a stochastic model?

In may situations, the effect of “randomness” cannot (or should not) be ignored.

- traffic in the Internet
- stock prices and their derivatives
- waiting times in healthcare queues
- reliability of multicomponent systems
- interacting populations
- epidemics
- physical properties of materials
- how a company/product establishes market share

Good mathematical models

- capture the non-trivial behaviour of a system (and replicate empirical observations), and
- are relatively simple (and can be analysed to derive the quantities of interest), and
- can be used to help make decisions.

Stochastic modelling

Stochastic modelling is about the study of *random experiments*, e.g.

- toss a coin once, or twice, or infinitely-many times
- the lifetime of a randomly selected battery (quality control)
- the size of an animal population (conservation)
- the length of a queue over the time interval $[0, \infty)$ (service)
- the infection status of a collection of individuals (health)
- the network "friend" structure of Facebook (ad revenue)

Making sense of stochastic models starts with an understanding of basic probability.

Basic Probability

Probability Space

We study a random experiment in the context of a Probability Space $(\Omega, \mathcal{F}, \mathbb{P})$. Here,

- Ω is the sample space - the set of all possible outcomes of our random experiment,
- $\mathcal{F}$ is a sigma-field - a collection (with certain properties) of subsets of Ω. We view these as events we can *see* or *measure*,
- $\mathbb{P}$ is a probability measure - a function (with certain properties) defined on the elements of $\mathcal{F}$ with certain properties.

The sample space Ω

The (nonempty) set of possible outcomes for the random experiment. Examples:

- coin tossing: $\{H, T\}$, $\{(H, H), (H, T), (T, H), (T, T)\}$, the set $\{H, T\}^{\mathbb{N}}$ of all infinite sequences of *H*s and *T*s.
- battery lifetime: $[0, \infty)$.
- animal population: $\mathbb{Z}_+$.
- queue length over time: the set of piecewise-constant functions from $[0, \infty)$ to $\mathbb{Z}_+$.
- infection status: $\{\mathit{susceptible}, \mathit{infected}, \mathit{immune}\}^n$ (if n is the number of individuals).
- facebook friend network: set of simple networks with number of vertices equal to the number of users: edges connect friends.

Elements of Ω are often written as ω.

Review of basic notions of set theory

- $A \subset B$.
  - A is a subset of B, or if A occurs then B occurs.
- $A \cup B = \{\omega \in \Omega : \omega \in A \text{ or } \omega \in B\} = B \cup A$.
  - Union of sets (events): at least one occurs.
  - $A_1 \cup A_2 \cup \cdots \cup A_n = \bigcup_{i=1}^{n} A_i$.
- $A \cap B = \{\omega \in \Omega : \omega \in A \text{ and } \omega \in B\} = B \cap A$.
  - Intersection of sets (events): all occur.
  - $A_1 \cap A_2 \cap \cdots \cap A_n = \bigcap_{i=1}^{n} A_i$.
- $A^c = \{\omega \in \Omega : w \notin A\}$.
  - Complement of a set/event: event doesn't occur.
- $\varnothing$: the empty set or impossible event.
- A and B are disjoint (or mutually exclusive) if $A \cap B = \varnothing$.

The *sigma-field* $\mathcal{F}$

A sigma-field $\mathcal{F}$ on Ω is a collection of subsets of Ω such that

- $\Omega \in \mathcal{F}$, and
- if $A \in \mathcal{F}$ then $A^c \in \mathcal{F}$, and
- if $A_1, A_2, \dots \in \mathcal{F}$ then $\cup_{i=1}^{\infty} A_i \in \mathcal{F}$.

- For countable sample spaces, $\mathcal{F}$ is typically the set of all subsets.

  Example: Toss a coin once, $\mathcal{F} = \big\{\emptyset, \{H\}, \{T\}, \{H, T\}\big\}$

- For uncountable sample spaces, the situation is more complicated - see later.

The probability measure $\mathbb{P}$

A probability measure $\mathbb{P}$ on $(\Omega, \mathcal{F})$ is a set function from $\mathcal{F}$ to $[0, 1]$ satisfying

P1. $\mathbb{P}(A) \geq 0$ for all $A \in \mathcal{F}$ [probabilities measure long run %'s or certainty]

P2. $\mathbb{P}(\Omega) = 1$ [There is a 100% chance something happens]

P3. Countable additivity: if A_1, $A_2 \cdots$ are disjoint events in $\mathcal{F}$, then $\mathbb{P}\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \mathbb{P}(A_i)$ [Think about it in terms of frequencies]

How do we specify $\mathbb{P}$?

The modelling process consists of

- defining the values of $\mathbb{P}(A)$ for some 'basic events' in $A \in \mathcal{F}$,
- deriving $\mathbb{P}(B)$ for the other 'unknown' more complicated events in $B \in \mathcal{F}$ from the axioms above.

Example: Toss a fair coin 1000 times. Any particular sequence (of length 1000) of H's and T's has chance 2^{-1000}.

- What is the chance there are more than 600 H's in the sequence?
- What is the chance the first time the proportion of heads exceeds the proportion of tails occurs after toss 20?

Properties of $\mathbb{P}$

- $\mathbb{P}(\emptyset) = 0$.
- $\mathbb{P}(A^c) = 1 - \mathbb{P}(A)$.
- $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$
- ...
- "Continuity":

$$\mathbb{P}(\cup_{i=1}^{\infty} A_i) = \lim_{n\to\infty} \mathbb{P}(\cup_{i=1}^{n} A_i)$$
$$\mathbb{P}(\cap_{i=1}^{\infty} A_i) = \lim_{n\to\infty} \mathbb{P}(\cap_{i=1}^{n} A_i)$$

Conditional probability

Let $A, B \in \mathcal{F}$ be events with $\mathbb{P}(B) > 0$. Supposing we know that B occurred, how likely is A given that information? That is, what is the conditional probability $\mathbb{P}(A|B)$?
We define

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}.$$

For a frequency interpretation, consider the situation where we have 10 tickets, numbered 1, 2, ..., 10, and we choose one ticket uniformly at random. Then $\Omega = \{1, 2, \dots, 10\}$.
Let $A = \{\omega \in \Omega : \omega \text{ is even}\} = \{2, 4, 6, 8, 10\}$,
and $B = \{\omega \in \Omega : \omega \text{ is a multiple of } 3\} = \{3, 6, 9\}$.
Then $\mathbb{P}(A) = 5/10$, $\mathbb{P}(B) = 3/10$, $\mathbb{P}(A \cap B) = \mathbb{P}(\{6\}) = 1/10$.
According to the definition, $\mathbb{P}(A|B) = \frac{1/10}{3/10} = 1/3$. This is nothing but the proportion of outcomes in B that are also in A.

Example:

Tickets are drawn consecutively and *without replacement* from a box of tickets numbered $1, \ldots, 0$. What is the chance the second ticket is even numbered given the first is

- even?
- labelled 3?

Law of total probability

Let $B_1, B_2, \cdots, B_n$ be disjoint events with $\mathbb{P}(B_i) > 0$ and $A \subset \bigcup_{j=1}^n B_j$, then

$$\mathbb{P}(A) = \sum_{j=1}^{n} \mathbb{P}(A|B_j)\mathbb{P}(B_j).$$

If also $\mathbb{P}(A) > 0$ then

$$\mathbb{P}(B_j|A) = \frac{\mathbb{P}(B_j \cap A)}{\mathbb{P}(A)} = \frac{\mathbb{P}(A|B_j)\mathbb{P}(B_j)}{\sum_{k=1}^n \mathbb{P}(A|B_k)\mathbb{P}(B_k)}.$$

Example:

A disease affects $1/1000$ newborns and shortly after birth a baby is screened for this disease using a cheap test that has a 2% false positive rate (the test has no false negatives). If the baby tests positive, what is the chance it has the disease?

Independent events

Events A and B are said to be independent if
$\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$.
If $\mathbb{P}(B) \neq 0$ then this is the same as $\mathbb{P}(A|B) = \mathbb{P}(A)$.
Events $A_1, \cdots, A_n$ are independent if, for each subset $\{i_1, ..., i_k\}$ of $\{1, ..., n\}$,

$$\mathbb{P}(A_{i_1} \cap \cdots \cap A_{i_k}) = \mathbb{P}(A_{i_1}) \times \cdots \times \mathbb{P}(A_{i_k}).$$

This is a stronger requirement than each pair of events being independent.

A disconcerting example

- Let S be the circle of radius 1.
- We say two points on S are in the same family if you can get from one to the other by taking steps of arclength 1 around the circle.
- Each family chooses a single member to be head.
- If X is a point chosen uniformly at random from the circle, what is the chance X is the head of its family?

A disconcerting example

- $A = \{X \text{ is head of its family}\}$.
- $A_i = \{X \text{ is } i \text{ steps clockwise from its family head}\}$.
- $B_i = \{X \text{ is } i \text{ steps counterclockwise from its family head}\}$.
- By uniformity, should have $\mathbb{P}(A) = \mathbb{P}(A_i) = \mathbb{P}(B_i)$, BUT
- law of total probability:

$$1 = \mathbb{P}(A) + \sum_{i=1}^{\infty} (\mathbb{P}(A_i) + \mathbb{P}(B_i)) \qquad !$$

The issue is that the set A is not one we can *measure* so should not be included in $\mathcal{F}$.

These kinds of issues are technical to resolve and are dealt with in later probability or analysis subjects which use *measure theory*.

Random variables

A random variable on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ is a function $X : \Omega \to \mathbb{R}$ such that

(Often we want to talk about the probabilities that random variables take values in $(-\infty, b] = \{x \in \mathbb{R} : x \leq b\}$*.*
When we write $\mathbb{P}(X \leq b)$*, we mean the probability of the set* $\{\omega : X(\omega) \leq b\}$*. In order for this to make sense, we need this set to be in* $\mathcal{F}$*!)*

$\{X \leq b\} := \{\omega \in \Omega : X(\omega) \leq b\} \in \mathcal{F}$ for every $b \in \mathbb{R}$.

(in measure theory, this is called a *Borel-measurable function*, after Emile Borel (1871-1956)).

Because $\mathcal{F}$ is a sigma-field, this tells us that things like $\{X > b\}$, $\{a < X < b\}$ etc. are also in $\mathcal{F}$.

In this course we will also allow our random variables to take the value $+\infty$ (or $-\infty$). So in general they will be functions from Ω to $\bar{\mathbb{R}} = \mathbb{R} \cup \{-\infty, \infty\}$.

Indicator random variables

The most important examples of random variables are the indicator random variables:

Let $A \in \mathcal{F}$ be an event. Then the function $\mathbb{1}_A : \Omega \to \mathbb{R}$ (or $I_A : \Omega \to \mathbb{R}$) given by

$$\mathbb{1}_A(\omega) = \begin{cases} 1, & \text{if } \omega \in A \\ 0, & \text{otherwise,} \end{cases}$$

is called the indicator of A.

Exercise: show that $\mathbb{1}_A$ is a random variable.

Distribution Functions

The function $F_X(t) = \mathbb{P}(X \leq t) = \mathbb{P}(\{\omega : X(\omega) \leq t\})$ that maps $\mathbb{R}$ to $[0, 1]$ is called the distribution function of the random variable X.

Any distribution function F

F1. is non-decreasing,

F2. is such that $F(x) \to 0$ as $x \to -\infty$ and $F(x) \to 1$ as $x \to \infty$,

F3. is 'right-continuous', that is $\lim_{h \to 0^+} F(t + h) = F(t)$ for all t.

Exercise: If $(\Omega, \mathcal{F}, \mathbb{P})$ is a probability space and $A \in \mathcal{F}$, find the distribution function of $\mathbb{1}_A$.

Distribution Functions

We say that a distribution function F is

- discrete if F only grows in jumps, i.e. there exists a countable set $A \subset \mathbb{R}$ and a function f mapping A to $[0,1]$ such that $F(x) = \sum_{y \in A : y \leq x} f(y)$.
- absolutely continuous if there exists a function f that maps $\mathbb{R}$ to $\mathbb{R}_+$ such that $F(t) = \int_{-\infty}^{t} f(u)du$.

A random variable X has a discrete distribution if F_X is discrete, and an (absolutely) continuous distribution if F is absolutely continuous.

The function f is called the probability mass function in the discrete case, and the probability density function in the absolutely continuous case.Note that there are distribution functions that are not even mixtures of the above (see e.g. Cantor Function). In other words there are random variables whose distributions are not mixtures of discrete and absolutely continuous distributions!

Examples of distributions

- Examples of discrete random variables: binomial, Poisson, geometric, negative binomial, discrete uniform http://en.wikipedia.org/wiki/Category:Discrete_distributions
- Examples of continuous random variables: normal, exponential, gamma, beta, uniform on an interval (a, b) http://en.wikipedia.org/wiki/Category:Continuous_distributions

Random Vectors

A random vector $\mathbf{X} = (X_1, ..., X_d)$ is a vector of random variables on the same probability space.
The distribution function $F_{\mathbf{X}}$ of $\mathbf{X}$ is

$$F_{\mathbf{X}}(\mathbf{t}) = \mathbb{P}(X_1 \leq t_1, \cdots, X_d \leq t_d), \ \mathbf{t} = (t_1, \cdots, t_d) \in \mathbb{R}^d.$$

One can also write this as

$$F_{\mathbf{X}}(\mathbf{t}) = \mathbb{P}(\cap_{i=1}^d \{X_i \leq t_i\}).$$

It follows that

$$\begin{aligned} &\mathbb{P}(s_1 < X_1 \leq t_1, s_2 < X_2 \leq t_2) \\ &\quad = \quad F(t_1, t_2) - F(s_1, t_2) - F(t_1, s_2) + F(s_1, s_2). \end{aligned}$$

Independent Random Variables

The random variables $X_1, \cdots, X_d$ are called independent if $F_{\mathbf{X}}(\mathbf{t}) = F_{X_1}(t_1) \times \cdots \times F_{X_d}(t_d)$ for all $\mathbf{t} = (t_1, \cdots, t_d)$.

This turns out to be equivalent to the statement that the events $\{X_1 \in I_1\}$, $\cdots$, $\{X_d \in I_d\}$ are independent for all intervals $I_1, \ldots I_d$.

If the random variables are all discrete, or the random vector is absolutely continuous (the latter is stronger than each of its coordinates being absolutely continuous) then this is equivalent to a relevant mass/density function $f_{\mathbf{X}}$ factorising as $f_{\mathbf{X}}(\mathbf{t}) = f_{X_1}(t_1) \cdots f_{X_d}(t_d)$.

Expectation

The expectation (or expected value) of a random variable X is

$$\mathbb{E}[X] = \int_{-\infty}^{\infty} x dF_X(x).$$

The integral on the right hand side is a Lebesgue-Stieltjes integral. It can be evaluated as

$$= \begin{cases} \sum_{i=1}^{\infty} x_i \mathbb{P}(X = x_i), & \text{if } X \text{ is discrete} \\ \int_{-\infty}^{\infty} x f_X(x) dx, & \text{if } X \text{ is absolutely continuous.} \end{cases}$$

In second year, we required that the integral be absolutely convergent. In this course we will allow the expectation to be infinite, provided that we never get in a situation where we have '$\infty - \infty$'.

Expectation of $g(X)$

If X is a random variable then for a measurable (nice) function g, $Y = g(X)$ is a random variable, so by definition

$$\mathbb{E}[Y] = \int_{-\infty}^{\infty} y dF_Y(y).$$

The law of the unconscious statistician says that also

$$\mathbb{E}[Y] = \mathbb{E}[g(X)] = \int_{-\infty}^{\infty} g(x) dF_X(x).$$

Properties of Expectation

- $\mathbb{E}[aX + bY] = a\mathbb{E}[X] + b\mathbb{E}[Y]$.
- If $\mathbb{P}(X \leq Y) = 1$, then $\mathbb{E}[X] \leq \mathbb{E}[Y]$.
- If $\mathbb{P}(X = c) = 1$, then $\mathbb{E}[X] = c$.
- ...
- If $0 \leq X_n \uparrow X$ pointwise in ω then $\mathbb{E}[X_n] \uparrow \mathbb{E}[X]$
- If $X_i \geq 0$ then $\mathbb{E}[\sum_{i=1}^{\infty} X_i] = \sum_{i=1}^{\infty} \mathbb{E}[X_i]$
- If $f \geq 0$ then $\mathbb{E}[\int_0^{\infty} f(X,t)dt] = \int_0^{\infty} \mathbb{E}[f(X,t)]dt$
- If $X \geq 0$ then $\mathbb{E}[X] = \int_0^{\infty} \mathbb{P}(X > t)dt = \int_0^{\infty} \mathbb{P}(X \geq t)dt$
- If $X \geq 0$ then $\mathbb{P}(X > x) \leq x^{-1}\mathbb{E}[X]$ (Markov's inequality)

Moments

- The kth moment of X is $\mathbb{E}[X^k]$.
- The kth central moment of X is $\mathbb{E}[(X - \mathbb{E}[X])^k]$.
- The variance $\mathrm{Var}(X)$ of X is the second central moment $\mathbb{E}[X^2] - (\mathbb{E}[X])^2$.
- $\mathrm{Var}(aX + b) = a^2\mathrm{Var}(X)$.
- The covariance of $\mathrm{Cov}(X, Y)$ of X and Y is $\mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])]$.
- If X and Y have finite means and are independent, then $\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$.
- $\mathrm{Var}(X + Y) = \mathrm{Var}(X) + \mathrm{Var}(Y) + 2\mathrm{Cov}(X, Y)$.

Conditioning on random variables

Let X, Y be random variables on $(\Omega, \mathcal{F}, \mathbb{P})$, and $A \in \mathcal{F}$.
The conditional probability of event A given a random variable X is written as $\mathbb{P}(A|X)$. The conditional expectation of Y given a random variable X is written as $\mathbb{E}[Y|X]$.
The official definitions are technical, but you should think of these as follows:

- $\mathbb{P}(A|X)$ is a function of X that takes the value $\mathbb{P}(A|X = x)$ on the event that $X = x$.
- $\mathbb{E}[Y|X]$ is a function of X that takes the value $\mathbb{E}[Y|X = x]$ on the event that $X = x$.

As X is a random variable, these two functions of X are also random variables.

Conditioning on random variables - technicality

A random variable has the property that $\{X \le x\} \in \mathcal{F}$ for every $x \in \mathbb{R}$. It could be however that $\mathcal{F}$ contains a lot more stuff than events regarding X. Let $\mathcal{G}_X$ denote the smallest σ-field on Ω such that X is a random variable on $(\Omega, \mathcal{G}_X, \mathbb{P})$. Suppose that $\mathbb{E}[|Y|] < \infty$. Then $\mathbb{E}[Y|X]$ is defined to be any random variable Z on $(\Omega, \mathcal{G}_X, \mathbb{P})$ such that $\mathbb{E}[Z\mathbb{1}_B] = \mathbb{E}[X\mathbb{1}_B]$ for each $B \in \mathcal{G}_X$.

- There is typically more than one random variable Z that satisfies this definition. Note that if Z and Z' both satisfy this definition then $\mathbb{P}(Z = Z') = 1$, so we generally dont't care which version of $\mathbb{E}[Y|X]$ we are working with.
- $\mathbb{P}(A|X)$ is defined to be $\mathbb{E}[\mathbb{1}_A|X]$.

Example: let $D \in \mathcal{F}$, and let $X = \mathbb{1}_D$. Suppose that Y is a random variable with $\mathbb{E}[|Y|] < \infty$ then the random variable

$$\mathbb{E}[Y|D]\mathbb{1}_D + \mathbb{E}[Y|D^c]\mathbb{1}_{D^c},$$

satisfies the definition of $\mathbb{E}[Y|X]$.

Conditioning on discrete random variables - the punchline

If X has a discrete distribution (taking values $x_1, x_2, \ldots$), and $\mathbb{E}[|Y|] < \infty$ then

$$\mathbb{E}[Y|X] = \sum_i \mathbb{E}[Y|X = x_i]\mathbb{1}_{\{X=x_i\}},$$

i.e. the right hand side satisfies the definition of $\mathbb{E}[Y|X]$.

This is nothing but the statement that if $\eta(x) = \mathbb{E}[Y|X = x]$, then $\eta(X)$ satisfies the definition of $\mathbb{E}[Y|X]$.

Conditional Distribution

The conditional distribution function $F_{Y|X}(y|X)$ of Y evaluated at the real number y is given by $\mathbb{P}(Y \leq y|X)$

Properties of Conditional Expectation

The following hold with probability 1:

- Linearity: $\mathbb{E}[aY_1 + bY_2|X] = a\mathbb{E}[Y_1|X] + b\mathbb{E}[Y_2|X]$,
- Monotonicity: If $\mathbb{P}(Y_1 \leq Y_2) = 1$, then $\mathbb{E}[Y_1|X] \leq \mathbb{E}[Y_2|X]$,
- $\mathbb{E}[c|X] = c$,
- $\mathbb{E}[\mathbb{E}[Y|X]] = \mathbb{E}[Y]$,
- For any nice (i.e. Borel measurable) function g, $\mathbb{E}[g(X)Y|X] = g(X)\mathbb{E}[Y|X]$
- $\mathbb{E}[Y|X]$ is the function of X that is closest to Y in the mean square sense. This means that $\mathbb{E}[(g(X) - Y)^2]$ is minimised when $g(X) = \mathbb{E}[Y|X]$ (see Borovkov, page 57).

Exercise

Let $\Omega = \{a, b, c, d\}$, and let $\mathcal{F}$ contain all subsets of Ω.
Let $\mathbb{P}$ be the probability measure satisfying
$\mathbb{P}(\{a\}) = \frac{1}{2}$, $\mathbb{P}(\{b\}) = \mathbb{P}(\{c\}) = \frac{1}{8}$ and $\mathbb{P}(\{d\}) = \frac{1}{4}$.

Define random variables,

$$\begin{aligned} Y(\omega) &= \begin{cases} 1, & \omega = a \text{ or } b, \\ 0, & \omega = c \text{ or } d, \end{cases} \\ X(\omega) &= \begin{cases} 2, & \omega = a \text{ or } c, \\ 5, & \omega = b \text{ or } d. \end{cases} \end{aligned}$$

Compute $\mathbb{E}[X]$, $\mathbb{E}[X|Y]$ and $\mathbb{E}[\mathbb{E}[X|Y]]$.

Example:

Suppose also that the number of individuals M entering a bank in a given day has a Poisson(λ) distribution.
Suppose that individuals entering the bank each hold an Australian passport with probability p, independently of each other and M.
Let N denote the number of individuals holding an Australian passport who enter the bank during that day.

- What is the distribution of N, given that $M = m$?
- Find $\mathbb{E}[N|M = m]$.
- Give an expression for $\mathbb{E}[N|M]$, simplifying where possible.
- Compute $\mathbb{E}[N]$.

Law of Large Numbers

The Law of Large Numbers (LLN) states that if $X_1, X_2, \cdots$ are independent and identically-distributed with mean μ, then with probability 1

$$\overline{X}_n := \frac{1}{n}\sum_{j=1}^{n} X_j \to \mu$$

as $n \to \infty$.

Central Limit Theorem

The Central Limit Theorem (CLT) states that if $X_1, X_2, \cdots$ are independent and identically-distributed with mean μ and variance $\sigma^2 \in (0, \infty)$, then for any x,

$$\mathbb{P}\left(\frac{\overline{X}_n - \mu}{\sigma/\sqrt{n}} \leq x\right) \to \Phi(x) \equiv \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-t^2/2} dt$$

as $n \to \infty$.

That is, a suitably-scaled variation from the mean approaches a standard normal distribution as $n \to \infty$.

(Note that writing $Z_n := \frac{\overline{X}_n - \mu}{\sigma/\sqrt{n}}$, this becomes

$$F_{Z_n}(x) \to F_Z(x),$$

for each x, where $Z \sim \mathcal{N}(0, 1)$.)

Limit Theorems

The Poisson Limit Theorem states that that if X_1, X_2, $\cdots$ are independent Bernoulli random variables with $\mathbb{P}(X_i = 1) = 1 - \mathbb{P}(X_i = 0) = p_i$, then $X_1 + X_2 + \cdots + X_n$ is well-approximated by a Poisson random variable with parameter $\lambda_n = p_1 + \cdots + p_n$.
Specifically, with $W_n = X_1 + X_2 + \cdots + X_n$, then, for each $x \in \mathbb{R}$

$$F_{W_n}(x) \approx F_{Y_n}(x)$$

where $Y_n \sim \mathrm{Poisson}(\lambda_n)$.
(There is, in fact, a bound on the accuracy of this approximation

$$|\mathbb{P}(W_n \in B) - \mathbb{P}(Y_n \in B)| \leq \frac{\sum_{i=1}^{n} p_i^2}{\max(1, \lambda)},$$

)

Example

Suppose there are three ethnic groups, A (20%), B (30%) and C (50%), living in a city with a large population. Suppose 0.5%, 1% and 1.5% of people in A, B and C respectively are over 200cm tall.

Suppose that we select at random 50, 50, and 200 people from A, B, and C respectively. What is the probability that at least four of the 300 people will be over 200cm tall?

Stochastic Processes

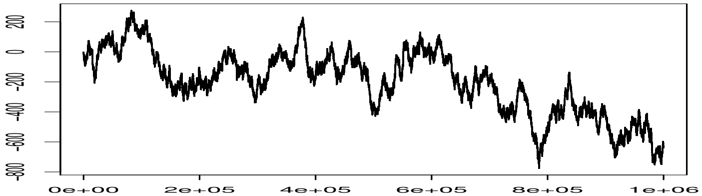
A collection of random variables $\{X_t,\ t \in I\}$ (or $\{X(t), t \in I\}$, or $(X_t)_{t \in I} \ldots$) on a common prob space $(\Omega, \mathcal{F}, \mathbb{P})$ is called a stochastic process. The index variable t is often called 'time'.

- If $I = \{0, 1, 2, \cdots\}$ or $\{\cdots, -2, -1, 0, 1, 2, \cdots\}$, the process is a discrete time process.
- If $I = \mathbb{R}$ or $[0, \infty)$, the process is a continuous time process.

Examples of Stochastic Processes

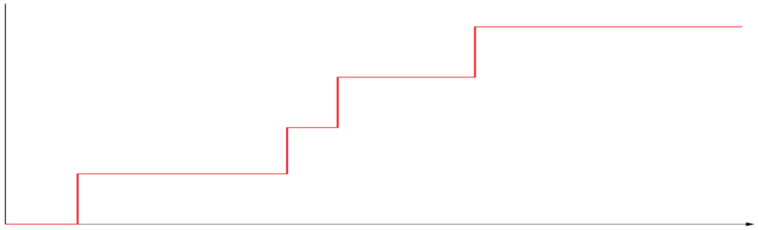
Standard Brownian Motion is a very special *Gaussian process* $(X_t)_{t \in [0,\infty)}$ where

- For each $t \geq 0$, $X_t \sim \mathcal{N}(0, t)$
- For any $0 \leq s_1 < t_1 \leq s_2 < \cdots \leq s_k < t_k$, $X_{t_1} - X_{s_1}$, ..., $X_{t_k} - X_{s_k}$ are independent.



Examples of Stochastic Processes

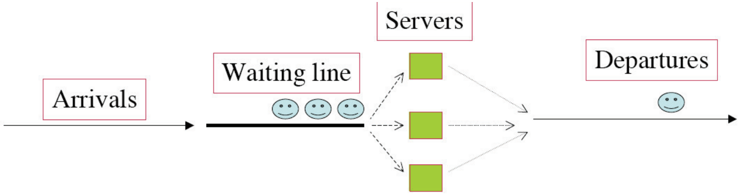
If e.g. X_t is the number of sales of an item up to time t,

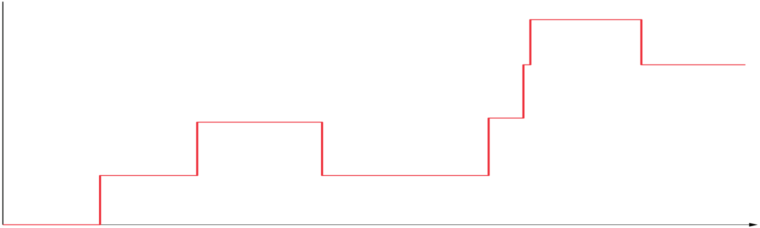


then $(X_t)_{t \geq 0}$ is called a counting process.

Examples of Stochastic Processes

X_t is the number of people in a queue at time t.





Interpretations

We can think of Ω as consisting of the set of sample paths $\omega = \{\omega_t : t \in I\}$, that is a set of sequences if I is discrete or a set of functions if I is continuous. Each $\omega \in \Omega$ has a value ω_t at each time point $t \in I$. With this interpretation,

- If both ω and t are fixed, then $X_t(\omega)$ is a real number.
- For a fixed t, the function $X_t : \Omega \to \mathbb{R}$ is a random variable.
- For a fixed ω, we can think of t as a variable, and $X_\cdot(\omega) : I \to \mathbb{R}$ as a function (realization, trajectory, sample path).
- If we allow ω to vary, we get a collection of trajectories.

Interpretations

If X_t is a counting process:

- For fixed ω and t, $X_t(\omega)$ is a non-negative integer.
- For fixed ω, $X_.(\omega)$ is a non-decreasing step function of t.
- For fixed t, X_t is a non-negative integer-valued random variable.
- For $s < t$, $X_t - X_s$ is the number of events that have occurred in the interval $(s, t]$.

If X_t is the number of people in a queue at time t, then $\{X_t : t \geq 0\}$ is a stochastic process where, for each t, X_t is a non-negative integer-valued random variable but it is NOT a counting process because, for fixed ω, $X_t(\omega)$ can decrease.

Finite-Dimensional Distributions

Knowing just the one-dimensional distributions (i.e. the distribution of X_t for each t) is not enough to describe a stochastic process.

To specify the complete distribution of a stochastic process $(X_t)_{t \in I}$, we need to know the finite-dimensional distributions , i.e. the family of joint distribution functions $F_{t_1, t_2, \cdots, t_k}(x_1, \cdots, x_k)$ of $X_{t_1}, \cdots, X_{t_k}$ for all $k \geq 1$ and $t_1, \cdots, t_k \in I$.

Discrete-Time Markov Chains

We are frequently interested in applications where we have a sequence $X_1, X_2, \cdots$ of outputs (which we model as random variables) in discrete time. For example,

- DNA: A (adenine), C (cytosine), G (guanine), T (thymine).
- Texts: X_j takes values in some alphabet, for example $\{A, B, \cdots, Z, a, \cdots\}$.
  - Developing and testing compression software.
  - Cryptology: codes, encoding and decoding.
  - Attributing manuscripts.

Independence?

Is it reasonable to assume that neighbouring letters are independent?

No, e.g. in English texts:

- *a* is very common, but *aa* is very rare.
- *q* is virtually always followed by *u* (and then another vowel).

The Markov Property

The Markov property embodies a natural first generalisation to the independence assumption. It assumes a kind of one-step dependence or memory. Specifically, for $I = \mathbb{Z}_+ = \{0, 1, 2, \dots\}$ (and discrete-valued) processes the Markov property takes the form

$$
\begin{aligned}
&\mathbb{P}(X_{n+1} = y | X_n = x_n, X_{n-1} = x_{n-1}, \cdots, X_0 = x_o) \\
=&\mathbb{P}(X_{n+1} = y | X_n = x_n)
\end{aligned}
$$

for each $n \in \mathbb{Z}_+$.

Discrete stopping times

Let $(X_t)_{t\in\mathbb{Z}_+}$ be a sequence of random variables. A random variable T taking values in $\mathbb{Z}_+ \cup \{+\infty\}$ is called a stopping time with respect to $(X_t)_{t\in\mathbb{Z}_+}$ if for each $t \in \mathbb{Z}_+$, there exists a non-random function h_t such that

$$\mathbb{1}_{\{T\le t\}} = h_t(X_0, \dots, X_t).$$

This says that we can determine whether or not $T \le t$, knowing only the values of $X_0, \dots, X_t$ (i.e. knowing about the past and the present but without knowing the future).

Example: First hitting time. For a sequence $(X_t)_{t\in\mathbb{Z}_+}$, let $T(x) = \inf\{t \in \mathbb{Z}_+ : X_t = x\}$ denote the first time that the sequence is equal to x. Then

$$\mathbb{1}_{\{T(x)\le t\}} = \sum_{m=0}^{t} \mathbb{1}_{\{T(x)=m\}} = \sum_{m=0}^{t} \mathbb{1}_{\{X_m=x\}} \prod_{i=0}^{m-1} \mathbb{1}_{\{X_i\neq x\}},$$

so $T(x)$ is a stopping time.

Examples:

We have seen above that for a sequence $(X_t)_{t\in\mathbb{Z}_+}$, the first hitting times $T(x) = \inf\{t \in \mathbb{Z}_+ : X_t = x\}$ are stopping times. The following are also stopping times:

- First strictly positive hitting times: $T'(x) = \inf\{t \in \mathbb{N} : X_t = x\}$.
- ith hitting times: $T_1(x) = T(x)$, $T_i(x) = \inf\{t > T_{i-1}(x) : X_t = x\}$.
- The maximum or minimum of two stopping times.
- A non-random time.

Something like $T - 1$ is not in general a stopping time. E.g. for a Bernoulli sequence, if T is the first time we see a 1 in the sequence then $T - 1$ is not a stopping time. Why?

Continuous stopping times:

A stopping time for a continuous-time process $(X_t)_{t \geq 0}$ is a random variable T (with values in $[0, \infty]$) such that for each $t < \infty$ there is a non-random (measureable) function h_t such that

$$\mathbb{1}_{\{T \leq t\}} = h_t((X_u)_{u \in [0,t]}).$$

As in the discrete setting, hitting times etc., are stopping times.

Discrete time Markov chains

Discrete-Time Markov Chains

A sequence $(X_t)_{t\in\mathbb{Z}_+}$ of discrete random variables forms a DTMC if

$$\begin{aligned}&\mathbb{P}(X_{t+1} = x_{t+1} | X_t = x_t, X_{t-1} = x_{t-1}, \cdots, X_0 = x_0) \\ =&\mathbb{P}(X_{t+1} = x_{t+1} | X_t = x_t),\end{aligned} \tag{1}$$

for all $t, x_0, \ldots, x_{t+1}$ such that the left hand side is well defined.

Let

$$p_{i,j}(t) := \mathbb{P}(X_{t+1} = j | X_t = i).$$

We will assume that (as long as they are well defined) the transition probabilities $p_{i,j}(t)$ do not depend on t, in which case the DTMC is called time homogeneous and we write $p_{i,j} := p_{i,j}(t)$.

The Markov property (1) above can then be rewritten as:

$$\mathbb{P}(X_{t+1} = j | X_t = i, X_{t-1} = x_{t-1}, \cdots, X_0 = x_0) = p_{i,j},$$

for all $t, i, j, x_0, \ldots, x_{t-1}$ such that the left hand side is well defined.

A more general picture

Henceforth all our DTMCs are time-homogeneous.

One can infer from the Markov property that

$$\mathbb{P}(\cap_{i=0}^{n}\{X_i = x_i\}) = \mathbb{P}(X_0 = x_0)\prod_{i=0}^{n-1} p_{x_i,x_{i+1}}.$$

If the left hand side is positive then this follows by recursive conditioning. If the left hand side is 0 then either $\mathbb{P}(X_0 = x_0) = 0$ or there is a smallest $m \in [1, n]$ such that $\mathbb{P}(\cap_{i=0}^{m}\{X_i = x_i\}) = 0$, and the recursive conditioning applies to this probability.
More generally the Markov property implies (assuming that these are well-defined)

$$\begin{aligned}&\mathbb{P}\big((X_{n+1}, X_{n+2}, \dots, X_{n+k}) \in B \big| X_n = x, (X_{n-1}, \dots, X_0) \in A\big)\\ =&\mathbb{P}\big((X_{n+1}, X_{n+2}, \dots, X_{n+k}) \in B \big| X_n = x\big).\end{aligned}$$

I.e. if you know the present state, then receiving information about the past tells you nothing more about the future.

Transition matrix

If the *state space* $\mathcal{S}$ (i.e. the set of possible values that the elements of the sequence can take) has $m \in \mathbb{N}$ elements, then by relabelling the values if necessary, we may assume that $\mathcal{S} = \{1, 2, \ldots, m\}$.

For a DTMC, we define the (one step)-transition matrix to be a matrix with rows and columns corresponding to the states of the process and whose *ij*-th entry is $p_{i,j}$. So

$$P = \begin{pmatrix} p_{11} & p_{12} & \cdots & p_{1m} \\ p_{21} & p_{22} & \cdots & p_{2m} \\ \cdots & \cdots & \cdots & \cdots \\ p_{m1} & p_{m2} & \cdots & p_{mm} \end{pmatrix}.$$

Transition matrix

For a transition matrix of a DTMC:

- Each entry is ≥ 0.
- Each row sums to 1.

Any square matrix having these two properties is called a stochastic matrix.

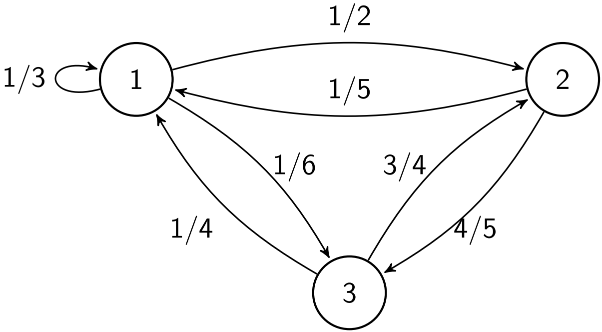
(If the state space is infinite we will still refer to the infinite matrix containing the $p_{i,j}$ as the transition matrix. It is also a stochastic matrix (albeit an infinite one).)

Transition diagram

We can associate a weighted "directed graph" (called the transition diagram) with a stochastic matrix by letting the nodes correspond to states and putting in an arc/edge jk with weight $p_{jk} > 0$ on it. Example: if

$$P = \begin{pmatrix} 1/3 & 1/2 & 1/6 \\ 1/5 & 0 & 4/5 \\ 1/4 & 3/4 & 0 \end{pmatrix},$$

then the transition diagram can be drawn as follows.



Examples

- Suppose that the $(X_t)_{t \in \mathbb{Z}_+}$ are i.i.d. random variables with $\mathcal{S} = \{1, \ldots, k\}$ and $\mathbb{P}(X_t = i) = p_i$. What does the transition matrix look like?
- A communication system transmits the digits 0 and 1 at discrete times. Let X_t denote the digit transmitted at time t. At each time point, there is a probability p that the digit will not change and prob $1 - p$ it will change. Find the transition matrix and draw the transition diagram.

Examples

- Suppose that whether or not it rains tomorrow depends on previous weather conditions only through whether or not it is raining today. Suppose also that if it rains today, then it will rain tomorrow with probability p and if it does not rain today, then it will rain tomorrow with probability q.If we say that $X_t = 1$ if it rains on day t and $X_t = 0$ otherwise, then $(X_t)_{t \in \mathbb{Z}_+}$ is a two-state Markov chain.
- Let $(Y_i)_{i \in \mathbb{N}}$ be i.i.d. random variables with $\mathbb{P}(Y_i = 1) = p$ and $\mathbb{P}(Y_i = -1) = 1 - p$. Let $S_0 = 0$ and $S_n = \sum_{i=1}^{n} Y_i$ for each $n \in \mathbb{N}$.Then $(S_n)_{n \in \mathbb{Z}_+}$ is a Markov chain called a simple random walk. If $p = 1/2$ it is called a symmetric (or unbiased) simple random walk.

n-step transition probabilities

The n-step transition probabilities $\mathbb{P}(X_{t+n} = j | X_t = i)$ of a (time-homogeneous) DTMC do not depend on t. For $n = 1, 2, \cdots$, we denote them by

$$p_{i,j}^{(n)} = \mathbb{P}(X_{t+n} = j | X_t = i).$$

It is also convenient to use the notation

$$p_{i,j}^{(0)} := \begin{cases} 1 & \text{if } j = i \\ 0 & \text{if } j \neq i. \end{cases}$$

Chapman-Kolmogorov equations

The Chapman-Kolmogorov equations show how we can calculate the n-step transition probabilities $p_{i,j}^{(n)}$ from smaller-step transition probabilities. For $n = 1, 2, \cdots$ and any $r = 1, 2, \cdots, n$,

$$p_{i,j}^{(n)} = \sum_{k \in \mathcal{S}} p_{ik}^{(r)} p_{kj}^{(n-r)}.$$

Interpretation: *In order to go from i to j in n steps, you have to be somewhere after r steps!*

n-step transition matrix

If we define the n-step transition matrix as

$$P^{(n)} = \begin{pmatrix} p_{11}^{(n)} & p_{12}^{(n)} & \ddots & \ddots \\ p_{21}^{(n)} & p_{22}^{(n)} & p_{23}^{(n)} & \ddots \\ \ddots & \ddots & \ddots & \ddots \end{pmatrix},$$

then the Chapman-Kolmogorov equations can be written in the matrix form

$$P^{(n)} = P^{(r)} P^{(n-r)}$$

with $P^{(1)} = P$.By mathematical induction, it follows that

$$P^{(n)} = P^n,$$

the nth power of P.

Distribution of a DTMC

How do we determine the distribution of a DTMC?

To uniquely determine the distribution we need 2 ingredients:

- the *initial distribution* $\boldsymbol{\pi}^{(0)} = (\pi_i^{(0)})_{i \in \mathcal{S}}$, where for each $j \in \mathcal{S}$, $\pi_j^{(0)} = \mathbb{P}(X_0 = j)$, and
- the transition matrix P.

In principle, we can use these and the Markov property to derive the finite dimensional distributions, e.g.

$$\mathbb{P}(X_0 = x_0, X_{t_1} = x_1, X_{t_2} = x_2, \cdots, X_{t_k} = x_k) = \pi_{x_0}^{(0)} p_{x_0,x_1}^{(t_1)} \cdots p_{x_{k-1},x_k}^{(t_k - t_{k-1})},$$

although the calculations are often intractable.

Example:

Suppose $\mathbb{P}(X_0 = 1) = 1/3$, $\mathbb{P}(X_0 = 2) = 0$, $\mathbb{P}(X_0 = 3) = 1/2$, $\mathbb{P}(X_0 = 4) = 1/6$ and

$$P = \begin{pmatrix} 1/4 & 0 & 1/4 & 1/2 \\ 1/4 & 1/4 & 1/4 & 1/4 \\ 0 & 0 & 2/3 & 1/3 \\ 1/2 & 0 & 1/2 & 0 \end{pmatrix}.$$

- Find the distribution of X_1.
- Find $\mathbb{P}(X_{n+2} = 2 | X_n = 4)$.
- Find $\mathbb{P}(X_3 = 2, X_2 = 3, X_1 = 1)$.

A fundamental question

What proportion of time does the chain spend in each state in the long run?
(and when does this question even make sense?)

To answer this appropriately we need to introduce a lot of concepts!

Classification of states/chains

Here are some definitions.

- State j is accessible from state i, denoted by $i \to j$, if there exists an $n \geq 0$ such that $p_{i,j}^{(n)} > 0$.That is, either $j = i$ or we can get from i to j in a finite number of steps.
- If $i \to j$ and $j \to i$, then states i and j communicate, denoted by $i \leftrightarrow j$.
- A state j is an absorbing state if $p_{jj} = 1$.

Example

Draw a transition diagram associated with the transition matrix below and determine which states communicate with each other. Are there any absorbing states?

$$P = \begin{pmatrix} 0 & 0.5 & 0.5 & 0 \\ 0.5 & 0 & 0 & 0.5 \\ 0 & 0 & 0.5 & 0.5 \\ 0 & 0 & 0.5 & 0.5 \end{pmatrix}$$

The communication relation

The communication relation $\leftrightarrow$ has the properties:

- $j \leftrightarrow j$ (reflexivity),
- $j \leftrightarrow k$ if and only if $k \leftrightarrow j$ (symmetry), and
- if $j \leftrightarrow k$ and $k \leftrightarrow i$, then $j \leftrightarrow i$ (transitivity).

A relation that satisfies these properties is known as an equivalence relation.

Communicating classes and irreducibility

Consider a set $\mathcal{S}$ whose elements can be related to each other via any equivalence relation $\Leftrightarrow$.

Then $\mathcal{S}$ can be partitioned into a collection of disjoint subsets $\mathcal{S}_1, \mathcal{S}_2, \mathcal{S}_3, \ldots, \mathcal{S}_M$ (where M might be infinite) such that $j, k \in \mathcal{S}_m$ if and only if $j \Leftrightarrow k$.

So the state space $\mathcal{S}$ of a DTMC is partitioned into communicating classes by the communication relation $\leftrightarrow$.

If a DTMC has only one communicating class (i.e., all states communicate) then it is called an irreducible DTMC. Otherwise it is called reducible.

Hitting and return probabilities

Let

$$h_{i,j} = \mathbb{P}(j \in \{X_0, X_1, \dots\} | X_0 = i),$$

which is the probability that we ever reach state j, starting from state i. The quantities $h_{i,j}$ are referred to as hitting probabilities. Note that

- $h_{i,i} = 1$
- $h_{i,j} > 0$ if and only if $i \to j$.

Let

$$f_i = \mathbb{P}(i \in \{X_1, X_2, \dots\} | X_0 = i),$$

which is the probability that we ever return to state i, starting from i. The quantities f_i are referred to as return probabilities.

Hitting and return times

Let $T(j) = \inf\{n \geq 0 : X_n = j\}$ denote the first hitting time of j (it is a stopping time). Then

$$m_{i,j} := \mathbb{E}[T(j)|X_0 = i],$$

is the expected time to reach state j starting from state i.By definition

- $m_{j,j} = 0$,
- if $h_{i,j} < 1$ then $\mathbb{P}(T(j) = \infty|X_0 = i) > 0$ so $m_{i,j} = \infty$,
- if $h_{i,j} = 1$ then $m_{i,j}$ may or may not be finite.

Let $T^+(i) = \inf\{n > 0 : X_n = i\}$. Then

$$\mu_i = \mathbb{E}[T^+(i)|X_0 = i],$$

is the expected time to return to state i.

The strong Markov property

The Markov property as defined, holds at each *fixed time*. We would like the same to be true at certain random times such as hitting times. The property that we need (which can be proved from the Markov property) is called the

Strong Markov property: Let $(X_t)_{t\in\mathbb{Z}_+}$ be a (time-homogeneous) DTMC, and T be a *stopping time* for the chain. Then

$$\begin{aligned}\mathbb{P}(X_{T+1} = j | T = t, X_0 = x_0, \ldots, X_T = i) \\ = \mathbb{P}(X_{T+1} = j | T < \infty, X_T = i) = p_{i,j}.\end{aligned}$$

This says that looking at the next step of a Markov chain at a stopping time is the same as starting the process from the random state X_T (provided that T is finite). As with the ordinary Markov property, this can be generalized to handle general future events etc.

Recurrence and transience

One can classify individual *states* as *recurrent* or *transient*. We will be mostly interested in applying such a classification to *irreducible chains* where we can avoid some technicalities.

Definition: An irreducible DTMC is recurrent if $h_{i,j} = 1$ for every $i, j \in \mathcal{S}$. Otherwise it is transient.

Characterizing recurrence

Let $\Delta_i(j)$ be the time between the $(i+1)$st and ith visit to state j, and let $N(j)$ be the number of visits to state j. Suppose that $X_0 = j$.

If $f_j = 1$ then each $\Delta_i(j)$ is finite since the chain is certain to return to j in finite time, and $N(j) = \infty$.

If $f_j < 1$ then with probability $1 - f_j > 0$ it will never return to j. From the Markov property we see that $N(j)$ has a geometric distribution.
Specifically, for $n \geq 1$,

$$\mathbb{P}(N(j) = n | X_0 = j) = f_j^{n-1}(1 - f_j).$$

This implies that $\gamma_j := \mathbb{E}[N(j)|X_0 = j] = \frac{1}{1-f_j} < \infty$.

Characterizing recurrence

Theorem: For an irreducible DTMC the following are equivalent:

(i) the chain is recurrent

(ii) $f_i = 1$ for every $i \in \mathcal{S}$

(iii) $f_i = 1$ for some $i \in \mathcal{S}$

(iv) $\gamma_i = \infty$ for some $i \in \mathcal{S}$

(v) $\gamma_i = \infty$ for every $i \in \mathcal{S}$

(Note that if an irreducible chain is transient, then it visits each state only a finite number of times, so the limiting proportion of time spent in each state is 0).

Characterizing recurrence

$(i) \Rightarrow (ii)$: $\qquad f_i = \sum_{j \in \mathcal{S}} p_{i,j} h_{j,i} = \sum_{j \in \mathcal{S}} p_{i,j} 1 = 1.$

$(iii) \Rightarrow (ii)$:Suppose that $f_i = 1$, and let $j \in \mathcal{S}$. Since $i \to j$, we have that for some n, $p_{i,j}^{(n)} > 0$. Let n_0 be the smallest such n. Then with probability $\zeta = p_{i,j}^{(n_0)} > 0$ the chain walks from i to j in n_0 steps (without revisiting i) . Since $f_i = 1$ we must have $h_{j,i} = 1$. Now start the chain from j, and since $h_{j,i} = 1$ the chain eventually hits i. Each time the chain visits i it has probability $\eta \geq \zeta$ of reaching j before returning to i. By the Markov property, the number of returns to i before hitting j has a geometric distribution with parameter η, so it is finite. Thus the chain must eventually return to j, so $f_j = 1$.

$(iii) \Leftrightarrow (iv)$: essentially done 2 slides ago.

others: exercise

Irreducible finite-state chains are recurrent

(We will state a stronger result later)

Corollary: Irreducible finite-state chains are recurrent.

Proof: Let $\mathcal{S} = \{1, \ldots, k\}$. Let $N(j)$ denote the number of visits to j.
Now $\sum_{j=1}^{k} N(j) = \infty$, so the expectation is also infinite.
If $\mathbb{E}[N(j)|X_0 = i] < \infty$ for each j then $\mathbb{E}[\sum_{j=1}^{k} N(j)|X_0 = i] < \infty$, which gives a contradiction.
Hence $\gamma_j \geq \mathbb{E}[N(j)|X_0 = i] = \infty$ for some j. $\square$

Simple Random Walk

Recall that $S_0 = 0$ and $S_n = \sum_{i=1}^n Y_i$, for $n \geq 1$, where $(Y_i)_{i \in \mathbb{N}}$ are i.i.d. random variables with $\mathbb{P}(Y_i = 1) = p = 1 - \mathbb{P}(Y_i = -1)$.

S_n is a sum of i.i.d. random variables. By the law of large numbers we have $S_n/n \to \mathbb{E}[Y_1] = 2p - 1$. Therefore:

If $p > 1/2$ then the random walk escapes to $+\infty$, and so the number of visits to any state is finite. This implies that the chain is transient (and similarly if $p < 1/2$).

Let us prove this another way (and also handle the case $p = 1/2$, for which the LLN does not give us the answer).

Simple Random Walk

Note that the number of visits to 0 satisfies

$$N(0) = \sum_{m=0}^{\infty} \mathbb{1}_{\{S_m = 0\}},$$

so (since the walk starts at 0)

$$\mathbb{E}[N(0)] = \sum_{m=0}^{\infty} \mathbb{E}[\mathbb{1}_{\{S_m=0\}}] = \sum_{m=0}^{\infty} \mathbb{P}(S_m = 0) = \sum_{m=0}^{\infty} p_{0,0}^{(m)}.$$

Note that $p_{j,j}^{(m)} = 0$ if m is odd. If $m = 2n$ then

$$p_{j,j}^{(m)} = p_{j,j}^{(2n)} = \binom{2n}{n} p^n (1-p)^n.$$

Simple Random Walk

Stirling's formula $n! \approx \sqrt{2\pi n} n^n e^{-n}$ gives us the fact that

$$p_{j,j}^{(2n)} \approx \frac{(4p(1-p))^n}{\sqrt{n\pi}},$$

and the series $\sum_{n=0}^{\infty} p_{j,j}^{(2n)}$

- diverges if $4p(1-p) = 1$ (i.e. $p = 1/2$), so the DTMC is recurrent
- converges if $4p(1-p) < 1$ (i.e. $p \neq 1/2$) (compare to geometric series), so the DTMC is transient.

We shall see another way of proving this same result later.

Periodicity

The simple random walk illustrates another phenomenon that can occur in DTMCs - periodicity.

Definition: For a DTMC, a *state* $i \in \mathcal{S}$ has period $d(i) \geq 1$ if $\{n \geq 1 : p_{i,i}^{(n)} > 0\}$ is non-empty and has greatest common divisor $d(i)$.

If state j has period 1, then we say that it is aperiodic (otherwise it is called periodic).

It turns out that if $i \leftrightarrow j$ then $d(i) = d(j)$, thus we can make the following definition:

Definition: An irreducible DTMC is periodic with period d if any (hence every) state has period $d > 1$. Otherwise it is aperiodic.

Examples

- The simple random walk is periodic with period $d = 2$.
- Which of the following are transition matrices for periodic chains?

$$P = \begin{pmatrix} 0 & 0.5 & 0.5 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix} \qquad P = \begin{pmatrix} 0 & 0 & 0.5 & 0.5 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix}$$

States in a communicating class have same period

Assume that state j has period $d(j)$ and $j \leftrightarrow k$. Then, as before, there must exist s and t such that $p_{j,k}^{(s)} > 0$ and $p_{k,j}^{(t)} > 0$. We know straight away that $d(j)$ divides $s + t$ since it is possible to go from j to itself in $s + t$ steps.

Now take a path from k to itself in r steps. If we concatenate our path from j to k in s steps, this r step path, and our path from from k to j in t steps, we have an $s + r + t$ step path from j to itself. So $d(j)$ divides $s + r + t$ which means that $d(j)$ divides r. So the $d(j)$ divides the period $d(k)$ of k.

Now we can switch j and k in the argument to conclude that $d(k)$ divides $d(j)$ which means that $d(j) = d(k)$, and all states in the same communicating class have a common period.

Computing hitting probabilities

For $i \in \mathcal{S}$ and $A \subset \mathcal{S}$, let $h_{i,A}$ denote the probability that the chain ever visits a state in A, starting from state i. If $A = \{j\}$ is a single state then we have seen this before: $h_{i,\{j\}} = h_{i,j}$.

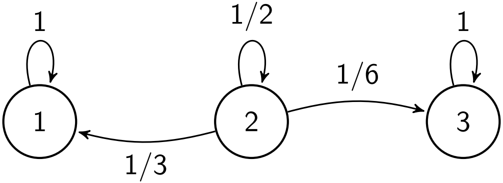
Let $T(A)$ denote the first time we reach a state in A. Then
$h_{i,A} = \mathbb{P}(T(A) < \infty | X_0 = i)$.
We have

- if $i \in A$ then $h_{i,A} = 1$
- $h_{i,A} > 0$ if and only if $i \to j$ for some $j \in A$
- $h_{i,A} = \sum_{j \in \mathcal{S}} p_{i,j} h_{j,A}$ if $i \notin A$.

This is a set of linear equations.

A simple example

Consider a Markov chain with $\mathcal{S} = \{1, 2, 3\}$, and $p_{11} = p_{33} = 1$ and $p_{22} = 1/2$, $p_{21} = 1/3$, $p_{23} = 1/6$. This Markov chain has transition diagram



The chain has (two) absorbing states. Find $h_{i,1}$ for each i.

Clearly $h_{1,1} = 1$, and $h_{3,1} = 0$. Finally,

$$h_{2,1} = \frac{1}{3}h_{1,1} + \frac{1}{2}h_{2,1} + \frac{1}{6}h_{3,1} = \frac{1}{3} + \frac{1}{2}h_{2,1}.$$

So $h_{2,1} = 2/3$.

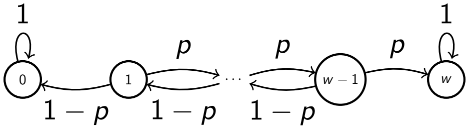
(There is another easy way to see this)

Example: Gambler's ruin

Starting with \$i, a gambler makes repeated bets of \$1 on a game of chance that she has probability p of winning on each attempt (independent of the past). She stops as soon as she reaches $\$w$ (where $w > i$) or \$0.

What is the probability that the gambler ends up with \$0?

Let $(X_n)_{n\geq 0}$ be a DTMC with state space $\mathcal{S} = \{0, 1, \ldots, w\}$, and transition probabilities $p_{i,i+1} = p$ and $p_{i,i-1} = 1 - p$ if $0 < i < w$ and $p_{0,0} = 1 = p_{w,w}$. The transition diagram is



Find the hitting probabilities $(h_{i,0})_{i\in\mathcal{S}}$.

Example: Gambler's ruin

We have $h_{0,0} = 1$, $h_{w,0} = 0$ and otherwise

$$h_{i,0} = ph_{i+1,0} + (1-p)h_{i-1,0}.$$

Rearrange to get

$$p(h_{i,0} - h_{i+1,0}) = (1-p)(h_{i-1,0} - h_{i,0}).$$

Let $u_i = h_{i-1,0} - h_{i,0}$, and $\alpha = (1-p)/p$. Then for $1 \leq i < w$

$$u_{i+1}p = u_i(1-p).$$

Thus, $u_{i+1} = \alpha u_i$ and $u_1 = 1 - h_{1,0}$. It follows that for $k \leq w$

$$u_k = \alpha^{k-1}u_1.$$

Example: Gambler's ruin

$$u_k = \alpha^{k-1} u_1.$$

On the other hand,

$$h_{i,0} = h_{0,0} + \sum_{m=1}^{i} (h_{m,0} - h_{m-1,0}) = 1 - u_1 \sum_{m=0}^{i-1} \alpha^m.$$

Since $h_{w,0} = 0$ we have that

$$u_1 = \left(\sum_{m=0}^{w-1} \alpha^m \right)^{-1},$$

and therefore

$$h_{i,0} = 1 - \frac{\sum_{r=0}^{i-1} \alpha^r}{\sum_{m=0}^{w-1} \alpha^m} = \frac{\sum_{r=i}^{w-1} \alpha^r}{\sum_{m=0}^{w-1} \alpha^m}.$$

Example: Gambler's ruin

So

$$h_{i,0}(w) = 1 - \frac{\sum_{r=0}^{i-1}\left(\frac{1-p}{p}\right)^r}{\sum_{m=0}^{w-1}\left(\frac{1-p}{p}\right)^m} = \frac{\sum_{r=i}^{w-1}\left(\frac{1-p}{p}\right)^r}{\sum_{m=0}^{w-1}\left(\frac{1-p}{p}\right)^m}.$$

If $p = 1/2$ we get

$$h_{i,0} = \frac{w-i}{w}.$$

Exercise: check that these do indeed satisfy the equations that we started with!

Exercise: find $\lim_{w\to\infty} h_{1,0}(w)$.

Example: Gambler's ruin with Martingales

When $p = 1/2$ in the Gambler's ruin problem, the chain is a *Martingale:*

$$\mathbb{E}[X_{n+1}|X_n] = X_n.$$

It is *bounded* because $0 \le X_n \le w$ for each n.
Let $T = T(\{0, w\})$ denote the first time that we hit 0 or w. Then since $(X_n)_{n\ge0}$ is a bounded Martingale

$$\mathbb{E}[X_T] = \mathbb{E}[X_0].$$

If we start with $X_0 = i$ then $\mathbb{E}[X_0] = i$. Also, $\mathbb{P}(X_T = 0) = h_{i,0}$ and $\mathbb{P}(X_T = w) = 1 - h_{i,0}$ so

$$i = \mathbb{E}[X_T] = 0 \times h_{i,0} + w \times (1 - h_{i,0}).$$

Solving gives $h_{i,0} = \frac{w-i}{w}$ as before.

Example: Simple random walk

Consider the simple random walk, which is irreducible. Let $A = \{0\}$. Then $h_{0,0} = 1$ and $h_{i,0} = ph_{i+1,0} + (1-p)h_{i-1,0}$ for $i \neq 0$. Then $h_{i,0} = 1$ for all i satisfies these equations, regardless of p. But we know that the walk is transient unless $p = 1/2$, so how is this possible?

If $\mathcal{S}$ is infinite then there need not be a unique solution to these equations.

Above we have not found the solution that we are looking for (except when $p = 1/2$).

Example: Simple random walk

If $i > 0$ then

$$h_{i,0} = ph_{i+1,0} + (1-p)h_{i-1,0}.$$

Let $x = h_{1,0}$. Then $x = ph_{2,0} + 1 - p$. Also

$$h_{i,0} = h_{i,i-1}h_{i-1,0} = h_{1,0}h_{i-1,0}.$$

In particular with $i = 2$ we get $h_{2,0} = h_{1,0}h_{1,0} = x^2$.
Therefore,

$$x = px^2 + (1-p).$$

Solving the quadratic gives solutions

$$x = 1, \quad \text{and } x = (1-p)/p.$$

If $p \leq 1/2$ then $h_{1,0} = 1$ is the only possible value for $h_{1,0}$, and in this case $h_{i,0} = 1$ for every $i > 0$.

If $p > 1/2$ then the walk is transient (to the right), so $h_{i,0}$ cannot be 1 for $i > 0$. Thus $h_{1,0} = (1-p)/p$ and $h_{i,0} = ((1-p)/p)^i$.

Do the above expressions look familiar?

The "correct" solution to the hitting probability equations

Theorem: The vector of hitting probabilities $(h_{i,B})_{i\in\mathcal{S}}$ is the unique minimal non-negative solution to the equations

$$h_{i,B} = \begin{cases} 1, & \text{if } i \in B \\ \sum_{j\in\mathcal{S}} p_{i,j} h_{j,B}. \end{cases}$$

Proof: Let $h_{i,B}^{(n)} = \mathbb{P}(T(B) \le n | X_0 = i)$. Let $(x_i)_{i\in\mathcal{S}}$ be a non-negative solution. We'll show that $h_{i,B}^{(n)} \le x_i$ for each $n \in \mathbb{Z}_+$, by induction. This is sufficient since $h_{i,B} = \lim_{n\to\infty} h_{i,B}^{(n)}$.

For $n = 0$, note that $h_{i,B}^{(0)} = 1$ if $i \in B$ and $h_{i,B}^{(0)} = 0$ if $i \notin B$. Since $(x_i)_{i\in\mathcal{S}}$ are non-negative and equal 1 for $i \in B$, we have $h_{i,B}^{(0)} \le x_i$ for all $i \in \mathcal{S}$.

Proceeding by induction, suppose that $h_{i,B}^{(n)} \le x_i$ for all $i \in \mathcal{S}$. Then

$$\begin{aligned} h_{i,B}^{(n+1)} = \sum_{j\in\mathcal{S}} p_{i,j} h_{j,B}^{(n)} &\le \sum_{j\in\mathcal{S}} p_{i,j} x_j \quad \text{(by the induction hypothesis)} \\ &= x_i \qquad\qquad ((x_i)_{i\in\mathcal{S}} \text{ is a solution}). \quad \square \end{aligned}$$

Exercise:

Check that what this theorem says about the simple random walk agrees with what we already proved.

A comment

Thus far (with the exception of the distribution at a fixed time) all of the properties of a DTMC that we have considered depend only on the transition matrix P (i.e. they are in fact properties of a stochastic matrix).

In what we study from now on, this will often not be the case. I.e. depending on P, we might also have to look at the initial distribution to find a solution to a given problem.

One of the interesting questions will be, when does a property of interest actually depend on the initial distribution?

Long run behaviour of DTMCs

Let $N_n(j) = \sum_{i=0}^{n-1} \mathbb{1}_{\{X_i=j\}}$ denote the number of visits to j before time n. Then $Y_n(j) = \frac{1}{n} N_n(j)$ is the proportion of time spent in state j before time n. Note that $Y_n(j)$ is a random variable.

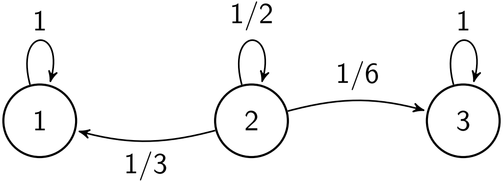
Recall our fundamental question: what is the long run proportion of the time spent in state j? I.e. what is $Y(j) = \lim_{n\to\infty} Y_n(j)$?

- We have argued that this will be zero if the chain is irreducible and transient.
- For an irreducible DTMC we will get an answer to this question that does not depend on the initial distribution.
- If the chain is reducible then the answer to this question may be random, and may depend on the initial distribution.

A related fundamental question is: what is $\lim_{t\to\infty} \mathbb{P}(X_t = j)$?

A reducible example

Consider a Markov chain with transition diagram



Find the limiting proportion of time spent in state 1.

If $X_0 = 1$ then $X_t = 1$ for all t so $Y(1) = 1, Y(2) = 0, Y(3) = 0$.
If $X_0 = 3$ then $X_t = 3$ for all t so $Y(1) = 0, Y(2) = 0, Y(3) = 1$.
If $X_0 = 2$ then $Y(1)$ is random, with $\mathbb{P}(Y(1) = 1) = h_{2,1} = 2/3$ and $\mathbb{P}(Y(1) = 0) = 1 - h_{2,1} = 1/3$
So in general, $Y(1)$ is a Bernoulli random variable with

$$\mathbb{P}(Y(1) = 1) = \mathbb{P}(X_0 = 1) + \frac{2}{3} \cdot \mathbb{P}(X_0 = 2).$$

Long run behaviour of irreducible DTMCs

For $i \in \mathcal{S}$ and $A \subset \mathcal{S}$ let $m_{i,A}$ denote the expected time to reach A starting from i. Note that $m_{i,\{j\}} = m_{i,j}$ is a special case that we have already seen before.
By definition and the Markov property we have that

$$m_{i,A} = \begin{cases} 0, & \text{if } i \in A \\ 1 + \sum_{j \in \mathcal{S}} p_{i,j} m_{j,A}, & \text{otherwise.} \end{cases}$$

Recall that an irreducible DTMC is recurrent if $h_{i,j} = 1$ for every $i, j \in \mathcal{S}$.

Definition: An irreducible DTMC is positive recurrent if $m_{i,j} < \infty$ for every $i, j \in \mathcal{S}$.
An irreducible DTMC that is recurrent but not positive recurrent is called null recurrent.

Computing expected hitting times

Theorem: The vector $(m_{iA})_{i\in\mathcal{S}}$ of mean hitting times is the minimal non-negative solution to

$$m_{i,A} = \begin{cases} 0, & \text{if } i \in A \\ 1 + \sum_{j\in\mathcal{S}} p_{i,j} m_{j,A}, & \text{otherwise.} \end{cases}$$

Example: the symmetric simple random walk is recurrent. Now,

$$m_{1,0} = 1 + \frac{1}{2} m_{0,0} + \frac{1}{2} m_{2,0} = 1 + \frac{1}{2} m_{2,0}.$$

But $m_{2,0} = m_{2,1} + m_{1,0} = 2m_{1,0}$ so

$$m_{1,0} = 1 + m_{1,0}.$$

This has no finite solution, so $m_{1,0} = \infty$, i.e. the simple symmetric random walk is null recurrent.

Finite irreducible DTMCs are positive recurrent

Lemma: Any irreducible DTMC with finite state space is positive recurrent.

Proof: Let $j \in \mathcal{S}$. Then there exists $n_0 \in \mathbb{N}$ such that $\mathbb{P}(T(j) \leq n_0 | X_0 = k) > 0$ for every $k \in \mathcal{S}$. Therefore there exists some $\varepsilon > 0$ such that $\mathbb{P}(T(j) \leq n_0 | X_0 = k) > \varepsilon$ for every $k \in \mathcal{S}$. Starting from any state i, observe whether the chain has reached j within n_0 steps. This has probability at least ε. If it has not reached j then observe it for the next n_0 steps... The number of blocks of n_0 steps that we have to observe is dominated by a geometric(ε) random variable. Therefore $m_{i,j} \leq n_0/\varepsilon < \infty$. $\square$

Long run behaviour of irreducible DTMCs

Recall that μ_j denotes the mean return time to state j. The definition of positive recurrence for irreducible chains is in fact equivalent to $\mu_j < \infty$ for every $j \in \mathcal{S}$ (try to see why as an exercise!)

Theorem: $(*)$ Let $(X_i)_{i\in\mathbb{Z}_+}$ be an irreducible (discrete-time, time-homogeneous) Markov chain with state space $\mathcal{S}$, and μ_i denote the mean return time to state i. Then

$$\mathbb{P}\left(\frac{N_n(i)}{n} \to \frac{1}{\mu_i}\right) = 1, \tag{2}$$

where we interpret the limit $\frac{1}{\mu_i}$ as $0 = 1/\infty$ if the chain is not positive recurrent.

Idea of proof: if the chain is transient then $N_n(i)$ is bounded so $N_n(i)/n$ converges to 0. Otherwise the chain visits state i about once in every μ_i steps, so the proportion of time spent at i is $1/\mu_i$. (We make this rigorous using the law of large numbers).

Computing mean return times

Note that

$$\mu_i = 1 + \sum_{j \in \mathcal{S}} p_{i,j} m_{j,i}.$$

So if you can find $(m_{j,i})_{j \in \mathcal{S}}$ then you are done!

Stationary distribution

A vector $\pi = (\pi_i)_{i \in \mathcal{S}}$ with non-negative entries is a stationary measure for a stochastic matrix P (or a DTMC with transition matrix P) if

$$\pi_i = \sum_{j \in \mathcal{S}} \pi_j p_{j,i}, \qquad \text{for every } i \in \mathcal{S}.$$

The above equations are called the full balance equations. In Matrix form this is $\pi P = \pi$ (with π as a row vector).

If π is a stationary measure with $\sum_{i \in \mathcal{S}} \pi_i = 1$ then π is called a stationary distribution for P.

Stationary distribution

Suppose that π is a stationary distribution for a DTMC $(X_n)_{n\in\mathbb{Z}_+}$ with transition matrix P, and suppose that $\mathbb{P}(X_0 = i) = \pi_i$ for each $i \in \mathcal{S}$. Then

$$\mathbb{P}(X_1 = i) = \sum_{j\in\mathcal{S}} \mathbb{P}(X_0 = j)p_{j,i} = \sum_{j\in\mathcal{S}} \pi_j p_{j,i} = \pi_i,$$

since π satisfies the full balance equations. This says that $\mathbb{P}(X_1 = i) = \pi_i$ too. By induction (exercise) we get:

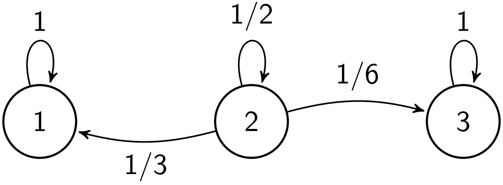
Lemma: Suppose that $(\pi_i)_{i\in\mathcal{S}}$ is a stationary distribution for P. Let $(X_n)_{n\in\mathbb{Z}_+}$ be DTMC with transition matrix P and initial distribution equal to π (i.e. $\mathbb{P}(X_0 = i) = \pi_i$ for each $i \in \mathcal{S}$), then

$$\mathbb{P}(X_n = i) = \pi_i, \qquad \text{for every } i \in \mathcal{S}, n \in \mathbb{N}.$$

So, if your initial distribution is a stationary distribution then your distribution at any time is the same (hence the use of the term stationary)!

Examples:

Find all stationary distributions for a Markov chain with the following transition diagram:



The full balance equations are:

$$\pi_1 = \pi_1 + \frac{1}{3} \cdot \pi_2, \qquad \pi_2 = \frac{1}{2}\pi_2, \qquad \pi_3 = \pi_3 + \frac{1}{6} \cdot \pi_2.$$

The second equation gives $\pi_2 = 0$. The other equations then reduce to $\pi_1 = \pi_1$ and $\pi_3 = \pi_3$. Thus, any vector $(\pi_1, \pi_2, \pi_3) = (a, 0, b)$ with $a, b \geq 0$ is a stationary measure. To get a stationary distribution, we require that $a + b = 1$, so the set of stationary *distributions* is the set of vectors of the form $(a, 0, 1 - a)$ with $a \in [0, 1]$.

Existence and uniqueness

We have just seen an example of a DTMC without a unique stationary distribution. The following is our main existence and uniqueness result.

Theorem: ($**$) An irreducible (time-homogeneous) DTMC with countable state space $\mathcal{S}$ has a stationary measure. It has a unique stationary distribution if and only if the chain is positive recurrent, and in this case $\pi_i = 1/\mu_i$ for each $i \in \mathcal{S}$.

Combining Theorems ($*$) and ($**$) we see that for a positive recurrent irreducible DTMC, the long run proportion of time spent in state i is the stationary probability π_i.

Limiting distribution

A distribution $(a_i)_{i\in\mathcal{S}}$ is called the limiting distribution for a DTMC $(X_t)_{t\in\mathbb{Z}_+}$ if

$$\lim_{n\to\infty}\mathbb{P}(X_n = i) = a_i, \quad \text{for each } i \in \mathcal{S}.$$

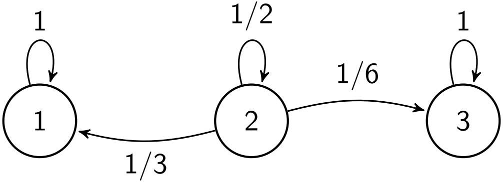
In general,

- A limiting distribution need not exist
- The limiting distribution (if it exists) is unique (by definition)
- The limiting distribution (if it exists) depends on both the initial distribution and P.

We have already seen that if π is a stationary distribution, and $\mathbb{P}(X_0 = i) = \pi_i$ for each $i \in \mathcal{S}$ then $\mathbb{P}(X_n = i) = \pi_i$ for each $i \in \mathcal{S}, n \in \mathbb{N}$, so in this case π is also the limiting distribution.

Examples:

Consider a DTMC with initial distribution (b_1, b_2, b_3) and transition diagram



Find the limiting distribution for the chain.

Note that if $b_1 = 1$ then the chain stays forever in state 1 so the limiting distribution is $(1, 0, 0)$ in this case. Similarly it is $(0, 0, 1)$ if $b_3 = 1$. If $b_2 = 1$ then we either hit state 1 (with probability $h_{2,3} = 2/3$) or 3 and then stay there, so the limiting distribution is $(2/3, 0, 1/3)$ in this case (see the next slide for details).

Examples:

To answer the question in general, note that if $X_k = 1$ then $X_n = 1$ for all $n \geq k$ and therefore $\{X_n = 1\} = \cup_{k=0}^{n} \{X_k = 1\}$. Thus,

$$\lim_{n\to\infty} \mathbb{P}(X_n = 1) = \lim_{n\to\infty} \mathbb{P}(\cup_{k=0}^{n} \{X_k = 1\}) = \mathbb{P}(\cup_{k=0}^{\infty} \{X_k = 1\}).$$

The last event $H_1 = \cup_{k=0}^{\infty} \{X_k = 1\}$ is the event that we ever reach state 1, so

$$\begin{aligned}
\mathbb{P}(H_1) &= \mathbb{P}(H_1 | X_0 = 1)\mathbb{P}(X_0 = 1) + \mathbb{P}(H_1 | X_0 = 2)\mathbb{P}(X_0 = 2) \\
&\qquad + \mathbb{P}(H_1 | X_0 = 3)\mathbb{P}(X_0 = 3) \\
&= 1 \cdot \mathbb{P}(X_0 = 1) + \frac{2}{3} \cdot \mathbb{P}(X_0 = 2) + 0 \qquad (3) \\
&= \mathbb{P}(X_0 = 1) + \frac{2}{3} \cdot \mathbb{P}(X_0 = 2).
\end{aligned}$$

Example:

Consider a DTMC with $\mathcal{S} = \{1, 2\}$, $\mathbb{P}(X_0 = 1) = p$ and transition matrix

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

This chain is periodic (with period 2). Note that $\mathbb{P}(X_{2m} = 1) = \mathbb{P}(X_0 = 1) = p$ and $\mathbb{P}(X_{2m+1} = 1) = \mathbb{P}(X_0 = 2) = 1 - p$. Thus

$$\mathbb{P}(X_n = 1) = \begin{cases} p, & \text{if } n \text{ is even} \\ 1 - p, & \text{if } n \text{ is odd.} \end{cases}$$

This converges if and only if $p = 1 - p$.

In other words, for this example, a limiting distribution exists if and only if $p = 1/2$.

Example: simple random walk

Recall that $S_0 = 0$ and $p_{i,i+1} = p > 1/2$ and $p_{i,i-1} = 1 - p$. We have seen (using Stirling's formula) that

$$\mathbb{P}(S_{2n} = 0) \approx \frac{C}{\sqrt{n}} \to 0,$$

while $\mathbb{P}(S_{2n+1} = 0) = 0$. So $\mathbb{P}(S_n = 0) \to 0$ as $n \to \infty$

Similarly one can show that $\mathbb{P}(S_n = i) \to 0$ for each i, so all of the limits exist, but they are all 0, so there is no limiting distribution in this example.

Limiting distribution results

Theorem: For a (time-homogeneous) DTMC, if a limiting distribution exists then it is a stationary distribution.

Theorem:$(***)$ Let $(X_n)_{n\in\mathbb{Z}_+}$ be an irreducible, aperiodic (time-homogeneous) DTMC with countable state space $\mathcal{S}$. Then for all $i, j \in \mathcal{S}$,

$$p_{i,j}^{(n)} = \mathbb{P}(X_n = j | X_0 = i) \to \frac{1}{\mu_j}, \quad \text{as } n \to \infty.$$

Ergodicity

We say that a DTMC is ergodic if the limiting distribution exists and does not depend on the starting distribution. This is equivalent to saying that $a_j = \lim_{n\to\infty} p_{i,j}^{(n)}$ exists for each $i, j \in S$, does not depend on i, and $\sum_{j\in\mathcal{S}} a_j = 1$.

Theorem: An irreducible (time-homogeneous) DTMC is ergodic if and only if it is aperiodic, and positive recurrent. For an ergodic DTMC the limiting distribution is equal to the stationary distribution.

Examples:

An $m \times m$ stochastic matrix P is called doubly-stochastic if all the column sums are equal to one.

Suppose that a finite-state DTMC has doubly stochastic transition matrix.
Suppose that $\mathcal{S}$ contains exactly m elements. We can easily verify that

$$(1/m, 1/m, \ldots, 1/m)P = (1/m, 1/m, \ldots, 1/m).$$

It follows that

$$\boldsymbol{\pi} = (1/m, 1/m, \ldots, 1/m),$$

and the stationary distribution is uniform on $\mathcal{S}$.
Conversely if the uniform distribution on $\mathcal{S} = \{1, \ldots, m\}$ is a stationary distribution for P then for each i,

$$\frac{1}{m} = \sum_{j \in \mathcal{S}} \frac{1}{m} p_{j,i} = \frac{1}{m} \sum_{j \in \mathcal{S}} p_{j,i},$$

so P is doubly stochastic.

Examples:

Let $(X_n)_{n\in\mathbb{Z}_+}$ be a DTMC with state space $\mathcal{S} = \{1, 2, 3\}$, and transition matrix

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

If the initial distribution is $x = (x_1, x_2, x_3)$, find:

(i) the limiting distribution

(ii) the limiting proportion of time spent in state 2

Reversibility

An irreducible DTMC is called reversible if there exists a probability distribution $\pi = (\pi_i)_{i\in\mathcal{S}}$ such that

$$\pi_i p_{i,j} = \pi_j p_{j,i}, \quad i, j \in \mathcal{S}.$$

The above equations are called the detailed balance equations.

Any solution to these equations is a stationary distribution since if π satisfies the detailed balance equations then

$$\sum_{j\in\mathcal{S}} \pi_j p_{j,i} = \sum_{j\in\mathcal{S}} \pi_i p_{i,j} = \pi_i \Big(\sum_{j\in\mathcal{S}} p_{i,j} \Big) = \pi_i.$$

Kolmogorov's reversibility criterion

A (time-homogeneous) irreducible DTMC with state space $\mathcal{S}$ is reversible if and only if it has a stationary distribution and satisfies

$$p_{j_n j_1} \prod_{i=1}^{n-1} p_{j_i j_{i+1}} = p_{j_1 j_n} \prod_{i=1}^{n-1} p_{j_{i+1} j_i},$$

for every n and every $\{j_1, j_2, \ldots, j_n\}$

Interpretation: Suppose I show you a short video clip of a Markov chain that starts and ends at the same state, and I choose to show it either forwards or in reverse. I tell you what P is. Then what you observe contains no information about whether I showed the process forwards or in reverse *if and only if the process is reversible*.

Example

Let P be an irreducible (so it has a stationary measure) stochastic matrix with $p_{i,j} = 0$ unless $j \in \{i-1, i, i+1\}$. (A Markov chain with such a stochastic matrix is called a *birth and death chain*.)

Then for any sequence of n transitions taking us from state i to state i: every occurrence of a transition $j \to j+1$ has a corresponding transition $j+1$ to j and vice versa.

Thus, reversing this sequence of n transitions we see exactly the same number of transitions from j to $j+1$ as the forward sequence. This is true for all j and therefore the Kolmogorov criterion is always satisfied for such a stochastic matrix, provided it also has a stationary distribution.

Thus a positive recurrent birth and death chain is always reversible.

Example

Let $(X_n)_{n\in\mathbb{Z}_+}$ be a DTMC with $\mathcal{S}=\mathbb{Z}_+$ and transition probabilities $p_{i,i+1}=p_i\in(0,1)$ for each $i\in\mathcal{S}$, and $p_{0,0}=1-p_{0,1}$ and $p_{i,i-1}=1-p_i$ for $i\geq 1$.

This is a birth and death chain, so if it has a stationary distribution then it satisfies the detailed balance equations.

Set $\pi_0=x$. The detailed balance equations are

$$\pi_i p_i=\pi_{i+1}(1-p_{i+1}),\qquad \text{i.e.}$$
$$\pi_{i+1}=p_i/(1-p_{i+1})\pi_i.$$

Letting $\rho_i=p_i/(1-p_{i+1})$, it follows that

$$\pi_i=x\prod_{j=0}^{i-1}\rho_j,$$

gives a solution to the detailed balance equations.

Thus if $\sum_{i=0}^{\infty}\left(\prod_{j=0}^{i-1}\rho_j\right)<\infty$ then there is a stationary distribution (otherwise not).

Example: Random walk with one barrier

Continuing the example above, if $p_i = p$, then $\rho_i = \rho := p/(1-p)$ for every i, and the sum is finite if and only if $\rho < 1$ (i.e. $p < 1/2$).

When $p < 1/2$ we have a stationary distribution, and the chain is irreducible and aperiodic (due to the boundary), so also ergodic. Otherwise we have no stationary distribution.

The limiting proportion of time spent in state 0 is π_0. We find this by

$$1 = \sum_{i=0}^{\infty} \pi_i = \sum_{i=0}^{\infty} \pi_0 \rho^i = \pi_0/(1-\rho).$$

In other words $\pi_0 = 1 - \rho$.

Summary:

We have introduced the Markov property, and notions of communicating classes (and irreducibility). We have seen how to calculate path probabilities, hitting probabilities and expected hitting times. We have discussed recurrence and null-recurrence, and periodicity.

For an irreducible, aperiodic and positive-recurrent DTMC, the stationary distribution $\boldsymbol{\pi}$ has a number of interpretations. It can be seen as:

- the initial distribution for which the process is a stationary process
- the limiting probability of being in each state
- the limiting proportion of time spent in each state

Tricks of the trade

Sometimes we have a process that is not a Markov chain, yet we can still use Markov chain theory to analyse its behaviour by being clever.

Consider the following example:

Let $(S'_t)_{t\in\mathbb{Z}_+}$ denote a random process with state space $\mathcal{S}' = \{1, 2, 3, 4\}$, $S'_0 = 1$ and $S'_1 = 2$, such that whenever the process is at i, having just come from j, it chooses uniformly at random from $\mathcal{S} \setminus \{i, j\}$. This process is a kind of non-backtracking random walk (on the set $\mathcal{S}$).

Exercise: Show that S' is not Markovian by showing that $\mathbb{P}(S'_4 = 4|S'_3 = 1) \neq \mathbb{P}(S'_4 = 4|S'_3 = 1, S'_2 = 4)$.

If we put $X_t = (S'_t, S'_{t+1})$ for $t \in \mathbb{Z}_+$ then $(X_t)_{t\in\mathbb{Z}_+}$ is a Markov chain on a state space $\mathcal{S} = \{(i,j) : i, j \in \mathcal{S}',\ i \neq j\}$ with 12 elements (can relabel them 1 to 12 if you wish), with

$$p_{(i_1,j_1),(i_2,j_2)} = \begin{cases} \frac{1}{2}, & \text{if } i_2 = j_1 \text{ and } j_2 \notin \{i_1, j_1\} \\ 0, & \text{otherwise.} \end{cases}$$

The Poisson Process

The Poisson distribution

A random variable N taking values in $\mathbb{Z}_+ = \{0, 1, 2, \dots\}$ has an Poisson distribution with a parameter $\lambda > 0$ (and we write $N \sim \mathrm{Pois}(\lambda)$), if its probability mass function is given by

$$\mathbb{P}(N = n) = \frac{e^{-\lambda}\lambda^n}{n!}, \qquad \text{for } n \in \mathbb{Z}_+.$$

The mean and variance of N are both equal to λ.

The exponential distribution

A random variable T has an exponential distribution with parameter $\lambda > 0$ (called the rate), denoted by $T \sim \exp(\lambda)$, if its distribution function is

$$F_T(t) = \begin{cases} 1 - e^{-\lambda t}, & \text{for } t \geq 0, \\ 0, & \text{for } t < 0. \end{cases}$$

It follows that the probability density function of T is

$$f_T(t) = \begin{cases} \lambda e^{-\lambda t}, & \text{for } t \geq 0, \\ 0, & \text{for } t < 0. \end{cases}$$

The mean of T is $1/\lambda$ and the variance of T is $1/\lambda^2$.

The law of rare events

The Poisson distribution arises as the limit of the binomial distribution.

- If $X_n \sim \mathrm{Bin}(n, \lambda/n)$ and $N_\lambda \sim \mathrm{Pois}(\lambda)$, then for $k = 0, 1, \ldots,$

$$\lim_{n\to\infty} \mathbb{P}(X_n = k) = \mathbb{P}(N_\lambda = k).$$

The exponential distribution arises as the limit of the geometric distribution.

- If $Y_n \sim \mathrm{Geo}(\lambda/n)$ and $T_\lambda \sim \mathrm{Exp}(\lambda)$, then for $t > 0$,

$$\lim_{n\to\infty} \mathbb{P}(Y_n/n \leq t) = \mathbb{P}(T_\lambda \leq t).$$

The Poisson Process

Definition:

A nonnegative integer-valued process $(N_t)_{t\geq 0}$ is a Poisson process with a rate λ if

- it has independent increments on disjoint intervals: for $k \geq 2$ and $0 \leq s_1 < t_1 \leq s_2 < \cdots < t_k$,

$$N_{t_1} - N_{s_1}, \ldots, N_{t_k} - N_{s_k}$$

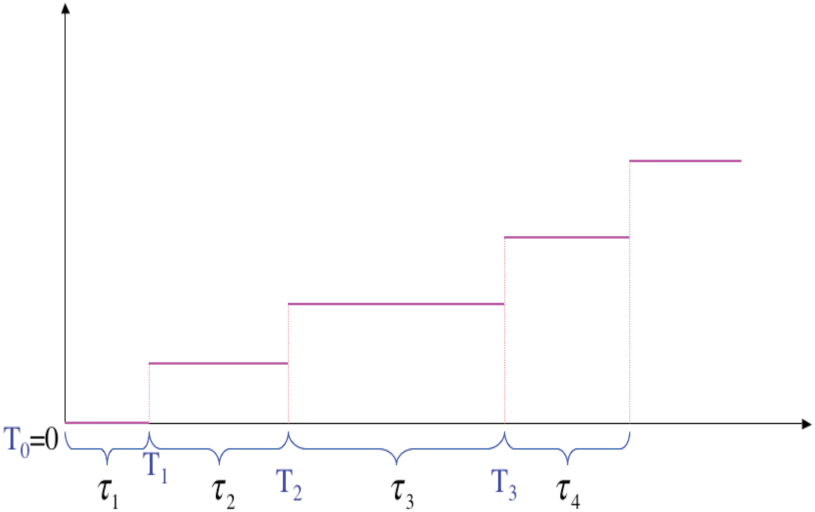
  are independent variables.
- For each $t > s \geq 0$, $N_t - N_s \sim \mathrm{Pois}(\lambda(t-s))$

Exercise

If $(N_t)_{t\geq 0}$ is a Poisson process with rate λ, show that for fixed $r > 0$, the process $(N^*_t)_{t\geq 0}$ defined by $N^*_t = N_{t+r} - N_r$, is also Poisson process with rate λ. [This is true even if r is a stopping time for the process.]

The Poisson Process

A trajectory:



Poisson Process Empirical Data

- Earthquakes
- Grazing animals head raises
- Horse kick deaths

The Poisson Process

Let $T_j = \min\{t : N_t = j\}$, the time of jth jump and define $\tau_j = T_j - T_{j-1}$ the time between the $(j-1)^{\mathrm{st}}$ jump and the j^{th} jump.

Theorem: $(N_t)_{t\geq 0}$ is a Poisson process with rate λ if and only if $\{\tau_j\}$ are independent $\exp(\lambda)$ random variables.

Proof: The key to the proof is to observe that the event $\{T_j \leq t\}$ is the same as $\{N_t \geq j\}$. That is the waiting time until the jth event is less than or equal to t if and only if there are j or more events up to (and including) time t.
Assume that $(N_t)_{t\geq 0}$ is a Poisson process. Then
$\mathbb{P}(T_1 \leq t) = \mathbb{P}(N_t \geq 1) = 1 - \mathbb{P}(N_t = 0) = 1 - e^{-\lambda t}$, so
$T_1 \sim \exp(\lambda)$.

The Poisson Process

Furthermore, we have

$$
\begin{aligned}
\mathbb{P}(T_j \leq t) &= \mathbb{P}(N_t \geq j) \\
&= \sum_{k=j}^{\infty} e^{-\lambda t} \frac{(\lambda t)^k}{k!} \\
&= 1 - \sum_{k=0}^{j-1} e^{-\lambda t} \frac{(\lambda t)^k}{k!}.
\end{aligned}
$$

This final expression is the distribution function for gamma distribution with parameter k and rate λ. You can check this by differentiating to get the density function
$f_{T_j}(t) = e^{-\lambda t} \lambda^j t^{j-1}/(j-1)!$.
So the waiting time until the jth event is the sum of j independent exponentially-distributed inter-event times with parameter λ.

The Poisson Process

This argument also holds in reverse.
Assuming that τ_1 is exponentially-distributed with parameter λ, we know that $\mathbb{P}(T_1 \leq t) = 1 - e^{-\lambda t}$, which tells us that $\mathbb{P}(N_t = 0) = e^{-\lambda t}$.
Furthermore, for $j > 1$, if $\{\tau_1, \ldots, \tau_j\}$ are independent and exponentially-distributed, then T_j has an Gamma distribution with parameters λ and j. So

$$\mathbb{P}(N_t \geq j) = \mathbb{P}(T_j \leq t) = 1 - \sum_{k=0}^{j-1} e^{-\lambda t} \frac{(\lambda t)^k}{k!},$$

which tells us that N_t has a Poisson distribution with parameter λt.

The Poisson Process

We still need to show that $N_{t_i} - N_{s_i}$ are independent over sets $[s_i, t_i)$ of disjoint intervals.
This follows from the memoryless property of the exponential distribution (so the remaining time from s_i doesn't depend on $s_i - T_{N_{s_i}}$) and the independence of the τ_j.

Order statistics

For random variables $\xi_1, \xi_2, \ldots, \xi_k$, denote by $\xi_{(i)}$ the ith smallest of them. Then $\xi_{(1)}, \xi_{(2)}, \ldots, \xi_{(k)}$ are called the order statistics associated with $\xi_1, \xi_2, \ldots, \xi_k$.

For example, if we sample these random variables and find that $\xi_1 = 1.3, \xi_2 = 0.9, \xi_3 = 0.7, \xi_4 = 1.1$ and $\xi_5 = 1.5$, then $\xi_{(1)} = 0.7, \xi_{(2)} = 0.9, \ldots, \xi_{(5)} = 1.5$.

Order statistics play a very important role in applications. For example, the maximum likelihood estimator of θ for a sample $\xi_1, \xi_2, \ldots, \xi_k$ from the uniform $[0, \theta]$ distribution is $\xi_{(k)}$.

Order Statistics

Examples

- If X_1, X_2 and X_3 are independent and identically-distributed random variables taking values 1, 2 and 3 each with probability 1/3, find the joint probability mass function of $(X_{(1)}, X_{(2)}, X_{(3)})$.
- If $Y_1, Y_2, \ldots, Y_k$ are i.i.d. random variables with distribution function F and density f, the distribution function of $Y_{(i)}$ is

$$F_{Y_{(i)}}(x) = \sum_{\ell=i}^{k} \binom{k}{\ell} F(x)^{\ell}(1 - F(x))^{k-\ell}.$$

- So, for the special case where $Y_1, Y_2, \ldots, Y_k$ are i.i.d.$\sim U[0, t]$, for $i \leq k$ and $x \leq t$, the distribution function of the order statistic $Y_{(i)}$ is given by

$$F_{Y_{(i)}}(x) = \sum_{\ell=i}^{k} \binom{k}{\ell} (x/t)^{\ell}(1 - x/t)^{k-\ell}.$$

Order Statistics

- Above we saw that if $Y_1, Y_2, \ldots, Y_k$ are i.i.d. with distribution function F the distribution function of $Y_{(i)}$ is

$$F_{Y_{(i)}}(x) = \sum_{\ell=i}^{k} \binom{k}{\ell} F(x)^{\ell} (1 - F(x))^{k-\ell}.$$

  If they are also absolutely continuous, with density f, then the density of the order statistic $Y_{(i)}$ is

$$\begin{aligned} f_{Y_{(i)}}(x) &= \binom{k}{i-1} (k - i + 1) F(x)^{i-1} f(x) (1 - F(x))^{k-i} \\ &= \binom{k}{i} i F(x)^{i-1} f(x) (1 - F(x))^{k-i}. \end{aligned}$$

Order Statistics

Similarly the joint densities for $1 \leq r \leq k$ and $x_1 < \cdots < x_r$ are

$$f_{Y_{(i_1)},\ldots,Y_{(i_r)}}(x_1,\ldots,x_r)$$
$$= \binom{k}{i_1-1,1,i_2-i_1-1,1\cdots,1,k-i_r}$$
$$\times \prod_{j=1}^{r} f(x_j) \prod_{j=0}^{r} (F(x_{j+1}) - F(x_j))^{i_{j+1}-i_j-1},$$

where $\binom{\ell}{a_1,\cdots,a_j}$ is the number of ways to choose subsets of sizes $a_1,\ldots,a_j$ from a set of size ℓ and for the sake of brevity we set $x_0 = -\infty$ and $x_{r+1} = \infty$ so $F(x_0) = 0$ and $F(x_{r+1}) = 1$.

Order Statistics

In particular for $r = k$, $x_1 < \cdots < x_r$,

$$f_{Y_{(1)},\ldots,Y_{(k)}}(x_1, \ldots, x_k) = k! \prod_{j=1}^{k} f(x_j).$$

Poisson Process and order statistics

Theorem: The conditional distribution of $(T_1, \cdots, T_k)$ given that $N_t = k$ is the same as the distribution of order statistics of a sample of k independent and identically-distributed random variables uniformly distributed on $[0, t]$. Thus,

$$(T_1, \cdots, T_k)|(N_t = k) \stackrel{d}{=} (U_{(1)}, \cdots, U_{(k)})$$

where $U_1, \cdots, U_k$ are independent Uniform $(0, t)$.
The same representation holds for the conditional distribution of $(T_1, \cdots, T_k)$ given that $T_{k+1} = t$.

Proof

According to our derivation for order statistics, $(U_{(1)}, \cdots, U_{(k)})$ has density $k!t^{-k}$ for $0 = x_0 < x_1 < \cdots < x_k < t$.
So we show the LHS has the same density:

$$
\begin{aligned}
&\mathbb{P}(T_1 \in dx_1, \cdots, T_k \in dx_k | N_t = k) \\
&\quad = \frac{\mathbb{P}(\tau_1 \in dx_1, \tau_2 \in d(x_2 - x_1), \ldots, \tau_k \in d(x_k - x_{k-1}), \tau_{k+1} > t - x_k)}{\mathbb{P}(N_t = k)} \\
&\quad = \frac{(\prod_{i=1}^k \lambda e^{-\lambda(x_i - x_{i-1})}) e^{-\lambda(t - x_k)}}{(\lambda t)^k e^{-\lambda t}/k!} dx_1 \ldots dx_k \\
&\quad = k! t^{-k} dx_1 \ldots dx_k.
\end{aligned}
$$

The proof of the last sentence is similar.

The Poisson Process

The theorem implies that if $\tau_1, \ldots, \tau_n$ are i.i.d. exponential variables, then

$$\left(\frac{\tau_1}{\sum_{j=1}^{n+1} \tau_j}, \frac{\tau_1 + \tau_2}{\sum_{j=1}^{n+1} \tau_j}, \ldots, \frac{\sum_{j=1}^{n} \tau_j}{\sum_{j=1}^{n+1} \tau_j} \right)$$

have the same distribution as uniform order statistics.

Superposition of Poisson processes

Let $(N_t)_{t\geq 0}$ and $(M_t)_{t\geq 0}$ be two independent Poisson processes with rates λ and μ respectively and $L_t = N_t + M_t$. Then $(L_t)_{t\geq 0}$ is a Poisson process with rate $\lambda + \mu$.

Proof

- By independence, $L_t - L_s \sim \mathrm{Po}(\lambda(t-s) + \mu(t-s))$.
- For disjoint $[s_1, t_1]$ and $[s_2, t_2]$,

$$\begin{aligned} L_{t_1} - L_{s_1} &= (N_{t_1} - N_{s_1}) + (M_{t_1} - M_{s_1}) \\ L_{t_2} - L_{s_2} &= (N_{t_2} - N_{s_2}) + (M_{t_2} - M_{s_2}) \end{aligned}$$

  which are independent because of the same property of $(N_t)_{t\geq 0}$ and $(M_t)_{t\geq 0}$. This argument extends to all finite collections of disjoint intervals.

Superposition example

A shop has two entrances, one from East St, the other from West St. Flows of customers through the two entrances are independent Poisson processes with rates 0.5 and 1.5 per minute, respectively.

- What is the probability that no new customers enter the shop in a fixed three minute time interval?
- What is the mean time between arrivals of new customers?
- What is the probability that a given customer entered from West St?

Thinning of a Poisson process

Suppose in a Poisson process $(N_t)_{t \geq 0}$ each 'customer' is 'marked' independently with probability p. Let M_t count the number of 'marked customers' that arrive on $[0, t]$.

Theorem:
The processes $(M_t)_{t \geq 0}$ and $(N_t - M_t)_{t \geq 0}$ are independent Poisson processes with rates λp and $\lambda(1 - p)$ respectively.

Thinning - proof

Proof:

$$\begin{aligned}
\mathbb{P}(M_t = j, N_t - M_t = k) &= \mathbb{P}(M_t = j, N_t = k + j) \\
&= \mathbb{P}(M_t = j | N_t = k + j)\mathbb{P}(N_t = k + j) \\
&= \binom{k+j}{j} p^j (1-p)^k \frac{e^{-\lambda t}(\lambda t)^{k+j}}{(k+j)!} \\
&= \frac{e^{-p\lambda t}(p\lambda t)^j}{j!} \frac{e^{-(1-p)\lambda t}((1-p)\lambda t)^k}{k!}.
\end{aligned}$$

Thinning - example

The flow of customers to a shop is a Poisson process with rate 25 customers per hour. Each of the customers independently has a probability $p = 0.8$ of making a purchase.

- What is the probability that all customers who enter the shop during the time interval from 11.00 am to 11.15 am make a purchase?
- What is the probability that, conditional on there being two customers that made a purchase during that period, all customers who enter the shop during the time interval make a purchase?

The Compound Poisson Process

Suppose that $(N_t)_{t \geq 0}$ is a Poisson process and $(X_i)_{i \in \mathbb{N}}$ are independent and identically-distributed random variables, which are also independent of $(N_t)_{t \geq 0}$.

For $t \geq 0$, define $Y_t = \sum_{j \leq N_t} X_j$. Then $(Y_t)_{t \geq 0}$ is called a compound Poisson process.

It can be shown that $(Y_t)_{t \geq 0}$ has independent increments and it is possible to compute the distribution of Y_t by conditioning on N_t.

Compound Poisson example

Suppose claims made to an insurance company arrive according to a Poisson process with rate λ, and each policy holder carries a policy for an amount X_k. Assume $X_1, X_2, \ldots$ are independent and identically-distributed, and the number of claims and the size of claims are independent.

Calculate the mean and variance of the total amount of claims on the company up to time t.

Continuous-Time Markov Chains

Continuous-Time Markov Chains

A stochastic process $(X_t)_{t\geq 0}$ in continuous time, taking values in a countable state space $\mathcal{S} \subset \mathbb{R}$ is said to be a Continuous-Time Markov Chain (CTMC) if, for all $k \geq 1$, $0 \leq t_1 < t_2 < \cdots < t_{k+1}$ and $i_1, i_2, \ldots, i_{k+1} \in \mathcal{S}$,

$$\begin{aligned} & \mathbb{P}(X_{t_{k+1}} = i_{k+1} | X_{t_1} = i_1, \cdots, X_{t_k} = i_k) \\ = \; & \mathbb{P}(X_{t_{k+1}} = i_{k+1} | X_{t_k} = i_k), \end{aligned}$$

whenever the left hand side is well defined.

As for DTMC, it is often convenient to assume that $\mathcal{S} = \{1, 2, \ldots, n\}$ for some $n \in \mathbb{N} \cup \infty$.

If $\mathbb{P}(X_{t+s} = k | X_s = j)$ does not depend on s then we say the CTMC is time homogeneous, and we can write it as $p^{(t)}_{j,k}$. We consider only time-homogeneous CTMCs in this course.

We can put the probabilities $p^{(t)}_{i,j}$ into a matrix $P^{(t)}$. We assume that $P^{(t)}$ is right continuous in the sense that $P^{(t+h)} \to P^{(t)}$ as $h \downarrow 0$.

DTMC vs CTMC

For DTMC, if $X_n = i$ then we waited for a geometric$(1 - p_{i,i})$ amount of time before jumping to a *new state*. At the time we jump to a *new state*, the probability of jumping to j is $b_{i,j} = p_{i,j}/(1 - p_{i,i})$. If $p_{i,i} = 0$ then $b_{i,j} = p_{i,j}$.

For a CTMC, if $X_t = i$, we wait an exponential(λ_i) time and then jump to a new state. The probability of jumping to j is $b_{i,j}$.

This is equivalent to the following: let $(T_{i,j})_{j \in \mathcal{S}}$ be independent exponential$(q_{i,j})$ random variables. We wait time $T_i' = \min_{\ell \in \mathcal{S}} T_{i,\ell}$ at state i and then jump to the state k such that $T_{i,k} = T_i'$.

To see the equivalence, set $q_{i,j} = \lambda_i b_{i,j}$ then $T_i' \sim$exponential$(\sum_{\ell \in \mathcal{S}} q_{i,\ell})$ (where $\sum_{\ell \in \mathcal{S}} q_{i,\ell} = \lambda_i$) and the probability that we jump to j is

$$\mathbb{P}(T_{i,j} < \min_{\ell \neq j} T_{i,\ell}) = \frac{q_{i,j}}{\sum_{\ell \in \mathcal{S}} q_{i,\ell}} = b_{i,j}.$$

Transition diagrams

For a DTMC we drew a diagram containing the transition probabilities $p_{i,j}$.

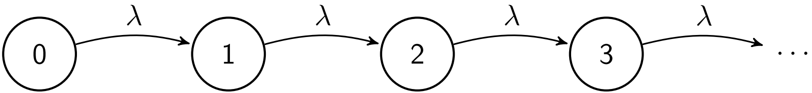
For a CTMC we draw a transition diagram containing the transition rates $q_{i,j}$.

The jump chain of a CTMC $(X_t)_{t\geq 0}$ is the DTMC $(X_n^J)_{n\in\mathbb{Z}_+}$ defined by $X_n^J = X_{T_n}$ where $T_0 = 0$, and $T_i = \inf\{t > T_{i-1} : X_t \neq X_{T_{i-1}}\}$ are the jump times of the CTMC. The transition probabilities of the jump chain are $b_{i,j}$. If $\lambda_i = 0$ we set $b_{i,i} = 1$ for the jump chain (otherwise $b_{i,i} = 0$).

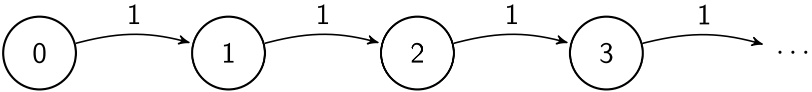
The transition diagram for the CTMC contains more information than that of its jump chain since the latter does not tell us how long the CTMC waits (on average) at each state.

Basic example: Poisson process

The transition diagram for the Poisson process $(N_t)_{t\geq 0}$ with rate λ is



The transition diagram for its jump process $(N_n^J)_{n\in\mathbb{Z}_+}$ is



Remarks:

Note that we have to wait for an *independent* exponential(λ_i) time on each successive visit to i

The waiting times must be exponential in order for the process to be memoryless/Markovian.

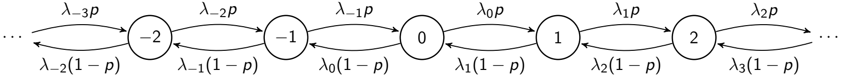
E.g. suppose $X_0 = j$ and T_1 is the first time the CTMC leaves j. Then the CTMC $(X_t)_{t\geq 0}$ must satisfy

$$\begin{aligned}
&\mathbb{P}(T_1 > t + s | T_1 > s) \\
&\quad = \mathbb{P}(X_v = j, 0 \leq v \leq t + s | X_u = j, 0 \leq u \leq s) \\
&\quad = \mathbb{P}(X_v = j, s < v \leq t + s | X_s = j) \text{ (Markov)} \\
&\quad = \mathbb{P}(X_v = j, 0 < v \leq t | X_0 = j) \text{ (homogeneous)} \\
&\quad = \mathbb{P}(T_1 > t)
\end{aligned}$$

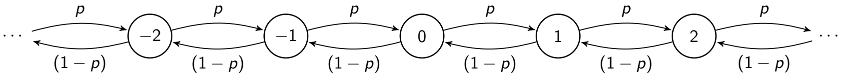
So T_1 must have an exponential distribution.

Example: continuous time random walk

Consider a CTMC with state space $\mathbb{Z}$ that waits for an exponential(λ_i) time in state $i \in \mathbb{Z}$ before jumping to $i+1$ with probability p and $i-1$ with probability $1-p$. Then the transition diagram is



The transition diagram of the jump chain is



If $\lambda_i = \lambda$ for each i we call the CTMC continuous time random walk.

Exercise:

Let $(X_t)_{t\geq 0}$ be a CTMC with state space $\mathcal{S} = \{1, 2, 3\}$ and transition rates $q_{i,j} = i$ for each i, j with $j \neq i$.

- Draw the transition diagram for $(X_t)_{t\geq 0}$
- Draw the transition diagram for the jump chain $(X_n^J)_{n\in\mathbb{Z}_+}$.

Classification of states and chains

As for DTMC we can ask about the following

(1) Whether a state i is absorbing (i.e. $\lambda_i = 0$)

(2) Whether $i \to j$ (i.e. $p_{i,j}^{(t)} > 0$ for some t)

(3) Communicating classes ($i \leftrightarrow j$ if $i \to j$ and $j \to i$)

(4) Irreducibility (i.e. $i \leftrightarrow j$ for every $i, j \in \mathcal{S}$)

(5) Hitting probabilities
(i.e. $h_{i,A} = \mathbb{P}(X_t \in A \text{ for some } t \geq 0 | X_0 = i)$)

(6) Recurrence (for irreducible chains: $h_{i,j} = 1$ for every $i, j \in \mathcal{S}$) and transience

(7) Expected hitting times (i.e. $m_{i,A}$, which is the expected time to reach a state in A starting from i)

(8) Positive recurrence (for irreducible chains: $m_{i,j} < \infty$ for every $i, j \in \mathcal{S}$)

(9) The long run behaviour of the chain (limiting proportion of time spent in state i etc.)

Some things are the same...

(1) A state i is an absorbing state for a CTMC if and only if $\lambda_i = 0$, if and only if $b_{i,i} = 1$, if and only if i is an absorbing state for the jump chain.

Similarly, items (2) to (6) above only depend on $(b_{i,j})_{i,j\in\mathcal{S}}$ and hence the CTMC has the given property if and only if its jump chain has that property.

On the other hand, items (7)-(9) depend on how long we wait (on average) at every state, so these properties will in general differ for the CTMC and its jump chain.

Exercise:

Consider a CTMC with state space $\mathcal{S} = \{1, 2, 3\}$ and transition rates $q_{2,1} = \lambda p$, $q_{2,3} = \lambda(1 - p)$, where $\lambda > 0$ and $p \in (0, 1)$, and all other transition rates are 0.

- Which states are absorbing?
- Find $m_{2,\{1,3\}}$.
- Find the hitting probability $h_{2,1}$.

Remarks

There are more general CTMCs that allow $\lambda_i = \infty$, so that some states are only visited for an instant - we will not talk about these.

One can construct CTMC that are explosive in the sense that the process can jump infinitely many times in a finite amount of time. E.g. if $\mathcal{S} = \mathbb{Z}_+$ and $\lambda_i = \lambda^i$ for some $\lambda > 1$ and $b_{i,i+1} = 1$ for each i then the CTMC is explosive. We henceforth assume that our CTMC is not explosive.

A similar argument to the one we saw in the DTMC setting shows that an irreducible finite-state CTMC is positive recurrent.

Long run behaviour

As for DTMC, the limiting proportion of time spent by a CTMC in a given state can be random (recall our reducible examples).

For an irreducible transient CTMC, the limiting proportion of time spent in each state is 0, and there is no limiting *distribution*.

An irreducible positive recurrent CTMC is *ergodic*, i.e. the limiting distribution exists and does not depend on the initial distribution. The limiting distribution can be specified in terms of a quantity similar to the "expected return time" that appeared in DTMC. This is also equal to the stationary distribution of the CTMC (see next slide)

Stationary distribution

Recall that for a DTMC, a distribution $\pi = (\pi_i)_{i \in \mathcal{S}}$ is a stationary distribution for (a DTMC with transition matrix) P if $\pi P = \pi$.

This can be rewritten as $\pi(P - I) = 0$, where I is the $|\mathcal{S}| \times |\mathcal{S}|$ identity matrix.

A distribution $\pi = (\pi_i)_{i \in \mathcal{S}}$ is called a *stationary distribution* for the family $(P^{(t)})_{t \geq 0}$ if $\pi P^{(t)} = \pi$ for each $t \geq 0$.
Let $q_{i,i} := -\sum_{j \neq i} q_{i,j} = -\lambda_i$ and let Q denote the matrix (called the (infinitesimal) generator, or rate matrix) whose i,jth entry is $q_{i,j}$.

For non-explosive CTMCs a distribution $\pi = (\pi_i)_{i \in \mathcal{S}}$ is a stationary distribution for (a Markov chain with rate matrix) Q if and only if $\pi Q = 0$.

This is equivalent to the set of equations $\pi_i \lambda_i = \sum_{j \neq i} \pi_j q_{j,i}$ for $i \in \mathcal{S}$ which are referred to as the full balance equations.

The main result

Theorem: An irreducible and positive recurrent CTMC has a unique stationary distribution π. For such a DTMC, the limiting proportion of time spent in state i is π_i and the limiting distribution is π (irrespective of the initial distribution).

Note that periodicity is not an issue for a CTMC.

For $i \in \mathcal{S}$, let $T_1^{(i)} = \inf\{t > 0 : X_t \neq i\}$ and $T^{i,i} = \inf\{t > T_1^{(i)} : X_t = i\}$. Then

$$\pi_i = \frac{\mathbb{E}[T_1^{(i)} | X_0 = i]}{\mathbb{E}[T^{i,i} | X_0 = i]}.$$

The stationary distribution continued

The quantity $\frac{\mathbb{E}[T_1^{(i)}|X_0=i]}{\mathbb{E}[T^{i,i}|X_0=i]}$ is a bit like the proportion of time spent at i up to the first time that we *return* to i (start from i initially).

The numerator of this quantity is $\frac{1}{\lambda_i} = -\frac{1}{q_{i,i}}$. By a first step analysis, the denominator is

$$\frac{1}{\lambda_i} + \sum_{j \in \mathcal{S}} b_{i,j} m_{j,i}.$$

This is because on average we take time $1/\lambda_i$ to escape from i, at which point we go to j with probability $b_{i,j}$ and then we have to get from j to i (which takes time $m_{j,i}$ on average).

Reversibility

Suppose that we find a distribution $\pi = (\pi_i)_{i \in \mathcal{S}}$ such that

$$\pi_i q_{i,j} = \pi_j q_{j,i}, \qquad \text{for all } i, j \in \mathcal{S}.$$

Then we say that Q is reversible.

The above equations are called the detailed balance equations.

Note that (exercise!) if π satisfies the detailed balance equations then it satisfies the full balance equations too. Thus any distribution satisfying the detailed balance equations is a stationary distribution.

Explosive CTMC revisited

For explosive CTMCs, it is possible to have a solution to

$$\boldsymbol{\pi} Q = 0,$$

with $\sum_j \pi_j = 1$ that is not the stationary distribution.

- Take the CTMC with $q_{i,i+1} = \lambda_i p$ for $i \geq 0$, and $q_{i,i-1} = (1-p)\lambda_i$ for $i \geq 1$ as the only non-zero transition rates.
- If $p > 1-p$, then the chain is transient, but choosing $\lambda_i = \lambda^i$ there is a solution to
$$\boldsymbol{\pi} Q = 0$$
of the form $\pi_i = \pi_0 \left(\frac{p}{(1-p)\lambda}\right)^i$. Thus we can get a solution that sums to 1 if $\lambda > p/(1-p)$.

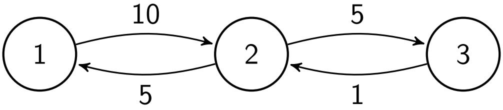
Expected hitting times

Theorem: For a CTMC with state space $\mathcal{S}$, and $A \subset \mathcal{S}$, the vector of mean hitting times $(m_{i,A})_{i\in\mathcal{S}}$ is the minimal non-negative solution to

$$m_{i,A} = \begin{cases} 0, & \text{if } i \in A, \\ \frac{1}{\lambda_i} + \sum_{j\in\mathcal{S}} b_{i,j} m_{j,A}, & \text{if } i \notin A. \end{cases}$$

Exercise:

Let $(X_t)_{t\geq 0}$ be a CTMC with the following transition diagram



- Find $m_{1,3}$.
- Find the stationary distribution for this chain.
- Find the stationary distribution for the jump chain.

The Chapman-Kolmogorov equations

Observe that

$$\begin{aligned} p_{i,j}^{(s+t)} &= \sum_k \mathbb{P}(X_{s+t} = j | X_s = k, X_0 = i)\mathbb{P}(X_s = k | X_0 = i) \\ &= \sum_k p_{i,k}^{(s)} p_{k,j}^{(t)}. \end{aligned}$$

These are the Chapman-Kolomogorov equations for a CTMC. In matrix form, the Chapman-Kolmogorov equations can be expressed as

$$P^{(t+s)} = P^{(s)} P^{(t)}.$$

Finding the transition probabilities

Thus far we have not actually computed $P^{(t)}$.

By analogy with the discrete-time case, we might hope that we can write $P^{(t)} = P^t$ for some matrix P.

If $t = m$ (a positive integer), the C-K equations tell us that $P^{(m)} = (P^{(1)})^m$ and our hope is fulfilled, but if $t < 1$?

We want a single object (like $P = P^{(1)}$ in the discrete case) that encodes the information of the chain.

It turns out that the generator Q is our single object. In fact we have the so-called Forward and Backward equations:

$$\frac{d}{dt} P^{(t)} = QP^{(t)} \quad (\ddagger) \qquad \text{backward equations}$$

and,

$$\frac{d}{dt} P^{(t)} = P^{(t)} Q. \quad (\dagger) \qquad \text{forward equations}$$

Solving the forward and backward equations

For (non-explosive) CTMCs, the matrix Q determines the transition probability completely by solving the backward or forward equations to get

$$\begin{aligned} P^{(t)} &= \exp(tQ) \\ &:= \sum_{k=0}^{\infty} \frac{1}{k!} t^k Q^k, \end{aligned}$$

subject to $P^{(0)} = I$.

Example: The Poisson process

The Poisson process is a CTMC with generator

$$Q = \begin{pmatrix} -\lambda & \lambda & 0 & 0 & 0 & 0 & \ddots \\ 0 & -\lambda & \lambda & 0 & 0 & 0 & \ddots \\ 0 & 0 & -\lambda & \lambda & 0 & 0 & \ddots \\ 0 & 0 & 0 & -\lambda & \lambda & 0 & \ddots \\ \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots \end{pmatrix}.$$

Poisson process transition probabilities

Can we derive the transition probabilities from Q? We could

- Compute $\exp(tQ)$.
- Solve the Kolmogorov backward or forward differential equations.

For the first case, one can show that $(Q^n)_{i,j} = 0$ if $j \notin \{i, i+1, \ldots, i+n\}$, and otherwise

$$(Q^n)_{i,j} = \lambda^n \binom{n}{j-i} (-1)^{n-(j-i)}.$$

Then for $j \geq i$,

$$\sum_{n=0}^{\infty} \frac{t^n}{n!} (Q^n)_{i,j} = \frac{(t\lambda)^{j-i}}{(j-i)!} e^{-t\lambda}.$$

Or we can directly solve e.g. the forward equations to find $p_{0,k}^{(t)}$.

Poisson process transition probabilities

Now,

$$\begin{aligned} \frac{d}{dt} p_{0,0}^{(t)} &= -\lambda p_{0,0}^{(t)} \\ \Longrightarrow p_{0,0}^{(t)} &= c e^{-\lambda t}. \end{aligned}$$

So with the condition that $p_{0,0}^{(0)} = 1$ we get $p_{0,0}^{(t)} = e^{-\lambda t}$.Similarly

$$\frac{d}{dt} p_{0,k}^{(t)} = \sum_{j=0}^{k} p_{0,j}^{(t)} q_{j,k} = \lambda(p_{0,k-1}^{(t)} - p_{0,k}^{(t)})$$

By induction, we can show that $p_{0,k}^{(t)} = e^{-\lambda t}(\lambda t)^k/k!$.

Interpretation of the generator

For small h,

$$\begin{aligned}
\mathbb{P}(X_{t+h} = k | X_t = j) &= p_{j,k}^{(h)} \\
&\approx (I + hQ)_{j,k} \\
&= \begin{cases} hq_{j,k}, & \text{if } j \neq k, \\ 1 + hq_{j,j}, & \text{if } j = k. \end{cases}
\end{aligned}$$

So indeed for $k \neq j$ we can think of $q_{j,k}$ as the rate of transition from j to k.

Example - birth and death processes

Let $(X_t)_{t \geq 0}$ be a CTMC on $\mathcal{S} = \mathbb{Z}_+$, where:

- X_t represents the number of 'people' in a system at time t.
- Whenever there are n 'people' in the system
  - new arrivals enter (by birth or immigration) the system at rate ν_n
  - 'people' leave (or die from) the system at rate μ_n
  - arrivals and departures occur independently of one another
- $(X_t)_{t \geq 0}$ is a birth-and-death process with arrival (or birth) rates $(\nu_n)_{n \in \mathcal{S}}$ and departure (or death) rates $(\mu_n)_{n \in \mathcal{S}}$.

Generator of a birth and death process

The generator of such a birth and death process has the form

$$Q = \begin{pmatrix} -\nu_0 & \nu_0 & 0 & 0 & 0 & \ddots \\ \mu_1 & -(\mu_1+\nu_1) & \nu_1 & 0 & 0 & \ddots \\ 0 & \mu_2 & -(\mu_2+\nu_2) & \nu_2 & 0 & \ddots \\ 0 & 0 & \mu_3 & -(\mu_3+\nu_3) & \nu_3 & \ddots \\ \ddots & \ddots & \ddots & \ddots & \ddots & \ddots \end{pmatrix}$$

The CTMC evolves by remaining in state k for an exponentially-distributed time with rate $\nu_k + \mu_k$, then it moves to state $k+1$ with probability $b_{k,k+1} = \nu_k/(\nu_k+\mu_k)$ and state $k-1$ with probability $b_{k,k-1} = \mu_k/(\nu_k+\mu_k)$, and so on.

Example

Consider a population in which there are no births ($\nu_i = 0$ for $i \geq 0$) and, for $i \geq 1$, the death rates are μ_i.
Suppose X_t is the population size at time t and so $(X_t)_{t \geq 0}$ is a CTMC.

1. Find expressions for $p_{i,i}^{(t)}$ and $p_{i,i-1}^{(t)}$ for $i \geq 1$.
2. Given that the population size is two at a particular time, calculate the probability that no more than one death will occur within the next one unit of time.

Birth and death stationary distribution

Assume $\nu_i > 0$ and $\mu_i > 0$ for all i. Assuming non-explosivity, we derive the stationary distribution (if it exists) by solving $\boldsymbol{\pi} Q = 0$. In fact we can solve the detailed balance equations:

$$\nu_k \pi_k = \mu_{k+1} \pi_{k+1}, \qquad k \in \mathbb{Z}_+$$

This has solution

$$\pi_k = \pi_0 \prod_{\ell=1}^{k} \frac{\nu_{\ell-1}}{\mu_\ell}.$$

So a stationary distribution exists if and only if

$$\sum_{k=0}^{\infty} \prod_{\ell=1}^{k} \frac{\nu_{\ell-1}}{\mu_\ell} < \infty$$

in which case

$$\pi_0 = \left(\sum_{k=0}^{\infty} \prod_{\ell=1}^{k} \frac{\nu_{\ell-1}}{\mu_\ell} \right)^{-1}.$$

Exercises

- Compute the stationary distribution of an 'M/M/1 queue', which is a birth and death process with $\nu_i = \nu$, $i \geq 0$, and $\mu_i = \mu$, $i \geq 1$.
- Compute the stationary distribution of a birth and death process with constant birth rate $\nu_i = \nu$, $i \geq 0$, and unit per capita death rate $\mu_i = i$, $i \geq 1$.

Finding the transition probabilities - the details

If t and h are nonnegative real numbers, we can write

$$\begin{aligned} \frac{P^{(t+h)} - P^{(t)}}{h} &= P^{(t)} \left[\frac{P^{(h)} - I}{h} \right] \\ &= \left[\frac{P^{(h)} - I}{h} \right] P^{(t)} \end{aligned}$$

This suggests that we should investigate the existence of the derivative

$$Q^* \equiv \lim_{h \to 0^+} \frac{P^{(h)} - I}{h}.$$

Under our assumptions about the chain (non-explosive, etc.), Q^* exists and equals Q.

Forward and backward equations

Since

$$\begin{aligned} \frac{P^{(t+h)} - P^{(t)}}{h} &= P^{(t)} \left[\frac{P^{(h)} - I}{h} \right] \\ &= \left[\frac{P^{(h)} - I}{h} \right] P^{(t)} \end{aligned}$$

if we can take the limits through the matrix multiplication then we get

$$\frac{d}{dt} P^{(t)} = Q P^{(t)} \quad (\ddagger) \qquad \text{backward equations}$$

and, similarly,

$$\frac{d}{dt} P^{(t)} = P^{(t)} Q. \quad (\dagger) \qquad \text{forward equations}$$

Why does $Q^* = Q$?

Let T_1 denote the time of the first jump of our process and T_2 denote the time of the second jump.
Then

$$\mathbb{P}(T_1 > h | X_0 = i) = e^{-\lambda_i h} = 1 - \lambda_i h + o(h) = 1 + q_{i,i} h + o(h).$$

$$\begin{aligned}
\mathbb{P}(T_2 < h | X_0 = i) &= \sum_{k \in \mathcal{S}} \mathbb{P}(T_1 < j, T_2 - T_1 < h - T_1, X_{T_1} = k | X_0 = i) \\
&\leq \sum_{k \in \mathcal{S}} \mathbb{P}(T_1 < j, T_2 - T_1 < h, X_{T_1} = k | X_0 = i) \\
&= \sum_{k \in \mathcal{S}} (1 - e^{-\lambda_k h})(1 - e^{-\lambda_i h}) b_{i,k} \\
&= \sum_{k \in \mathcal{S}} (\lambda_k h + o(h))(\lambda_i h + o(h)) b_{i,k} \\
&= h^2 \sum_{k \in \mathcal{S}} (\lambda_k + o(1))(\lambda_i + o(1)) b_{i,k} = o(h),
\end{aligned}$$

provided that λ_k do not grow too fast with k.

Why does $Q^* = Q$?

So, $\mathbb{P}(T_1 > h | X_0 = i) = 1 - \lambda_i h + o(h)$. Therefore $\mathbb{P}(T_1 < h | X_0 = i) = \lambda_i h + o(h)$.
Also $\mathbb{P}(X_{T_1} = j | X_0 = i) = b_{i,j} = q_{i,j}/\lambda_i$.
Also $\mathbb{P}(T_2 < h | X_0 = i) = o(h)$.
So

$$p_{i,i}^{(h)} = \mathbb{P}(T_1 > h | X_0 = i) + \mathbb{P}(T_2 < h, X_h = i | X_0 = i) = 1 - \lambda_i h + o(h).$$

Similarly, for $j \neq i$,

$$\begin{aligned} p_{i,j}^{(h)} &= \mathbb{P}(X_{T_1} = j, T_1 < h, T_2 > h | X_0 = i) + \mathbb{P}(T_2 < h, X_h = j | X_0 = i) \\ &= \frac{q_{i,j}}{\lambda_i} \lambda_i h + o(h). \end{aligned}$$

Now you can see that $(P^{(h)} - I)_{i,j} = q_{i,j} h + o(h)$. Divide by h and take the limit...

Justifying the forward and backward equations

We write "hope that" since we need to justify pushing the limits through the (possibly infinite) sums

$$\lim_{h\to 0}\sum_{k\in\mathcal{S}} p_{i,k}^{(t)} \left[\frac{P^{(h)}-I}{h}\right]_{k,j} = \sum_{k\in\mathcal{S}} p_{i,k}^{(t)} \lim_{h\to 0} \left[\frac{P^{(h)}-I}{h}\right]_{k,j}$$

$$\lim_{h\to 0}\sum_{k\in\mathcal{S}} \left[\frac{P^{(h)}-I}{h}\right]_{i,k} p_{k,j}^{(t)} = \sum_{k\in\mathcal{S}} \lim_{h\to 0} \left[\frac{P^{(h)}-I}{h}\right]_{i,k} p_{k,j}^{(t)}$$

for each $i, j \in S$.

If $\mathcal{S}$ is finite, then there is no problem and both (‡) and (†) hold.

Justifying the backward equations

In fact, we know from Fatou's Lemma that, for $j, k \in S$,

$$\begin{aligned}
\liminf_{h \to 0^+} \frac{p_{jk}^{(t+h)} - p_{jk}^{(t)}}{h} &= \liminf_{h \to 0^+} \sum_{i \in S} \frac{(p_{ji}^{(h)} - \delta_{ji}) p_{ik}^{(t)}}{h} \\
&\geq \sum_{i \in S} \lim_{h \to 0^+} \frac{(p_{ji}^{(h)} - \delta_{ji}) p_{ik}^{(t)}}{h} \\
&= \sum_{i \in S} a_{ji} p_{ik}^{(t)}.
\end{aligned}$$

Similarly,

$$\liminf_{h \to 0^+} \frac{p_{jk}^{(t+h)} - p_{jk}^{(t)}}{h} \geq \sum_{i \in S} p_{ji}^{(t)} a_{ik}.$$

Justifying the backward equations

We can show that the inequality in the first expression is, in fact, an equality, as follows. For $N > j$,

$$\begin{aligned}
\sum_{i \in S} \frac{\left[p_{ji}^{(h)} - \delta_{ji}\right] p_{ik}^{(t)}}{h} &= \sum_{i=1}^{N} \frac{\left[p_{ji}^{(h)} - \delta_{ji}\right] p_{ik}^{(t)}}{h} + \sum_{i=N+1}^{\infty} \frac{p_{ji}^{(h)} p_{ik}^{(t)}}{h} \\
&\leq \sum_{i=1}^{N} \frac{\left[p_{ji}^{(h)} - \delta_{ji}\right] p_{ik}^{(t)}}{h} + \sum_{i=N+1}^{\infty} \frac{p_{ji}^{(h)}}{h} \\
&= \sum_{i=1}^{N} \frac{\left[p_{ji}^{(h)} - \delta_{ji}\right] p_{ik}^{(t)}}{h} + \frac{1 - \sum_{i=1}^{N} p_{ji}^{(h)}}{h} \\
&= \sum_{i=1}^{N} \frac{\left[p_{ji}^{(h)} - \delta_{ji}\right] \left[p_{ik}^{(t)} - 1\right]}{h}.
\end{aligned}$$

Justifying the backward equations

Therefore

$$\begin{aligned}
\limsup_{h\to 0^+} \sum_{i\in S} \frac{\left[p_{ji}^{(h)} - \delta_{ji}\right] p_{ik}^{(t)}}{h} &\leq \sum_{i=1}^{N} \lim_{h\to 0^+} \frac{\left[p_{ji}^{(h)} - \delta_{ji}\right] \left[p_{ik}^{(t)} - 1\right]}{h} \\
&= \sum_{i=1}^{N} a_{ji} p_{ik}^{(t)} - \sum_{i=1}^{N} a_{ji}.
\end{aligned}$$

Now we let $N \to \infty$ and use the fact that $\sum_{i=1}^{\infty} a_{ji} = 0$ to derive

$$\limsup_{h\to 0^+} \frac{p_{jk}^{(t+h)} - p_{jk}^{(t)}}{h} \leq \sum_{i\in S} a_{ji} p_{ik}^{(t)},$$

which proves that $p_{jk}^{(t)}$ is differentiable (since $\liminf \geq \limsup$) and

$$\frac{dp_{jk}^{(t)}}{dt} = \sum_{i\in S} a_{ji} p_{ik}^{(t)}.$$

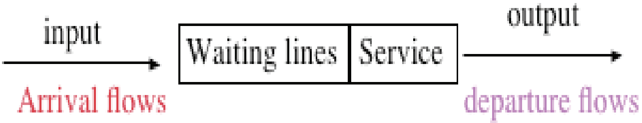
Justifying the forward and backward equations

So we have justified the interchange leading to (‡). However, the interchange leading to (†) does not hold for explosive CTMCs.

Queueing systems

Introduction

Queueing theory is the mathematical study of the operation of stochastic systems describing processing of flows of jobs. Queues occur when current demand for service exceeds the capacity of the service facility.



Arrivals

- We use the terminology 'customers', but they could be telephone calls, computer jobs, information packets, etc.
- Arrival times T_1, T_2, T_3, $\cdots$. The inter-arrival times are $\tau_1 = T_1 - T_0, \tau_2 = T_2 - T_1, \tau_3 = T_3 - T_2 \ldots$
- The inter-arrival times are assumed to be i.i.d.
- Alternatively, we could use the counting process N_t giving the number of arrivals in $[0, t]$, $t \geq 0$.

Service

- There is a total of m spaces for both receiving service and waiting for it.
- If there is an idle server, an arriving customer is serviced immediately.
- The service time $S_i^{(j)}$ of the ith customer at the jth server is a random variable.
- The service times are assumed to be i.i.d.
- When a server is serving a customer, it cannot provide any service to other customers.
- If all servers are busy, then the arriving customers join a queue if there is enough space, otherwise, the customer is rejected.

Service Disciplines

This could be

- FIFO: First In - First Out (FCFS: First Come - First Served).
- Last Come - First Served (with or without pre-emption).
- Processor Sharing.
- Priority (with or without pre-emption).
- more complicated disciplines?

We will consider only FIFO in this course.
One can use such queueing systems to construct queueing networks by forwarding customers departing from one queue to other queues.

Quantities of interest

X_t is the number of customers in the system at time t (including those in service and those waiting to begin service).

waiting times: length of time a customer spends in the queue before her/his service commences.

sojourn times: total length of time spent in the system (waiting time plus service time).

Kendall's notation

This was devised by David Kendall in 1953. It takes the form $A/B/n/m$ where

- A describes the arrival process
  - $A = M$ (Markov) inter-arrival times are exponentially-distributed.
  - $A = GI$ or (G) inter-arrival times have some arbitrary distribution.
  - $A = D$ inter-arrival times are deterministic.

Kendall's notation

- B describes the service process
  - $B = M$ service times are exponentially-distributed.
  - $B = GI$ or (G) service times have some arbitrary distribution.
  - $B = D$ service times are deterministic.
- n gives the number of servers.
- m gives the capacity of the system. When $m = \infty$, this is usually omitted.

Questions

- Does a queueing system have a steady-state regime or does the queue increase unboundedly?
- What is the steady-state queue length distribution if it exists?
- What is the steady-state waiting time distribution if it exists?
- What is the average load on the server?
- What fraction of time is the server idle?

$M/M/1$ queue

- Arrival stream: Poisson process with intensity λ
- Service: $n = 1$ server, service times $\sim \exp(\mu)$
- Infinite space for waiting: $m = \infty$
- The state X_t gives the number of customers at time t:
  - If $X_t = 0$ the server is idle.
  - If $X_t = k \geq 1$ one customer is being served and $k - 1$ customers are waiting in the queue.

This is a CTMC (in fact a birth and death process) with non-zero transition rates $q_{i,i+1} = \lambda$ and $q_{i+1,i} = \mu$ for all $i \in \mathbb{Z}_+$.

Exercise: draw the transition diagram for this CTMC

$M/M/1$ an interpretation

If $X_t = 0$, the process remains at 0 for an $\exp(\lambda)$ time τ_+ until a new customer arrives, so $X_{t+\tau_+} = 1$.
If $X_t = k > 0$, the process remains at k for a time $\tau = \min(\tau_+, \tau_-)$ where

- $\tau_+ \sim \exp(\lambda)$ is the time until the next arrival after t
- $\tau_- \sim \exp(\mu)$ is the time until the end of service of the customer in service at t.

$X_{t+\tau} = k + 1$ if $\tau_+ < \tau_-$ and $X_{t+\tau} = k - 1$ if $\tau_+ > \tau_-$.

$M/M/1$ stationary distribution

Using our results from CTMCs, we see that a stationary distribution for $(X_t)_{t\geq 0}$ exists if (and only if) the chain is positive recurrent. This is equivalent to $\rho \equiv \lambda/\mu < 1$, in which case, for $n \in \mathbb{Z}_+$,

$$\pi_n = (\lambda/\mu)^n \pi_0.$$

Using the normalisation condition $\sum_{i=0}^{\infty} \pi_n = 1$, we see that

$$\pi_0 \sum_{i=0}^{\infty} (\lambda/\mu)^i = 1$$

which tells us that

$$\pi_0 = 1 - \rho$$

and, for $n \geq 0$,

$$\pi_n = (1-\rho)\rho^n.$$

So the stationary distribution for the number of customers in the system is geometric*$(1-\rho)$. (Note that this geometric takes values in $\mathbb{Z}_+$).

$M/M/1$ further questions

- What is the stationary expected number ℓ of customers in the whole system?
- What is the stationary expected number ℓ_q of customers in just the queue?
- What is the expected waiting time of a customer in stationarity?
- What is the distribution of the waiting time?

We might guess that $\ell_q = \ell - 1$. However this is not right because the queue might be empty.

$M/M/1$ waiting times in stationarity

Assume that an $M/M/1$ queue is operating under a FCFS discipline.

- In the stationary regime, a tagged arriving customer will find a random number N of customers where $N \sim (\pi_k)_{k \in \mathbb{Z}_+}$ (PASTA).
- If $N = 0$, then the customer will go straight into service.
- If $N > 0$, the remaining service time S_1 for the customer being served $\sim \exp(\mu)$.
- The service times $S_2, S_3, \ldots, S_N$, for those in the queue are independent $\exp(\mu)$ random variables, also independent of N.
- So the waiting time for our tagged customer is $W = \sum_{j=1}^{N} S_j$, where we interpret the empty sum as equal to 0.

$M/M/1$ waiting times in stationarity

The distribution of a non-negative random variable Y is characterized by its Laplace transform $M_Y(-s) = \mathbb{E}[e^{-sY}]$ for $s > 0$.
We can write

$$\begin{aligned}
\mathbb{E}_\pi[e^{-sW}] &= \mathbb{E}_\pi[\mathbb{E}_\pi[e^{-sW}|N]] = \mathbb{E}_\pi[\mathbb{E}_\pi[e^{-s\sum_{j=1}^N S_j}|N]] \\
&= \mathbb{E}_\pi[(\mathbb{E}_\pi[e^{-sS_1}])^N] \\
&= \mathbb{E}_\pi\left[\left(\frac{\mu}{s+\mu}\right)^N\right] = M_N(\log(\mu/(s+\mu))) \\
&= (1-\rho)\sum_{n=0}^{\infty}\rho^n\left(\frac{\mu}{s+\mu}\right)^n \\
&= \frac{(1-\rho)(s+\mu)}{s+\mu-\lambda} \\
&= (1-\rho) + \rho\frac{\mu-\lambda}{s+\mu-\lambda},
\end{aligned}$$

and we see that the distribution of W is a mixture of a 0-random variable and an exponential$(\mu-\lambda)$ random variable. To be precise,

$$\mathbb{P}(W=0) = 1-\rho, \qquad \mathbb{P}(W > x) = \rho e^{-x(1-\rho)\mu}, \quad \text{for } x > 0.$$

$M/M/1$ waiting times in stationarity

It follows that the expected waiting time is

$$\mathbb{E}[W] = \frac{\rho}{\mu - \lambda}.$$

Once we have the expected waiting time, we can calculate the expected total time d in the system via the formula

$$d = \mathbb{E}[W] + \frac{1}{\mu} = \frac{1}{\mu - \lambda}.$$

Little's law:

Recall that ℓ is the expected number of customers in the system, while ℓ_q is the expected number of customers waiting for service (both at stationarity).

Little's law says that

$$\ell = \lambda d,$$

and

$$\ell_q = \lambda \mathbb{E}[W].$$

Sketch proof of Little's law

Let N_t and D_t denote the number of customers who have *entered* and *departed* from the system in $[0, t]$ respectively. So the number in the system at time t is $X_t = N_t - D_t$. Denoting the area under the function X_s for $s \leq t$ by A_t, we calculate $\mathbb{E}[A_t/t]$ in two different ways. First

$$\mathbb{E}\left[\frac{A_t}{t}\right] = \mathbb{E}\left[\frac{1}{t}\int_0^t X_u du\right]$$

which approaches the average number ℓ in the system as $t \to \infty$.

Sketch proof of Little's law

Second, we have,

$$\frac{A_t}{t} = \frac{1}{t}\sum_{n=1}^{N_t} D_n + o(1)$$

where D_i is the time spent in the system by the ith customer.

Now

$$\begin{aligned}\mathbb{E}\left[\frac{A_t}{t}\right] &= \frac{1}{t}\mathbb{E}[\sum_{n=1}^{N_t} D_n] + o(1)\\ &= \frac{1}{t}\lambda t d + o(1)\\ &= \lambda d + o(1)\end{aligned}$$

So taking t large we have $\ell = \lambda d$.

Example

A repairperson is assigned to service a bank of machines in a shop. Assume that failure times occur according to a Poisson process with rate $\lambda = 1/12$ per minute and the repair rate is $\mu = 1/8$ per minute.

Example

- The traffic intensity is $\rho = 2/3 < 1$, so a stationary distribution exists.
- For $k \geq 0$, the stationary distribution is $\pi_k = (1-\rho)\rho^k$.
- The repairperson is idle with prob $1 - \rho = 1/3$.
- The expected number of machines under repair is $\ell = \rho/(1-\rho) = 2$.
- The expected time that a machine spends with the repair person is $d = \ell/\lambda$ (or $1/(\mu - \lambda)$) $= 24$ minutes.
- The expected time waiting for repair is $\mathbb{E}[W] = d - 1/\mu = 16$ minutes.
- Also, e.g. $\mathbb{P}(W > 10) = \rho e^{-(\mu-\lambda)10} \approx 0.44$.

Example

- Suppose that the failure rate of machines increases (e.g. due to aging) by 16% to $\lambda' = 1/10$, then the new traffic intensity is $\rho' = 4/5$, and $\ell' = 4$ with $d' = 40$ and $\mathbb{E}[W'] = 32$.
- A 16% increase in arrival time rate has drastically increased the expected number of failed machines and doubled the time that they have to wait before getting repaired.
- We see that, when ρ is close to 1, the effect of small changes of ρ is profound: if a queueing system has long waiting times and lines, a rather modest increase in the service rate can bring about a dramatic reduction in waiting times.

Costs example

- Suppose there is a new piece of equipment that will increase the repair rate from $\mu = 1/8$ to $\mu^* = 1/6$, that is, decrease the expected repair time from 8 minutes to 6 minutes.
- The increase in maintenance cost for the new equipment is $c_M = \$6$ per minute.
- The cost of lost production when a machine is out of order is $c_D = \$5$ per minute.
- Should we purchase the new equipment?

Costs example solution

Without the new equipment

- The expected number of failed machines is $\ell = \rho/(1-\rho) = 2$.
- The expected cost of lost production is $\ell c_D = \$10$ per minute.
- With the new equipment, $\rho^* = 0.5$, and the expected cost is $\ell^* c_D + c_M = \$11$ per minute.
- We should buy the equipment if $\ell^* c_D + c_M < \ell c_D$, so we should not buy the equipment.

Another costs example

At a service station the rate of service is μ cars per hour, and the rate of arrivals of cars is λ per hour. The cost incurred by the service station due to delaying cars is $\$c_1$ per car per hour and the operating and service costs are $\$\mu c_2$ per hour. The rate of service μ is a control parameter. Determine the value of μ so that the least expected cost is achieved and find the value of the latter.

Another costs example solution

- If there are Y cars in the service station, the cost is \$$c_1 Y + \mu c_2$ per hour.
- In the stationary regime, $\mathbb{E}[Y] = \rho/(1-\rho)$.
- The expected total cost per hour is \$$c(\mu) = c_1\rho/(1-\rho) + \mu c_2$
- To find the minimum, we find μ such that $c'(\mu) = 0$ and since $\mu > \lambda$, we have a solution $\mu_0 = \lambda + \sqrt{\lambda c_1/c_2}$.
- We can check that μ_0 achieves the minimum and $c(\mu_0) = \lambda c_2 + 2\sqrt{\lambda c_1 c_2}$.

$M/M/a$ Queue

This system has the following properties

- $a \geq 1$ servers,
- a Poisson arrival process with rate λ,
- a FIFO service discipline,
- independent $\exp(\mu)$ service times,
- when an arrival finds more than one idle server, it chooses one at random,
- when k servers are working, the total service rate is $k\mu$.

$M/M/a$ Queue

The transition rates are $q_{i,i+1} = \lambda$, for $i \geq 0$ and $q_{i,i-1} = \mu \times \min(a, i)$ for $i \geq 1$.

Exercise: draw the transition diagram

This is a birth-and-death process with $\nu_i = \lambda$ for $i = 0, 1, 2, \ldots$ and $\mu_i = i\mu$ for $i = 1, 2, \ldots, a$ and $\mu_i = a\mu$ for $i > a$.

$M/M/a$ ergodicity

$$\kappa_j \equiv \frac{\nu_0 \cdots \nu_{j-1}}{\mu_1 \cdots \mu_j} = \begin{cases} (\lambda/\mu)^j / j! & \text{if } j \leq a \\ \frac{(\lambda/\mu)^j}{a! a^{j-a}} & \text{if } j > a. \end{cases}$$

We know that

$$\sum_{j=0}^{\infty} \kappa_j < \infty \iff \sum_{j=a}^{\infty} \kappa_j < \infty.$$

This occurs if $\lambda < a\mu$, in which case

$$\sum_{j=0}^{\infty} \kappa_j = \sum_{k=0}^{a-1} \frac{\lambda^k}{k! \mu^k} + \frac{\lambda^a}{a! \mu^a} \frac{a\mu}{a\mu - \lambda}.$$

$M/M/a$ ergodicity

So the $M/M/a$ queue is ergodic if and only if arrival rate λ is less than the maximum service rate $a\mu$.
In this case, the stationary distribution is given by

$$\pi_k = \begin{cases} \pi_0(\lambda/\mu)^k/k! & \text{if } k < a \\ \pi_0(\lambda/\mu)^k/(a!a^{k-a}) & \text{if } k \geq a, \end{cases}$$

where

$$\pi_0 = \left(\sum_{j=0}^{\infty} \kappa_j \right)^{-1}.$$

$M/M/a$ busy servers

For what proportion of time δ_q are all the servers busy? This is the same as the probability that an arriving customer will have to wait. (Why?)

We have

$$\delta_q = \sum_{k=a}^{\infty} \pi_k = \pi_0 \frac{\lambda^a}{\mu^a a!} \frac{a\mu}{a\mu - \lambda}.$$

The expected queue length is

$$\ell_q = \mathbb{E}[\max(X_t - a, 0)] = \frac{\lambda}{a\mu - \lambda} \delta_q.$$

$M/M/a$ busy servers

In stationarity, the expected number b_s of busy servers is

$$\mathbb{E}_\pi[\min(X_t, a)] = \frac{\lambda}{\mu}.$$

Note that, provided that $\lambda < a\mu$, this does not depend on a.
The expected number of customers in the system is

$$\ell = b_s + \ell_q = \frac{\lambda}{\mu} + \frac{\lambda}{a\mu - \lambda}\delta_q$$

$M/M/a$ waiting times

By Little's Law, the expected waiting time is

$$\mathbb{E}_\pi[W] = \frac{\ell_q}{\lambda} = \frac{\delta_q}{a\mu - \lambda}.$$

(can also get the above directly from π)
The expected delay is

$$d = \mathbb{E}_\pi[W] + \frac{1}{\mu} = \frac{\delta_q}{a\mu - \lambda} + \frac{1}{\mu}.$$

$M/M/a$ Example

An insurance company has 3 claim adjusters in its branch office. Claims against the company arrive according to a Poisson process at an average rate of 20 per 8 hour day. The amount of time an adjuster spends with a claimant is exponentially-distributed with mean service time 40 minutes.

- How many hours a week can an adjuster expect to spend with claimants?
- How much time, on average, does a claimant spend in the branch office?

$M/M/a$ example solution

- The arrival rate is $\lambda = 20/8 = 2.5$ per hour.
- The service rate is $\mu = 1.5$ per hour.
- $\lambda/(a\mu) = 5/9 < 1$, so a stationary distribution exists.
- We get $\mathbb{P}_\pi(\text{adjuster is busy})$ by noticing that

$$\mathbb{E}_\pi[\text{number of busy adjusters}] = \sum_{i=1}^{3} \mathbb{E}_\pi[\mathbb{1}_{\{i\text{th adjuster is busy}\}}]$$

So, by symmetry,

$$\begin{aligned} \mathbb{E}_\pi[\text{number of busy adjusters}] &= 3\mathbb{E}_\pi[\mathbb{1}_{\{\text{a given adjuster is busy}\}}] \\ &= 3\mathbb{P}_\pi(\text{a given adjuster is busy}). \end{aligned}$$

$M/M/a$ example solution

Substituting the parameter values, we calculate that each adjuster spends 22.2 hours a week on claims and that $\pi_0 = 24/139$, $\delta_q = 125/417$ and $d = 0.817$ hours (which corresponds to 49 minutes).

If there were only two adjusters, we could similarly calculate

- An adjuster will spend on average 33.3 hours with claimants.
- We can calculate that $\pi_0 = 1/11$, $\delta_q = 25/33$ and $d = 2.18$ hours.

We can use this information to quantify the trade-off between the cost of an extra adjuster and the extra level of service that is produced.

Single or multiple servers?

Which is better? A single fast server or several smaller ones with the same "total productivity"? Assume that the arrival process is Poisson with rate λ, and compare:

- A single server with service rate $a\mu$, and
- a servers with service rate μ each.

A heuristic argument tells us that

- if $X_t \geq a$, both systems work with the same rate, but
- if $X_t = k < a$ the rate for the a server queue is $k\mu$, which is less than the rate $a\mu$ for the single server.

So we might conclude that the single server is better. This is "easy" to prove via the technique of *coupling*.

Single or multiple servers?

We saw that, for the $M/M/a$ queue, the expected number in the system is

$$= \frac{\lambda}{\mu} + \frac{\lambda}{a\mu - \lambda}\delta_q$$

and the expected time in the system is

$$= \frac{1}{\mu} + \frac{1}{a\mu - \lambda}\delta_q.$$

For the $M/M/1$ queue with service rate $a\mu$, the expected number in the system is

$$= \frac{\lambda}{a\mu - \lambda}$$

and the expected time in the system is

$$= \frac{1}{a\mu - \lambda}.$$

Single or multiple servers?

With some work, we can show that $\delta_q + (a\mu - \lambda)/\mu > 1$, so both the expected number in the system and the expected time in the system are smaller for the $M/M/1$ queue, which proves our conjecture.
As an exercise, think about the waiting time, rather than the time in the system, for each of the systems.

What if our queue is not Markovian??

Then we need renewal theory!

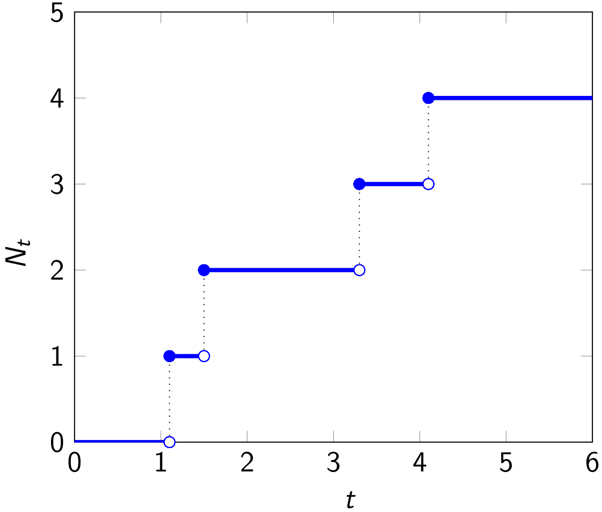
Renewal theory

Renewal process

A renewal process $(N_t)_{t \geq 0}$ is a counting process for which the times $\tau_j \geq 0$ between successive events, called renewals, are i.i.d. non-negative-real-valued random variables with an arbitrary common distribution function F.

- We assume $F(0) < 1$.
- The mean of τ_1 is $\mu > 0$ (which may or may not be finite).
- A Poisson process is a renewal process where the τ_i have an exponential distribution.
- A renewal process that is not a Poisson process is not Markovian.

A picture of N_t



$T_n = \sum_{i=1}^{n} \tau_i$ is time of nth jump.

Counting vs waiting representations

When we looked at the Poisson process, we saw that we could use a counting process description in terms of the number N_t of points in the interval $[0, t]$ or a waiting time description in terms of the time T_n until the nth event. This carries over to the study of renewal processes. Specifically

- $\{N_t \geq n\} = \{T_n \leq t\}$
- $\{N_t < n\} = \{T_n > t\}$
- $\{N_t = n\} = \{T_n \leq t < T_{n+1}\}$
- $T_{N_t} \leq t < T_{N_t+1}$.

Example

Light bulbs have a lifetime that has distribution function F. If a light bulb burns out, it is immediately replaced. Let N_t be the number of bulbs that have failed by time t. Then N_t is a renewal process.

N_t goes to ∞ as $t \to \infty$

For any fixed n,

$$\begin{aligned}\lim_{t\to\infty} \mathbb{P}(N_t > n) &= \lim_{t\to\infty} \mathbb{P}(T_n < t) \\ &= 1.\end{aligned}$$

This implies that with probability 1, $N_t \to \infty$ as $t \to \infty$.

Questions

- Can there be an explosion (that is an infinite number of renewals in a finite time)?
- What is the distribution of N_t?
- What is the average renewal rate? That is, at which rate does $N_t \to \infty$?

Explosion?

For any fixed $t < \infty$, $\mathbb{P}(N_t = \infty) = 0$. To see this, write

$$\begin{aligned}
\mathbb{P}(N_t = \infty) &= \lim_{n\to\infty} \mathbb{P}(N_t \geq n) = \lim_{n\to\infty} \mathbb{P}(T_n \leq t) \\
&= \lim_{n\to\infty} \mathbb{P}(\sum_{i=1}^{n} \tau_i \leq t) \\
&= \lim_{n\to\infty} \mathbb{P}(e^{-\sum_{i=1}^{n} \tau_i} \geq e^{-t}).
\end{aligned}$$

Using Markov's inequality ($\mathbb{P}(X \geq a) \leq \mathbb{E}[X]/a$ for $X \geq 0$ and $a > 0$) we have

$$\mathbb{P}(e^{-\sum_{i=1}^{n} \tau_i} \geq e^{-t}) \leq e^t \mathbb{E}[e^{-\sum_{i=1}^{n} \tau_i}] = e^t \left(\mathbb{E}[e^{-\tau_1}]\right)^n,$$

which goes to 0 as $n \to \infty$ since $\mathbb{E}[e^{-\tau_1}] < 1$ (why?)

Distribution of N_t

$$\begin{aligned}
\mathbb{P}(N_t = n) &= \mathbb{P}(T_n \leq t < T_{n+1}) \\
&= \mathbb{P}(T_n \leq t) - \mathbb{P}(T_n \leq t, T_{n+1} \leq t) \\
&= \mathbb{P}(T_n \leq t) - \mathbb{P}(T_{n+1} \leq t) \\
&= F_{T_n}(t) - F_{T_{n+1}}(t)
\end{aligned}$$

where F_n is the distribution function of T_n.
There are very few F for which this is easy to evaluate (can you name one?)

Growth of N_t

Above, we saw that $T_{N_t} \leq t < T_{N_t+1}$. It follows that

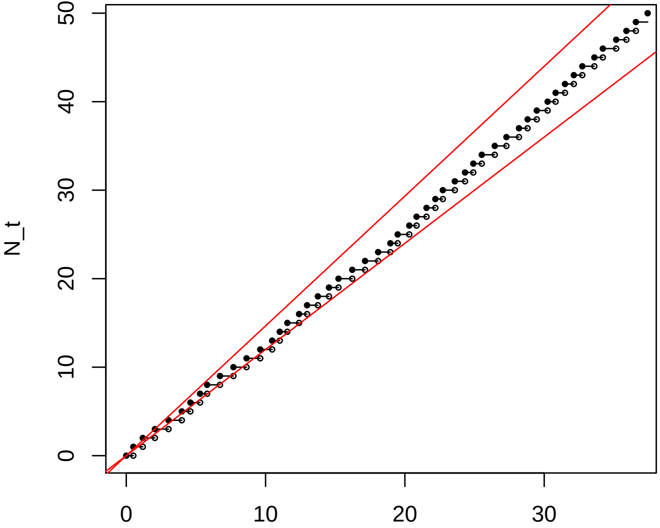
$$\frac{N_t}{N_t+1} \cdot \frac{N_t+1}{T_{N_t+1}} = \frac{N_t}{T_{N_t+1}} < \frac{N_t}{t} \leq \frac{N_t}{T_{N_t}}$$

Since $N_t \to \infty$ as $t \to \infty$, the Law of Large Numbers tells us that, with probability one, both the first and last terms approach μ^{-1}. Therefore, with probability one,

$$\lim_{t\to\infty} \frac{N_t}{t} = \mu^{-1},$$

and we see that, for large t, if $\mu < \infty$ then N_t grows like t/μ.

The LLN for N_t



Example

Jenny uses her smart phone a lot, so she carries a powerful portable charger with her. Whenever her phone gives her a low energy warning she immediately charges her phone for 30 minutes and that charge lasts for a $U(30, 60)$ (minutes, independent of previous charges) amount of time before she receives a warning. On average how many times per hour does Jenny charge her phone?

- $\mu = \mathbb{E}[\tau_1] = (45 + 30)/60$, so the rate is

$$\lim_{t \to \infty} \frac{N_t}{t} = \frac{1}{\mu} = \frac{60}{75} = \frac{4}{5} \text{ per hour}$$

$M/G/1/1$ queue

The arrival process is Poisson with parameter λ.

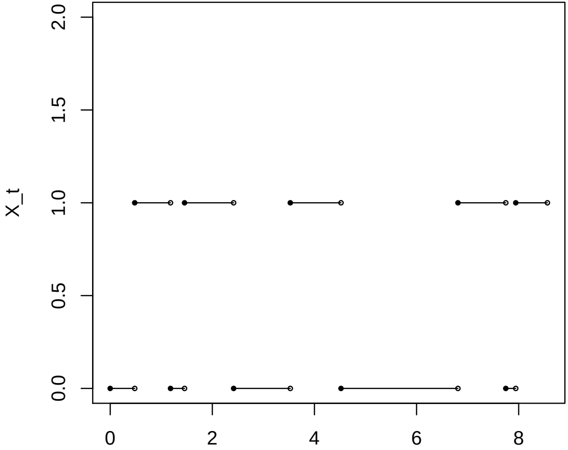
There is no "queue": when an arriving customer finds the server busy, s/he does not enter. Service times are independent and identically-distributed with distribution function G (with mean μ_G).

- What is the rate at which customers actually enter the system?
- What proportion of potential customers actually receive service?

$M/G/1/1$ queue

Let N_t be the number of customers who have been admitted by t. Then the times between successive entries of customers are made up of:

- a service time, and then
- a waiting time from the end of service until the next arrival.



$M/G/1/1$ queue

The mean time between renewals is $\mu = \frac{1}{\lambda} + \mu_G$. So the rate at which customers actually enter the system is

$$\frac{1}{\mu} = \frac{1}{\frac{1}{\lambda} + \mu_G} = \frac{\lambda}{1 + \lambda\mu_G}.$$

Customers arrive at rate λ, and so the *proportion* that actually enter the system is

$$\frac{\text{entry rate}}{\text{arrival rate}} = \frac{\frac{\lambda}{1+\lambda\mu_G}}{\lambda} = \frac{1}{1 + \lambda\mu_G}.$$

If $\lambda = 10$ per hour and $\mu_G = 0.2$ hours, then on average only 1 out of 3 customers will actually get served!

$G/G/n/m$ queue

In the very general setting of the $G/G/n/m$ queue, the beginnings of busy periods (i.e. times at which a customer arrives to find the system empty) constitute renewal times.

The time of the first "renewal" will typically have a different distribution though

The Renewal Central Limit Theorem

If $\mathbb{E}[\tau_j] = \mu$, $\mathrm{Var}(\tau_j) = \sigma^2 < \infty$, then

$$\frac{N_t - \frac{t}{\mu}}{\sqrt{t\sigma^2/\mu^3}} \xrightarrow{d} \mathcal{N}(0,1) \text{ as } t \to \infty.$$

So for each x,

$$\lim_{t\to\infty} \mathbb{P}\left(\frac{N_t - \frac{t}{\mu}}{\sqrt{t\sigma^2/\mu^3}} \le x \right) = \Phi(x),$$

where Φ is the normal distribution function.

Residual lifetime

The residual lifetime R_t at time t is the amount of time until the next renewal time.

Since $T_{N_t} \leq t < T_{N_t+1}$, the residual lifetime at time t is $R_t = T_{N_t+1} - t > 0$.

If τ_i are exponential(δ) then what is the distribution of R_t?

Let F be the c.d.f. of τ_1. We will study the large t behaviour of R_t assuming that F is non-lattice (that is, it does not concentrate its mass at multiples of a fixed amount), and has finite mean μ.

Residual lifetime for large t

Theorem: If F is non-lattice with finite mean μ and $x \geq 0$ then

$$\lim_{t\to\infty} \mathbb{P}(R_t \leq x) = \frac{1}{\mu}\int_0^x (1 - F(y))dy.$$

Recall that for $Z \geq 0$,

$$\mathbb{E}[Z] = \int_0^\infty (1 - F_Z(z))dz,$$

so $\frac{1-F(y)}{\mu}$, $y \geq 0$, is a probability density function.

Example

A computer receives packets of information whose sizes are uniformly distributed between 1 and 5 GB. It saves them on hard drives of total size 700GB, until the a hard drive is full.

- For the first file for which there is not enough space on a hard drive, find the approximate distribution and the mean of the length of the residual part that the hard drive does not have space for.

- Give a (symmetric) interval to which, the total number of saved files belongs with probability ≈ 0.95.

Example solution

Here, "time" is measured in GB! We are looking for the residual lifetime at time $t = 700$, where t is considered to be large (compared to the size of a packet).

- The limiting distribution of the residual part has density

$$\frac{1}{\mu}(1 - F(x)) = \begin{cases} \frac{1}{3} & \text{if } x \in [0, 1) \\ \frac{5-x}{12} & \text{if } x \in [1, 5]. \end{cases}$$

- The mean of the residual part is $31/18$, which is greater than half of the mean interval length, which is $3/2$.
- We have

$$\frac{N_t - \frac{t}{\mu}}{\sqrt{\frac{t}{\mu} \times \frac{\sigma^2}{\mu^2}}} \overset{d}{\approx} \mathcal{N}(0, 1).$$

  With $t = 700$, $\mu = 3$, $\sigma^2 = 4/3$, the desired (symmetric) interval is $233.33 \pm 5.88 \times 1.96 = (221.81, 244.85)$.

Age

The age of the renewal process at time t is the time since the most recent renewal, i.e. $A_t = t - T_{N_t}$.

Theorem: If F is non-lattice with finite mean μ and $x \geq 0$ then

$$\lim_{t\to\infty} \mathbb{P}(A_t \leq x) = \frac{1}{\mu}\int_0^x (1 - F(y))dy.$$

Age - some intuition

- Consider the process after it has been in operation for a long time.
- When we look backwards in time, the times between successive renewals are still (approximately) independent and identically-distributed with distribution F.
- Looking backwards, the residual lifetime at t is exactly the age at t of the original process.

Example

Suppose $(N_t)_{t\geq 0}$ is a Poisson process with rate λ, find the (approximate) distributions of R_t, A_t and (R_t, A_t) when t is large. What is the expected duration of the inter-event time $T_{N_t+1} - T_{N_t}$?

Renewal CLT - sketch proof

Recall that $T_n = \sum_{i=1}^{n} \tau_i$.

Let $Z_n = \frac{T_n - n\mu}{\sqrt{n\sigma^2}}$ and let $Z \sim \mathcal{N}(0,1)$. Then

$$\begin{aligned}
\mathbb{P}(N_t \geq n) &= \mathbb{P}(T_n \leq t) \\
&= \mathbb{P}\left(Z_n \leq \frac{t - n\mu}{\sqrt{n\sigma^2}}\right) \\
&\approx \mathbb{P}\left(Z \leq \frac{t - n\mu}{\sqrt{n\sigma^2}}\right) \\
&= \mathbb{P}\left(Z \geq \frac{n\mu - t}{\sqrt{n\sigma^2}}\right).
\end{aligned}$$

Renewal CLT - sketch proof

Now, we choose $n(x)$ such that $\frac{n\mu - t}{\sqrt{n\sigma^2}} \approx x$. That is, we put $n(x) \approx \frac{t}{\mu} + x\sqrt{\frac{t}{\mu} \cdot \frac{\sigma^2}{\mu^2}}$.

Then, reversing the above argument, we have

$$\begin{aligned} \mathbb{P}\left(Z \geq x\right) &\approx \mathbb{P}(N_t \geq n(x)) \\ &\approx \mathbb{P}\left(\frac{N_t - \frac{t}{\mu}}{\sqrt{t\sigma^2/\mu^3}} \geq x\right). \end{aligned}$$

Residual lifetime for large t - sketch of proof

Consider a period of n renewals. The proportion of time that the residual lifetime is greater than x is (by the Law of Large Numbers),

$$\begin{aligned} \frac{\sum_{i=1}^{n}(\tau_i - x)\mathbb{1}_{\{\tau_i > x\}}}{\sum_{i=1}^{n}\tau_i} &= \frac{\frac{1}{n}\sum_{i=1}^{n}(\tau_i - x)\mathbb{1}_{[\tau_i > x]}}{\frac{1}{n}\sum_{i=1}^{n}\tau_i} \\ &\to \frac{\mathbb{E}\left[(\tau_1 - x)\mathbb{1}_{[\tau_1 > x]}\right]}{\mathbb{E}[\tau_1]}. \end{aligned}$$

as n approaches infinity.

Residual lifetime for large t - sketch of proof

Under the stated conditions, it can also be shown that

$$\frac{\sum_{i=1}^{n}(\tau_i - x)\mathbb{1}_{[\tau_i > x]}}{\sum_{i=1}^{n}\tau_i} \to \lim_{t\to\infty}\mathbb{P}(R_t > x).$$

Hence

$$\begin{aligned}
\lim_{t\to\infty}\mathbb{P}(R_t > x) &= \frac{\mathbb{E}\left[(\tau_1 - x)\mathbb{1}_{[\tau_1 > x]}\right]}{\mathbb{E}[\tau_1]} \\
&= \frac{1}{\mu}\int_0^\infty \mathbb{P}((\tau_1 - x)\mathbb{1}_{[\tau_1 > x]} > y)dy \\
&= \frac{1}{\mu}\int_x^\infty \mathbb{P}(\tau_1 > u)du.
\end{aligned}$$

Age proof

For $x, y \geq 0$ the events $\{R_t > x, A_t > y\}$ and $\{R_{t-y} > x + y, A_{t-y} > 0\}$ are equal since

$$\begin{aligned} R_t > x \text{ and } A_t > y &\iff \text{no renewal in } [t-y, t+x] \\ &\iff R_{t-y} > x + y \text{ and } A_{t-y} > 0. \end{aligned}$$

Since τ is non-lattice, $\mathbb{P}(A_{t-y} = 0) \to 0$ as $t \to \infty$, so

$$\begin{aligned} \lim_{t\to\infty} \mathbb{P}(R_t > x, A_t > y) &= \lim_{t\to\infty} \mathbb{P}(R_{t-y} > x + y) \\ &= \frac{1}{\mu} \int_{x+y}^{\infty} [1 - F(z)] dz. \end{aligned}$$

Setting $x = 0$ we get

$$\begin{aligned} \lim_{t\to\infty} \mathbb{P}(A_t \leq y) &= \frac{1}{\mu} \int_0^y [1 - F(z)] dz \\ &= \lim_{t\to\infty} \mathbb{P}(R_t \leq y). \end{aligned}$$

Example

For large t, find the joint probability density function of (R_t, A_t) in the computer packets example.

- First,

  $$\mathbb{P}(A_t \le x, R_t \le y) = \mathbb{P}(A_t \le x) - \mathbb{P}(R_t > y) + \mathbb{P}(A_t > x, R_t > y),$$

  so
  $$\frac{\partial^2 \mathbb{P}(A_t \le x, R_t \le y)}{\partial x \partial y} = \frac{\partial^2 \mathbb{P}(A_t > x, R_t > y)}{\partial x \partial y}.$$

- When t is large, $\mathbb{P}(A_t > y, R_t > x) \approx \int_{x+y}^{\infty} \frac{1-F(z)}{\mu} dz$.
- Hence, the joint pdf is $1/12$ if $1 < x + y < 5$ and 0 otherwise.