MA3H2 Markov Processes and Percolation Theory

Peter Windridge, last update April 10, 2012.

Email: p dot windridge at warwick.

Lecture times: Mon 10am in B3.03, Tue 3pm in MS.03 and Thu 5pm in MS.04.

5 exercise sheets. First due Monday 3pm week 3. Rest are due Friday 3pm weeks 4,6,8,10. Marks from the best 4 count for 15%.

Remaining 85% is from the final summer exam.

Problem classes: taken by M. Eyers on Thu 11am in B1.01, beginning week 2.

Exercise sheets and book recommendations will be posted at `http://warwick.ac.uk/~masjah/ma3h2/`.

Part I (discrete time Markov chains) will be weeks 1–4/5. Part II (continuous time Markov chains) weeks 5/6–7. Part III (percolation theory) till the end of term.

Warning: (1) Sections marked with a '†' are not examinable.

(2) these notes are being typed up rather quickly and may contain small mistakes. Please ask if there is any doubt. Typos will be corrected asap.

Thanks to Dr N. Grinberg for creating the diagrams.

1 Markov chains in discrete time

1.1 Introduction

A stochastic process is a model for a random quantity changing over time, like monthly rainfalls, FTSE100 closing values, daily bus journey time to campus.

We consider memoryless ('gold fish') processes whose future evolution is independent of the past given the present.

1.2 Preliminaries

All random variables are defined on an underlying probability space denoted $(\Omega, \mathcal{F}, \mathbb{P})$–

- Ω is a set of 'possible realisations'
- $\mathcal{F} \subset 2^\Omega$ is a σ-algebra for Ω: $\Omega \in \mathcal{F}$, $A \in \mathcal{F} \Rightarrow A^c \in \mathcal{F}$ and $A_1, A_2, \ldots \in \mathcal{F} \Rightarrow \cup_i A_i \in \mathcal{F}$.
- $\mathbb{P} : \mathcal{F} \to [0,1]$ is a probability measure: $\mathbb{P}(\Omega) = 1$, $\mathbb{P}(\cup_i A_i) = \sum_i \mathbb{P}(A_i)$ for any disjoint $A_1, A_2, \ldots \in \mathcal{F}$.

$X : \Omega \to \mathbb{R}$ is a random variable (henceforth RV) if $\{\omega \in \Omega : X(\omega) \leq t\} \in \mathcal{F}$ for any $t \in \mathbb{R}$, i.e. X is measurable with respect to the Borel σ-algebra $\mathcal{B}(\mathbb{R})$.

If all this is unfamiliar, check your ST111 notes for a reminder.

Definition 1.1. *A stochastic process with state space $I \subset \mathbb{R}$ is a family, $(X_t)_{t \in T}$ say, of random variables indexed by a set $T \subset [0, \infty)$ ('time') taking values in I.*

The state space is assumed countable.

Typically $I = \{1, 2, \ldots, n\}$, $\mathbb{Z}$ or $\mathbb{N}_0 = \{0, 1, 2, \ldots\}$.

More exotic examples are $I = \{C, A, G, T\}^n$ (DNA sequence evolution) or $\{-1, +1\}^V$ (Glauber dynamics).

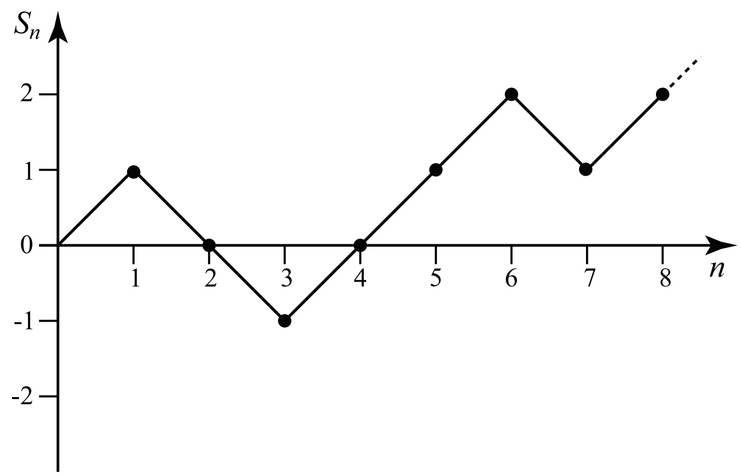
For now, $T = \mathbb{N}_0$. The index 'n' is used rather than 't'.

To save keystrokes and chalk '$(X_n)_{n \in \mathbb{N}_0}$' is frequently shortened to 'X', if it is clear that 'X' refers to a stochastic process.

Example 1.2 (Simple random walk)**.** *Let $p \in [0,1]$ and $\xi_1, \xi_2, \ldots$ be independent and identically distributed (IID) random variables taking value -1 or $+1$ with $\mathbb{P}(\xi_i = +1) = p$.*

Define $S_n = S_{n-1} + \xi_n$, $n \geq 1$ and $S_0 = x_0 \in \mathbb{Z}$ is chosen arbitrarily.

The stochastic process $S = (S_n)_{n \in \mathbb{N}_0}$ is called a simple random walk on $\mathbb{Z}$ starting from x_0.



- *A realisation of S is a sequence $x_0, x_1, \ldots \in \mathbb{Z}$ with $x_{n+1} - x_n \in \{-1, +1\}$. Such a sequence is called a nearest neighbour path (NNP).*
- *If $p = 1/2$ then S is called symmetric, otherwise it is called biased.*

1.3 Markov property

Why use SRW as an example? It is a widely studied process with the memoryless property referred to in the introduction–

Let S be a SRW on $\mathbb{Z}$ started at x_0.

Let $x_0, x_1, \ldots \in \mathbb{Z}$. Then

$$\mathbb{P}(S_{n+1} = x_{n+1}, S_n = x_n, \ldots, S_0 = x_0) = \mathbb{P}(S_{n+1} - S_n = x_{n+1} - x_n, S_n = x_n, \ldots, S_1 = x_1, S_0 = x_0)$$

which equals

$$\mathbb{P}(S_{n+1} - S_n = x_{n+1} - x_n)\mathbb{P}(S_n = x_n, \ldots, S_1 = x_1, S_0 = x_0)$$

since $S_{n+1} - S_n = \xi_{n+1}$ is independent of $S_k = x_0 + \sum_{j=1}^k \xi_j$ for $k \leq n$.

Recall Bayes Law. If $A, B \in \mathcal{F}$, $\mathbb{P}(B) > 0$, then

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

So, if $\mathbb{P}(S_n = x_n, \ldots, S_0 = x_0) > 0$ (which it is if $x_0, x_1, \ldots$ is a NNP, then

$$\mathbb{P}(S_{n+1} = x_{n+1}|S_n = x_n, \ldots, S_0 = x_0) = \mathbb{P}(S_{n+1} - S_n = x_{n+1} - x_n) = \mathbb{P}(S_{n+1} = x_{n+1}|S_n = x_n).$$

That is, given the present and the past, only the present is relevant for the next step.

Now we want to formalise the idea that S has no memory and starts afresh at each step.

It is helpful to explicitly identify the starting point S_0 with the notation $\mathbb{P}(\cdot|S_0 = x_0)$.

If this notation bothers you then let S_0 be random, independent of $\xi_1, \xi_2, \dots$ and $\mathbb{P}(S_0 = x_0) > 0$ for every x_0. Then $\mathbb{P}(\cdot|S_0 = x_0)$ is a well defined conditional distribution.

The transition probability

$$\mathbb{P}(S_{n+1} = j|S_n = i) = \mathbb{P}(S_{n+1} - S_n = j - i) = \mathbb{P}(\xi_{n+1} = j - i) = \mathbb{P}(S_1 = j|S_0 = i).$$

does not depend on n for any $i, j \in \mathbb{Z}$. This property is called time-homogeneity.

Loosely, memoryless processes are characterised by their initial distribution and the transition probability of jumping to state j when the present state is i.

Transition probabilities are conveniently written in a matrix.

Definition 1.3. *$P = (p_{i,j})_{i,j\in I}$ is called a transition matrix if its entries satisfy*

- *$p_{i,j} \geq 0$ for every $i, j \in I$ and*
- *$\sum_j p_{i,j} = 1$ for any $i \in I$.*

I.e., $(p_{i,j})_{j\in I}$ is a probability distribution on I for every $i \in I$.

Example 1.4.

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

is a transition matrix.

Definition 1.5. *A stochastic process $X = (X_n)_{n\in\mathbb{N}_0}$ on I is called a Markov chain with transition matrix P and initial probability distribution $(\lambda_i)_{i\in I}$ if it*

1. *satisfies the Markov property: given X_n, X_{n+1} is independent of $X_1, \dots, X_{n-1}$. I.e.*

   $$\mathbb{P}(X_{n+1} = i_{n+1}|X_n = i_n, \dots, X_0 = i_0) = \mathbb{P}(X_{n+1} = i_{n+1}|X_n = i_n)$$

   for any $i_0, \dots, i_n, i_{n+1} \in I$ and $n \in \mathbb{N}_0$ with $\mathbb{P}(X_0 = i_0, \dots, X_n = i_n) > 0$.
2. *is time-homogenous: $\mathbb{P}(X_{n+1} = j|X_n = i) = p_{i,j}$ for any $i, j \in I$ with $\mathbb{P}(X_n = i) > 0$.*
3. *has initial distribution λ: $\mathbb{P}(X_0 = i) = \lambda_i$ for all i.*

We write 'X is a (P, λ)-MC'.

In the case $\lambda_j = \delta_{ij}$ for every j and a fixed $i \in I$, we say X starts from i.

It is often helpful to draw a graph with nodes $V = I$ and a directed edge (i, j) if $p_{i,j} > 0$.

[Example with P as above].

Recap definition of a (λ, P)-MC.

Let $(S_n)_{n\in\mathbb{N}_0}$ be a SRW started from $x_0 \in \mathbb{Z}$ with up jump probability $\mathbb{P}(S_{n+1} - S_n = +1) = p \in [0,1]$.

Then S is (λ, P)-MC on $I = \mathbb{Z}$ with transition matrix $P = p_{i,j}$ given by

$$p_{i,i+1} = p,\ p_{i,i-1} = 1 - p$$

and all other entries equal to zero.

The initial distribution $\lambda_i = \delta_{i,x_0}$.

Joint distribution of $(S_0, \ldots, S_n)$: Let $x_0, x_1, \ldots \in \mathbb{Z}$ be a nearest neighbour path (NNP), i.e. $x_m - x_{m-1} \in \{-1, +1\}$ for all m.

Then, as before, the probabilities factorise into

$$\mathbb{P}(S_1 = x_1, \ldots, S_n = x_n) = \prod_{m=1}^{n} \mathbb{P}(S_m - S_{m-1} = x_m - x_{m-1}).$$

Each of the probabilities in the product is p or $(1-p)$ depending on whether they correspond to an up jump or down jump. Thus,

$$\mathbb{P}(S_1 = x_1, \ldots, S_n = x_n) = p^u (1-p)^d$$

where $u = \#\{m \le n : x_m - x_{m-1} = +1\}$ is the number of up jumps and d is the number of down jumps.

Note that

$$u + d = n,\ u - d = \sum_{m=1}^{n} ((x_m - x_{m-1})^+ + (x_m - x_{m-1})^-) = x_n - x_0.$$

For convenience write $x_n = a \in \mathbb{Z}$.

Remembering $x_0 = 0$, we have $u + d = n$, $u - d = a$, i.e. $u = (n+a)/2$ and $d = (n-a)/2$ (note that $n + a$ must be even).

Hence

$$\mathbb{P}(S_1 = x_1, \ldots, S_n = a) = p^{(n+a)/2} (1-p)^{(n-a)/2}$$

does not depend on $x_1, \ldots, x_{n-1}$ and every possible realisation has equal probability.

Thus the probability of some event occurring involving SRW is computed by counting NNPs, e.g.

$$\mathbb{P}(S_n = a | S_0 = 0) = \binom{n}{(n+a)/2} p^{(n+a)/2} (1-p)^{(n-a)/2}.$$

In general the transition probabilities of a MC determine the joint distributions–

Theorem 1.6. *Let X be a stochastic process. Then X is a (λ, P)-MC iff*

$$\mathbb{P}(X_n = i_n, \ldots, X_0 = i_0) = \lambda_{i_0} p_{i_0,i_1} \ldots p_{i_{n-1},i_n}. \tag{1}$$

for any $i_0, \ldots, i_n \in I$.

Proof. Suppose X is a (λ, P)-MC. $\mathbb{P}(X_0 = i_0) = \lambda_{i_0}$ is immediate so (1) holds with $n = 0$.

Suppose for induction that the equality (1) holds for some $n \in \mathbb{N}$. If both sides vanish then

$$\mathbb{P}(X_{n+1} = i_{n+1}, X_n = i_n, \ldots, X_0 = i_0) = 0 = (\lambda_{i_0} p_{i_0,i_1} \ldots p_{i_{n-1},i_n}) p_{i_n,i_{n+1}}.$$

Otherwise,

$$\mathbb{P}(X_{n+1} = i_{n+1}, X_n = i_n, , \ldots, X_0 = i_0) = \mathbb{P}(X_{n+1} = i_{n+1} | X_n = i_n) \mathbb{P}(X_n = i_n, \ldots, X_0 = i_0)$$

by the Markov property. But

$$\mathbb{P}(X_{n+1} = i_{n+1} | X_n = i_n) \mathbb{P}(X_n = i_n, \ldots, X_0 = i_0) = (\lambda_{i_0} p_{i_0,i_1} \ldots p_{i_{n-1},i_n}) p_{i_n,i_{n+1}}$$

by the inductive hypothesis. Thus (1) holds with n replaced by $n + 1$.

Conversely, suppose X is a stochastic processes and satisfies (1). Again, $\mathbb{P}(X_0 = i_0) = \lambda_{i_0}$ is immediate. Suppose $\mathbb{P}(X_0 = i_0, \ldots, X_n = i_n) > 0$. We have

$$\mathbb{P}(X_{n+1} = i_{n+1} | X_n = i_n, \ldots, X_0 = i_0) = \frac{\lambda_{i_0} p_{i_0,i_1} \ldots p_{i_{n-1},i_n} p_{i_n,i_{n+1}}}{\lambda_{i_0} p_{i_0,i_1} \ldots p_{i_{n-1},i_n}}.$$

□

Theorem 1.7. *Suppose X is a (λ, P)-MC and $n \in \mathbb{N}$. Then*

1. *The distribution of X_n is given by*

$$\mathbb{P}(X_n = j) = (\lambda P^n)_j = \sum_i \lambda_i (P^n)_{ij}$$

   for every $j \in I$.

2. *Let $\lambda_i > 0$. The n-step transition probability is given by $\mathbb{P}(X_n = j | X_0 = i) = (P^n)_{ij}$ for any $i, j \in I$.*

Here P^n is defined as for finite matrices, i.e. $(P^{n+1})_{ij} = (PP^n)_{ij} = \sum_{k \in I} p_{ik} (P^n)_{kj}$.

Proof. Using the law of total probability,

$$\mathbb{P}(X_n = j) = \sum_{i_0 \in I} \ldots \sum_{i_{n-1} \in I} \mathbb{P}(X_n = j, X_{n-1} = i_{n-1}, \ldots, X_1 = i_1, X_0 = i_0).$$

So the previous theorem gives

$$\mathbb{P}(X_n = j) = \sum_{i_0,\ldots,i_{n-1}\in I} \lambda_{i_0} p_{i_0,i_1} \ldots p_{i_{n-1},j} = (\lambda P^n)_j.$$

The second identity is similar;

$$\mathbb{P}(X_n = j|X_0 = i) = \sum_{i_1,\ldots,i_{n-1}} \mathbb{P}(X_n = j, X_{n-1} = i_{n-1}, \ldots, X_1 = i_1, X_0 = i)/\lambda_i.$$

□

Thus calculating the n-step transition probabilities boils down to some linear algebra.

Example 1.8. *Consider again the transition matrix*

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

The eigenvalues are 1, $-1/4$, $-1/6$. *They are all distinct. Hence* P *has independent eigenvectors and can be diagonalised, i.e. there is an invertible matrix* V *with*

$$P = V \begin{pmatrix} 1 & 0 & 0 \\ 0 & -1/4 & 0 \\ 0 & 0 & -1/6 \end{pmatrix} V^{-1},\ P^n = V \begin{pmatrix} 1 & 0 & 0 \\ 0 & (-1/4)^n & 0 \\ 0 & 0 & (-1/6)^n \end{pmatrix} V^{-1},$$

It follows that for any i, j,

$$(P^n)_{ij} = A_{ij} + B_{ij}(-1/4)^n + C_{ij}(-1/6)^n.$$

Suppose $i = 2$, $j = 3$. *Drop the* ij *subscripts on* A, B, C. *We get*

$$A + B + C = (P^0)_{23} = 0,\ A - B/4 - C/6 = p_{23} = 1/3$$

and

$$A + B/16 + C/36 = (P^2)_{23} = p_{21}p_{13} + p_{22}p_{23} + p_{23}p_{33} = 13/36.$$

Solving gives $A = 12/35$,$B = 4/5$,$C = -8/7$. *Hence if* X *is a* (δ_2, P)*-MC then*

$$\mathbb{P}(X_n = 3) = 12/35 + (4/5)(-1/4)^n - (8/7)(-1/6)^n.$$

1.4 Existence and uniqueness in distribution †

(This is not examinable).

Given (λ, P), existence of a corresponding Markov chain is not trivial. We must appeal to a result from measure theory.

Let $\Omega = \mathbb{R}^{\mathbb{N}} = \{(\omega_1, \omega_2, \ldots) : \omega_i \in \mathbb{R}\}$ and $\mathcal{F}$ the σ-algebra generated by the finite dimensional rectangles

$$(a_1, b_1] \times (a_2, b_2] \times \ldots \times (a_n, b_n] \times \mathbb{R} \times \mathbb{R} \ldots,$$

where $-\infty \le a_i < b_i \le \infty$ and $n \in \mathbb{N}$.

Theorem 1.9 (Kolmogorov Extension Theorem)**.** *Suppose that μ_n is a probability measure on $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$ for every n and the measures are consistent in that*

$$\mu_n((a_1, b_1] \times \ldots \times (a_n, b_n]) = \mu_{n+1}((a_1, b_1] \times \ldots \times (a_n, b_n] \times \mathbb{R})$$

for all choices a, b and n.

Then there exists a unique probability measure on $(\Omega, \mathcal{F})$ with

$$\mu((a_1, b_1] \times \ldots \times (a_n, b_n]) = \mu_n((a_1, b_1] \times \ldots \times (a_n, b_n])$$

for all choices a, b and n.

Proof omitted. Consult e.g. Appendix 3 of Durrett's Probability Theory and Examples book if interested.

Corollary 1.10. *Let P be a transition matrix and λ a distribution on $I \subset \mathbb{R}$. A (λ, P)-MC exists. The induced measure on $(\Omega, \mathcal{F})$ is unique.*

Proof. Assume without loss of generality $I \subset \mathbb{R}$, say $I = \mathbb{N}$.

Define atomic measures μ_n according to Theorem 1.6 –

$$\mu_{n+1}(\{i_0\} \times \{i_1\} \times \ldots \{i_n\}) = \lambda_{i_0} p_{i_0,i_1} \ldots p_{i_{n-1},i_n}$$

for any $i_0, \ldots, i_n \in I$.

The family is consistent since for any $i_0, \ldots, i_n$,

$$\mu_{n+2}(\{i_0\} \times \{i_1\} \times \ldots \{i_n\} \times \mathbb{R}) = \sum_{j \in I} \mu_{n+2}(\{i_0\} \times \{i_1\} \times \ldots \{i_n\} \times \{j\})$$

which equals $\mu_{n+1}(\{i_0\} \times \{i_1\} \times \ldots \{i_n\})$ using the fact $\sum_{j \in I} p_{ij} = 1$ for any $i \in I$.

The Kolmogorov extension theorem gives a unique measure μ on $(\mathbb{R}^{\mathbb{N}}, \mathcal{F})$ consistent with μ_n.

The canonical variables $X_n(\omega) = \omega_{n+1}$, $n \in \mathbb{N}_0$ are measurable and

$$\mu(\{\omega \in \Omega : X_0(\omega) = i_0, \ldots, X_n(\omega) = i_n\}) = \mu_{n+1}(\{i_0\} \times \ldots \times \{i_n\}) = \lambda_{i_0} p_{i_0,i_1} \ldots p_{i_{n-1},i_n}.$$

for any $i_0, i_1, \ldots, i_n$.

So $(X_n)_{n\in\mathbb{N}_0}$ is a (λ, P)-MC by Theorem 1.6.

Now for the second claim in the statement of the corollary– in the lecture I didn't define what the 'induced measure' of the process actually is and possibly it sounded somewhat mysterious. Let's clear that up.

Let $(Y_n)_{n\in\mathbb{N}}$ be a stochastic process. Assume Y is defined on an arbitrary probability space $(\tilde{\Omega}, \tilde{\mathcal{F}}, \tilde{\mathbb{P}})$.

Viewed as a sequence of numbers in $\mathbb{R}$, Y gives a map from $\tilde{\Omega}$ to $\mathbb{R}^{\mathbb{N}}$.

It is actually a random variable in the measurable space $(\mathbb{R}^{\mathbb{N}}, \mathcal{F})$ with $\mathcal{F}$ as defined above.

To see this, we first check measurability for finite dimensional rectangles. For any $\tilde{\omega} \in \tilde{\Omega}$, and $-\infty \leq a_i < b_i \leq \infty$, $1 \leq i \leq n$,

$$Y(\tilde{\omega}) \in A = (a_1, b_1] \times (a_2, b_2] \times \ldots \times (a_n, b_n] \times \mathbb{R} \times \ldots$$

iff $Y_i(\tilde{\omega}) \in (a_i, b_i]$ for $1 \leq i \leq n$. That is

$$\{Y \in A\} = \{Y_1 \in (a_1, b_1]\} \cap \ldots \cap \{Y_n \in (a_n, b_n]\} \in \tilde{\mathcal{F}}.$$

Now $\{B \in \mathcal{F} : Y^{-1}(B) \in \tilde{\mathcal{F}}\}$ is a σ-algebra. We just checked it contains finite dimensional rectangles. The latter, however, generate $\mathcal{F}$. It follows that $Y^{-1}(B) \in \tilde{\mathcal{F}}$ for every $B \in \mathcal{F}$.

The **induced measure** $\mu_Y : \mathcal{F} \to [0, 1]$ on $(\Omega, \mathcal{F})$ is defined by $\mu_Y(B) = \tilde{\mathbb{P}}(Y \in B)$ for any $B \in \mathcal{F}$.

Suppose now Y is a (λ, P)-MC on state space $I = \mathbb{N}$. Then, by definition of μ_Y,

$$\mu_Y(\{i_0\} \times \{i_1\} \times \ldots \times \{i_n\}) = \tilde{\mathbb{P}}(X_0 = i_0, X_1 = i_1, \ldots, X_n = i_n) = \lambda_{i_0} p_{i_0,i_1} \ldots p_{i_{n-1},i_n}.$$

for any $i_0, \ldots \in I$, $n \in \mathbb{N}$.

But the measure μ above from Kolmogorov extension theorem is the only measure on $(\Omega, \mathcal{F})$ with this property.

The point is- (λ, P)-MCs have finite dimensional distributions determined by λ and P. Finite dimensional distributions determine the law of the whole process.

□

1.4.1 Construction from uniform random variables †

A more direct construction uses uniform random variables. In practice this might also be a way to simulate a Markov process on the computer.

Assume $I = \mathbb{N}$. Define a function $G : I \times [0, 1) \to I$ as follows. Fix $(i, u) \in I \times [0, 1)$.

Since

$$\sum_{k=1}^{\infty} p_{ik} = 1,$$

there exists $J \in \mathbb{N}$ with $\sum_{k=1}^{J} p_{ik} > u$. Let j be the smallest such j. Thus

$$\sum_{k=1}^{j-1} p_{ik} \leq u < \sum_{k=1}^{j} p_{ik}.$$

Define $G(i,u) = j$. Then,

- for any RVs X on I and U on $[0,1)$,

$$\{G(X,U) = j\} = \bigcup_i \{X = i\} \cap \{\sum_{k=1}^{j-1} p_{ik} \leq U < \sum_{k=1}^{j} p_{ik}\}.$$

  So $G(X,U)$ is a bonda-fide RV.

- If U is uniform on $[0,1)$ then

$$\mathbb{P}(G(i,U) = j) = \mathbb{P}(\sum_{k=1}^{j-1} p_{ik} \leq U < \sum_{k=1}^{j} p_{ik}) = p_{ij}.$$

Hence, if

- $U_1, U_2, \ldots$ are IID uniform RVs
- X_0 has distribution $\mathbb{P}(X_0 = i) = \lambda_i$
- X_n is defined to be $G(X_{n-1}, U_n)$ for $n \in \mathbb{N}$

then, using independence of $X_0, U_1, U_2, \ldots$,

$$\mathbb{P}(X_n = i_n, \ldots, X_0 = i_0) = \mathbb{P}(G(i_{n-1}, U_n) = i_n, \ldots, G(i_0, U_1) = i_1, X_0 = i_0) = \lambda_{i_0} p_{i_0,i_1} \ldots p_{i_{n-1},i_n}$$

for any $i_0, i_1, \ldots \in I$ and $n \in \mathbb{N}$. Applying Theorem 1.6 confirms $(X_n)_{n\in\mathbb{N}_0}$ is a (λ, P)-MC.

1.5 Back to SRW: the reflection principle and ballot theorem.

Back to the running example, a SRW $(S_n)_{n\in\mathbb{N}_0}$ on $\mathbb{Z}$.

Notation: for $a, b \in \mathbb{Z}$, $n \in \mathbb{N}$, let $\mathcal{N}_n(a,b)$ denote the set of nearest neighbour paths that go from a to b in n steps, i.e.

$$\mathcal{N}_n(a,b) = \left\{(x_0, \ldots, x_n) \in \mathbb{Z}^{n+1} : x_0 = a, x_n = b, x_m - x_{m-1} \in \{-1, +1\}, 1 \leq m \leq n\right\}.$$

Note that $\mathcal{N}_n(a,b)$ is empty if $n+(b-a)$ is odd. Otherwise, each path is an allocation of $u = (n+(b-a))/2$ 'up' steps to the n possible steps, i.e.

$$\#\mathcal{N}_n(a,b) = \binom{n}{u}.$$

By convention $\binom{n}{\alpha} = 0$ when α is non-integer, $\alpha < 0$ or $\alpha > n$.

Further, any $x \in \mathcal{N}_n(a,b)$ has

$$\mathbb{P}(S_n = b, S_{n-1} = x_{n-1}, \ldots, S_1 = x_1 | S_0 = a) = p^u(1-p)^{n-u}.$$

Thus calculating probabilities means counting nearest neighbour paths with certain properties.

In both applications and theory, events that S should be reach a particular value, $\{S_m \geq 10$ for some $m \leq n\}$ say, are of interest.

A useful trick to count paths in this situation is to reflect the path after it first hits that level.

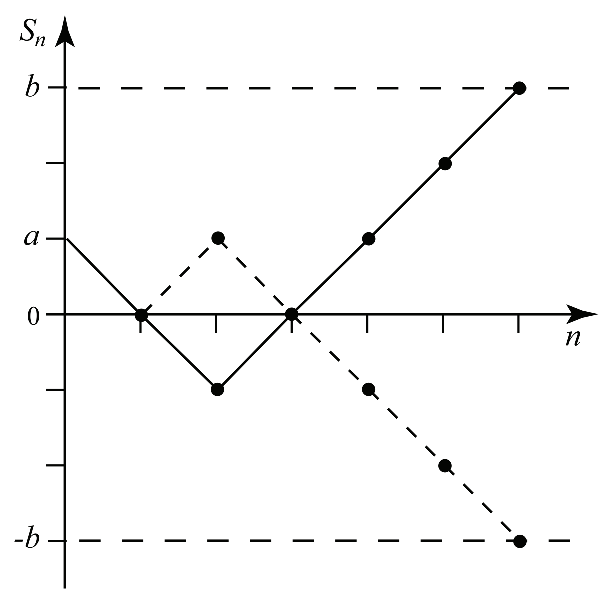
This technique makes use of something called the reflection principle.

To state it, more notation is needed: let $\mathcal{N}_n^0(a,b) \subset \mathcal{N}_n(a,b)$ be the set of nearest neighbour paths that go from a to b in n steps *and hit 0*, i.e.

$$\mathcal{N}_n^0(a,b) = \{x \in \mathcal{N}_n(a,b) : x_m = 0 \text{ for some } 1 \leq m \leq n\}.$$

Theorem 1.11 (Reflection principle)**.** *Suppose $a, b \in \mathbb{N}$ are positive. Then $\#\mathcal{N}_n^0(a,b) = \#\mathcal{N}_n(a,-b)$.*

Proof. Intuitively this is clear from the diagram



More formally we define a bijection $\mathcal{N}_n^0(a,b) \leftrightarrow \mathcal{N}_n(a,-b)$ as follows.

Suppose $x \in \mathcal{N}_n^0(a,b)$.

By definition there exists $0 < k < n$ with $x_k = 0$. Let k' be the smallest such k and $\tilde{x}$ be the path with everything after step k' reflected in the origin.

I.e. $\tilde{x}_m = x_m$ for $0 \leq m \leq k'$, and $\tilde{x}_m = -x_m$ for $k' + 1 \leq m \leq n$.

Then $\tilde{x}_0 = x_0 = a$ and $\tilde{x}_n = -x_n = -b$. Further, for $1 \leq m \leq k'$,

$$\tilde{x}_m - \tilde{x}_{m-1} = x_m - x_{m-1} \in \{-1, +1\}$$

and for $k' + 1 \leq m \leq n$,

$$\tilde{x}_m - \tilde{x}_{m-1} = -x_m + x_{m-1} \in \{-1, +1\}.$$

So $\tilde{x} \in \mathcal{N}_n(a, -b)$. Further, the reflection procedure is injective.

Going the other way, any $\tilde{x} \in \mathcal{N}_n(a, -b)$ must hit 0. Repeating the above reflection procedure gives a path in $\mathcal{N}_n^0(a, b)$.

□

Here is an example of the reflection principle in use.

Theorem 1.12 (The Ballot Theorem)**.** *Let $b > 0$. The number of nearest neighbour paths $x \in \mathcal{N}_n(0, b)$ that are positive after the first step (i.e. $x_m > 0$ for $1 \leq m \leq n$) and end at b is*

$$\frac{b}{n}\binom{n}{(n+b)/2}.$$

Consequently, if A and B are candidates in an election, A gets $\alpha > 0$ votes and B gets $0 \leq \beta < \alpha$ votes then the probability that A led B from the start of counting is $(\alpha - \beta)/(\alpha + \beta)$.

Proof. Every path $x \in \mathcal{N}_n(0, b)$ with $x_m > 0$ for $m \geq 1$ satisfies $x_1 = 1$.

Thus, the number N of such paths is

$$N = \#\mathcal{N}_{n-1}(1, b) - \#N^0_{n-1}(1, b) = \#\mathcal{N}_{n-1}(1, b) - \#N^0_{n-1}(1, -b)$$

by the reflection principle.

We have

$$\#\mathcal{N}_{n-1}(1, b) = \binom{n-1}{(n-1+b-1)/2}, \; \#\mathcal{N}_{n-1}(1, -b) = \binom{n-1}{(n-1+b+1)/2}.$$

So, letting $u = (n + b)/2$,

$$N = \binom{n-1}{u-1} - \binom{n-1}{u} = \frac{1}{n}\binom{n}{u}(2u - n) = \frac{b}{n}\binom{n}{(n+b)/2}.$$

For the ballot part: Initially both candidates have zero votes. After each vote is counted, A's lead either increases by $+1$ or decreases by 1. Thus A's lead (=the number of A votes minus number of B votes) after m votes have been counted is a nearest neighbour walk.

The difference in votes ends at $b = \alpha - \beta$ after $n = \alpha + \beta$ steps.

Candidate A being ahead throughout corresponds to the walk not hitting zero after the first step.

Each count is equally likely. Using the first part, the desired probability is

$$\left(\frac{b}{n}\binom{n}{(n+b)/2}\right) / \#\mathcal{N}_n(0, b) = \frac{\alpha - \beta}{\alpha + \beta}$$

as claimed.

□

1.6 Hitting probabilities and times

Suppose X is a Markov chain on I with transition matrix $P = (p_{ij})_{i,j \in I}$.

Notation: $\mathbb{P}_i$ denotes a probability measure under which $X_0 = i \in I$.

For $A \subset I$, let $H^A = \inf\{n \geq 0 : X_n \in A\}$ be the first time that X enters state in A. H^A is called the hitting time of A.

It is agreed $H^A = +\infty$ if X never hits A. So H^A is a RV in the extended space $\mathbb{N}_0 \cup \{+\infty\}$.

Consider $h_i^A = \mathbb{P}_i(H^A < \infty)$, the probability that X hits A when started at $i \in I$.

Let $i \in A$. Under $\mathbb{P}_i$, $X_0 = i \in A$ and $H^A = 0$. It follows that $h_i^A = 1$.

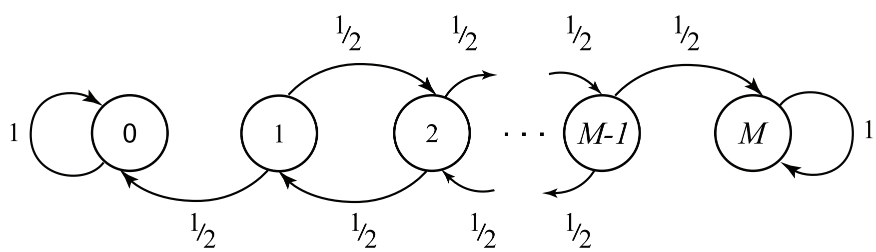
Now suppose $X_0 = i \notin A$. Consider what happens after one step. Intuitively, X moves to a new state, $X_1 = j$ say, with probability p_{ij}. After that, X continues evolving as though it had started in j.

The vector h_i *should* solve

$$h_i^A = \begin{cases} 1 & i \in A, \\ \sum_j p_{ij} h_i^A & i \notin A. \end{cases} \tag{2}$$

Before proving this is the case here is an example.

Example 1.13. *Suppose $I = \{0, 1, \ldots, M\}$ ($M \in \mathbb{N}$) and $P = (p_{ij})$ is given by $p_{00} = 1$, $p_{MM} = 1$ and $p_{i,i+1} = p_{i,i-1} = 1/2$ for $0 < i < M$ (all other entries are zero).*



X behaves as a SRW away from 0 and a. When X hits 0 or M it gets stuck there.

Consider $h_i = \mathbb{P}_i(H_{\{M\}} < \infty)$, the probability that X hits and gets stuck at M.

Firstly- $h_0 = 0$ since X never moves from zero when it starts there, and $h_M = 1$.

For $0 < i < M$,

$$h_i = p_{i,i+1}h_{i+1} = p_{i,i-1}h_{i-1} = (1/2)h_{i+1} + (1/2)h_i - (1/2)h_i + (1/2)h_{i-1},$$

It follows

$$h_{i+1} - h_i = h_i - h_{i-1} = \ldots = h_1 - h_0 = h_1,$$

and

$$h_{i+1} = h_i + (h_i - h_{i-1}) = h_i + h_1 = (i+1)h_1.$$

So $1 = h_M = Mh_1$ gives $h_i = i/M$.

In the example above X is closely related to SRW.

More specifically, if S is a SRW started from i under $\mathbb{P}_i$ then X can be realised as $X_n = S_{n \wedge H_0 \wedge H_M}$. Here, $u \wedge v = \min\{u, v\}$.

Hence $\{H_M < \infty\}$ is the event that S hits a before hitting 0.

The hitting probability formula (2) from yesterday requires that X starts afresh after its first step.

The Markov property of definition 1.5(i) only concerns one step into the future. An extra few lines are required to justify that the entire future is conditionally independent of the past given the present.

Theorem 1.14 (Markov property)**.** *Let X be a (λ, P)-MC on I. Suppose $i \in I$ and $m \in \mathbb{N}$ are such that $\mathbb{P}(X_m = i) > 0$.*

Conditional on $X_m = i$, $(X_{m+n})_{n\in\mathbb{N}_0}$ is a (δ_i, P)-MC independent of $X_0, X_1, \ldots, X_{m-1}$.

(Recall that $\delta_i = (\delta_{ij})_{j\in I}$ where $\delta_{ij} = 1$ if $i = j$ and 0 otherwise.)

Proof. This a straightforward application of Theorem 1.6 from last Tuesday.

It is sufficient to check conditional independence

$$\mathbb{P}(A \cap B|X_m = i) = \mathbb{P}(A|X_m = i)\mathbb{P}(B|X_m = i) \tag{3}$$

for events of the form $A = \{X_0 = i_0, \ldots, X_{m-1} = i_{m-1}\}$ and $B = \{X_{m+1} = j_1, X_{m+1} = j_1, \ldots, X_{m+n} = j_n\}$, where $i_0, i_1, \ldots, i_{m-1} \in I$ and $j_1, \ldots, j_n \in I$.

We have

$$\mathbb{P}(A \cap B \cap \{X_m = i\}) = \mathbb{P}(X_0 = i_0, \ldots, X_m = i, X_{m+1} = j_1, \ldots, X_{m+n} = j_n)$$

which is

$$\lambda_{i_0} p_{i_0,i_1} \ldots p_{i_{m-1},i} p_{i,j_1} \ldots p_{j_{n-1},j_n} = \mathbb{P}(A \cap \{X_m = i\}) p_{j_0,j_1} \ldots p_{j_{n-1},j_n}.$$

Dividing by $\mathbb{P}(\{X_m = i\})$ gives

$$\mathbb{P}(A \cap B|X_m = i) = \mathbb{P}(A|X_m = i) p_{j_0,j_1} \ldots p_{j_{n-1},j_n}.$$

The product $p_{i,j_1} \ldots p_{j_{n-1},j_n}$ does not depend on A. So (3) holds with

$$\mathbb{P}(B|X_m = i) = p_{i,j_1} \ldots p_{j_{n-1},j_n}.$$

Use Theorem 1.6 to conclude $(X_{m+n})_{n\in\mathbb{N}_0}$ is a (δ_i, P)-MC. □

Theorem 1.15. *Let X be a Markov chain on I with transition matrix P. Suppose $A \subset I$ and $H^A = \inf\{n \in \mathbb{N}_0 : X_n \in A\}$. The vector $h^A = (\mathbb{P}_i(H^A < \infty))_{i\in I}$ is the minimal non-negative solution to*

$$h_i^A = \begin{cases} 1 & i \in A, \\ \sum_j p_{ij} h_i^A & i \notin A. \end{cases} \tag{4}$$

Proof. For notational brevity drop the A superscript in h_i^A.

For $i \in A$, $h_i = 1$ is immediate.

Suppose $i \notin A$. Break the hitting event down according to the first step,

$$\mathbb{P}_i(H^A < \infty) = \sum_{j\in I} \mathbb{P}_i(H^A < \infty, X_1 = j).$$

For $j \in A$,

$$\mathbb{P}_i(H^A < \infty, X_1 = j) = \mathbb{P}_i(X_1 = j) = p_{ij}.$$

Consider $j \notin A$. Only those j with $p_{ij} > 0$ contribute to the sum. In that case,

$$\mathbb{P}_i(H^A < \infty, X_1 = j) = \mathbb{P}_i(H^A < \infty | X_1 = j) p_{ij}.$$

But, $\{H^A < \infty, X_1 = j, X_0 = i\} = \{X_{k+2} \in A \text{ for some } k \geq 0\} \cap \{X_1 = j, X_0 = i\}$.

So theorem 1.14 gives

$$\mathbb{P}_i(H^A < \infty | X_1 = j) = \mathbb{P}_j(H^A < \infty) =: h_j.$$

In either case, $\mathbb{P}_i(H^A < \infty, X_1 = j) = p_{ij} h_j$, establishing (4).

Let $g_i \geq 0$, $i \in I$, be another solution to (4).

For $i \in A$, $g_i = 1 = h_i$.

Suppose $i \notin A$. Then splitting the sum in (4) gives

$$g_i = \sum_{j \in A} p_{ij} g_j + \sum_{j \notin A} p_{ij} g_j = \sum_{j_1 \in A} p_{i,j_1} + \sum_{j_1 \notin A} p_{i,j_1} g_{j_1}.$$

Substituting in $g_{j_1} = \sum_{j_2 \in A} p_{j_1,j_2} + \sum_{j_2 \notin A} p_{j_1 j_2} g_{j_2}$ gives

$$g_i = \sum_{j_1 \in A} p_{i,j_1} + \sum_{j_1 \notin A} \sum_{j_2 \in A} p_{i,j_1} p_{j_1,j_2} + \sum_{j_1, j_2 \notin A} p_{i,j_1} p_{j_1,j_2} g_{j_2}.$$

Repeating a total of $n \in \mathbb{N}$ times gives

$$g_i = \sum_{k=1}^{n} \sum_{j_1, j_2, \ldots, j_{k-1} \notin A, j_k \in A} p_{i,j_1} p_{j_1,j_2} \ldots p_{j_{k-1},j_k} + \sum_{j_1, j_2, \ldots, j_n \notin A} p_{i,j_1} p_{j_1,j_2} \ldots p_{j_{n-1},j_n} g_{j_n}.$$

The first sum on the RHS is $\sum_{k=1}^{n} \mathbb{P}_i(H^A = k)$. Thus,

$$g_i \geq \mathbb{P}_i(H^A \leq n) \nearrow \mathbb{P}_i(H^A < \infty) = h_i \text{ as } n \to \infty$$

by the monotone convergence theorem. □

Here is an example where (4) has many non-negative solutions.

Example 1.16 (A birth-death chain). *Suppose $I = \mathbb{N}_0$ and $P = (p_{ij})$ is given by $p_{01} = 1$ and for $i \geq 1$, $p_{i,i+1} = p_i$, $p_{i,i-1} = q_i$ where $p_i + q_i = 1$ and $p_i = (1 + i^{-1})^2 q_i$ (i.e. up jumps are slightly favoured).*

The hitting probabilities $h_i = \mathbb{P}_i(H^0 < \infty)$ of zero satisfy

$$h_0 = 1, \quad h_i = p_i h_{i+1} + q_i h_{i-1}, i \geq 1.$$

Using $p_i + q_i = 1$ *leads and rewriting gives*

$$p_i(h_{i+1} - h_i) = q_i(h_i - h_{i-1}).$$

Hence the differences $u_i = h_i - h_{i-1}$, $i \geq 1$, *satisfy*

$$u_{i+1} = \left(\frac{q_i}{p_i}\right) u_i = \left(\frac{q_i q_{i-1}}{p_i p_{i-1}}\right) u_{i-1} = \ldots = \left(\frac{q_1 q_2 \ldots q_i}{p_1 p_2 \ldots p_i}\right) u_1.$$

Taking $\gamma_k = \frac{q_1 q_2 \ldots q_i}{p_1 p_2 \ldots p_i}$, $k \geq 1$,

$$h_{i+1} - h_0 = (h_{i+1} - h_i) + \ldots + (h_1 - h_0) = u_1 + u_2 + \ldots + u_{i+1} = u_1(1 + \gamma_1 + \ldots + \gamma_i).$$

So,

$$h_{i+1} = 1 + u_1 \sum_{k=0}^{i} \gamma_k$$

(with $\gamma_0 = 1$*) and it remains to find the value of* u_1 *which gives the minimal non-negative solution.*
Notice $\gamma_k = (1+k)^{-2}$ *and* $\sum_{k=0}^{\infty} \gamma_k = \pi^2/6$.
If $u_1 < -6/\pi^2$, *then* $0 < -u_1^{-1} < \pi^2/6$. *Choose* $0 < \epsilon < \pi^2/6 + u_1^{-1}$. *There exists* $n \in \mathbb{N}$ *with*

$$\sum_{k=0}^{n} \gamma_k > \pi^2/6 - \epsilon > -u_1^{-1}.$$

But that means

$$h_{n+1} = 1 + u_1 \sum_{k=0}^{n} \gamma_k < 1 - 1 = 0.$$

Thus $u_1 \geq -6/\pi^2$ *and the minimal non-negative solution is*

$$h_{n+1} = 1 - (6/\pi^2) \sum_{k=0}^{n} \gamma_k.$$

In particular, the chain has probability $1 - h_1 = 6/\pi^2 = 0.607..$ *of never hitting zero if it starts at 1.*

Thursday 19/01/12

Initial course feedback

Some people felt the aims of the course weren't clear. A handout will follow to clarify this. The rough list is,

1. Basics of Markov chains in discrete and continuous time (transition probabilities, first step analysis, hitting times, strong Markov property).
2. Asymptotic properties (does the chain visit a state infinitely often? does it converge in distribution?).
3. Specific examples and applications (SRW, birth-death chains, card shuffling, PageRank, Ehrenfest's urn, MCMC, ...).
4. Introductory percolation theory (iid bond percolation, critical probability p_c, correlation inequalities, RSW method, $p_c = 1/2$ in $\mathbb{Z}^2$, subcritical exponential decay)
5. General techniques used in applied probability.
6. Proofs, assuming a couple of results from measure theory.

Pace– mostly considered 'just right'. Will include more optional exercises for those who feel it is too slow.

Handwriting and organisation– will make more effort.

For those who want to register for the course but didn't take MA359 Measure Theory– if you did ST213 or ST318, or can demonstrate sufficient understanding of the material then it is fine.

Reminder: first exercise sheet due in Monday. In qn. 2, $(S_n)_{n\in\mathbb{N}_0}$ is a symmetric simple random walk started from 0. Qn. 2(b) is false if S is not symmetric– consider $\mathbb{P}(S_1 > 0, S_2 > 0) + \mathbb{P}(S_1 < 0, S_2 < 0) = p^2 + q^2 \neq 2pq = \mathbb{P}(S_2 = 0)$ for $p = 1 - q \neq 1/2$.

Hint for qn. 3: notice that the probability in question does not change if you modify S so that 0 and N absorbing.

[Admin ends]

The technique from Monday and Tuesday (looking at what happens after X takes one step) is called a first step analysis.

It also works for the expected hitting times $k_i^A = \mathbb{E}_i[H^A]$, where $\mathbb{E}_i$ means expectation under measure $\mathbb{P}_i$.

k_i^A can be $+\infty$. In particular this is true if $\mathbb{P}(H^A = \infty) > 0$. (n.b. $\mathbb{P}(H^A = \infty) = 0 \not\Rightarrow k_i^A < \infty$, see exercise sheet 1).

If $i \in A$ then $k_i^A = 0$. Otherwise, k_i^A is comprised of the time for the first step and the expected time thereafter–

Theorem 1.17. *Let X be a Markov chain on I with transition matrix P. Suppose $A \subset I$ and $H^A = \inf\{n \in \mathbb{N}_0 : X_n \in A\}$. The expected hitting times $k_i^A = \mathbb{E}_i(H^A < \infty)$, $i \in I$ form the minimal non-negative solution to*

$$k_i^A = \begin{cases} 0 & i \in A, \\ 1 + \sum_{j \notin A} p_{ij} k_i^A & i \notin A. \end{cases} \tag{5}$$

Proof. For $X_0 \in A$ we have $H^A = 0$ hence $k_i^A = 0$ for $i \in A$.

Suppose $i \notin A$ and $j \in I$. Claim:

$$\mathbb{E}_i[H^A \, \mathbb{1}_{X_1=j}] = p_{ij}(1 + \mathbb{E}_j[H^A]). \tag{6}$$

(Recall, the indicator of event $B \in \mathcal{F}$ is $\mathbb{1}_B(\omega) = 1$ if $\omega \in B$ and zero otherwise).

If $p_{ij} = 0$ both sides vanish so suppose $p_{ij} > 0$.

Since $X_0 \notin A$, $H^A \geq 1$. Further,

$$H^A = \inf\{n \geq 0 : X_n \in A\} = \inf\{n \geq 1 : X_n \in A\} = 1 + \inf\{n \geq 0 : X_{n+1} \in A\} = 1 + \tilde{H}^A,$$

where $\tilde{H}^A = \inf\{n \geq 0 : X_{n+1} \in A\}$ only depends on $(X_{n+1})_{n \geq 0}$.

Hence, using $\mathbb{E}[\xi] = \sum_{k \geq 1} \mathbb{P}(\xi \geq n)$ for a non-neg RV ξ on $\mathbb{N}_0 \cup \{\infty\}$ (on forthcoming exercise sheet 2),

$$\begin{aligned} \mathbb{E}_i[\tilde{H}^A \, \mathbb{1}_{X_1=j}] &= \sum_{n \geq 1} \mathbb{P}_i(\tilde{H}^A \, \mathbb{1}_{X_1=j} \geq n) \\ &= \sum_{n \geq 1} \mathbb{P}_i(\tilde{H}^A \geq n, X_1 = j) \\ &= \sum_{n \geq 1} \mathbb{P}_i(\tilde{H}^A \geq n | X_1 = j) p_{ij} \\ &= \sum_{n \geq 1} \mathbb{P}_j(H^A \geq n) p_{ij} \\ &= \mathbb{E}_j[H^A] p_{ij} \end{aligned}$$

by the Markov property Theorem 1.14.

THe claim then follows from

$$\mathbb{E}_i[H^A \, \mathbb{1}_{X_1=j}] = \mathbb{E}_i[(1 + \tilde{H}^A) \, \mathbb{1}_{X_1=j}].$$

From here,

$$\begin{aligned} k_i^A &= \sum_{j \in I} \mathbb{E}_i[H^A \, \mathbb{1}_{X_1=j}] \\ &= \sum_{j \in A} (1 + \mathbb{E}_j[H^A]) p_{ij} \\ &= 1 + \sum_{j \notin A} p_{ij} k_j^A \end{aligned}$$

using $k_j^A = 0$ for $j \in A$ and $\sum_j p_{ij} = 1$.

Now for the minimality part. Suppose $y_i \geq 0$, $i \in I$ is another solution. Let $i \notin A$. Using (5) twice,

$$y_i = 1 + \sum_{j \notin A} p_{ij} y_j = 1 + \sum_{j \notin A} p_{ij}(1 + \sum_{k \notin A} p_{jk} y_k),$$

which is

$$= \mathbb{P}_i(H^A \geq 1) + \mathbb{P}_i(H^A \geq 2) + \sum_{j \notin A} \sum_{k \notin A} p_{ij} p_{jk} y_k.$$

Repeatedly using (5) gives

$$y_i = \sum_{k=1}^{n} \mathbb{P}(H^A \geq k) + \sum_{j_1 \notin A} \cdots \sum_{j_n \notin A} p_{i,j_1} \dots p_{j_{n-1},j_n} y_{j_n}.$$

All the summands are non-negative by assumption $y_{j_k} \geq 0$.

So

$$y_i \geq \sum_{k=1}^{n} \mathbb{P}_i(H^A \geq k) \nearrow \mathbb{E}_i[H^A] = k_i^A.$$

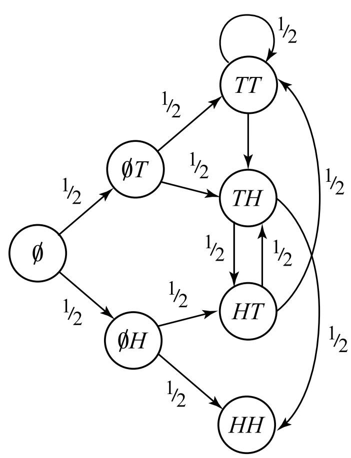
□

Example 1.18. *A coin is tossed repeatedly until two consecutive heads are seen. Find the expected number of flips required.*

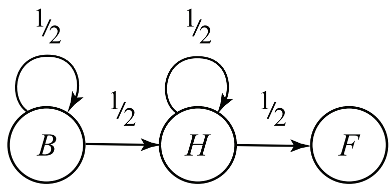
To begin we identify an appropriate Markov chain. There are several possibilities.

A first attempt is to model the last two coin flips as a Markov chain.

The chain has state space $\{\emptyset, \emptyset T, \emptyset H, TT, TH, HT, HH\}$.



This can be simplified. In particular, states $\emptyset, \emptyset T, TT, HT$ *collapse into a single state which we denote* B *(corresponding to the last flip, if any, not being a head).* $TH, \emptyset H$ *collapse to a state* H *(for the last flip being head) and* $F = HH$ *for having finished.*



We have

$$k_B = 1 + (1/2)k_B + (1/2)k_H, \;\; k_H = 1 + (1/2)k_B + (1/2)k_F$$

which give

$$(1/2)k_B = 1 + (1/2) + (1/4)k_B.$$

Hence the expected number of tosses required is $k_B = 6$.

1.7 Stopping times, strong Markov property

In the previous section we broke the hitting event down according to the first step taken by the chain.

For this we required the Markov property Theorem 1.14 that, after conditioning on the first step, the MC evolves independently.

The *strong Markov property* says this holds for certain *random* times too. A sufficient condition on the random time is that it only depends on the past observations of the process.

Such times are called optional or stopping times.

Definition 1.19 (Stopping time)**.** *A RV $T : \Omega \to N_0 \cup \{+\infty\}$ is called a stopping for a stochastic process $(X_n)_{n\in\mathbb{N}_0}$ if $\{T = n\}$ ($\in \mathcal{F}$) depends only on $X_0, X_1, \ldots, X_n$ for any $n \in \mathbb{N}_0$.*

That means we can check whether $T = n$ by observing X only upto time n.

Example 1.20. 1. *The first hitting time H^A is a stopping time because*

$$\{H^A = n\} = \{X_0 \notin A, X_1 \notin A, \ldots X_n \in A\}.$$

2. *The last hitting time $L^A = \sup\{n \in \mathbb{N}_0 : X_n \in A\}$ is* not *a stopping time because $\{L^A = n\}$ requires $X_{n+1} \notin A$, $X_{n+2} \notin A$, ...*

Theorem 1.21 (Strong Markov property)**.** *Let X be a (λ, P)-MC, T a stopping time for X and $i \in I$ such that $\mathbb{P}(T < \infty, X_T = i)$.*

Then, conditional on $T < \infty$ and $X_T = i$, the process $(X_{T+n})_{n\geq 0}$ is a (δ_i, P)-MC, and is independent of $X_0, X_1, \ldots, X_{T-1}$.

The proof is similar to Thereom 1.14 with conditioning $\{T = m\}$.

Proof. Let $m \in \mathbb{N}$. The event $\{T = m\}$ depends only on $X_0, \ldots, X_m$. In particular

$$\mathbb{1}_{T=m}(\omega) = \tilde{T}_m(X_0(\omega), \ldots, X_m(\omega))$$

for some measurable function $\tilde{T}_m$.

Suppose $i_0, i_1, \ldots \in I$ are such that $\tilde{T}_m(i_0, \ldots, i_m) = 1$.

Then

$$\{X_0 = i_0, \ldots, X_{T-1} = i_{T-1}, X_T = i_T\} \subseteq \{T = m\}.$$

Hence for any $j_1, \ldots \in I$, $n \in \mathbb{N}$,

$$\mathbb{P}(T = m, X_0 = i_0, \ldots, X_{T-1} = i_{T-1}, X_T = i, X_{T+1} = j_1, \ldots, X_{T+n} = j_n)$$

$$= \mathbb{P}(X_0 = i_0, \ldots, X_{m-1} = i_{m-1}, X_m = i, X_{m+1} = j_1, \ldots, X_{m+n} = j_n)$$

$$= \mathbb{P}(X_0 = i_0, \ldots, X_{m-1} = i_{m-1}, X_m = i, T = m) p_{i,j_1} p_{j_1,j_2} \ldots p_{j_{n-1},j_n}.$$

If $\tilde{T}_m(i_0, \ldots, i_m) = 0$ then both sides vanish.

Summing over m and dividing by $\mathbb{P}(T < \infty, X_T = i) > 0$ yields

$$\mathbb{P}(X_0 = i_0, \ldots, X_{T-1} = i_{T-1}, X_{T+1} = j_1, \ldots, X_{T+n} = j_{T+n} | T < \infty, X_T = i)$$

$$= \left(\sum_m \mathbb{P}(X_0 = i_0, \ldots, X_{m-1} = i_{m-1}, X_m = i, T = m) / \mathbb{P}(T < \infty, X_T = i) \right) p_{i,j_1} p_{j_1,j_2} \ldots p_{j_{n-1},j_n}.$$

$$= \mathbb{P}(X_0 = i_0, \ldots, X_{m-1} = i_{T-1} | X_T = i, T < \infty) p_{i,j_1} p_{j_1,j_2} \ldots p_{j_{n-1},j_n}.$$

□

Example 1.22. *Suppose X is MC on $\mathbb{N}_0$ with transition matrix $p_{i,i+1} = p \in (0,1)$, $p_{i,i-1} = 1 - p = q$ for $i \geq 0$.*

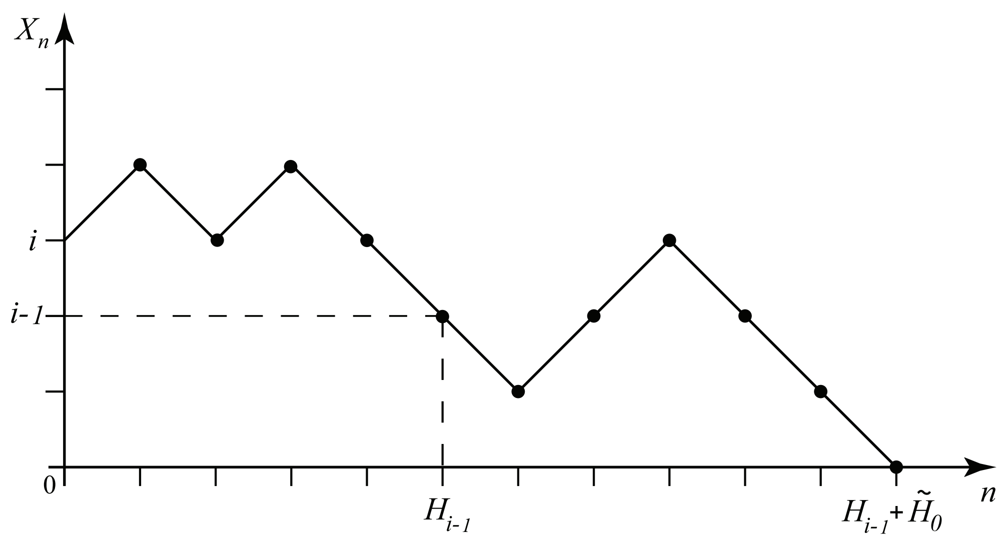
Let $H_i = \inf\{n \in \mathbb{N}_0 : X_n = i\}$.

The distribution of H_0 under $\mathbb{P}_i$ is determined by the generating function $\phi_i(s) = \mathbb{E}_i[s^{H_0}]$.

For $i \geq 1$, break H_0 down into the time to hit $i-1$ and the time $\tilde{H}_0$ thereafter, i.e.

$$H_0 = H_{i-1} + \tilde{H}_0$$

where $\tilde{H}_0 = \inf\{n \geq 0 : X_{n+H_{i-1}} = 0\}$.



Then

- *H_{i-1} is a stopping time with $\mathbb{P}_i(H_{i-1} < \infty) \geq q > 0$.*
- *If $H_{i-1} < \infty$ then $X_{H_{i-1}} = i - 1$.*

- *the distribution of* $\tilde{H}_0$ *depends only on* $(X_{n+H_{i-1}})_{n\geq 0}$.

Hence for $0 < s < 1$,

$$\phi_i(s) = \mathbb{E}_i[s^{H_0}\,\mathbb{1}_{H_0<\infty}] = \mathbb{E}_i[s^{H_{i-1}}s^{\tilde{H}_0}; H_0 < \infty, X_{H_{i-1}} = i-1]$$

($\mathbb{E}_i[\xi; B]$ *is shorthand for* $\mathbb{E}_i[\xi\,\mathbb{1}_B]$*), using the strong Markov property*

$$= \mathbb{E}_i[s^{H_{i-1}}]\,\mathbb{E}_{i-1}[s^{H_0}\,\mathbb{1}_{H_0<\infty}] = \phi_1(s)\phi_{i-1}(s) = \ldots = (\phi_1(s))^i.$$

So it suffices to calculate

$$\phi(s) = \phi_1(s) = \mathbb{E}_1[s^{H_0}\,\mathbb{1}_{X_1=0}] + \mathbb{E}_1[s^{H_0}\,\mathbb{1}_{X_1=2}] = qs + sp\,\mathbb{E}_2[s^{H_0+1}] = qs + ps\phi(s)^2,$$

which is the solution to the quadratic equation

$$-\phi + qs + ps\phi^2 = 0.$$

Using the old GCSE favourite

$$\phi(s) = \frac{1}{2ps}\left(1 - \sqrt{1-4pqs^2}\right).$$

(the solution with a '+' sign blows up as $s \to 0$ whereas $\phi(s) \leq 1$)

Recall that probabilities are recovered from generating functions by differentiation.

By Analysis II the power series $\sum_{n\geq 0} s^n \mathbb{P}_i(H_0 = n)$ can be differentiated term by term (at least for $s < 1$) to get

$$\phi_i^{(k)}(s) = \sum_{n=k} n(n-1)\ldots(n-k+1)s^{n-k}\mathbb{P}(H_0 = n) \searrow \mathbb{P}(H_0 = k)$$

as $s \searrow 0$.

1.8 Recurrence and transcience

Next we consider asymptotic properties of Markov chains. The most basic question is– does a chain visit a given state infinitely often or only finite many times?

Definition 1.23 (Recurrent, transient)**.** *Suppose* X *is a MC on* I*. State* $i \in I$ *is* recurrent *if*

$$\mathbb{P}_i(X \text{ visits } i \text{ infinitely often}) = \mathbb{P}(\forall N \in \mathbb{N}_0\ \exists n > N \text{ s.t. } X_n = i) = 1.$$

State $i \in I$ *is called* transient *if the probability above is 0, which is the same as*

$$\mathbb{P}_i(\exists N \in \mathbb{N}_0 \text{ s.t. } X_n \neq i \text{ for } n > N) = 1.$$

In the definition there is no mention of intermediate probabilities. That is because the probabilities in question are always 0 or 1.

We prove this next lecture by looking at hitting probabilities.

Actually we consider *return* probabilities. This is just a technical difference- H^A defined above is zero if we start in A. Here it is more convenient to consider hits after the first step.

Theorem 1.24. *Suppose X is a MC on I. Let $i \in I$, $T_i = \inf\{n \geq 1 : X_n = i\}$, and $f_i = \mathbb{P}_i(T_i < \infty)$. State i is recurrent iff $f_i = 1$. Otherwise, if $f_i < 1$, then i is transient.*

Proof. Let $T_i^{(1)} = T_i$ and

$$T_i^{(k+1)} = T_i^{(k)} + \inf\{n \geq 1 : X_{n+T_i^{(k)}} = i\}$$

be the $(k+1)$-st return time. Denote the inf on the right by $\tilde{T}_i^{(k+1)}$.

Remark: the $\tilde{T}_i^{(k)}$ are called excursion times.

$T_i^{(k)}$ is a stopping time and $B_i^{(k)} = \{T_i^{(k)} < \infty\}$ is the event that X visits i at least k times.

By the strong Markov property,

$$\mathbb{P}_i(T_i^{(k+1)} < \infty) = \mathbb{P}_i(T_i^{(k)} < \infty, \tilde{T}_i^{(k+1)} < \infty, X_{T_i^{(k)}} = i) = \mathbb{P}_i(T_i^{(k)} < \infty)\mathbb{P}_i(T_i < \infty)$$

Hence, $\mathbb{P}_i(B_i^{(k)}) = \mathbb{P}_i(T_i^{(k)} < \infty) = f_i^k$.

Since $B_i^{(1)} \supset B_i^{(2)} \supset \ldots$ the Monotone convergence theorem applies to give

$$\mathbb{P}(X_n = i \text{ i.o}) = \mathbb{P}\left(\bigcap_n B_i^{(n)}\right) = \lim_{k\to\infty} f_i^k.$$

The limit on the RHS is 1 if $f_i = 1$ and zero for $0 \leq f_i < 1$.

On the other hand, i recurrent implies $f_i = 1$ immediately. □

To check whether a state is recurrent or transient it is often easier to look at the expected number of visits instead of the probabilities above.

Theorem 1.25. *Suppose X a MC on I. Let $V_i = \sum_{n\geq 0} \mathbb{1}_{X_n=i}$ be the total number of visits to $i \in I$. State i is recurrent iff*

$$\mathbb{E}_i[V_i] = +\infty. \tag{7}$$

Remark 1.26. *the condition (7) is equivalent to*

$$\mathbb{E}_i[V_i] = \mathbb{E}_i\left[\sum_{n\geq 0} \mathbb{1}_{X_n=i}\right] = \sum_{n\geq 0} \mathbb{P}_i(X_n = i) = +\infty$$

by Fubini's Theorem and the Theorem is often stated this way.

Proof. Let $B_i^{(k)}$ and f_i be as in Theorem 1.24.

$B_i^{(k)}$ is the event $\{V_i \geq k\}$, so

$$\mathbb{E}_i[V_i] = \sum_{k\geq 1} \mathbb{P}_i(V_i \geq k) = \sum_{k\geq 1} \mathbb{P}_i(B_i^{(k)}) = \sum_{k\geq 1} f_i^k.$$

If i is recurrent then $f_i = 1$ and the sum is infinite.

If i is transient then $f_i < 1$ and the sum is $f_i/(1-f_i) < \infty$.

□

Definition 1.27. *A MC X on I is called recurrent if every state $i \in I$ is recurrent.*

1.8.1 Recurrence of Random walks in $\mathbb{Z}$

Theorem 1.28. *Simple random walk on $\mathbb{Z}$ is recurrent iff it is symmetric.*

Proof. Let S be a SRW on $\mathbb{Z}$ and $p = \mathbb{P}(S_{n+1} - S_n = +1) = 1 - q$.

It is sufficient to check whether 0 is recurrent (SRW is spatially homogeneous). So take $S_0 = 0$. Then

$$\mathbb{P}(S_{2k} = 0) = \binom{2k}{k} p^k q^k = \frac{(2k)!}{k!k!} p^k q^k.$$

Stirling's formula $n!/\sqrt{2\pi} n^{n+1/2} \exp(-n) \to 1$ implies that for big enough $N \in \mathbb{N}$,

$$(1 - 1/2)\sqrt{2\pi} k^{k+1/2} \exp(-k) < k! < (1 + 1/2)\sqrt{2\pi} k^{k+1/2} \exp(-k)$$

for $k > N$. So, for some constants $C, \tilde{C} > 0$,

$$C(4pq)^k/\sqrt{k} < \mathbb{P}(S_{2k} = 0) < \tilde{C}(4pq)^k/\sqrt{k}.$$

for $k > N$.

Hence

$$\sum_{k=0}^{\infty} \mathbb{P}(S_{2k} = 0) = \sum_{k=0}^{N} \mathbb{P}(S_{2k} = 0) + \sum_{k=N+1}^{N} \mathbb{P}(S_{2k} = 0)$$

is infinite iff $\sum_k (4pq)^k/\sqrt{k}$ is.

Now $4pq = 4p(1-p) \leq 1$ with equality iff $p = 1/2$, in which case the summand is $k^{-1/2}$ and the series diverges.

If $p \neq 1/2$ then $4p(1-p) < 1$ and the series converges. □

So far only SRW has been discussed. General RWs have an arbitrary increment distribution.

Definition 1.29 (Random Walk). *A stochastic process $(S_n)_{n\in\mathbb{N}_0}$ on $\mathbb{Z}^d$ ($d \in \mathbb{N}$) is called a random walk (RW) if the increments $\xi_n = S_n - S_{n-1}$, $n \geq 1$ are IID random variables.*

The (common) distribution of the increments ξ_i is called the step distribution.

- The Markov property of a general random walk is checked similarly to SRW.
- The transition probability of a RW with step distribution p is

$$\mathbb{P}(S_{n+1} = x | S_n = y) = \mathbb{P}_y(S_1 = x) = \mathbb{P}(\xi_1 = x - y) = p(x-y),\ x, y \in \mathbb{Z}^d.$$

Here is a sufficient condition for recurrence of a RW in 1D.

Theorem 1.30. *Suppose S is a RW in $\mathbb{Z}$. If the weak law of large numbers*

$$\mathbb{P}(|S_n|/n > a) \to 0 \tag{8}$$

for any $a > 0$ holds then S is recurrent.

From ex. sheet 1, (8) certainly holds if the step distribution has mean zero and finite variance.

Proof. This proof needs only three small steps. Suppose S starts from zero.

1. Suppose $x \in \mathbb{Z}$ and $H^x = \inf\{n \geq 0 : S_n = x\}$. Then for $n > l$, the Markov property and spatial homogeneity give

$$\mathbb{P}(H^x = l, S_n = x) = \mathbb{P}(H^x = l)\mathbb{P}_x(S_{n-l} = x) = \mathbb{P}(H^x = l)\mathbb{P}(S_{n-l} = 0).$$

2. We now show

$$\sum_{n=0}^{\infty} \mathbb{P}(|S_n| < m) \leq 2m \sum_{n=0}^{\infty} \mathbb{P}(S_n = 0)$$

for any $m \in \mathbb{N}$:

Breaking the LHS down according to the first hitting of level x,

$$\sum_{n=0}^{\infty} \mathbb{P}(|S_n| < m) = \sum_{n=0}^{\infty} \sum_{|x|<m} \sum_{l=0}^{n} \mathbb{P}(H^x = l, S_n = x)$$

Using part 1., exchanging sums (which is fine as everything is non-negative)

$$= \sum_{n=0}^{\infty} \sum_{|x|<m} \sum_{l=0}^{n} \mathbb{P}(H^x = l)\mathbb{P}(S_{n-l} = 0) = \sum_{|x|<m} \sum_{l=0}^{\infty} \mathbb{P}(H^x = l) \sum_{n=l}^{\infty} \mathbb{P}(S_{n-l} = 0).$$

Finally, $\sum_{l=0}^{\infty} \mathbb{P}(H^x = l) \le 1$ for any x. There are $2m - 1$ possible choices for x.

3. Fix $0 < \epsilon < 1$. Suppose $a > 0$. By hypothesis there exists $N \in \mathbb{N}$ such that

$$\mathbb{P}(|S_n| > na) < \epsilon$$

for $n \ge N$.

Suppose $m \in \mathbb{N}$ and $N < n < m/a$. Then $m > na$ and $\mathbb{P}(|S_n| < m) \ge \mathbb{P}(|S_n| \le na) \ge 1 - \epsilon$.

Using previous part,

$$\begin{aligned}
\sum_{n=0}^{\infty} \mathbb{P}(S_n = 0) &\ge \frac{1}{2m} \sum_n \mathbb{P}(|S_n| < m) && (9)\\
&\ge \frac{1}{2m} \sum_{n=N+1}^{m/a} \mathbb{P}(|S_n| < m) && (10)\\
&\ge \left(\frac{1}{2m}\right)(m/a - N - 1)(1 - \epsilon), && (11)
\end{aligned}$$

as all the terms are non-negative.

Letting $m \to \infty$ shows $\sum_n \mathbb{P}(S_n = 0) \ge (1/2)a^{-1}(1 - \epsilon)$.

The choice of $a > 0$ was arbitrary, hence $\sum_n \mathbb{P}(S_n = 0) = \infty$. Apply Theorem 1.25. □

1.8.2 Symmetric simple random walk in $d \ge 2$

Definition 1.31 (Symmetric simple RW on $\mathbb{Z}^d$)**.** *A RW S on $\mathbb{Z}^d$, $d \in \mathbb{N}$ is called simple if it makes only nearest neighbour steps, i.e. $|S_{n+1} - S_n| = 1$ for every n, or equivalently, $S_{n+1} - S_n = \pm e_i$ for some $1 \le i \le d$. Here $(e_i)_j = \delta_{ij}$ is the i^{th} standard basis vector.*

It is called symmetric if steps in any direction are equally likely-

$$\mathbb{P}(S_{n+1} - S_n = -e_i) = \mathbb{P}(S_{n+1} - S_n = +e_i) = \frac{1}{2d},\ 1 \le i \le d.$$

Theorem 1.32. *Symmetric simple RW in $\mathbb{Z}^d$ is recurrent if $d = 1, 2$ and transient if $d > 3$.*

Proof. Case $d = 1$: Dealt with in Theorem 1.28.

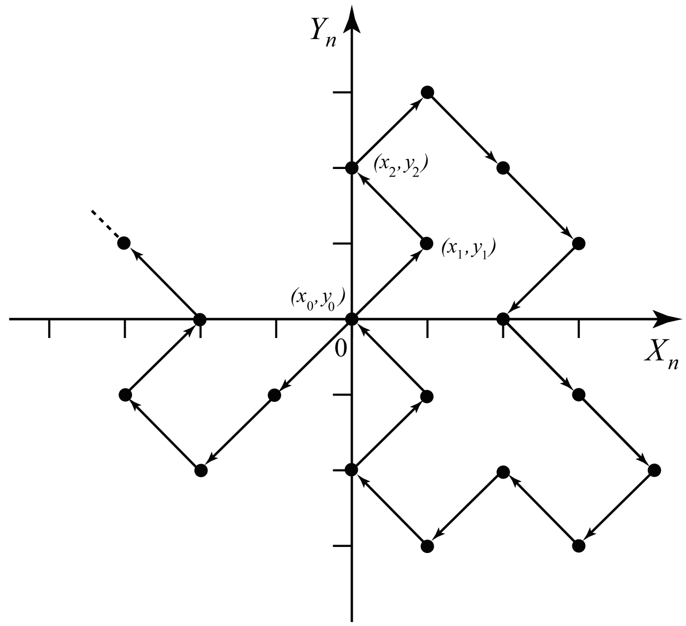
Case $d = 2$: Let $(X_n)_{n \in \mathbb{N}_0}$, $(Y_n)_{n \in \mathbb{N}_0}$ be independent SSRWs in $\mathbb{Z}$ starting from 0.

Observe that rotating the joint process $(X_n, Y_n)^T_{n \ge 0}$ by $\pi/4$ gives a SSRW on $\sqrt{2}\mathbb{Z}^2$.

I.e.

$$S_n = \begin{pmatrix} S_n^{(1)} \\ S_n^{(2)} \end{pmatrix} = \begin{pmatrix} 1/2 & 1/2 \\ -1/2 & 1/2 \end{pmatrix} \begin{pmatrix} X_n \\ Y_n \end{pmatrix},\ n \ge 0$$

has IID increments in $\{\pm e_1, \pm e_2\}$.



Further, $S_n = (0,0)^T$ iff $X_n = Y_n = 0$, and so

$$\mathbb{P}(S_{2k} = 0) = \mathbb{P}(X_{2k} = 0, Y_{2k} = 0) = \mathbb{P}(X_{2k} = 0)\mathbb{P}(Y_{2k} = 0).$$

Using the bound from Stirling's formula found in 1.28, there exists $N \in \mathbb{N}$ such that

$$\mathbb{P}(X_{2k} = 0) \geq Ck^{-1/2},$$

for $k > N$. For such k,

$$\mathbb{P}(S_{2k} = 0) = \mathbb{P}(X_{2k} = 0)\mathbb{P}(Y_{2k} = 0) \geq Ck^{-1}$$

and

$$\sum_k \mathbb{P}(S_{2k} = 0) = +\infty.$$

Now apply Theorem 1.25.

Case $d \geq 3$ (†): Assume $d = 3$. The proof below extends to $d > 3$, but you can do (optional) exercise:

If SSRW in $d \geq 2$ dimensions is recurrent then SSRW in $d - 1$ dimensions is reccurent (hint: consider the first $d - 1$ components).

Let S be a SSRW in $\mathbb{Z}^3$ starting from 0. As before S_n can be zero only if $n = 2k$ is even.

Each step has probability $1/2d = 1/6$, so any possible path has probability 6^{-2k}.

To count the number of possible paths with $S_0 = 0$, $S_{2k} = 0$ observe that each component of S must make an equal number of $+1$ and -1 steps.

So the number of paths with a total of $2i$ steps in the first direction, $2j$ in the second and $2l$ in the third $(i + j + l = k)$ is $2k!/(i!j!l!)^2$ (we choose i of them to be $+1$ steps in the first direction, i to be -1 steps in the first direction, and so on).

Hence, summing over the possible ways of distributing the k up steps to the 3 different dimensions,

$$\mathbb{P}(S_{2k}=0) = \sum_{i,j,l: i+j+l=k} \left(\frac{(2k)!}{(i!j!l!)^2}\right) 6^{-2k} = \left(\frac{(2k)!}{(k!)^2}\right) 2^{-2k} 3^{-k} 3^{-k} \sum_{i+j+l=k} \left(\frac{(k!)}{(i!j!l!)}\right)^2.$$

Note that

$$\sum_{i+j+l=k} \frac{(k!)}{(i!j!l!)} = 3^k \tag{12}$$

is the number of ways of putting k objects into 3 piles (of any size).

Now suppose $k = 3m$ for some m (i.e. $n = 6m$).

For any i, j, l with $i + j + l = k = 3m$, either $i = j = l = m$ or WLOG $l > m$. In the latter case,

$$i!j!l! = i!j!m!(m+1)(m+2)\ldots l.$$

On the right hand side there is a product $(m+1)(m+2)\ldots l$ of $l - m = 2m - i - j = (m-i) + (m-j)$ terms which are all bigger than m. If $i < m$, put $m - i$ of them with the $i!$, and put the remaining terms with $j!$ to see $i!j!l! > (m!)^3$. It follows that

$$\frac{(k!)}{(i!j!l!)} \leq (3m)!/(m!)^3. \tag{13}$$

Using (12) and (13),

$$\mathbb{P}(S_{6m}=0) \leq \frac{(6m)!}{((3m)!)^2} \frac{(3m)!}{(m!)^3} 2^{-6m} 3^{-3m}.$$

From here it is a matter of applying Stirling's formula. Instead of writing everything out like we did in Theorem 1.28, introduce the following labour saving notation:

Suppose $(a_n), (b_n)$ are positive sequences. Write $a_n \simeq b_n$ if $a_n/b_n \to$ a positive constant as $n \to \infty$. Thus $a_n \sim b_n$ implies that there exist $c, C > 0$ such that $cb_n < a_n < Cb_n$ for any n.

Stirling's formula is $n! \simeq n^{n+1/2}e^{-n}$. Plugging that in above,

$$\mathbb{P}(S_{6m}=0) \simeq m^{-3/2}.$$

For n not divisible by 6:

$$\mathbb{P}(S_{6m}=0) \geq \mathbb{P}(S_{6m-2}=0, S_{6m-1}=+e_1, S_{6m}=0) = (1/6)^2 \mathbb{P}(S_{6m-2}=0)$$

by the Markov property. Similarly, $\mathbb{P}(S_{6m-2}=0) \geq (1/6)^2 \mathbb{P}(S_{6m-4}=0)$.

I.e., $\mathbb{P}(S_{6m-2}=0), \mathbb{P}(S_{6m-4}=0) \leq 6^4 \mathbb{P}(S_{6m}=0)$ and $\mathbb{P}(S_{2k}=0) \leq Ak^{-3/2}$ (for some A). But $\sum_k k^{-3/2} < \infty$. $\square$

1.9 Communicating classes

A week last Thursday we saw an example of consolidating several states into a single state in order to simplify a MC when calculating expected hitting times.

The idea was to break the MC into smaller chunks, each of which is easy to understand and which together capture properties of the original chain.

This technique is also useful for analysing limiting distributions. The chunks are called communicating classes-

Notation: if $P = (p_{ij})_{i,j \in I}$ is a transition matrix and $n \in \mathbb{N}_0$, $p_{ij}^{(n)} := (P^n)_{ij}$.

Definition 1.33. *Suppose P is a transition matrix on I. We say $i \in I$ leads to $j \in I$ (under P) if $p_{ij}^{(n)} > 0$ for some $n \in \mathbb{N}_0$. We denote this $i \stackrel{(P)}{\to} j$.*

I.e. if X has transition matrix P, then $i \stackrel{P}{\to} j$ iff $\mathbb{P}_i(X_n = j) > 0$ for some $n \in \mathbb{N}_0$.

We say i communicates with j (written $i \leftrightarrow j$) if $i \to j$ and $j \to i$.

Observe that

1. $i \leftrightarrow i$ (reflexivity)
2. $i \leftrightarrow j$ iff $i \leftrightarrow j$ (symmetry)
3. $i \to j$, $j \to k$ implies $p_{ij}^{(m)} p_{jk}^{(n)} > 0$ for some $m, n \in \mathbb{N}_0$ implies $i \to k$ (transitivity)

for any $i, j, k \in I$.

Thus $\leftrightarrow$ defines an equivalence relationship and partitions I up into equivalence classes $[i] = \{j \in I : i \leftrightarrow j\}$, $i \in I$.

Definition 1.34. *The equivalence classes of $\leftrightarrow$ are called communicating classes, i.e. $\mathcal{C} \subset I$ is a communicating class if, for any $i, j \in \mathcal{C}$ and $k \in I \setminus \mathcal{C}$, $i \leftrightarrow j$ but $i \not\leftrightarrow k$.*

Example 1.35. *Let $I = \{1, 2, 3, 4\}$ and*

$$P = \begin{pmatrix} 2/3 & 0 & 1/3 & 0 \\ 1/2 & 0 & 0 & 1/2 \\ 0 & 7/8 & 0 & 1/8 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

has two communicating classes– $\{1, 2, 3\}$ and $\{4\}$.

We may refer to the communicating classes of a MC. In this case we of course mean the communicating classes of its transition matrix.

Theorem 1.36. *Suppose X is a MC, $\mathcal{C}$ a communicating class of X and $i, j \in \mathcal{C}$.*

If i is recurrent then so is j.

Proof. Let P be the transition matrix of X. Since $i \leftrightarrow j$ there exist $m, n \in \mathbb{N}_0$ with $p_{ij}^{(m)} > 0$ and $p_{ji}^{(n)} > 0$.

Then, for any $k \in \mathbb{N}$

$$\mathbb{P}_j(X_{k+m+n} = j) \geq \mathbb{P}_j(X_{k+m+n} = j, X_n = i) = p_{ji}^{(n)} \mathbb{P}_i(X_{k+m} = j) \geq p_{ji}^{(n)} \mathbb{P}_i(X_k = i) p_{ij}^{(m)}.$$

So,

$$\sum_k \mathbb{P}_j(X_k = j) \geq p_{ji}^{(n)} \left(\sum_k \mathbb{P}_i(X_k = i) \right) p_{ij}^{(m)} = +\infty$$

if i is recurrent. □

Continuing to think of communicating classes as little chunks of a MC, we see that the 'building blocks' MCs have just one communicating class, i.e. you can get from any state to another with positive probability. They are called irreducible.

Definition 1.37. *A MC X on I is called irreducible if it has only one communicating class, i.e. $i \leftrightarrow j$ for every $i, j \in I$.*

1.10 Invariant measures

The next step towards limiting distributions is the concept of an invariant measure. These are a kind of 'fixed point' in the measures on I.

Definition 1.39. *Let P be a transition matrix on I. A measure μ on I ($\mu = (\mu_i)_{i \in I}$, $\mu_i \geq 0$) is called invariant if*

$$\sum_i \mu_i p_{ik} = \mu_k,$$

for any $k \in I$, or in more compact notation, $\mu P = \mu$.

Theorem 1.40. *Suppose X is a MC on I with transition matrix P and $T_k = \inf\{n \geq 1 : X_n = k\}$, $k \in I$.*

Let

$$\gamma_i^k = \mathbb{E}_k \left[\sum_{n=0}^{T_k - 1} \mathbb{1}_{X_n = i} \right] \geq 0$$

be the expected number of visits to $i \in I$ on an excursion from k.

Then

(a) For any $j \neq k$,

$$(\gamma^k P)_j := \sum_{i\in I} \gamma_i^k p_{ij} = \gamma_j^k, \tag{14}$$

and

$$(\gamma^k P)_k := \sum_{i\in I} \gamma_i^k p_{ik} = \mathbb{P}_k(T_k < \infty) \leq \gamma_k^k \ (= 1).$$

So $(\gamma^k P)_k = 1 = \gamma_k^k$ iff k is recurrent by Theorem 1.24.

(b) If P is irreducible and recurrent then $0 < \gamma_i^k < \infty$ for every $i, k \in I$.

Proof. (a) First off, for any $i \in I$,

$$\gamma_i^k = \mathbb{E}_k\left[\sum_{n=0}^{T_k-1} \mathbb{1}_{X_n=i}\right] = \mathbb{E}_k\left[\sum_{n=0}^{\infty} \mathbb{1}_{X_n=i,T_k=+\infty}\right] + \sum_{m=1}^{\infty} \mathbb{E}_k\left[\sum_{n=0}^{m-1} \mathbb{1}_{X_n=i,T_k=m}\right],$$

and after swapping the order of summation this becomes

$$\sum_{n=0}^{\infty} \mathbb{P}_k(X_n = i, T_k = +\infty) + \sum_{n=0}^{\infty}\sum_{m=n+1}^{\infty} \mathbb{P}_k(X_n = i, T_k = m) = \sum_{n=0}^{\infty} \mathbb{P}(X_n = i, T_k > n).$$

This shows

$$\gamma_i^k = \sum_{n=0}^{\infty} \mathbb{P}(X_n = i, T_k > n). \tag{15}$$

Let us prove (14) beginning from the RHS. Let $j \neq k$. Using (15),

$$\gamma_j^k = \sum_{n=0}^{\infty} \mathbb{P}_k(X_n = j, T_k > n) = p_{kj} + \sum_{n=2}^{\infty} \mathbb{P}_k(X_n = j, T_k > n).$$

Breaking down the sum according to possible states for X_{n-1},

$$p_{kj} + \sum_{n=2}^{\infty} \mathbb{P}_k(X_n = j, T_k > n) = p_{kj} + \sum_{i\neq k}\sum_{n=2}^{\infty} \mathbb{P}_k(X_{n-1} = i, X_n = j, T_k > n).$$

But

$$\mathbb{P}_k(X_{n-1} = i, X_n = j, T_k > n) = \mathbb{P}_k(X_{n-1} = i, X_n = j, T_k > n-1) = \mathbb{P}_k(X_{n-1} = i, T_k > n-1)p_{ij}$$

using the Markov property. Substituting this and $\gamma_k^k = 1$ in the penultimate expression and using (15) again shows

$$\gamma_j^k = \gamma_k^k p_{kj} + \sum_{i\neq k}\sum_{n=1}^{\infty} \mathbb{P}_k(X_n = i, T_k > n)p_{ij} = \sum_{i\in I} \gamma_i^k p_{ij} = (\gamma^k P)_j.$$

Now for $j = k$: Using (15),

$$(\gamma^k P)_k = \gamma^k_k p_{kk} + \sum_{i \neq k} \gamma^k_i p_{ik} = p_{kk} + \sum_{i \neq k} \sum_{n=1}^{\infty} \mathbb{P}(X_n = i, T_k > n) p_{ik}$$

Using the Markov property and $\{T_k = n+1, X_{n+1} = k\} = \{T_k > n, X_{n+1} = k\}$,

$$p_{kk} + \sum_{i \neq k} \sum_{n=1}^{\infty} \mathbb{P}_k(X_n = i, T_k > n) p_{ik} = p_{kk} + \sum_{n=1}^{\infty} \sum_{i \neq k} \mathbb{P}_k(X_n = i, T_k = n+1, X_{n+1} = k).$$

$$= p_{kk} + \sum_{n=1}^{\infty} \mathbb{P}_k(T_k = n+1, X_{n+1} = k) = \sum_{n=1}^{\infty} \mathbb{P}_k(T_k = n) = \mathbb{P}_k(T_k < \infty) \leq 1 = \gamma^k_k.$$

$\square$

Proof of part (b) tomorrow.

Proof of Theorem 1.40 ctd. (b) Since k recurrent, part (a) gives $\gamma^k P^m = \gamma^k P P^{m-1} = \ldots = \gamma^k$ for any $m \in \mathbb{N}$.

Let $i \in I$. By irreducibility there exists $m \in \mathbb{N}$ such that $p_{ki}^{(m)} > 0$. Hence,

$$\gamma_i^k = (\gamma^k P^m)_i = \sum_{j \in I} \gamma_j^k p_{ji}^{(m)} \geq \gamma_k^k p_{ki}^{(m)} > 0.$$

On the other hand, there exists m' with $p_{ik}^{(m')} > 0$, whence

$$1 = \gamma_k^k = (\gamma^k P^{m'})_k = \sum_{j \in I} \gamma_j p_{jk}^{(m')} \geq \gamma_i^k p_{ik}^{(m')}$$

and $\gamma_i^k \leq 1/p_{ik}^{(m')}$.

□

Theorem 1.41 (Uniqueness of invariant meaures)**.** *Suppose μ is an invariant measure (IM) for P ($\mu_i \geq 0$ all i, $\mu P = \mu$) with $\mu_k = 1$ for some $k \in I$.*

Let γ^k be as defined in Theorem 1.40. Then

(a) $\mu_i \geq \gamma_i^k$ for all $i \in I$.

(b) If P is irreducible and recurrent then $\mu = \gamma^k$.

Proof. (a) Let $l \neq k$. Using invariance repeatedly,

$$\mu_l = \sum_{j \in I} \mu_j p_{jl} = \mu_k p_{kl} + \sum_{j \neq k} \mu_j p_{jl} = p_{kl} + \sum_{j_1 \neq k} \sum_{j_2 \in I} \mu_{j_2} p_{j_2, j_1} p_{j_1, l} \ldots .$$

After repeating $n+1$ times,

$$= p_{kl} + \sum_{j_1 \neq k} p_{k,j_1} p_{j_1,l} + \ldots + \sum_{j_1 \neq k} \cdots \sum_{j_n \neq k} p_{k,j_n} p_{j_n,j_{n-1}} \cdots p_{j_1,l} + \sum_{j_1 \neq k} \cdots \sum_{j_{n+1} \neq k} \mu_{j_{n+1}} p_{j_{n+1},j_n} p_{j_n,j_{n-1}} \cdots p_{j_1,l}.$$

Using non-negativity of μ,

$$\geq p_{kl} + \sum_{j_1 \neq k} p_{k,j_1} p_{j_1,l} + \ldots + \sum_{j_1 \neq k} \cdots \sum_{j_n \neq k} p_{k,j_n} p_{j_n,j_{n-1}} \cdots p_{j_1,l}.$$

But $p_{kl} = \mathbb{P}_k(X_1 = l) = \mathbb{P}_k(X_1 = l, T_k > 1)$ and

$$\sum_{j_1 \neq k} \cdots \sum_{j_n \neq k} p_{k,j_n} p_{j_n,j_{n-1}} \cdots p_{j_1,l} = \mathbb{P}_k(X_{n+1} = l, T_k > n+1).$$

Hence,

$$\mu_l \geq \sum_{m=1}^{n+1} \mathbb{P}_k(X_m = l, T_k > m) \nearrow \sum_{m=1}^{\infty} \mathbb{P}_k(X_m = l, T_k > m) = \gamma_l^k.$$

(b) By hypothesis

- P is irreducible and recurrent, so $\gamma^k P = \gamma^k$ is invariant by Theorem 1.40.
- $\mu P = \mu$.

So by linearity $\tilde{\mu} = \mu - \gamma^k$ has $\tilde{\mu}P = \mu P - \gamma^k P = \tilde{\mu}$. Thus $\tilde{\mu}$ is an IM. From part (a), $\tilde{\mu}$ is non-negative.

Now, $\tilde{\mu}_k = \mu_k - \gamma_k^k = 1 - 1 = 0$. Suppose $i \neq k$ is such that $\tilde{\mu}_i > 0$. By irreducibility there exists $m \in \mathbb{N}$ with $p_{ik}^{(m)} > 0$. Hence, by invariance of $\tilde{\mu}$,

$$0 = \tilde{\mu}_k = \sum_{j \in I} \tilde{\mu}_j p_{jk}^{(m)} \geq \tilde{\mu}_i p_{ik}^{(m)} > 0.$$

Contradiction. □

1.10.1 Invariant distributions, null and positive recurrence

Theorem 1.40 says irreducible recurrent MCs have an IM, μ say, but we want an invariant *probability* measure, i.e. total mass = 1.

Well, if μ is an IM for P and $c > 0$, then

$$(c\mu)P = c\mu P = c\mu.$$

Hence if $\sum_i \mu_i < \infty$ then the normalised measure $(\sum_i \mu_i)^{-1}\mu$ is a an invariant probabity measure.

This suggests making a distinction according to whether $\sum_i \mu_i$ is finite or infinite

For the IM γ^k from Theorem 1.40, the total mass is

$$\sum_i \gamma_i^k = \sum_i \mathbb{E}_k \left[\sum_{n=0}^{T_k - 1} \mathbb{1}_{X_n = i} \right] = 1 + \sum_{i \neq k} \mathbb{E}_k \left[\sum_{n=1}^{T_k - 1} \mathbb{1}_{X_n = i} \right] = 1 + \mathbb{E}_k[T_k - 1] = \mathbb{E}_k[T_k],$$

the expected return time from k.

This motivates the definition-

Definition 1.42. *Suppose X is a MC, state $k \in I$ is recurrent and $T_k = \inf\{n \geq 1 : X_n = k\}$ is the return time.*

Then k is called positive *recurrent if $\mathbb{E}_k[T_k] < \infty$ and* null *recurrent if $\mathbb{E}_k[T_k] = +\infty$.*

Example 1.43. *Symmetric SRW on $\mathbb{Z}$ is recurrent. By ex. 8 sheet 1, the return time T_0 satisfies $\mathbb{E}_0[T_0^\alpha] < \infty$ iff $\alpha < 1/2$. So SSRW is only null recurrent.*

Theorem 1.44. *Suppose X is an irreducible MC with transition matrix P.*

(i) if X is positive recurrent-

(a) X has a unique invariant probability measure π and $\pi_i > 0$ for every $i \in I$.

(b) The expected return time for state k is given by $\mathbb{E}_k[T_k] = \frac{1}{\pi_k}$.

(ii) Conversely, if X has an invariant probability measure, π, then X is positive recurrent (and π is as in (i)).

Part (c) already looks like it says something about limiting distributions– intuitively you expect to visit state k every $\mathbb{E}_k[T_k]$ units of time.

Proof. (i)(a) Following the discussion above-

Let γ^k be as in Theorem 1.40.

Since P is irreducible and recurrent, Theorem 1.40 ensures $\gamma^k P = \gamma^k$ and $0 < \gamma_i^k < \infty$.

P is positive recurrent so $\sum_i \gamma_i^k = \mathbb{E}_k[T_k] < \infty$.

So $\pi = (\sum_i \gamma_i^k)^{-1}\gamma^k$ is an invariant probability measure, with $\pi_j = (\sum_i \gamma_i^k)^{-1}\gamma_j^k > 0$.

Suppose $\tilde{\pi}$ is another invariant probability measure.

Then $\tilde{\pi}_k > 0$ some $k \in I$ since $\sum_i \tilde{\pi}_i = 1$.

The measure $\mu = \tilde{\pi}/\pi_k$ has $\mu_k = 1$. So Theorem 1.41 (b) guarantees

$$\mu_j = \gamma_j^k = \pi_j \sum_i \gamma_i^k, \ \forall j \in I.$$

Summing over j and using $\sum_i \tilde{\pi}_i = \sum_i \pi_i = 1$ gives $\tilde{\pi}_k = (\sum_i \gamma_i^k)^{-1}$.

But $\tilde{\pi}/\tilde{\pi}_k = \gamma^k$. So

$$\tilde{\pi} = (\sum_i \gamma_i^k)^{-1}\gamma^k = \pi.$$

□

Rest of proof on Thursday.

Proof of Theorem 1.44 ctd. (b) From end of part (a), $\pi_k = (\sum_i \gamma_i^k)^{-1}$. But $\sum_i \gamma_i^k = \mathbb{E}_k[T_k]$.

(ii) Let π be an invariant probability measure and P irreducible.

Since $\sum_i \pi_i = 1$, there exists k such that $\pi_k > 0$. Then $\mu = \pi/\pi_k$ has $\mu_k = 1$.

Part (a) of Theorem 1.41 implies $\mu_i \geq \gamma_i^k$ for every $i \in I$.

Hence $\mathbb{E}_k[T_k] = \sum_i \gamma_i^k \leq 1/\pi_k < \infty$ and k is positive recurrence.

□

Example 1.45 (A birth-death chain)**.** *Let $I = \mathbb{N}_0$ and $P = (p_{ij})$ be*

$$p_{i,i+1} = p \in (0,1),\ p_{i,i-1} = 1-p = q,\ p_{01} = r \in (0,1].$$

Then P is irreducible.

Show positive recurrence for $p < 1/2$.

By Theorem 1.44 it is sufficient to find an invariant distribution, π say.

The invariance equations are

$$\pi_i = \pi_{i-1}p + \pi_{i+1}q,\ i \geq 1,$$

and $\pi_0 = \pi_0(1-r) + q\pi_1$.

The general solution is

$$\pi_i = A + B(p/q)^i, i \geq 1.$$

If $A \neq 0$, then $\sum_i \pi$ diverges. So look for a solution with $A = 0$.

Then $r\pi_0 = q\pi_1 = qB(p/q) = pB$.

I.e. $\pi_i = \pi_0(r/p)(p/q)^i$, $i \geq 1$.

We must choose π_0 so that π is a probability measure,

$$1 = \sum_i \pi_i = \pi_0\left(1 + (r/p)\sum_{i=1}^{\infty}(p/q)^i\right) = \pi_0\left(1 + (r/p)(p/q)\frac{1}{(1-p/q)}\right) = \pi_0\left(\frac{1-2p+r}{1-2p}\right).$$

So taking

$$\pi_0 = \left(\frac{1-2p+r}{1-2p}\right)^{-1} (< 1)$$

does the trick.

Also by Theorem 1.44 the expected excursion time is

$$\mathbb{E}_k[T_k] = 1/\pi_k = (q/p)^k(p/r)\pi_0^{-1},\ k \geq 1,$$

and the number of visits to 0 on an excursion from $k > 0$ is

$$\gamma_i^k = \pi_0/\pi_k = (q/p)^k(p/r).$$

Here is a warning: in Theorem 1.44 existence of a finite mass IM for an irreducible MC implies positive recurrence. One may wonder whether existence of an IM with infinite mass implies null recurrence. This is false–

Example 1.46 (Irreducible + existence of IM $\not\Rightarrow$ recurrence)**.** *Recall SSRW on $\mathbb{Z}^d$, $d \geq 3$ is transient. Let $\pi_x = 1$ for any $x \in \mathbb{Z}^d$. Then*

$$\pi_x = 2d(1/2d) = \sum_{j=1}^{d} (\pi_{x-e_i} + \pi_{x+e_i})(1/2d),$$

so π is an IM and has infinite mass.

1.11 Convergence to invariant probability measure

Suppose X is a MC on $I \ni i, j$. Does $\mathbb{P}_i(X_n = j)$ converge as $n \to \infty$? What can go wrong? Does the limit depend on the initial state i?

There are three things that can go wrong.

(1) Consider

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

Then $P^2 = I$, $P^{2k} = (P^2)^k = I$ and $P^{2k+1} = P^{2k}P = P$ for any k. So $\mathbb{P}_i(X_n = j) = (P^n)_{ij}$ oscillates from 0 and 1.

This phenomenon is called periodicity. It already came up for a SRW S– if $S_0 = 0$ then S_n is even iff n even.

Definition 1.47. *Suppose P is a transition matrix on I. The period of state $i \in I$ is defined as*

$$period(i) = gcd\{n \geq 1 : p_{ii}^{(n)} > 0\},$$

where gcd=greatest common divisor is undefined if the set is empty.

If period(i) = 1 then i is called aperiodic.

(2) The chain may be reducible– an easy example is having two absorbing states. If the chain starts in either, it stays there forever, so $\mathbb{P}_i(X_n = j) = \delta_{ij}$. To avoid this problem we only consider irreducible chains.

(3) As in ex. sheet 2, $\mathbb{P}_i(X_n = j) \to 0$ for all j is possible (mass disappears). This occurs in transient or null recurrent chains.

Theorem 1.48 (Convergence to equilibrium)**.** *Let X be an irreducible, aperiodic and positive recurrent MC. Then X has a unique invariant probability measure π (Theorem 1.44), and*

$$\mathbb{P}_i(X_n = j) \to \pi_j$$

for any initial state i. I.e. X_n converges weakly to its invariant measure π no matter where it starts.

Proof on Monday.

In preparation for proving the convergence Theorem 1.48, we need to show that aperiodic irreducible chains can jump from any state to another after any sufficiently large number of steps.

Forthcoming sheet 3 exercise: all states of a communicating class have the same period.

So it makes sense to talk of the period of an irreducible MC (or its transition matrix).

Theorem 1.49. *Let P be irreducible and aperiodic. For any $i, j \in I$ there exists $N \in \mathbb{N}$ such that $p_{ij}^{(n)} > 0$ for any $n > N$.*

Proof. † Let $S = \{n \geq 1 : p_{ii}^{(n)} > 0\}$. Enumerate the elements of S as $s_1 < s_2 < \ldots$.

Consider the sequence of integers $a_n = \gcd\{s_1, \ldots, s_n\}$. It is decreasing $1 \leq a_{n+1} \leq a_n$ (since the gcd of $s_1, \ldots, s_{n+1}$ must divide $s_1, \ldots s_n$.

Further, $a_n \searrow 1 = \gcd(S)$ – suppose $a_n \searrow d \in \mathbb{N}$. Then $\exists N' \in \mathbb{N}$ such that $d \leq a_n \leq d + 1/2$ for $n > N'$.

Thus $a_n = d$ eventually, and d divides s_n for every $n \in \mathbb{N}$. In particular $\gcd(S) \geq d$.

It follows that $1 = a_M = \gcd\{s_1, \ldots, s_M\}$ for some $M \in \mathbb{N}$.

Now, since $p_{ii}^{(s_k)} > 0$,

$$p_{ii}^{(\sum_{k=1}^{m} l_k s_k)} \geq \prod_{k=1}^{m} (p_{ii}^{(s_k)})^{l_k} > 0$$

for any non-negative integers $l_1, \ldots, l_k \geq 0$.

Hence it is sufficient to find $N \in \mathbb{N}$ such that any $n > N$ can be written $n = \sum_{k=1}^{m} l_k s_k$.

By Euclid's algorithm (see e.g. MA132) there exist $\alpha_1, \ldots, \alpha_M \in \mathbb{Z}$ such that

$$\sum_{k=1}^{M} \alpha_k s_k = \gcd\{s_1, \ldots, s_M\} = 1.$$

In particular

$$\sum_{k=1}^{M} (j\alpha_k) s_k = j, \quad j \in \mathbb{Z}.$$

The α_k here may be negative. To account for this, let $N = s_1 \sum_{k=1}^{M} |\alpha_k| s_k$.

For any $0 \leq j < s_1$. Then any $n > N$ can be written as

$$n = N + ls_1 + j = ls_1 + \sum_{k=1}^{M} (s_1|\alpha_k| + j\alpha_k) s_k,$$

where l and $s_1|\alpha_k| + j\alpha_k$ are non-negative integers.

So $p_{ii}^{(n)} > 0$ for $n > N$. By irreducibility there exists $m \in \mathbb{N}$ with $p_{ij}^{(m)} > 0$. Let $n' > N' = N + m$. Then

$$p_{ij}^{(n')} = p_{ij}^{(m+(n'-m))} \geq p_{ii}^{(n'-m)} p_{ij}^{(m)}.$$

□

Proof of Theorem 1.48 †. We'll prove the Theorem for an arbitrary initial distribution, $\mathbb{P}(X_0 = i) = \lambda_i$, any $i \in I$.

Denote the transition matrix of X by P. Let Y be a (π, P)-MC independent of X. Fix $b \in I$ and

$$H = \inf\{n \geq 0 : X_n = Y_n = b\}$$

Overview of proof Step (1): at the first time H that both X and Y hit b, we use the Strong Markov property to exchange $(X_{m+H})_{m\geq 0}$ and $(Y_{n+H})_{n\geq 0}$.

This defines a new process $Z_n = X_n$ for $n < H$ and $Z_n = Y_n$ for $n \geq H$. We'll see Z is a (λ, P)-MC.

Step (2): Using the definition of Z and that $\mathbb{P}(Y_n = j) = \pi_j$ we show $|\mathbb{P}(X_n = j) - \pi_j| \leq \mathbb{P}(H > n)$.

Step (3): Show $\mathbb{P}(H > n) \to 0$ as $n \to \infty$.

This is called a coupling argument. J. Michael Steele says: Coupling is one of the most powerful of the "genuinely probabilistic" techniques. Here by "genuinely probabilistic" we mean something that works directly with random variables rather than with their analytical co-travelers (like distributions, densities, or characteristic functions).

<u>Step (1)</u>

$(X_n, Y_n)_{n\in\mathbb{N}_0}$ is a Markov chain on $I \times I$.

It has transition matrix $\tilde{P}$ with entries

$$\tilde{p}_{(i,k),(j,l)} = p_{ij}p_{kl}$$

and initial distribution $\mu_{(i,k)} = \lambda_i \pi_k$ for any $i, j, k, l \in I$.

The hitting time H is a stopping time. Hence the Strong Markov property implies $(X_{T+n}, Y_{T+n})_{n\geq 0}$ is a $(\delta_{(b,b)}, \tilde{P})$-MC independent of $(X_m, Y_m)_{0\leq m<T}$.

The transition probabilities $\tilde{P}$ are symmetric in the two components. So, if $(X'_n, Y'_n)_{n\geq 0}$ is a $(\delta_{(b,b)}, \tilde{P})$-MC then $(Y'_n, X'_n)_{n\geq 0}$ is.

In particular, $(X_{T+n}, Y_{T+n})_{n\geq 0}$ is a $(\delta_{(b,b)}, \tilde{P})$-MC independent of $(X_m, Y_m)_{0\leq m<T}$ and

$$W_n = \begin{cases} (X_n, Y_n) & n < H \\ (Y_n, X_n) & n \geq H. \end{cases}$$

is a $(\mu, \tilde{P})$-MC. But the first component of a $(\mu, \tilde{P})$-MC, which here is

$$Z_n = \begin{cases} X_n & n < H \\ Y_n & n \geq H, \end{cases}$$

is a (λ, P)-MC.

Step (2) Z and X have the same distribution, so

$$\mathbb{P}(X_n = j) = \mathbb{P}(Z_n = j, H > n) + \mathbb{P}(Z_n = j, H \leq n) = \mathbb{P}(Z_n = j, H > n) + \mathbb{P}(Y_n = j, H \leq n)$$

since $Z_n = Y_n$ on $\{H \leq n\}$.

But

$$\mathbb{P}(Y_n = j, H \leq n) = \mathbb{P}(Y_n = j) - \mathbb{P}(Y_n = j, H > n) = \pi_j - \mathbb{P}(Y_n = j, H > n)$$

since $\mathbb{P}(Y_n = j) = (\pi P^n)_j = \pi_j$ by invariance of π.

So,

$$\mathbb{P}(X_n = j) - \pi_j = \mathbb{P}(Z_n = j, H > n) - \mathbb{P}(Y_n = j, H > n)$$

which has absolutely value at most $\mathbb{P}(H > n)$.

Step (3)

We show $\mathbb{P}(H > n) \to 0$ as $n \to \infty$.

P is irreducible and aperiodic by assumption. So, for any $i, j, k, l \in I$, the previous Theorem implies that there exists $N \in \mathbb{N}$ such that

$$p_{ij}^{(n)} > 0, \;\; p_{kl}^{(n)} > 0$$

for $n > N$. So $\tilde{p}_{(i,k),(k,l)}^{(n)} > 0$ for such n. In particular $\tilde{P}$ is irreducible.

Next, $\tilde{\pi}_{(i,k)} = \pi_i \pi_k$ is invariant for $\tilde{P}$ –

$$(\tilde{\pi}\tilde{P})_{j,l} = (\pi P)_j (\pi P)_l = \pi_j \pi_l = \tilde{\pi}_{j,l}.$$

It follows from Theorem 1.44 that (X, Y) is (positive) recurrent.

So $\mathbb{P}(H > n) \to 0$ by exercise A2 on sheet 2,

□

1.12 Reversibility and detailed balance

Before moving to continuous time chains we briefly touch on the connection between a MC watched backwards in time and its invariant measure.

Definition 1.50. *A MC X on I is called reversible if $(X_n)_{0\leq n\leq N}$ and the reversal $(X_{N-n})_{0\leq n\leq N}$ have the same distribution for all $N\in\mathbb{N}$, i.e.*

$$\mathbb{P}(X_0=i_0,\dots X_n=i_n)=\mathbb{P}(X_0=i_n,X_1=i_{n-1},\dots,X_n=i_0)$$

for any $i_0,i_1,\dots\in I$ and $n\in\mathbb{N}$.

Theorem 1.51. *Let X be a MC with transition matrix P.*

(i) X is reversible if and only if there exists a probability measure $\lambda=(\lambda_i)_{i\in I}$ such that

$$\lambda_i p_{ij}=\lambda_j p_{ji} \tag{16}$$

for every $i,j\in I$. These are called the detailed balance equations (DBEs) for P and λ.

(ii) If the DBEs (16) *hold above then λ is an invariant probability measure.*

Proof. (i) Let $\lambda_i=\mathbb{P}(X_0=i)$ for every $i\in I$. Since X is reversible,

$$\mathbb{P}(X_0=i,X_1=j)=\mathbb{P}(X_0=j,X_1=i) \tag{17}$$

for any i,j. Summing over j shows

$$\lambda_i=\mathbb{P}(X_0=i)=\sum_j\mathbb{P}(X_0=i,X_1=j)=\sum_j\mathbb{P}(X_0=j,X_1=i)=\mathbb{P}(X_1=i).$$

Suppose $\lambda_i>0$. Then, conditioning on $X_0=i$,

$$\mathbb{P}(X_0=i,X_1=j)=\mathbb{P}(X_1=j|X_0=i)\mathbb{P}(X_0=i)=\lambda_i p_{ij}.$$

Similarly, $\mathbb{P}(X_0=i,X_1=j)=\lambda_j p_{ji}$. Hence $\lambda_i p_{ij}=\lambda_j p_{ji}$ using (17).

If $\lambda_i=0$ then $\lambda_j p_{ji}=\mathbb{P}(X_0=j,X_1=i)=0$. Thus the DBEs (16) hold.

The converse is left as an exercise (sheet 3).

(ii) If the DBEs hold for λ and P then

$$\lambda_j=\lambda_j\sum_{i\in I}p_{ji}=\sum_{i\in I}\lambda_i p_{ij}=(\lambda P)_j$$

for any $j\in I$. □

Thus to find an invariant probability measure for an irreducible chain we can start by trying to solve the DBEs. When they have a solution, the DBEs are usually easier to solve than the invariance equation $\pi P = \pi$.

Warning exercise (sheet 3): not every MC that admits an invariant distribution is reversible!

Example 1.52 (MCMC)**.** *A very helpful and popular use of the convergence theorem for MCs is to sample from complicated probability distributions. Indeed the popularity has supported a whole research field for the last few decades.*

The idea is to construct a Markov chain which converges to the target distribution.

Denote the target distribution $\pi = (\pi_i)_{i\in I}$, wlog $\pi_i > 0$ for all $i \in I$. It is assumed difficult to sample from π directly. In some cases perhaps it is even impossible to calculate the normalising constant.

For a specific example– let $\Lambda = \{1, 2, \ldots, n\}^d$ ($d \in \mathbb{N}$) and $I = \{-1, +1\}^\Lambda$. In the Ising model for magnetism, an outcome $s = (s_x)_{x\in\Lambda} \in I$ has probability

$$\pi_s = \frac{1}{Z} \exp\left(-T^{-1} \sum_{x,y\in\Lambda:|x-y|=1} s_x s_y\right)$$

where $T > 0$ is a parameter (temperature) and $Z > 0$ is chosen so that $\sum_s \pi_s = 1$.

MCMC Algorithm

As an ingredient, the algorithm takes a 'proposal' distribution $(q_{ij})_{i,j}$ ($q_{ij} \geq 0$, $\sum_j q_{ij} = 1$).

For now the only requirement for q is $q_{ij} > 0$ iff $q_{ji} > 0$ for any i, j. Other requirements for q will become apparent below.

- *(0) Suppose X_0 is chosen arbitrarily.*
- *(n) Suppose $X_n = i$.*

  Sample a proposal Y according to $\mathbb{P}(Y = j) = q_{ij}$.

  Accept $X_{n+1} = j$ with probability

$$\alpha(j|i) = \min\left\{1, \frac{\pi_j q_{ji}}{\pi_i q_{ij}}\right\}$$

  Otherwise, keep $X_{n+1} = i$.

This defines a MC with transition probabilities

$$p_{ij} = q_{ij}\alpha(j|i)$$

for $i \neq j$ and

$$p_{ii} = q_{ii}\alpha(i|i) + \sum_{j\in I} q_{ij}(1 - \alpha(j|i)).$$

To check P is a bona-fide transition matrix: $p_{ij} \geq 0$ is immediate and

$$\sum_j p_{ij} = \sum_{j \neq i} q_{ij}\alpha(j|i) + q_{ii}\alpha(i|i) + \sum_j q_{ij}(1 - \alpha(j|i)) = 1.$$

Now, X is reversible: for $i \neq j$ with $q_{ij}, q_{ji} > 0$,

$$\pi_i p_{ij} = \pi_i q_{ij}\alpha(j|i) = \min\{\pi_i q_{ij}, \pi_j q_{ji}\} = \pi_j q_{ji}\alpha(i|j) = \pi_j p_{ji}$$

so the detailed balance equations hold.

Thus, if q is chosen so that X is recurrent and aperiodic, then

$$\mathbb{P}(X_n = i) \to \pi_i$$

as $n \to \infty$. I.e. if we run the algorithm for many steps we get something that is nearly distributed according to π.

Main questions: How long do we have to run the chain before $\sup_i |\mathbb{P}(X_n = i) - \pi_i| < \epsilon$ (for some $\epsilon > 0$)? What choices of q minimise the amount of time to wait? Is it possible to tune q while running the simulation? and so on.

1.13 Summary of discrete time MCs

A MC $(X_n)_{n\geq 0}$ on countable I is characterised entirely by its initial distribution $\mathbb{P}(X_0 = i), i \in I$ and transition probability matrix P (Theorem 1.6).

X is irreducible if for any i, j, $(P^n)_{ij} > 0$ for some n.

A state i is recurrent if $\mathbb{P}_i$(X visits i i.o) $= 1$, or equivalently the return probability $\mathbb{P}_i(T_i < \infty) = 1$ (Theorem 1.24), or expected visits to i, $\sum_i \mathbb{P}_i(X_n = i) = \infty$ (Theorem 1.25).

Otherwise i is transient.

A recurrent state i is positive recurrent if $\mathbb{E}_i[T_i] < \infty$, null recurrence otherwise.

A measure $\mu = (\mu_i)_{i\in I}, \mu_i \geq 0$ is called invariant if $\mu P = \mu$.

If X is positive recurrent irreducible it has a unique invariant probability measure $\pi_i = (\mathbb{E}_i[T_i])^{-1} > 0$ (Theorem 1.44).

State i is called aperiodic if $gcd\{n \geq 1 : p_{ii}^{(n)} > 0\} = 1$.

If X is aperiodic positive recurrent irreducible then $\mathbb{P}_i(X_n = j) \to \pi_j$ for any $i, j \in I$ (Theorem 1.48).

2 Markov chains in continuous time

Recall from definition 1.1 that a continuous stochastic process with state space I is a family, $(X_t)_{t\geq 0}$ say, of random variables taking values in I.

As in the previous section, I is assumed discrete.

We only consider the case X is right continuous, i.e. $\lim_{t'\searrow t} X_{t'} = X_t$ for any $t \geq 0$, with probability one. This limit is with respect to the discrete topology (i.e. think of $I = \mathbb{Z} \subset \mathbb{R}$ with ϵ-δ definition of convergence from Analysis I– there are no limit points).

This assumption guarantees that certain interesting events are actually measurable.

For example, let $H^A = \inf\{t \geq 0 : X_t \in A\}$ and consider the event $\{H^A < \infty\}$ that X hits $A \subset I$.

The discrete time analogue of this event can be written as a countable union of events, but now

$$\{H^A < \infty\} = \bigcup_{t\geq 0} \{X_t \in A\}$$

is an uncountable union. Thus it may not be measurable.

However, if X is right continuous in discrete topology, and $X_t = i \in I$ then there exists $\delta > 0$ such that $X_s = i$ for $t \leq s < \delta$.

But $[t, t + \delta)$ contains a rational number $r \in \mathbb{Q}_+ = \mathbb{Q} \cap [0, \infty)$.

So,

$$\{X_t = i\} \subseteq \bigcup_{r \in \mathbb{Q}_+} \{X_r = i\} \subseteq \bigcup_{t \geq 0} \{X_t = i\}.$$

In particular $\{H^A < \infty\}$ is measurable.

Another useful consequence of right continuity is the applicability of a FACT from measure theory: if $(X_t)_{t \geq 0}$ is right continuous then its (joint) distribution is determined by the finite dimensional distributions

$$\mathbb{P}(X_{t_0} = i_0, X_{t_1} = i_1, \ldots, X_{t_n} = i_n),$$

$i_0, i_1, \ldots \in I$, $0 \leq t_1 \leq t_2 \leq \ldots$ and $n \in \mathbb{N}$.

Definition 2.1. *A stochastic process $(X_t)_{t \geq 0}$ is a Markov chain if*

$$\mathbb{P}(X_{t_{n+1}} = i_{n+1} | X_{t_0} = i_0, \ldots, X_{t_n} = i_n) = \mathbb{P}(X_{t_{n+1}} = i_{n+1} | X_{t_n} = i_n)$$

for any $i_0, i_1, \ldots \in I$, $0 \leq t_1 \leq t_2 \leq \ldots$ and $n \in \mathbb{N}$ with

$$\mathbb{P}(X_{t_0} = i_0, \ldots, X_{t_n} = i_n) > 0.$$

It is time homogenous if

$$\mathbb{P}(X_{t+s} = j | X_s = i)$$

does not depend on $s \geq 0$ for any $i, j \in I$ and $t \geq 0$.

All our Markov chains will be time homogenous.

2.1 Anatomy of a continuous time Markov chain I

Let X be a Markov chain with $X_0 = i \in I$ and

$$J_1 = \inf\{t > 0 : X_t \neq i\}$$

be the first jump time. Let's explain why J_1 has exponential distribution.

First of all, for any $s, t \geq 0$,

$$\mathbb{P}(J_1 > s + t | J_1 > s) = \mathbb{P}(X_u = i,\ s \leq u \leq s + t | X_u = i,\ 0 \leq u \leq s),$$

(assuming $\mathbb{P}(J_1 > s) > 0$). Using the Markov property and then time-homogeneity, the RHS is

$$\mathbb{P}(X_u = i,\ s \leq u \leq s + t | X_s = i) = \mathbb{P}(X_u = i,\ 0 \leq u \leq t | X_0 = i) = \mathbb{P}(J_1 > t).$$

That is, J_1 is memoryless – $\mathbb{P}(J_1 > s + t | J_1 > s) = \mathbb{P}(J_1 > t)$.

Recall that T has Exponential(λ), $\lambda > 0$ distribution if $\mathbb{P}(T > t) = \exp(-\lambda t)$. We can also allow $\lambda = 0$, in which case $T = \infty$ almost surely.

Theorem 2.2. *Let T be a possibly infinite non-negative random variable with $\mathbb{P}(0 < T < \infty) > 0$. Then T has Exponential(λ) distribution (for some $\lambda > 0$) if and only if*

$$\mathbb{P}(T > t + s | T > s) = \mathbb{P}(T > t), \ \ \forall s, t \geq 0,$$

(when $\mathbb{P}(T > s) > 0$).

Proof. If T has Exponential($\lambda > 0$) distribution, then $\mathbb{P}(T > s) = \exp(-\lambda s) > 0$ and

$$\mathbb{P}(T > t + s | T > s) = \exp(-(t+s)\lambda)/\exp(-\lambda s) = \exp(-t\lambda)$$

for any $s, t \geq 0$.

Now, for the converse. Define $g(t) = \mathbb{P}(T > t)$, $t \geq 0$.

Suppose $s, t \geq 0$ and $g(s) > 0$. Since T is memoryless,

$$g(t) = g(s + t - s) = \mathbb{P}(T > s + t - s) = \mathbb{P}(T > s + t | T > s) = g(t + s)/g(s).$$

If $g(s) = 0$, then $g(t + s) = 0$ as g is monotone decreasing. I.e.

$$g(s + t) = g(s)g(t) \tag{18}$$

for any $s, t \geq 0$.

Since g is right continuous (by monotone convergence Theorem) and $g(0) = \mathbb{P}(T > 0) > 0$ by hypothesis, there exists $n \in \mathbb{N}$ with $g(1/n) > 0$.

By (18)

$$g(1) = g(1/n + 1/n + \ldots + 1/n) = g(1/n)g((n-1)/n) = \ldots = g(1/n)^n > 0.$$

On the other hand, $g(1) \leq 1$. If $g(1) = 1$ then for any $m \in \mathbb{N}$,

$$g(m) = g(1 + (m-1)) = g(1)g(m-1) = \ldots = g(1) = 1.$$

This implies that $\mathbb{P}(T = \infty) = 1$. But by hypothesis $\mathbb{P}(T < \infty) > 0$.

Hence $0 < g(1) < 1$. Let $\lambda = -\ln g(1)$. Then

$$g(1/n) = g(1)^{1/n} = \exp(-\lambda/n).$$

For any non-negative integer j,

$$g(j/n) = g(1/n)^j = \exp(-\lambda j/n),$$

and thus for $r \in \mathbb{Q}_+$,

$$g(r) = \exp(-\lambda r).$$

For any $t > 0$, choose rationals $r, r' \in Q_+$ such that $r < t < r'$. By monotonicity of g,

$$\exp(-\lambda r') \leq g(t) \leq \exp(-\lambda r).$$

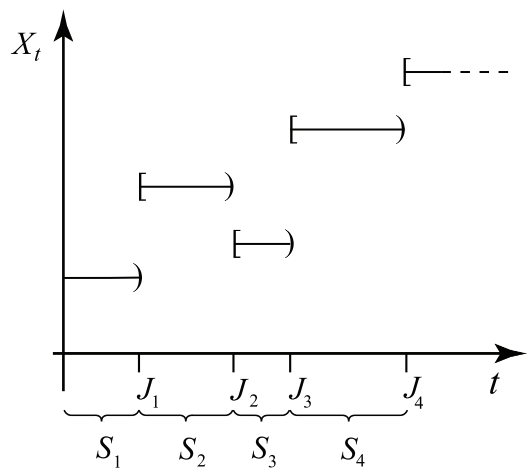
Let $r' \searrow t$, $r \nearrow t$ and use continuity of exp and the sandwich theorem.

□

Suppose $(X_t)_{t \geq 0}$ is a right-continuous stochastic process on the discrete space I. The jump times are defined by

$$J_0 = 0, J_{n+1} = \inf\{t \geq J_n : X_t \neq X_{J_n}\}. \tag{19}$$

By convention $\inf \emptyset$ is infinite and J_n infinite implies J_{n+1} infinite.



The n^{th} holding time is $S_n = J_n - J_{n-1}$ if $J_{n-1} < \infty$. Right continuity forces $S_n > 0$.

If $J_{n-1} = \infty$ then S_n is undefined.

Yesterday we argued that if X is Markov then $S_1 = J_1 > 0$ satisfies the memoryless hypothesis of Theorem 2.2. So there are two possibilities:

- $\mathbb{P}(J_1 < \infty) > 0$ and J_1 is Exponential(q_i) for some $q_i > 0$. We call q_i the rate of leaving state i. Note that actually $J_1 < \infty$ almost surely in this case.
- $J_1 = \infty$ almost surely. This corresponds to state i being absorbing. The rate of leaving is zero.

This observation is an early warning that exponential random variables will feature prominently.

Now we continue defining terms for right-continuous processes.

The jump process is defined by $Y_n = X_{J_n}$, $n \in \mathbb{N}_0$. By convention $X_\infty = J_k$ where $k = \inf\{k \geq 0 : J_k < \infty\}$, the state at which X is absorbed.

2.1.1 Explosion and minimality

Right-continuous processes can jump infinitely many times in a finite time. If this happens the process is said to explode.

Definition 2.3. *With J_n as in* (19)*, the random variable $\zeta = \sup_n J_n = \lim_{n\to\infty} J_n$ is called the (first) explosion time of X. X is called explosive if $\mathbb{P}(\zeta < \infty) > 0$.*

After explosion, a process can respawn in many different ways. It can go on to explode infinitely many times. For simplicity we ignore this possibility and assume that when a process explodes it dies and gets buried in a special 'cemetry' state ∂.

This corresponds to X having a minimal life span.

Definition 2.4. *Suppose $I \ni \partial$. X is called minimal if $X_t = \partial$ for $t \geq \zeta$.*

Any process can be modified to be minimal by adjoining ∂ to I and redefining $X_t = \partial$ for $t \geq \zeta$.

A minimal process is determined by the holding times $S_1, S_2, \dots$ and jump process Y. In particular, for any $t \geq 0$,

$$X_t = \begin{cases} Y_n & \text{if } \sum_{i=1}^n S_i \leq t < \sum_{i=1}^{n+1} S_i, \\ \partial & t \geq \sum_{i=1}^\infty S_i. \end{cases} \tag{20}$$

2.2 Poisson processes

The recipe (20) also allows a right-continuous process to be constructed from holding times and a jump process.

A particular example is the case $I = \mathbb{N}_0$, $S_1, S_2, \dots$ are IID Exponential($\lambda > 0$) and $Y_n = n$, $n \in \mathbb{N}_0$.

The resulting process,

$$N_t = \begin{cases} n & \text{if } \sum_{i=1}^n S_i \leq t < \sum_{i=1}^{n+1} S_i, \\ \partial & t \geq \sum_{i=1}^\infty S_i \end{cases}$$

is called a Poisson process with rate (or intensity) λ (or PP(λ) for short).

Theorem 2.5. *PP(λ) is non-explosive for any $\lambda > 0$.*

Proof. Let $S_1, \dots$ be as above. We must show the explosion time

$$\zeta = \sum_{n=1}^\infty S_n$$

is infinite almost surely.

For any $N \in \mathbb{N}$,

$$\mathbb{E}[\exp(-\zeta)\, \mathbb{1}_{\zeta<\infty}] \leq \mathbb{E}\left[\exp\left(-\sum_{n=1}^N S_n\right)\right] = \prod_{n=1}^N \mathbb{E}[\exp(-S_n)]$$

using independence of the S_n. But

$$\mathbb{E}[\exp(-S_n)] = \int_0^\infty \exp(-t)\lambda\exp(-\lambda t)dt = \lambda/(1+\lambda) < 1,$$

so $\prod_{n=1}^N \mathbb{E}[\exp(-S_n)] \to 0$ as $N \to \infty$.

Thus $\mathbb{E}[\exp(-\zeta)\,\mathbb{1}_{\zeta<\infty}] = 0$, which implies $\exp(-\zeta)\,\mathbb{1}_{\zeta<\infty} = 0$ almost surely.

□

Recall the Gamma(λ, n), $\lambda > 0$, $n \in \mathbb{N}$ distribution on $(0, \infty)$ has density

$$\frac{\lambda^n}{(n-1)!} t^{n-1} e^{-\lambda t}.$$

Theorem 2.6. *Suppose $S_1, S_2, \ldots$ are IID Exponential($\lambda > 0$) random variables. Then $S_1 + \ldots + S_n$ has Gamma(λ, n) distribution for any $n \in \mathbb{N}$.*

Proof. The Gamma$(\lambda, 1)$ density is $\lambda \exp(-\lambda t)$, $t > 0$ which is the same as the Exp(λ) density.

Suppose $S_1 + \ldots + S_n$ is Gamma(λ, n) distributed.

Then

$$\mathbb{P}(S_1 + \ldots + S_n + S_{n+1} > t) = \int_0^\infty \mathbb{P}(S_1 + \ldots + S_n \in da, S_{n+1} > t - a)$$

$$= \int_0^t e^{-\lambda(t-a)} \frac{\lambda^n}{(n-1)!} a^{n-1} e^{-\lambda a} da + \int_t^\infty \frac{\lambda^n}{(n-1)!} a^{n-1} e^{-\lambda} da = e^{-\lambda t} \frac{(\lambda t)^n}{n!} + \int_t^\infty \frac{\lambda^n}{(n-1)!} a^{n-1} e^{-\lambda a} da.$$

Differentiate to get (minus) the density,

$$-\lambda e^{-\lambda t} \frac{(\lambda t)^n}{n!} + n e^{-\lambda t} \frac{\lambda^n t^{n-1}}{n!} - \frac{\lambda^n}{(n-1)!} t^{n-1} e^{-\lambda t}.$$

This is (minus) the Gamma$(\lambda, n+1)$ density. □

Theorem 2.7. *Let N be a PP(λ). Then N_t has Poisson(λt) distribution.*

Proof: see exercise sheet 3.

Theorem 2.8. *Suppose N a PP($\lambda > 0$). For any $t \geq 0$, $(N_{t+s} - N_t)_{s \geq 0}$ is also a PP(λ) independent of $(N_u)_{0 \leq u \leq t}$.*

Proof. Let $k \in \mathbb{N}_0$. We condition on the event $\{N_t = k\}$, which has positive probability.

Let $S_1, S_2, \ldots$ be the exponential holding times of N, i.e. $S_{n+1} = J_{n+1} - J_n$ where $J_n = \inf\{t \geq 0 : N_t = n\}$, $n \in \mathbb{N}_0$.

On the event $\{N_t = k\} = \{J_n \leq t < J_{n+1}\}$

- $(N_u)_{0 \leq u \leq t}$ is determined by $S_1, \ldots, S_k$.
- $(N_{s+t})_{s \geq 0}$ is determined by $\tilde{S} = J_{n+1} - t = \sum_{i=1}^{k+1} S_i - t$, and $\tilde{S}_i = S_{k+i}$, $i \geq 2$.

Hence it is sufficient to show that conditional on $\{J_n \leq t < J_{n+1}\}$, the random variables $S_1, \ldots, S_k, \tilde{S}_1, \tilde{S}_2, \ldots$ are IID Exponential$(\lambda > 0)$ random variables.

First off, $\tilde{S}_i = S_{i+k}$, $i \geq 2$ are Exp(λ) independent of $S_1, \ldots, S_{k+1}$ (and hence of $\tilde{S}_1$).

Now, the joint density of $(S_1, \dots, S_{k+1})$ is just a product of $\mathrm{Exp}(\lambda)$ densities. Hence for any $a_1, \dots, a_k \geq 0$, $s > 0$,

$$\mathbb{P}(S_1 \leq a_1, \dots, S_k \leq a_k, \tilde{S}_1 > s, J_n \leq t < J_{n+1})$$

is

$$\int_0^{a_1} \dots \int_0^{a_k} \lambda e^{-\lambda s_1} \dots \lambda e^{-\lambda s_k} \mathbb{1}_{\sum_{i=1}^k s_i \leq t} \mathbb{P}(S_{k+1} > s + t - \sum_{i=1}^k s_i) ds_1 \dots ds_k.$$

$$= \exp(-\lambda s) \mathbb{P}(S_1 \leq a_1, \dots, a_k, \sum_{i=1}^k S_i \leq t, S_{k+1} > t - \sum_{i=1}^k s_i).$$

So $\tilde{S}_1$ is independent of $S_1, \dots, S_k$ and has $\mathrm{Exp}(\lambda)$ distribution. □

Theorem 2.9 (Characterisation of PP)**.** *Suppose $(N_t)_{t \geq 0}$ is a right cts stochastic process with state space $\mathbb{N}_0$, $N_0 = 0$ and $\lambda > 0$.*

TFAE

(a) *The holding times $S_n = J_n - J_{n-1}$, $n \in \mathbb{N}$ (where $J_0 = 0$, $J_{n+1} = \inf\{t \geq J_n : N_t \neq N_{J_n}\}$ are the jump times) are IID Exp(λ) random variables. Also, the jump chain $Y_n = X_{J_n}$ is given by $Y_n = n$, $n \in \mathbb{N}_0$.*

(b) *N has independent increments, i.e. for any $0 \leq s_1 \leq t_1 \dots$*

$$N_{t_k} - N_{s_k}, \quad k \in \mathbb{N}$$

are independent random variables. Also, uniformly in $t \geq 0$,

$$\mathbb{P}(N_{t+h} - N_t = 0) = 1 - \lambda h + o(h), \quad \mathbb{P}(N_{t+h} - N_t = 1) = \lambda h + o(h).$$

(c) *N has independent increments and $N_t - N_s$ has Poisson($\lambda(t-s)$) distribution for any $0 \leq s \leq t$.*

Proof. (a) implies (b): Suppose that (a) holds. Let $0 \leq s_1 \leq t_1 \leq s_2 \leq t_2 \leq \dots$.

Suppose for induction that $N_{t_k} - N_{s_k}$, $k = 1, \dots, n$ are independent.

By the Markov property Theorem 2.8, $(N_{s_{n+1}+t} - N_{s_{n+1}})_{t \geq 0}$ is independent of $(N_u)_{0 \leq u \leq s_{n+1}}$. In particular, $(N_{t_{n+1}} - N_{s_{n+1}})$ is independent of every $N_{t_k} - N_{s_k}$, $k = 1, \dots, n$.

□

Proof of Theorem 2.9 ctd. (a) implies (b) ctd. Yesterday: (a) implies N has independent increments by Theorem 2.8.

Let $t, h \geq 0$. Again by Theorem 2.8, $N_{t+h} - N_t$ has the same distribution has $N_h - N_0 = N_h$, and $\mathbb{P}(N_{t+h} - N_t = k) = \mathbb{P}(N_h = k)$ does not depend on t for any k.

Using the fact that J_1 has $\text{Exp}(\lambda)$ distribution,

$$\mathbb{P}(N_h \geq 1) = \mathbb{P}(J_1 \leq h) = 1 - \exp(-\lambda h) = \lambda h + o(h),$$

and

$$\mathbb{P}(N_h \geq 2) \leq \mathbb{P}(S_1, S_2 \leq h) = (\lambda h + o(h))^2 = o(h).$$

So,

$$\mathbb{P}(N_h = 1) = \mathbb{P}(N_h \geq 1) - \mathbb{P}(N_h \geq 2) = \lambda h + o(h).$$

Recall that (a) stipulates the jump chain is $Y_n = n$ for all n. Hence, $\{N_h = 0\} = \{N_h \geq 1\}^c$

$$\mathbb{P}(N_h = 0) = 1 - \mathbb{P}(N_h \geq 1) = 1 - \lambda h + o(h).$$

(b) implies (c): Suppose (b) holds. Then N has independent increments. It remains to show $N_t - N_s$ has $\text{Poisson}(\lambda(t-s))$ distribution for any $s, t \geq 0$.

The bounds in part (b) apply uniformly in t, so it is enough to show $N_t - N_0 \sim \text{Poisson}(\lambda t)$ for any $t \geq 0$.

Define $p_i(t) = \mathbb{P}(N_t = i)$ for any $t \geq 0$, $i \in \mathbb{N}_0$. Then, for $h \geq 0$,

$$p_j(t+h) = \sum_{i=0}^{\infty} \mathbb{P}(N_t - N_0 = i, N_{t+h} - N_t = j - i) \sum_{i=0}^{\infty} \mathbb{P}(N_t = i)\mathbb{P}(N_{t+h} - N_t = j - i).$$

Suppose $i \geq 1$ and use the infinitesimal probabilities from (b). The right hand above is

$$p_i(t)(1 - \lambda h + o(h)) + p_{i-1}(t)(\lambda h + o(h)) + \sum_{i \neq j, i \neq j-1} p_i(t)\mathbb{P}(N_{t+h} - N_t = j - i).$$

Using $p_i(t) \leq 1$,

$$0 \leq \sum_{i \neq j, i \neq j-1} p_i(t)\mathbb{P}(N_{t+h} - N_t = j - i) \leq \mathbb{P}(N_{t+h} - N_t \notin \{0, 1\}) = o(h).$$

Thus

$$p_j(t+h) - p_j(t) = -\lambda h p_i(t) + \lambda h p_{i-1}(t) + o(h), \tag{21}$$

and so

$$|p_j(t+h) - p_j(t)| \leq 2h\lambda + o(h).$$

Hence p_j is right continuous. Let $s \geq h \geq 0$ and take $t = s - h$. This gives

$$p_j(s) - p_j(s-h) = -\lambda h p_i(s-h) + \lambda h p_{i-1}(s-h) + o(h), \tag{22}$$

so p_j is also left continuous.

Dividing (21) and (22) by $h > 0$ gives

$$h^{-1}(p_j(t+h) - p_j(t)) = -\lambda p_i(t) + \lambda p_{i-1}(t) + o(h)/h$$

and

$$h^{-1}(p_j(s) - p_j(s-h)) = -\lambda p_i(s-h) + \lambda p_{i-1}(s-h) + o(h)/h.$$

Letting $h \searrow 0$ and using continuity shows for any $t \geq 0$, $j \geq 1$,

$$p_j'(t) = -\lambda p_j(t) + \lambda p_{i-1}(t) \tag{23}$$

(right sided derivative at zero).

For the case $j = 0$, using a similar argument

$$p_0(t+h) = p_0(t)(1 - \lambda h + o(h)) + o(h),$$

and so $p_0'(t) = -\lambda p_0(t)$.

The latter has solution $p_0(t) = Ae^{-\lambda t}$ for some constant A. Since $N_0 = 0$, $A = p_0(0) = 1$.

Suppose for induction that $p_k(t) = e^{-\lambda t}(\lambda t)^k/k!$ for some $k \in \mathbb{N}$. Using (23) with $j = k+1$,

$$\int_0^T \frac{d}{dt}(p_{k+1}(t)e^{\lambda t})dt = \int_0^T \frac{\lambda^{k+1}t^k}{k!}dt = \frac{(\lambda T)^{k+1}}{(k+1)!},\ T \geq 0.$$

The LHS is $p_{k+1}(T)e^{\lambda T} - p_{k+1}(0)e^0 = p_{k+1}(T)e^{\lambda T}$, and the induction step follows.

□

Proof of Theorem 2.9 conclusion. (c) implies (a): Suppose N is any right continuous process with $N_0 = 0$ satisfying (c). Then, for any $0 \leq t_1 \leq t_2 \leq \ldots$, $i_1, i_2, \ldots \in \mathbb{N}_0$ and $n \in \mathbb{N}$,

$$\mathbb{P}(N_{t_1} = i_1, \ldots, N_{t_n} = i_n) = f_{\lambda t_1}(i_1) \ldots f_{\lambda(t_n - t_{n-1})}(i_n - i_{n-1}),$$

where $f_\mu(k) = \exp(-\mu)\mu^k/k!$ ($\mu > 0$, $k \in \mathbb{N}_0$).

More to the point– the finite dimensional distributions of N are specified and hence the distribution of $(N_t)_{t\geq 0}$ is determined entirely. In turn, this means the (joint) distribution the jump times J_n, holding times S_{n+1} and jump chain Y_n, $n \in N_0$ is determined uniquely.

I.e. all processes satisfying (c) have the same holding time and jump chain distribution. But we constructed *a* right continuous process satisfying (a) and showed it satisfies (c). Therefore all processes satisfying (c) also satisfy (a). □

2.3 Markov chains on finite spaces

The Poisson process is a widely studied and used CTMC, but is rather dull. This section begins the theory for more general processes, starting with I finite (although all definitions apply for I countably infinite).

An important object is the matrix exponential. It has the following properties

Lemma 2.10. *Suppose I is finite and $M = (m_{ij})_{i,j\in I}$ any matrix with real entries.*

1. $\sum_{n=0}^{\infty} M^k/k!$ *converges component-wise. We denote the limit* e^M.
2. *If* $M' = (m'_{ij})_{i,j\in I}$ *commutes with* M *then* $e^{M+M'} = e^M e^{M'}$.
3. $(e^{tM})_{ij} = \sum_{k=0}^{\infty} t^k (M^k)_{ij}/k!$ *has infinite radius of convergence as a power series of* $t \in \mathbb{R}$.

Proof. Omitted. Optional exercise. □

Theorem 2.11. *Suppose I finite and $M = (m_{ij})_{i,j\in I}$. Let $P(t) = e^{tM}$ for all $t \geq 0$.*

1. $P(t+s) = P(t)P(s)$ *all* $s, t \geq 0$.
2. P *is the unique solution to*
$$\frac{d}{dt}P(t) = P(t)Q,\ t \geq 0,\ P(0) = I,$$
(the 'forwards equation').
3. P *is the unique solution to*
$$\frac{d}{dt}P(t) = QP(t),\ t \geq 0,\ P(0) = I,$$
(the 'backwards equation').

Proof. (1) This follows from Lemma 2.10 part 2.

(2) Using Lemma 2.10 part 3, the derivative of P can be found with term by term differentiation (result from Analysis II).

$$(\frac{d}{dt}P(t))_{ij} = \sum_{k=0}^{\infty} \frac{d}{dt} t^k (Q^k)_{ij}/k! = \sum_{k=1}^{\infty} t^{k-1}(Q^{k-1}Q)_{ij}/(k-1)!,$$

for any $i, j \in I$. Expanding the matrix product, the RHS is,

$$\sum_{k=1}^{\infty} t^{k-1} \sum_l (Q^{k-1})_{il} q_{lj}/(k-1)! = \sum_{l \in I} \left(\sum_{k=1}^{\infty} t^{k-1} \frac{(Q^{k-1})_{il}}{(k-1)!} \right) q_{lj} = (P(t)Q)_{ij},$$

where the exchange of sums is justified because I is finite. The initial condition $P(0) = I$ is immediate from the definition of $\exp(0Q)$.

Suppose $\tilde{P}$ is another solution to the forwards equation. Differentiating entry by entry,

$$\frac{d}{dt}(\tilde{P}(t)e^{-tQ}) = \tilde{P}'(t)e^{-tQ} + \tilde{P}(t)(-Qe^{-tQ}), \tag{24}$$

where $(d/dt)e^{-tQ} = -Qe^{-tQ}$ using the argument above.

If $\tilde{P}'(t) = \tilde{P}(t)Q$ then the RHS of (24) vanishes. If also $\tilde{P}(0) = I$, then $\tilde{P}(t)e^{-tQ} = I$ for all $t \geq 0$ is constant. Right multiplying by e^{tQ} and applying Lemma 2.10 part 2 again gives $\tilde{P}(t) = e^{tQ}$.

3. Similar to 2.

□

Definition 2.12 (Q-Matrix)**.** *Let I be countable. A matrix $Q = (q_{ij})_{i,j \in I}$ is called a Q-matrix if*

1. $q_{ij} \geq 0$ *for distinct* $i, j \in I$.
2. $\sum_j q_{ij} = 0$ *for any* $i \in I$.

Note part 2 of the definition implies $q_{ii} \leq 0$ for any $i \in I$.

In my opinion, calling a matrix object that is usually denoted Q a Q-matrix isn't so helpful. A more suggestive name is 'rate matrix', or generator. We'll see why in a moment.

Example 2.13. $I = \{1, 2, 3\}$,

$$Q = \begin{pmatrix} -3 & 2 & 1 \\ 1 & -2 & 1 \\ 0 & 1 & -1 \end{pmatrix}.$$

Looking ahead, we'll see this Q corresponds to a Markov chain that waits in state 1 for an Exp(3) units of time then jumps to 2 with probability $2/3$ and 3 otherwise. In state 2 it waits for an Exp(2) amount of time then jumps to 1 or 3, with equal probability. In state 3 it waits for Exp(1) then jumps to 2.

Recall from the discrete time section, a transition matrix $P = (p_ij)_{i,j \in I}$ satisfies $p_{ij} \geq 0$ and sum $\sum_k p_{ik} = 1$ for any $i, j \in I$.

Theorem 2.14. *Suppose I finite, $Q = (q_{ij})_{i,j \in I}$. Then Q is a Q-matrix if and only if $P(t) = e^{tQ}$ is a transition matrix for every $t \geq 0$.*

Proof. '⇒': Suppose Q is a Q-matrix and $i \in I$. Then $\sum_j q_{ij} = 0$. Further, Q^k also has zero row sums since

$$\sum_j q_{ij}^{(k+1)} = \sum_j (Q^k Q)_{ij} = \sum_j \sum_l Q^k_{il} q_{lj} = \sum_l q_{il}^{(k)} \sum_j q_{lj} = 0,$$

where $q_{ij}^{(k)} = (Q^k)_{ij}$.

But $P(t) = I + Q + Q^2/2 + \ldots$. So

$$\sum_j p_{ij}(t) = 1 + \sum_{k=1}^{\infty} \sum_j t^k q_{ij}^{(k)}/k! = 1.$$

□

Proof of 2.14 ctd. It remains to show $P(t)$ has non-negative entries. It is sufficient to show this is the case for t small enough, say $0 \le t \le T$. Then for arbitrary $s > 0$,

$$P(s) = P(s/n)^n$$

has non-negative entries once $n > s/T$ is big enough that $P(s/n)$ has all non-negative entries.

The diagonal entry $p_{ii}(t) = 1 + q_{ii}t + O(t^2) > 0$ for $t \ge 0$ small enough.

For the off diagonal entries, write $p_{ij}(t) = q_{ij}t + q_{ij}^{(2)}t^2/2 + \ldots = q_{ij}t + O(t^2)$.

In the case $q_{ij} > 0$ it is immediate that $p_{ij}(t) > 0$ for t small enough.

By definition of a Q-matrix, the only other possibility is $q_{ij} = 0$. In this case,

$$q_{ij}^{(2)} = (Q^2)_{ij} = \sum_l q_{il}q_{lj} \ge 0$$

as all summands are non-negative for $l \ne i, j$. If either $l = i$ or $l = j$, the summand involves $q_{ij} = 0$ and vanishes.

If $q_{ij}^{(2)} > 0$ then positivity of P follows. For $q_{ij}^{(2)} = 0$, repeat the argument. Generally, if $q_{ij}^{(n)} = 0$ and $q_{ij} = 0$ then $q_{ij}^{(n+1)} \ge 0$.

Hence either, $q_{ij}^{(n)} = 0$ for every n (in which case $p_{ij}(t) = 0$ for every $t \ge 0$), or $q_{ij}^{(1)} = 0, \ldots, q_{ij}^{(n-1)} = 0$ and $q_{ij}^{(n)} > 0$. In the latter case

$$p_{ij}(t) = q_{ij}^{(n)}t^n + O(t^{n+1}) > 0$$

for $t \ge 0$ small enough.

'$\Leftarrow$': Suppose $P(t) = e^{tQ}$ is a transition matrix for every $t \ge 0$. Using the forwards equation (Theorem 2.11),

$$\sum_j q_{ij} = \sum_j (QP(0))_j = \frac{d}{dt}\sum_j p_{ij}(t)\bigg|_{t=0} = 0$$

since $\sum_j p_{ij}(t) = 1$ is constant for any $i \in I$.

Also, if $q_{ij} < 0$ for distinct i, j then

$$p_{ij}(t) = q_{ij}t + O(t^2) < 0$$

for $t > 0$ small enough. Thus $p_{ij}(t) \ge 0$ for every $t \ge 0$ implies $q_{ij} \ge 0$.

□

Remark 2.15. *It is not true that every transition matrix P has the form e^Q! E.g.*

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

(not invertible).

2.4 Jump-hold construction of Markov chains on finite state space

Theorem 2.14 says a Q-matrix Q on finite I defines 'transition' probabilities via $P(t) = e^{tQ}$. This section concerns construction of a Markov chain X with $\mathbb{P}(X_t = j | X_0 = i) = (P(t))_{ij}$.

The first step is to define the embedded jump chain corresponding to Q.

Definition 2.16. *Let I be countable and $Q = (q_{ij})_{i,j \in I}$ a Q-matrix. The jump chain transition matrix, H say, of Q is given by*

$$H_{ij} = \begin{cases} -q_{ij}/q_{ii} & \text{if } q_{ii} < 0, \\ 0 & \text{if } q_{ii} = 0 \end{cases}$$

for distinct $i, j \in I$ and

$$H_{ii} = \begin{cases} 0 & \text{if } q_{ii} < 0, \\ 1 & \text{if } q_{ii} = 0. \end{cases}$$

Now, a useful Lemma which allows the holding times to be conveniently specified only using Exp(1) random variables.

Lemma 2.17. *Suppose T is Exp(1) distributed and $\lambda > 0$. Then $\lambda^{-1}T$ has Exp(λ) distribution.*

Proof. Exercise sheet 4. □

Here is some standard notation– $q(i) = q_i = -q_{ii} \geq 0$ is called the rate of leaving state $i \in I$ in Q-matrix $Q = (q_{ij})_{i,j \in I}$.

The rate of leaving is so called because it is the rate parameter for the Exponential holding time at the state.

Accordingly, and in light of Lemma 2.17, let $T_1, T_2, \dots$ be IID Exp(1) random variables and $(Y_n)_{n \in \mathbb{N}_0}$ a discrete time Markov chain with transition matrix H. Define

$$\begin{cases} S_n = q(Y_n)^{-1} T_n & \text{if } q(Y_{n-1}) > 0 \\ \infty & \text{otherwise,} \end{cases}$$

and $J_0 = 0$, $J_n = J_{n-1} + S_n$, $n \geq 1$.

A right continuous process with jump times J and jump chain Y may be defined via

$$X_t = \begin{cases} Y_n & \text{if } J_n \leq t < J_{n+1} \\ \partial & t \geq \sup_n J_n. \end{cases}$$

The latter case above is that X explodes. This does not happen a.s. on a finite state space (Exercise – wlog $0 < q(i) \leq C$ for all $i \in I$, using fact that I is finite, then apply proof of non-explosion of Poisson process with $\lambda = 1$).

Theorem 2.18. *Suppose I finite and $(X_t)_{t \geq 0}$ a right continuous process on I. Let Q be a Q-matrix. TFAE*

1. *(Jump-hold) Let* H *be the jump chain transition matrix of* Q*,* $\mu = (\mathbb{P}(X_0 = i))_{i \in I}$*, and* $J_n, n \in \mathbb{N}_0$ *the jump times of* X*.*

   The jump chain $Y = (X_{J_n})_{n \in \mathbb{N}_0}$ *of* X *is a* (μ, H)*-MC and the holding times* $S_n = J_n - J_{n-1}$*,* $n \in \mathbb{N}$ *are independent random variables with* $S_n \sim$ *Exp*$(q(Y_n))$*.*

2. *(Infinitesimal) For any* $t, h \geq 0$*,* $i, j \in I$*, and* $0 \leq t_1 \leq t_2 \leq \ldots \leq t_n \leq t$*,* $i_1, i_2, \ldots \in I$*,* $n \in \mathbb{N}$*,*

$$\mathbb{P}(X_{t+h} = j | X_t = i, X_{t_1} = i_1, \ldots, X_{t_n} = i_n) = \delta_{ij} + q_{ij}h + o(h)$$

   as $h \searrow 0$*.*

   (As usual it is implicit that $\mathbb{P}(X_{t_1} = i_1, \ldots, X_{t_n} = i_n) > 0$*).*

3. *(transition probabilities) For any* $t, s \geq 0$*,* $i, j \in I$*,* $0 \leq t_1 \leq t_2 \leq \ldots \leq t_n \leq t$*,* $i_1, i_2, \ldots \in I$ *and* $n \in \mathbb{N}$

$$\mathbb{P}(X_{t+s} = j | X_t = i, X_{t_1} = i_1, \ldots, X_{t_n} = i_n) = (P(s))_{ij},$$

   where $P(s) = e^{sQ}$*.*

A process satisfying any of these conditions is called a continuous time Markov chain (CTMC) with Q-matrix (or rates) Q.

Proof of Theorem 2.18 on Thursday.

Proof of Theorem 2.18 †. This is very similar to the proof of Theorem 2.9.

(1) implies (2):

Let X be as in (1). Then X is Markov, using an argument similar to that of Theorem 2.8. Here is a sketch proof.

Fix $i \in I$. Conditional on $\{X_t = i\}$ and $\{J_k \leq t < J_{k+1}\}$ $(k \in \mathbb{N}_0)$,

1. $(X_{t+s})_{s\geq 0}$ is determined entirely by $(Y_{k+m})_{m\in\mathbb{N}_0}$ and $\tilde{S}_1 = J_{k+1} - t$, $\tilde{S}_m = S_{k+m}, \ldots$, $m \geq 2$, and $(X_s)_{0\leq s\leq t}$ is determined by $S_1, \ldots, S_k$ and $(Y_m)_{0\leq m\leq k}$.
2. $\tilde{S}_1, \tilde{S}_2, \ldots$ are independent of $S_1, \ldots, S_k$, and $\tilde{S}_m$ has Exponential$(q(Y_{k+m-1}))$ distribution, $m \geq 1$.
3. $(Y_{k+m})_{m\in\mathbb{N}_0}$ is a (δ_i, H)-MC independent of $(Y_m)_{0\leq m\leq k}$ (Markov property of jump chain).

I.e. $\tilde{X}_s = X_{t+s}$, $s \geq 0$ has $\tilde{X}_0 = i$ and embedded jump chain $\tilde{Y}_m = Y_{k+m}$, $m \in \mathbb{N}_0$ with transition matrix H and holding times $\tilde{S}_m \sim$Exp$(q(\tilde{Y}_{m-1}))$, independently of $(X_s)_{0\leq s\leq t}$.

So for any $t, h \geq 0$, $i, j \in I$, $0 \leq t_1 \leq t_2 \leq \ldots \leq t_n \leq t$ and $i_1, i_2, \ldots \in I$,

$$\mathbb{P}(X_{t+h} = j | X_t = i, X_{t_1} = i_1, \ldots, X_{t_n} = i_n) = \mathbb{P}(X_h = j | X_0 = i),$$

and the RHS does not depend on $t \geq 0$.

For $i = j$,

$$\mathbb{P}(X_h = j | X_0 = i) = \mathbb{P}(J_1 > h | X_0 = i) = e^{-hq(i)} = 1 + hq_{ii} + o(h),$$

since $J_1 \sim$Exp$(q(i))$ when $Y_0 = X_0 = i$.

For $i \neq j$,

$$\mathbb{P}(X_h = j | X_0 = i) \geq \mathbb{P}(Y_1 = j, J_1 \leq h, J_2 > h | Y_0 = i) \geq \mathbb{P}(Y_1 = j, S_1 \leq h, S_2 > h | Y_0 = i).$$

The RHS is

$$h_{ij}(1 - e^{-q_i h})e^{-q_j h} = (q_{ij}/q_i)(1 - 1 + q_i h + o(h))(1 - q_j h + o(h)) = q_{ij} h + o(h).$$

Thus $\mathbb{P}(X_h = j | X_0 = i) \geq \delta_{ij} + q_{ij} h + o(h)$ for any i, j with equality if $j = i$.

It remains to show equality when $j \neq i$. Suppose $\alpha > 0$ and

$$\mathbb{P}(X_h = j | X_0 = i) \geq (\alpha + q_{ij})h + o(h)$$

as $h \to 0$.

For any $h > 0$,

$$1 = \sum_{k\in I} \mathbb{P}(X_h = k | X_0 = i) \geq 1 + \sum_k q_{ik} h + \alpha h + o(h),$$

which shows, after using $\sum_k q_{ik} = 0$, that

$$\alpha h + o(h) < 0$$

for all $h > 0$, giving a contradiction. It follows that

$$q_{ij}h + o(h) \leq \mathbb{P}(X_h = j|X_0 = i) < (\alpha + q_{ij})h + o(h)$$

for any $\alpha > 0$ and (2) follows.

(2) implies (3). The tactic is to show the forwards equation holds.

Define $p_{ij}(t) = \mathbb{P}(X_s = j|X_0 = i)$ for any $t \geq 0$, $i, j \in I$.

Suppose (2) holds and $h \geq 0$. Then

$$\begin{aligned} p_{ij}(t+h) = \sum_{k \in I} \mathbb{P}(X_t = k, X_{t+h} = j|X_0 = i) &= \sum_{k \in I} \mathbb{P}(X_t = k|X_0 = i)\mathbb{P}(X_{t+h} = j|X_t = k, X_0 = i) \\ &= \sum_k p_{ik}(t)(\delta_{kj} + q_{kj}h) + o(h) \\ &= p_{ij}(t) + h\sum_k p_{ik}(t)q_{kj} + o(h), \end{aligned}$$

uniformly in $t \geq 0$.

Hence,

$$p_{ij}(t+h) - p_{ij}(t) = h\sum_k p_{ik}(t)q_{kj} + o(h).$$

It follows p_{ij} is right continuous and right differentiable–

$$(p_{ij}(t+h) - p_{ij}(t))/h = \sum_k p_{ik}(t)q_{kj} + o(h)/h \to \sum_k p_{ik}(t)q_{kj}$$

as $h \searrow 0$.

Taking $s \geq h > 0$ and $t = s - h$, we see p_{ij} is left continuous and

$$(p_{ij}(s) - p_{ij}(s-h))/h = \sum_k p_{ik}(s-h)q_{kj} + o(h)/h \to \sum_k p_{ik}(s)q_{kj}$$

as $h \searrow 0$ using left continuity.

So p_{ij} is differentiable and satisfies

$$(d/dt)p_{ij}(t) = \sum_k p_{ik}(t)q_{kj} = (P(t)Q)_{ij},\ p_{ij}(0) = \delta_{ij}$$

where $P(t) = (p_{lk}(t))_{l,k \in I}$.

But the unique solution to this equation is $P(t) = e^{tQ}$ by Theorem 2.11.

Now, suppose $p_{ij}(s) = \mathbb{P}(X_{t+s} = j|X_t = i, X_{t_1} = i_1, \ldots, X_{t_n} = i_n)$ for an arbitrary sequence of states $i_1, i_2, \ldots$ and times $0 \leq t_1 \leq \ldots \leq t$.

The same argument works, since, for any $h > 0$,

$$\begin{aligned} p_{ij}(s+h) &= \sum_{k\in I} \mathbb{P}(X_{t+s+h} = j, X_{t+s} = k | X_t = i, X_{t_1} = i_1, \ldots, X_{t_n} = i_n) \\ &= \sum_{k\in I} \mathbb{P}(X_{t+s+h} = j | X_{t+s} = k, X_t = i, X_{t_1} = i_1, \ldots, X_{t_n} = i_n) \times \\ &\quad \mathbb{P}(X_{t+s} = k | X_t = i, X_{t_1} = i_1, \ldots, X_{t_n} = i_n) \\ &= \sum_{k\in I} (\delta_{jk} + q_{kj} h) p_{ik}(s) + o(h). \end{aligned}$$

(3) implies (2): this is exactly the same argument as in Theorem 2.9 – (3) determines the finite dimensional distributions. These specify the right continuous X entirely, and in particular determine the holding times and jump chain. But the process from (1) satisfies (3). Thus any right continuous process satisfying (3) also satisfies (1). □

2.5 Markov chains on countable state spaces

In the previous section the fact that the state space was finite helped in several places where sums had to be exchanged etc.

The case of a countably infinite state space requires a bit more work.

Due to time constraints we will only state results here.

As before the chain will be specified by a Q-matrix. It is no longer the case that we can simply write down the exponential of the Q-matrix to get the transition probabilities. For example, if

$$Q = \begin{pmatrix} -1 & 1/2 & 1/4 & 1/8 & \dots \\ 2 & -2 & 0 & 0 & \dots \\ 4 & 0 & -4 & 0 & \dots \\ 8 & 0 & 0 & -8 & \dots \end{pmatrix}$$

then $(Q^2)_{11} = (-1)^2 + 2(1/2) + \ldots = \infty$.

However, the transition probabilities do solve the forwards and backwards equations. This is an countable system of differential equations and it is not trivial that they have a solution.

This is the subject of the next Theorem, which is the analogue of 2.11.

Theorem 2.19. *Let I be countable and $Q = (q_{ij})_{i,j\in I}$ a Q-matrix. Then*

1. *The backwards equation*
$$P'(t) = QP(t), t \geq 0, \quad P(0) = I$$
has a minimal non-negative solution. I.e., there are non-negative differentiable $(p_{ij}(t))_{i,j\in I}$, $t \geq 0$ such that
$$p'_{ij}(t) = \sum_k q_{ik}p_{kj}(t), \; t \geq 0, \; p_{ij}(0) = \delta_{ij}, \quad i, j \in I,$$
and if $\tilde{p}$ is another non-negative solution then $p_{ij}(t) \leq \tilde{p}_{ij}(t)$ for every $i, j \in I$, $t \geq 0$.

2. *$P(t) = (p_{ij}(t))_{i,j} \in I$ satisfies $\sum_j p_{ij}(t) \leq 1$ for any $i \in I$ and $t \geq 0$. Also,*
$$P(t + s) = P(t)P(s), \; s, t \geq 0,$$
(the 'semigroup' property of Theorem 2.11).

3. *P is also the minimal non-negative solution to the forwards equation $P'(t) = P(t)Q$, $t \geq 0$, $P(0) = I$.*

P is called the semigroup of Q.

The p_{ij} from the last Theorem can be thought as transition probabilities for a Markov chain. The possibility that $\sum_j p_{ij}(t) < 1$ for some $i \in I$ corresponds to mass disappearing when the chain starts at i– it explodes and burns up. Indeed some authors *define* explosion in this way.

One way to accommodate the situation is to add an absorbing cemetery site $\partial \notin I$ for explosion debris, i.e. take $\bar{I} = I \cup \{\partial\}$, and extend p to $\bar{I}$ via

$$p_{i\partial}(t) = 1 - \sum_j p_{ij}(t), \; p_{\partial\partial}(t) = 1.$$

Then $\sum_{j \in \bar{I}} p_{ij} = 1$ for any $i \in \bar{I}$.

Modifying p in this way yields a minimal chain.

[See 2.1.1 for reminder of definition of explosion and minimal].

Theorem 2.20. *Suppose $I \not\ni \partial$ countable and $Q = (q_{ij})_{i,j \in I}$ a Q-matrix. Suppose $(X_t)_{t \geq 0}$ is a right continuous minimal process on $\bar{I} = I \cup \{\partial\}$. TFAE*

1. *(Jump-hold) The embedded jump chain $Y_n = X_{J_n}, n \geq 0$ of X is a discrete time MC with transition matrix H (as defined in 2.16).*

   The holding times $S_n = J_n - J_{n-1}$, $n \geq 1$ are independent random variables with $S_n \sim$Exp$(q(Y_{n-1}))$.

2. *(Transition probabilities) Let $P = ((p_{ij}(t))_{i,j \in I}, t \geq 0)$ be the minimal solution to the forwards/backwards equation from Theorem 2.19*

   For any $s, t \geq 0$, $i, j \in I$,

   $$\mathbb{P}(X_{s+t} = j | X_t = i, X_{t_1} = i_1, \ldots, X_{t_n} = i_n) = p_{ij}(s),$$

   where $i_1, i_2, \ldots \in I$, $0 \leq t_1 \leq t_2 \leq \ldots \leq t$ has $\mathbb{P}(X_t = i, X_{t_1} = i_1, \ldots, X_{t_n} = i_n) > 0$.

Definition 2.21. *Suppose Q and X are as in Theorem 2.20. If either (1) or (2) hold, X is called a continuous time Markov chain with Q-matrix (or rates, or generator) Q and initial distribution $\lambda = (\mathbb{P}(X_0 = i))_{i \in I}$.*

Write 'X a (λ, Q)-CTMC' for short.

The previous two Theorems are usually proved together. You begin from the jump hold construction in 2.20(1), show that it gives a Markov chain, then that its transition probabilities satisfy the backwards equation, then show minimality. Finally, you show that the minimal solution to the backwards equation is also the minimal solution for the forwards equation. Interested students can consult Norris §2.8.

The jump hold construction from Exponential random variables and a discrete time Markov chain gives existence of arbitrary (λ, Q)-CTMCs.

As before, $\mathbb{P}_i$ denotes the measure under which the process starts from $i \in I$.

2.6 Hitting probabilities and times

CTMCs can be analysed by applying results from §1 to the embedded jump chain. The first example of this concerns hitting probabilities.

Theorem 2.22. *Suppose X is a CTMC with Q-matrix Q and $A \subset I$. Let*

$$D^A = \inf\{t \geq 0 : X_t \in A\}$$

be the hitting of A. Then the hitting probabilities $h_i^A = \mathbb{P}_i(D^A < \infty)$ form the minimal non-negative solution to $h_i^A = 1$ for $i \in A$,

$$h_i^A = \sum_{j \neq i} \frac{q_{ij}}{q_i} h_j^A, \ i \notin A, \ q_i > 0$$

and $h_i^A = 0$ if $i \notin A$ and $q_i = 0$.

Proof. X is minimal so $\{D^A < \infty\} = \{H^A < \infty\}$, where $H^A = \inf\{n \in \mathbb{N}_0 : Y_n \in A\}$ is the hitting time for the embedded jump chain of X.

Hence $h_i^A = \mathbb{P}_i(H^A < \infty)$ and the result follows from Theorem 1.15. □

The expected hitting times are given by the expected time holding time for the starting state plus everything thereafter. The Markov property of the embedded jump chain makes the proof straightforward.

Theorem 2.23. *Suppose X is a CTMC with Q-matrix Q, $A \subset I$ and*

$$D^A = \inf\{t \geq 0 : X_t \in A\}$$

as above.

Then the expected hitting times $k_i^A = \mathbb{E}_i[D^A]$ form the minimal non-negative solution to $k_i^A = 0$ for $i \in A$,

$$k_i^A = q_i^{-1} + \sum_{j \neq i} \frac{q_{ij}}{q_i} k_j^A, \ i \notin A, \ q_i > 0$$

and $k_i^A = \infty$ if $i \notin A$ and $q_i = 0$.

Proof. The non-trivial case is $i \notin A$, $q_i > 0$. Then $D^A > J_1$, the first jump time. Conditioning on the first jump of Y,

$$k_i^A = \mathbb{E}_i[J_1] + \mathbb{E}_i[D^A - J_1] = q_i^{-1} + \sum_j (q_{ij}/q_i)\,\mathbb{E}_i[D^A - J_1 | Y_1 = j].$$

$D^A - J_1 = \sum_{m=1}^{H^A} q(Y_m)T_m$ (where $T_1, T_2, \ldots$ are IID Exp(1)) depends only on Y through Y_m, $m \geq 1$, and so the Markov property of Y, the RHS above becomes

$$= q_i^{-1} + \sum_j (q_{ij}/q_i)\,\mathbb{E}_j[D^A].$$

The result follows. Proof of minimality omitted. □

2.7 Recurrence and transience

Definition 2.24. *Suppose X a CTMC. State $i \in I$ is called recurrent if*

$$\mathbb{P}_i(\sup\{t \geq 0 : X_t = i\} = \infty) = 1,$$

i.e. with probability one, X_t visits i at arbitrarily large times.

State i is called transient if the probability above is zero.

Transience/recurrence of a state is a dichotomy and corresponds exactly to the jump chain.

Theorem 2.25. *Suppose X a CTMC. State $i \in I$ is transient (recurrent) for X if it is transient (recurrent) in the embedded jump chain of X.*

In particular every state is recurrent or transient.

Proof. As usual let $Y_n = X_{J_n}, n \in \mathbb{N}_0$ be the embedded jump chain.

Suppose $i \in I$, $X_0 = Y_0 = i$ and i recurrent for Y. Then (exercise sheet 4), X is non-explosive, i.e. $J_n \nearrow \infty$ as $n \to \infty$.

Also, recurrence implies that, almost surely, for every N there exists $n > N$ such that $Y_n = i$. But $Y_n = i$ implies $X_{J_n} = i$. Hence

$$\sup\{t \geq 0 : X_t = i\} \geq J_n \geq J_N \nearrow \infty$$

as $N \to \infty$.

Suppose state i is transient for Y. Then there exists N such that $Y_n \neq i$ for $n \geq N$.

It is possible that Y_N is at an absorbing state. Without loss of generality suppose Y is not absorbed at time $N-1$. Then $q(Y_i) > 0$ for $0 \leq i \leq N-1$ and consequently $J_N < \infty$.

Thus,

$$\sup\{t \geq 0 : X_t = i\} \leq J_N < \infty.$$

□

Definition 2.26. *A CTMC is called recurrent if all states are recurrent.*

Definition 2.27. *A CTMC X on I is called irreducible if, for every $i, j \in I$, $\mathbb{P}_i(X_t = i) > 0$ for some $t > 0$.*

Theorem 2.28. *Let X be a CTMC X. TFAE*

1. *X is irreducible.*
2. *the embedded jump chain Y is irreducible.*
3. *$\mathbb{P}_i(X_t = i) > 0$ for every $t > 0$.*

Proof. (3) implies (1): by definition.

(1) implies (2): Let $i, j \in I$. If X is irreducible, then for some $t > 0$,

$$0 < \mathbb{P}_i(X_t = j) = \sum_{n=0}^{\infty} \mathbb{P}_i(J_n \leq t < J_{n+1}, X_t = Y_n = j) \leq \sum_{n=0}^{\infty} \mathbb{P}_i(Y_n = j).$$

So $\mathbb{P}_i(Y_n = j) > 0$ for some $n \in \mathbb{N}_0$.

(2) implies (3): Suppose $t > 0$, $i, j \in I$ are distinct. Let Q denote the Q-matrix of X. Recall that the transition matrix of Y is $H = (h_{ij})_{i,j \in I}$ with $h_{ij} = q_{ij}/q_i$ when $i \neq j$ for i non-absorbing (see 2.16).

If Y is irreducible, all states are non-absorbing so there exists a sequence of distinct states $i_0, i_1, \ldots, i_n \in I$ with $i_0 = i$, $i_n = j$ and

$$q_{i_0,i_1} \cdots q_{i_{n-1},i_n} > 0.$$

For any pair, i_m, i_{m+1}, we have

$$p_{i_m,i_{m+1}}(t) = \mathbb{P}_{i_m}(X_t = i_{m+1}) \geq \mathbb{P}_{i_m}(J_1 \leq t < J_2, Y_1 = i_{m+1}) \geq (q_{i_m,i_{m+1}}/q_{i_m})(1 - e^{-q_{i_m}t})e^{-q_{i_{m+1}}t} > 0.$$

Finally,

$$\mathbb{P}_i(X_t = j) \geq p_{i_0,i_1}(t/n)p_{i_1,i_2}(t/n)\dots p_{i_{n-1},i_n}(t/n) > 0.$$

Also,

$$\mathbb{P}_i(X_t = i) \geq \mathbb{P}_i(J_1 > t) = e^{-q_i t} > 0.$$

□

2.8 Invariant measures

Definition 2.29. *Suppose X is a non-explosive CTMC with Q-matrix Q, and corresponding semigroup $P(t)$, $t \geq 0$.*

A measure $\lambda = (\lambda_i)_{i\in I}$ is called invariant for X (or Q, or P) if

$$\lambda P(t) = \lambda,\ t \geq 0.$$

If $\sum_i \lambda_i = 1$ then λ is called an invariant probability measure, or distribution.

Note the definition only applies to non-explosive chains.

Theorem 2.30. *Suppose X is a irreducible and recurrent CTMC with Q-matrix Q. Then λ is invariant iff $\lambda Q = 0$.*

Proof of Theorem 2.30. Due to time constraints we consider only I finite.

Using the backwards equation,

$$(d/dt)(\lambda P(t)) = \lambda(d/dt)P(t) = \lambda QP(t).$$

So $\lambda Q = 0$ implies $\lambda P(t) = \lambda P(0) = \lambda$ for any $t \geq 0$.

On the other hand, $\lambda P(t) = \lambda$ for every $t \geq 0$ implies

$$0 = (d/dt)(\lambda P(t))\big|_{t=0} = \lambda QP(t)\big|_{t=0} = \lambda Q.$$

Lemma 2.31. *Suppose that Q is a Q-matrix, H the corresponding jump chain transition matrix.*

1. *For any measure λ on I, $\lambda Q = 0$ iff $\mu H = H$, where $\mu_i = q_i \lambda_i$ for every $i \in I$.*
2. *Suppose Q is irreducible and recurrent, and λ, λ' are any non-negative non-zero measures with $\lambda' Q = 0$, $\lambda Q = 0$. Then $\lambda' = c\lambda$ for some $c > 0$.*

Proof. (1) $\mu H = \mu$ iff $\mu(H - I) = 0$ iff

$$\sum_i \mu_i (H - I)_{ij} = 0$$

for any $i \in I$. Using $\mu_i = \lambda_i q_i$ and $q_i(h_{ij} - \delta_{ij}) = q_{ij}$ for any $i, j \in I$ the latter holds iff

$$\sum_i \lambda_i q_{ij} = 0$$

for all i. That is, $\lambda Q = 0$.

(2) Suppose λ' a non-zero measure with $\lambda' Q = 0$. Then part (1) implies $\mu_i = q_i \lambda_i$ and $\mu'_i = q_i \lambda'_i$ are both invariant for H.

The embedded jump chain is recurrent and irreducible (Theorem 2.25, 2.28). Hence Theorem 1.41 [all non-negative solutions to $\mu H = \mu$ are all proportional] implies $\mu' = c\mu$ for some $c > 0$

I.e.,

$$\mu'_i = q_i \lambda'_i = c\mu_i = cq_i \lambda_i$$

for every $i \in I$.

Ignoring the trivial case $I = \{i\}$, $q_i > 0$ for every i. The result follows. □

□

Definition 2.32. *Suppose X a CTMC, $i \in I$ is recurrent, and*

$$T_i = \inf\{t \geq J_1 : X_t = i\}$$

is the first return time to i.

State i is called positive recurrent if $\mathbb{E}_i[T_i] < \infty$ and null recurrent if $\mathbb{E}_i[T_i] = \infty$.

The continuous time analogue of Theorem 1.44 is–

Theorem 2.33. *Suppose Q irreducible. TFAE*

1. *All states are positive recurrent.*
2. *There exists a positive recurrent state.*
3. *Q is non-explosive and there exists a probability measure with $\lambda Q = 0$.*

Remark 2.34. *if (iii) applies then all states are positive recurrent by (i). Hence Theorem 2.30 applies, and λ is a bona-fide invariant distribution.*

Proof. Later, if time. □

2.9 Reversibility, detailed balance

The last topic is reversibility. There is a small technical gotcha when looking a CTMC going backwards in time– it is only *left* continuous at jumps. We agree to ignore this issue.

Definition 2.35. *A CTMC X is called reversible if $(X_t)_{0\leq t\leq T}$ has the same distribution as $(X_{T-t})_{0\leq t\leq T}$ for any $T > 0$.*

I.e. the finite dimensional distributions are equal– for any $0 \leq t_1 \leq t_2 \leq \ldots$, and $i_1, i_2, \ldots \in I$, $n \in \mathbb{N}$,

$$\mathbb{P}(X_{t_1} = i_1, \ldots, X_{t_n} = i_n) = \mathbb{P}(X_{t_1} = i_n, \ldots, X_{t_n} = i_1).$$

Note– a reversible CTMC is necessarily non-explosive according to this definition, assuming minimality.

Theorem 2.36. *Suppose Q is a irreducible, non-explosive Q-matrix and λ is a probability measure. Let X be a (λ, Q)-CTMC. TFAE,*

1. *X is reversible.*
2. *Q and λ satisfy the detailed balance equations*

   $$\lambda_i q_{ij} = \lambda_j q_{ji},$$

   for any $i, j \in I$.
3. *λ is invariant for Q.*

Proof. (2) implies (3): using the detailed balance equations,

$$(\lambda Q)_i = \sum_{j \in I} \lambda_j q_{ji} = \lambda_i \sum_{j \in I} q_{ij} = 0.$$

Now apply Theorem 2.30. □

Example 2.37. *The last Theorem can be used to prove explosion.*

Consider a CTMC on $I = \mathbb{N}_0$ with transition rates given by

$$q_{i,i+1} = pq_i, \; q_{i,i-1} = (1-p)q_i$$

$p = 2/3$, $q_i = i!$, *for* $i \geq 1$. *Take* $q_{01} = q_0 > 0$.

The chain is irreducible. The DBEs are

$$\pi_i p q_i = \pi_{i+1}(1-p)q_{i+1}, \; i \geq 1$$

and $\pi_0 q_0 = \pi_1(1-p)q_1$.

So, for $i \geq 1$,

$$\pi_{i+1} = \frac{\pi_i p q_i}{(1-p)q_{i+1}} = \pi_{i-1}\left(\frac{p}{1-p}\right)^2 \frac{q_i q_{i-1}}{q_{i+1}q_i} = \ldots = A(p/(1-p))^{i+1} q_{i+1}^{-1}$$

for some $A > 0$.

Hence, $\sum_i \pi_i < \infty$ *iff*

$$\sum_i (p/(1-p))^i q_i^{-1} < \infty.$$

But with our choice of p *and* q_i, *the summand is* $2^i/i!$ *and the series converges to* e^2.

If the chain is non-explosive then Theorem 2.36 shows that π *is an invariant distribution. In particular the chain is recurrent. But the embedded jump chain is transient!*

Conclusion– the chain is explosive.

3 Percolation

3.1 Introduction

The final two weeks of the course concerns a simple model for connectivity in a random graph.

Real world applications of the theory include oil seeping through a porous rock, friend connections in social networks and electrical conduction in metals with impurities.

3.2 Basic definitions and preliminary results

As in the previous section, we study a countable family of random variables.

The difference here is that the random variables are no longer indexed by time but by edges in an undirected graph $G = (V, \mathcal{E})$. Here, V is a set of vertices and $\mathcal{E} \subset V \times V$ are edges. Undirected-ness means $(u, v) \in \mathcal{E}$ iff $(v, u) \in \mathcal{E}$.

Each random variable indicates the connectivity of the corresponding edge.

In this course, edges are either open or closed, independently of all the other edges.

The graph considered here the integer lattice with nearest neighbour edges – $V = \mathbb{Z}^d$ and $(u, v) \in \mathcal{E} = \mathcal{E}(\mathbb{Z}^d)$ iff $|u - v| = 1$.

By default all random variables live on a probability space denoted $(\Omega, \mathcal{F}, \mathbb{P})$.

It will be convenient to take

- $\Omega = \{0, 1\}^{\mathcal{E}}$. An element $\omega \in \Omega$ is called an edge configuration– $\omega_e = 0$ means edge $e \in \mathcal{E}$ is closed and $\omega_e = 1$ means the edge is open.
- $\mathcal{F}$ is the σ-algebra of Ω generated by subsets of the form
$$\{\omega \in \Omega : \omega_{e_1} = i_1, \ldots, \omega_{e_n} = i_n\},$$
for $e_1, e_2, \ldots, e_n \in E$ and $i_1, \ldots, i_n \in \{0, 1\}$.

  These are called cylinder sets.
- The probability measure is defined by letting each edge be open with probability $p \in [0, 1]$, independently of the others.

  Note it is not possible to define the infinite product measure directly on $\mathcal{F}$ (any event $\omega \in \Omega$ has probability zero).

  However it is sufficient to define the measure on 'cylinder sets';
$$\mathbb{P}(\{\omega \in \Omega : \omega_{e_1} = i_1, \ldots, \omega_{e_n} = i_n\}) = \prod_{m=1}^{n} p^{i_m}(1 - p)^{1 - i_m}.$$

A variant of the Kolmogorov extension Theorem guarantees there is a unique probability measure $\mathbb{P}$ on $(\Omega, \mathcal{F})$ with this property.

Subscripts may be added to $\mathbb{P}$ to explicitly indicate the value of p and the underlying graph.

Definition 3.1. *Suppose $x, y \in \mathbb{Z}^d$ and $\omega \in \Omega$. Then x is connected to y (written $x \leftrightarrow y$) if there is a path of open edges in ω from x to y.*

I.e. there exists a sequence of edges $(a_1, b_1), \ldots, (a_n, b_n) \in \mathcal{E}$ with

- *$a_1 = x$, $b_m = a_{m+1}$ for $1 \le m \le n-1$ and $b_n = y$ and*
- *$\omega_{(a_m, b_m)} = 1$ for $1 \le m \le n$.*

Write $0 \leftrightarrow \infty$ if, for every $N \in \mathbb{N}$, there exists x with $|x| > N$ and $0 \leftrightarrow x$.

Definition 3.2. *The percolation probability is $\theta(p) = \mathbb{P}_p(0 \leftrightarrow \infty)$.*

When $p = 0$ all edges are closed $\theta(0) = 0$, while $p = 1$ means all edges are open so $\theta(1) = 1$.

Exercise sheet 5: when $d = 1$, $\theta(p) = 0$ for $0 \le p < 1$.

Lemma 3.3. *$\theta(p)$ is increasing in p.*

The proof uses a coupling argument.

Proof. Let U_e, $e \in \mathcal{E}$ be IID Uniform[0,1] random variables defined on a space $(\tilde{\Omega}, \tilde{\mathcal{F}}, \tilde{\mathbb{P}})$. Then $\omega^p = \mathbb{1}_{U_e \le p}$, $e \in \mathcal{E}$ has the same distribution as ω_e, $e \in \mathcal{E}$ under $\mathbb{P}_p$ and for any x, y,

$$\tilde{\mathbb{P}}(x \overset{\omega^p}{\leftrightarrow} y) = \mathbb{P}_p(x \leftrightarrow y),$$

where $\overset{\omega^p}{\leftrightarrow}$ means existence of a path using open edges in ω^p.

Suppose $p' \ge p$. Then $\omega_e^p \le \omega_e^{p'}$ for any edge e. So, any open path from $x \in \mathbb{Z}^d$ to $y \in \mathbb{Z}^d$ using open edges in ω^p the same path is open in $\omega^{p'}$.

In particular $0 \overset{\omega^p}{\leftrightarrow} \infty$ implies $0 \overset{\omega^{p'}}{\leftrightarrow} \infty$ and

$$\theta(p) = \tilde{\mathbb{P}}(0 \overset{\omega^p}{\leftrightarrow} \infty) \le \tilde{\mathbb{P}}(0 \overset{\omega^{p'}}{\leftrightarrow} \infty) = \theta(p').$$

□

The main question of percolation theory: what is the smallest p for which $\theta(p) > 0$?

Definition 3.4. *The critical probability is $p_c = \sup\{p \ge 0 : \theta(p) = 0\}$.*

By exercise, $p_c(d = 1) = 1$. This is rather boring. Let us rule out this possiblity for $d \ge 2$.

Theorem 3.5. *For $d \ge 2$,*

$$(2d-1)^{-1} \le p_c(d) \le p_c(2) \le 7/8.$$

Definition 3.6. *A nearest neighbour path* $(x_0, x_1, \dots, x_n) \in (\mathbb{Z}^d)^{n+1}$ *of length* n *is called a self avoiding walk if* $x_0, \dots, x_n$ *are all distinct.*

Let SAW(n) denote the set of such paths starting from the origin.

Lemma 3.7. *The number* $r_d(n) = |SAW(n)|$ *of self avoiding walks with length* n *satisfies* $r_d(n) \leq 2d(2d-1)^{n-1}$ *for any* $n \geq 1$.

Proof. Firstly, $r_d(1) = 2d$. Let $x = (x_0, \dots, x_{n+1}) \in \text{SAW}(n+1)$. Then $x' = (x_0, \dots, x_n) \in \text{SAW}(n)$ and $x_{n+1} \neq x_{n-1}$.

There are $r_d(n)$ such x' and $2d-1$ possible x_{n+1}, i.e. $r_d(n+1) \leq r_d(n)(2d-1)$ for any $n \geq 1$. □

Proof of Theorem 3.5 lower bound. We must show that $\theta_d(p) = 0$ for $p < (2d-1)^{-1}$.

Suppose that $0 \leftrightarrow \infty$. Then by definition, for any $n \in \mathbb{N}$ there is an $x \in \mathbb{Z}^d$ with $|x| > n$ connected to the origin by an open path.

Erasing any loops of this open path from 0 to x gives a self avoiding walk starting from the origin of length at least n.

Any path of length m has probability p^m of being open under $\mathbb{P}_p$, $p > 0$.

So, with $r_d(m) = |\text{SAW}(m)|$,

$$\theta(p) = \mathbb{P}(0 \leftrightarrow \infty) \leq \sum_{m=n}^{\infty} p^m r_d(m) \leq (2d)p \sum_{m=n}^{\infty} (p(2d-1))^m \to 0$$

as $n \to \infty$ if $p(2d-1) < 1$. □

Exercise: use a coupling argument to show $p_c(d) \leq p_c(2)$.

Suppose $d = 2$. We must show $\theta(p) > 0$ for $p \geq 7/8$. Again it is helpful to count paths. However, this time the paths are in the 'dual' of $\mathbb{Z}^2$.

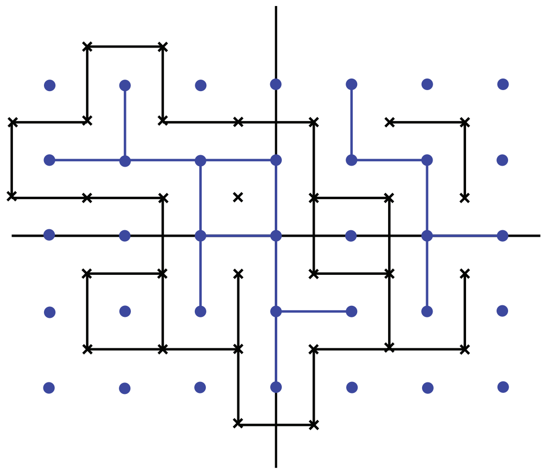
Definition 3.8. *The dual graph* $\mathbb{Z}^2_\star$ *of* $\mathbb{Z}^2$ *is is the graph with vertices* $\mathbb{Z}^2 + (1/2, 1/2)$ *and nearest neighbour edges* $\mathcal{E}(\mathbb{Z}^2_\star)$ *(i.e.* $(x, y) \in \mathcal{E}(\mathbb{Z}^2_\star)$ *iff* $|x-y| = 1$*).*

$\mathbb{Z}^2$ *is called the primal graph.*

Thus each dual edge crosses exactly one primal edge. For $e \in \mathcal{E}(\mathbb{Z}^2)$ *write* $e^\star$ *for the dual edge crossed by* e.

For any edge configuration $\omega \in \Omega$ *in the primal graph* $\mathbb{Z}^2$ *define the dual edge configuration* $\omega^\star \in \{0,1\}^{\mathbb{Z}^2_\star}$ *by* $\omega^\star_{e^\star} = 1$ *iff* $\omega_e = 0$.

Definition 3.9. *A nearest neighbour path* $(x_0, \dots, x_n) \in \mathbb{Z}^2_\star$ *of length* n *is called a simple loop if* $x_0 = x_n$ *and* $x_1, \dots, x_n$ *are distinct.*



Let $C_0 = \{x \in \mathbb{Z}^2 : 0 \leftrightarrow x\}$ be the connected component containing the origin.

Note that $0 \leftrightarrow \infty$ iff $|C_0| = \infty$.

The external boundary of $C_0 \subset \mathbb{Z}^2$ is a set of open edges in the dual graph $Z^2_\star$.

The following Lemma is considered 'obvious by drawing a picture' –

Lemma 3.10. *If $C \subset \mathbb{Z}^2$ is a finite connected subgraph of $\mathbb{Z}^2$ then the external boundary of C forms a simple loop in $\mathbb{Z}^2_\star$.*

For proof see Bollobas and Riordan Lemma 1.

Proof of Theorem 3.5 upper bound. We show $\theta(p) > 0$ for $p \geq 7/8$. By previous Lemma, $|C_0| < \infty$ implies that in $\mathbb{Z}^2_\star$ there is an open simple loop containing the origin.

Let's bound the probability for this to happen. Any dual loop of length n enclosing the origin must hit $\{(k + 1/2, 1/2) : 0 \leq k < n\}$. This gives n possibilities. For each possibility, there are at most $r_2(n-1) \leq 4^n$ simple loops.

Any loop of length n has probability $(1-p)^n$ of being open in the dual. Further, the smallest loop has length 4.

Thus,

$$1 - \theta(p) = \mathbb{P}(|C_0| < \infty) \leq \sum_{n=4}^{\infty} n(4(1-p))^n.$$

If $p \geq 7/8$, the right hand side is at most

$$\sum_{n=4}^{\infty} n(4(1-7/8))^n = \sum_{l=4} n2^{-n} = 2 - (1/2) - 2(1/2)^2 - 3(1/2)^3 < 1.$$

□

Exercise: sharpen to give $p_c(2) \le 1 - 25^{-1/3}$.

This contour argument is useful for several models in statistical mechanics. It named after Peierls', who was 'a major player in the drama of the eruption of nuclear physics into world affairs' (Physics Today) and was also at Birmingham University for some time.

3.3 Useful correlation inequalities

We're heading towards $p_c(2) = 1/2$ and an investigation of how big the connected component containing 0 is. To get there we need to show events involving open paths are positively correlated.

Intuitively this is obvious– if $0 \leftrightarrow x$ then there is an open path from 0 to x. The path may wander around a bit or go straight to x, but in either case $0 \leftrightarrow y$ seems more likely.

Loosely, the events of interest are those that are more likely to occur when more edges are open.

Definition 3.11. *Let $\omega, \tilde{\omega} \in \Omega$. We write $\tilde{\omega} \geq \omega$ if $\tilde{\omega}_e \geq \omega_e$ for every $e \in \mathcal{E}$.*

An event $A \in \mathcal{F}$ is called increasing if $\omega \in A$ and $\tilde{\omega} \geq \omega$ implies $\tilde{\omega} \in A$.

A is called decreasing if $\omega \in A$ and $\omega \geq \tilde{\omega}$ implies $\tilde{\omega} \in A$.

Example: $\{0 \leftrightarrow x\}$ is an increasing event.

Exercise: show A increasing iff A^c decreasing.

Theorem 3.12 (Harris-FKG Inequality)**.** *Suppose $A, B \in \mathcal{F}$ are both increasing or both decreasing. Then*

$$\mathbb{P}_p(A \cap B) \geq \mathbb{P}_p(A)\mathbb{P}_p(B)$$

for any p.

Proof. Exercise: show it is sufficient to prove the statement for increasing events.

We consider events that depend only finitely edges. The extension to infinitely many edges is a standard argument which we omit.

For notational convenience take space $S_n = \{0,1\}^n$ and

$$\mu_n(s) = \prod_{i=1}^{n} p^{s_i}(1-p)^{1-s_i}$$

for any $s = (s_1, \ldots, s_n) \in S_n$.

The proof is by induction on n. Let $n = 1$. $A \subset S_1$ increasing means $A = \{1\}$ or $A = \{0,1\}$. Similarly B increasing means $B = \{1\}$ or $B = \{0,1\}$.

In all cases $\mu_1(A \cap B) \geq \mu_1(A)\mu_1(B)$.

Let $n \geq 2$ and suppose for induction that

$$\mu_n(A \cap B) \geq \mu_n(A)\mu_n(B)$$

for any A, B increasing.

For any $E \subset S_{n+1}$ define

$$E_t = \{s \in S_n : (s_1, \ldots, s_n, t) \in E\},$$

for $t \in \{0, 1\}$.

Note that

- using the fact that μ is product measure, the probability for E can be decomposed as

$$\mu_{n+1}(E) = \mu_{n+1}(E, s_{n+1} = 0) + \mu_{n+1}(E, s_{n+1} = 1) = \mu_n(E_0)(1-p) + \mu_n(E_1)p.$$

- If E is increasing, $s \in E_t$, and $\tilde{s} \geq s$ then

$$(\tilde{s}_1, \ldots, \tilde{s}_n, t) \geq (s_1, \ldots, s_n, t) \in E$$

  so $\tilde{s} \in E_t$. I.e. E_t also increasing.

- If E increasing and $s \in E_0$ then $(s_1, \ldots, s_n, 1) \geq (s_1, \ldots, s_n, 0) \in E$ and so $s \in E_1$. I.e. $E_0 \subset E_1$.
- If $E' \subset S_{n+1}$ is another set then

$$(E \cap E')_t = \{s \in S_n : (s_1, \ldots, s_n, t) \in E \cap E'\} = E_t \cap E'_t.$$

Let $A, B \subset S_{n+1}$ be increasing. Let's use the properties the above.

$$\mu_{n+1}(A)\mu_{n+1}(B) = (p\mu_n(A_1) + (1-p)\mu_n(A_0))(p\mu_n(B_1) + (1-p)\mu_n(B_0)).$$

$A_0 \subset A_1$ and $B_0 \subset B_1$ so

$$(\mu_n(A_1) - \mu_n(A_0))(\mu_n(B_1) - \mu_n(B_0)) \geq 0.$$

Expanding brackets in the previous two equations and rearranging gives

$$\mu_{n+1}(A)\mu_{n+1}(B) \leq p\mu_n(A_1)\mu_n(B_1) + (1-p)\mu_n(A_0)\mu_n(B_0) \leq p\mu_n(A_1 \cap B_1) + (1-p)\mu_n(A_0 \cap B_0)$$

using the inductive hypothesis. But the RHS is $\mu_{n+1}(A \cap B)$.

□

The Harris-FKG inequality is often used with something called the n^{th} root trick.

Corollary 3.13. *Let $A_1, \ldots, A_n \in \mathcal{F}$ be increasing. Then for some $1 \leq i \leq n$ we have*

$$\mathbb{P}_p(A_i) \geq 1 - (1 - \mathbb{P}_p(A_1 \cup \ldots \cup A_n))^{1/n}.$$

for any p.

Proof. A_i increasing implies A_i^c decreasing. Also the intersection of decreasing events is decreasing so, Harris-FKG gives

$$\mathbb{P}_p(A_1^c \cap A_2^c \cap \ldots \cap A_n^c) \geq \prod_{i=1}^{n} \mathbb{P}_p(A_i^c).$$

But

$$\mathbb{P}_p(A_i^c) > \mathbb{P}_p(A_1^c \cap A_2^c \cap \ldots \cap A_n^c)^{1/n}$$

for every i gives

$$\prod_{i=1}^{n} \mathbb{P}_p(A_i^c) > \mathbb{P}_p(A_1^c \cap A_2^c \cap \ldots \cap A_n^c).$$

So

$$\mathbb{P}_p(A_i^c) \leq \mathbb{P}_p(A_1^c \cap A_2^c \cap \ldots \cap A_n^c)^{1/n}$$

for some i.

The desired inequality follows from

$$1 - \mathbb{P}_p(A_1 \cup \ldots \cup A_n) = \mathbb{P}_p(A_1^c \cap A_2^c \cap \ldots \cap A_n^c), \;\; \mathbb{P}_p(A_i^c) = 1 - \mathbb{P}_p(A_i).$$

$\square$

Last Thursday considered the Harris-FKG inequality, which states that increasing events are correlated.

This lecture concerns a kind of converse inequality for increasing events that can occur 'disjointly'.

Take two increasing events $A = \{0 \leftrightarrow x\}$ and $B = \{0 \leftrightarrow y\}$.

These are correlated by the Harris-FKG inequality.

What about the event that the path connecting 0 to x and the path from 0 to y are disjoint?

Intuitively this is less likely than, independently, there being paths to x and y from 0.

Definition 3.14. *Suppose that $F \subset \mathcal{E}$ and $\omega, \omega' \in \Omega$. Write $\omega'|_F = \omega|_F$ if $\omega'_e = \omega_e$ for any $e \in F$. For events $A, B \in \mathcal{F}$ define*

$$A \Box B = \{\omega \in \Omega : \exists F \subset \mathcal{E} \text{ s.t. } \omega'|_F = \omega_F \implies \omega' \in A,\ \omega'|_{\mathcal{E}\setminus F} = \omega_{\mathcal{E}\setminus F} \implies \omega' \in B \text{ for any } \omega' \in \Omega\}$$

This is the event that A and B occur disjointly.

The set F in the definition above can depend on ω!

Exercise: $A \Box B \subset A \cap B$ and A, B both increasing implies $A \Box B$ increasing.

Theorem 3.15 (BK-inequality)**.** *Suppose that $A, B \in \mathcal{F}$ are increasing events that depend only on finitely many edges. Then*

$$\mathbb{P}(A \Box B) \leq \mathbb{P}(A)\mathbb{P}(B).$$

The BK here stands for van den Berg and Kesten.

Proof. Induction proof similarish to Harris-FKG. Again it is convenient to consider the space $S_n = \{0,1\}^n$ and μ_n = product measure.

First consider $n = 1$. Let $A, B \subset \{0,1\}$ and check $\mu_1(A \Box B) \leq \mu_1(A)\mu_1(B)$ by examining all cases (use $A \Box B \subset A \cap B$).

Let $n \geq 1$ and for induction suppose that

$$\mu_n(A' \Box B') \leq \mu_n(A')\mu_n(B')$$

for any increasing $A', B' \subset S_n$.

Take $A, B \subset\subset S_{n+1}$ be increasing and $C = A \Box B$.

Exercise: show that $C_0 = A_0 \Box B_0$ and $C_1 = (A_0 \Box B_1) \cup (A_1 \Box B_0)$, where $C_t = \{s \in S_n : (s_1, \ldots, s_n, t) \in C\} \subset S_n$ as in proof of Harris-FKG.

Furthermore,

$$A_0 \subset A_1,\ B_0 \subset B_1,\ C_0 \subset (A_0 \Box B_1) \cap (A_1 \Box B_0)$$

and $C_1 \subset A_1 \Box B_1$.

So, using the inductive hypothesis,

$$\mu(C_0) = \mu(A_0 \square B_0) \le \mu(A_0)\mu(B_0), \quad \mu(C_1) \le \mu(A_1 \square B_1) \le \mu(A_1)\mu(B_1).$$

(To save typing the 'n' subscript is dropped).

Further,

$$\mu(C_0) + \mu(C_1) \le \mu((A_0 \square B_1) \cap (A_1 \square B_0)) + \mathbb{P}((A_0 \square B_1) \cup (A_1 \square B_0)) \le \mathbb{P}(A_0 \square B_1) + \mathbb{P}(A_1 \square B_0)$$

using the inclusion-exclusion rule.

Multiply the last three inequalities by $(1-p)^2$, p^2 and $p(1-p)$ respectively and use the results to see

$$\mu_{n+1}(A \square B) = (1-p)\mu(C_0) + p\mu(C_1) = (1-p)^2\mu(C_0) + (1-p)p\mu(C_0) + (1-p)p\mu(C_1) + p^2\mu(C_1)$$

$$\le (1-p)^2\mu(A_0)\mu(B_0) + p(1-p)\mu(A_0)\mu(B_1) + p(1-p)\mu(A_1)\mu(B_0) + p^2\mu(A_1)\mu(B_1) = \mu_{n+1}(A)\mu_{n+1}(B).$$

□

3.4 Number of infinite clusters

Recall $\theta(p) = \mathbb{P}_p(0 \leftrightarrow \infty) = \mathbb{P}(|C_0| = +\infty)$ is the probability that the origin is in an infinite connected component.

Is it possible that there is another infinite cluster (not connected to 0)?

Note that, for $d \ge 2$ and any $k \in \mathbb{N}_0 \cup \{\infty\}$ there is an $\omega \in \Omega$ with k infinite components.

Theorem 3.16. *The number N_∞ of infinite connected components is almost surely (a) $N_\infty = 0$ if $\theta(p) = 0$ and (b) $N_\infty = 1$ if $\theta(p) > 0$.*

For part (b), the infinite component may not be connected to 0 if $\theta(p) < 1$!

Proof of part (a). Let $C_x = \{y \in \mathbb{Z}^d : x \leftrightarrow y\}$. Then

$$\{N_\infty \ge 1\} = \bigcup_x \{|C_x| = \infty\},$$

so by countable subadditivity,

$$\mathbb{P}_p(N_\infty \ge 1) \le \sum_x \mathbb{P}_p(|C_x| = \infty).$$

Now, there was nothing special about the labelling of $\mathbb{Z}^d$ we used to define $\theta(p)$. In particular, $\mathbb{P}(|C_x| = \infty) = \theta(p)$ for any x. So $\theta(p) = 0$ implies the RHS of the last inequality vanishes and hence $\mathbb{P}(N_\infty \ge 1) = 0$. □

For part (b), the following 0-1 law is needed. It is stated without proof.

Lemma 3.17. *Suppose $A \in \mathcal{F}$ is translation invariant (i.e. $\{\tau_x\omega : \omega \in A\} = A$ where $(\tau_x\omega)_e = \omega_{e+x}$). Then $\mathbb{P}(A) \in \{0,1\}$.*

The strategy of the proof is to show that A is independent of itself.

Partial proof of Theorem 3.16 (b). The event $\{N_\infty = k\}$ is translation invariant for any k. Hence $\mathbb{P}(N_\infty = k) \in \{0,1\}$ by the last Lemma and N_∞ is almost surely equal to a constant.

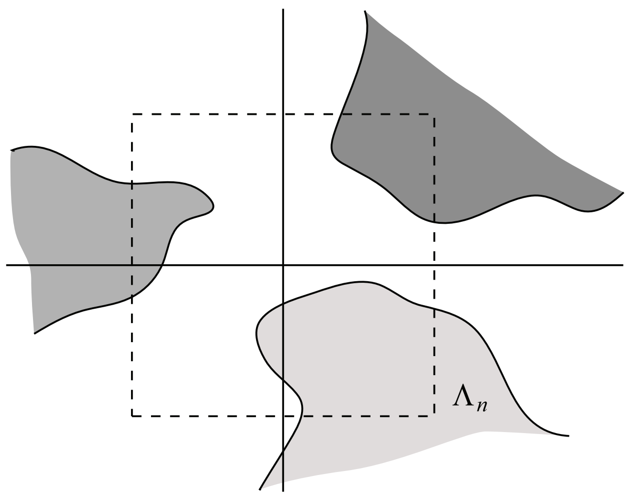
If $\theta(p) > 0$ then $\mathbb{P}_p(N_\infty = 0) < 1$. By the 0-1 law it is zero.

We must rule out $\mathbb{P}(N_\infty = k) = 1$ for $k = 2, 3, \ldots$ and $k = \infty$.

The case $k = \infty$ uses a different argument to finite k. It is omitted due to time constraints.

Suppose $N_\infty = k \geq 2$ a.s. We derive a contradiction by showing $N_\infty = 1$ with positive probability.

Let $\Lambda_n = [-n, n]^d$ and A_n be the event that there are k infinite clusters and Λ_n intersects all of them. A_n is increasing and $\cup_{n\geq 1} A_n = \{N_\infty = k\}$. THe monontone convergence theorem implies $\mathbb{P}_p(A_n) \nearrow \mathbb{P}_p(N_\infty = k) = 1$ as $n \to \infty$. Choose N big enough that $\mathbb{P}_p(A_N) > 0$.



The diagram above has $k = 3$.

Denote the nearest neighbour edges in Λ_N by $\mathcal{E}(\Lambda_N)$ and suppose $s \in \{0,1\}^{\mathcal{E}(\Lambda_N)}$. Define

$$T_{N,s} = A_N \cap \{\omega_e = s_e,\ \forall e \in \mathcal{E}(\Lambda_N)\}.$$

Then

$$\mathbb{P}(\bigcup_s T_{N,s}) = \mathbb{P}(A_N) > 0$$

hence $\mathbb{P}(T_{N,s}) > 0$ for some s.

Using the fact that edges $\mathcal{E}(\Lambda_N)$ inside Λ_N are independent of those outside,

$$0 < \mathbb{P}(T_{N,s}) = \mathbb{P}(T_{N,s} | \omega_e = s_e,\ e \in \mathcal{E}(\Lambda_N))\mathbb{P}(\omega_e = s_e,\ e \in \mathcal{E}(\Lambda_N))$$

$$\leq \mathbb{P}(T_{N,s}^{\Lambda_N^c}) = p^{-|\mathcal{E}(\Lambda_N)|}\mathbb{P}(T_{n,s}^{\Lambda_N^c})\mathbb{P}(\omega_e = 1,\ e \in \mathcal{E}(\Lambda_N)),$$

where $T_{n,s}^{\Lambda_N^c}$ is the set of ω whose restriction to $\mathcal{E}(\Lambda_N)^c$ lies in $T_{n,s}$.

Now let $\tilde{A}_N$ be the event that every infinite cluster touches the edge of Λ_N. $\tilde{A}_N$ differs from A_N by not requiring that there are exactly k infinite clusters, or rather $\tilde{A}_N$ doesn't care about the edges inside Λ_N. Most usefully $T_{n,s}^{\Lambda_N^c} \subset \tilde{A}_N$. This gives

$$0 < p^{-|\mathcal{E}(\Lambda_N)|}\mathbb{P}(\tilde{A}_N)\mathbb{P}(\omega_e = 1,\ e \in \mathcal{E}(\Lambda_N)) = p^{-|\mathcal{E}(\Lambda_N)|}\mathbb{P}(\tilde{A}_N, \omega_e = 1,\ e \in \mathcal{E}(\Lambda_N)).$$

[draw another diagram]

The event appearing on the RHS is a subset of $\{N_\infty = 1\}$.

□

Excercise: explain why the argument fails $N_\infty = \infty$!

3.5 Harris Theorem: $\theta_2(1/2) = 0$

The previous section shows that there is exactly one infinite connected cluster if $\theta(p) > 0$.

In this section we use the 'self duality' of percolation on $\mathbb{Z}^2$ with $p = 1/2$ to show that, if $\theta(1/2) > 0$, then there is a positive probability of there being TWO infinite clusters. Contradiction!

This gives a proof for Harris' Theorem–

Theorem 3.18 (Harris)**.** *For percolation on $\mathbb{Z}^2$ the probability that there is an infinite cluster when $p = 1/2$ is $\theta_2(1/2) = 0$. Consequently $p_c(2) \geq 1/2$.*

Proof. There are four steps, the first and last are similar in flavour to the proof of Theorem 3.16.

Suppose that $\theta(1/2) > 0$. We always work with $p = 1/2$ and so can drop the subscript from $\mathbb{P}_{1/2}$.

Step 1. Take $\Lambda_n = [-n, n]^2$. Let A_n be the event that there is one infinite cluster and it intersects Λ_n. As before we have $\mathbb{P}(A_n) \nearrow \mathbb{P}(N_\infty = 1) = 1$.

Hence $\mathbb{P}(A_n) \geq 1 - 10^{-4}$ for large enough n.

Step 2. Let $C^\uparrow$ be the event that there is an infinite open path of edges in $\mathcal{E}(\mathbb{Z}^2) \setminus \mathcal{E}(\Lambda_n)$ (edges with at least one endpoint outside of Λ_n) which starts from the northern boundary $[-n, n] \times \{n\}$ of Λ_n.

Likewise, $C^\rightarrow, C^\downarrow, C^\leftarrow$ are the events that infinite open paths emanate from the eastern boundary $\{n\} \times (-n, n)$, southern boundary $[-n, n] \times \{-n\}$ and west $\{-n\} \times (-n, n)$, respectively.

Now, $A_n \subset C^\uparrow \cup C^\rightarrow \cup C^\downarrow \cup C^\leftarrow$. Further, all the $C^\cdot$ events are increasing. The nth root trick Corollary 3.13 thus shows $\mathbb{P}(C^i) \geq 1 - (10^{-4})^{1/4} = 9/10$ for some $i \in \{\uparrow, \rightarrow, \downarrow, \leftarrow\}$ and n large enough.

But all $\mathbb{P}(C^i)$ are equal. Thus, both

$$\mathbb{P}(C^{\uparrow}), \mathbb{P}(C^{\downarrow}) \geq 9/10.$$

Step 3. We repeat step 2 for dual percolation. Recall $\mathbb{Z}^2_\star = \mathbb{Z}^2 + (1/2, 1/2)$ and the (nearest neighbour) dual edge $e^\star \in \mathcal{E}(\mathbb{Z}^2_\star)$ is defined open iff the primal edge $e \in \mathcal{E}(\mathbb{Z}^2)$ that it crosses is closed.

Thus, under $\mathbb{P} \equiv \mathbb{P}_{1/2}$ each dual edge is also open with probability $1 - 1/2 = 1/2$ independently of the others (not surprising $d = 2$, $p = 1/2$ percolation is called self-dual).

Let $\Lambda^\star_n = [-n + 1/2, n - 1/2]^2$ be a box in $\mathbb{Z}^2_\star$ lying just inside Λ_n.

Define events $C^{\uparrow}_\star, C^{\rightarrow}_\star, C^{\downarrow}_\star, C^{\leftarrow}_\star$ analogously to the $C^{\cdot}$– $C^{\leftarrow}_\star$ is the event that an infinite open path begins from the eastern boundary $\{-n + 1/2\} \times (-n + 1/2, n - 1/2)$ of $\Lambda^\star_n$ without going inside $\Lambda^\star_n$, etc.

Repeating the argument in step 1 shows that for large enough n,

$$\mathbb{P}(C^{\rightarrow}_\star), \mathbb{P}(C^{\leftarrow}_\star) \geq 9/10.$$

Step 4. Let n be big enough for the previous steps. We have

$$1 - \mathbb{P}(C^{\uparrow} \cap C^{\downarrow} \cap C^{\rightarrow}_\star \cap C^{\leftarrow}_\star) \leq 4 \times (1 - 9/10) = 4/10.$$

Hence,

$$\mathbb{P}(C^{\uparrow} \cap C^{\downarrow} \cap C^{\rightarrow}_\star \cap C^{\leftarrow}_\star) \geq 6/10 > 0.$$

Now, the intersection $C^{\uparrow} \cap C^{\downarrow} \cap C^{\rightarrow}_\star \cap C^{\leftarrow}_\star$ is independent of the finitely many edges inside Λ_n.

Thus, the probability that both $C^{\uparrow} \cap C^{\downarrow} \cap C^{\rightarrow}_\star \cap C^{\leftarrow}_\star$ holds and also all edges inside Λ_n are closed (i.e. all dual edges in $\Lambda^\star_n$ are open) is strictly positive.

However, on this event we have $N_\infty \geq 2$– draw a diagram to convince yourself that the infinite open paths from the northern and southern boundary of Λ_n cannot be connected.

But $\mathbb{P}(N_\infty \geq 2) > 0$ contradicts Theorem 3.16! □

Remark 3.19. † *For percolation on $\mathbb{Z}^2$ we just proved $p = p_c$ then there is almost surely no infinte connected component, i.e. $\theta_d(p_c(d)) = 0$ for $d = 2$.*

Somewhat surprisingly, the result has not been proved for $2 < d < 19$. The proof for $d \geq 19$ uses some surprisingly heavy machinery (lace expansion).

Most people think that θ_d is continuous as a function of p, which entails $\theta_d(p_c(d)) = 0$.

It's tempting to think the analogous result for percolation on general graphs should hold. In fact there is a counter-example for this.

3.6 The connected component containing the origin when $p < p_c$

The case $p < p_c$ is called sub-critical. Here, the connected component containing the origin, $C_0 = \{y \in \mathbb{Z}^d : 0 \leftrightarrow y\}$ is finite a.s. But how big is it? Is $\chi(p) = \mathbb{E}_p[|C_0|] < \infty$?

One way to measure the size of C_0 is to see if it extends to touch a big box–

As before let $\Lambda_n = [-n, n]^d$, $\partial\Lambda_n = \Lambda_n \setminus \Lambda_{n-1}$. For $x \in \mathbb{Z}^d$ and $\Lambda \subset \mathbb{Z}^d$ write $x \leftrightarrow \Lambda$ if there exists $y \in \Lambda$ with $x \leftrightarrow y$.

$\mathbb{P}_p(0 \leftrightarrow \partial\Lambda_n)$ must go to zero when $p < p_c$, but how quickly?

Theorem 3.20. *Suppose $0 < p < p_c$. Then there exists $\psi(p) > 0$ s.t. $\mathbb{P}_p(0 \leftrightarrow \partial\Lambda_n) \leq e^{-n\psi(p)}$ for any $n \geq 1$.*

The Theorem was proven independently in two papers in the mid-80s, one by Menshikov and the other by Aizenman and Barsky.

Note that the Theorem gives

$$\chi(p) = \mathbb{E}_p[|C_0|] = \sum_{n=1}^{\infty} \mathbb{E}_p[|C_0 \cap \partial\Lambda_n|] \leq \sum_{n=1}^{\infty} |\partial\Lambda_n| \mathbb{P}(0 \leftrightarrow \partial\Lambda_n) \leq C \sum_{n=1}^{\infty} n^{d-1} e^{n\psi(p)} < \infty.$$

In fact Aizenman-Barsky's proof began from the stepping stone $\chi(p) < \infty$ for $p < p_c$, which had already been proved by Hammersley. Let's see their proof.

Theorem 3.21. *Suppose p is such that $\chi(p) = \mathbb{E}_p[|C_0|] < \infty$. Then there exists $\psi(p) > 0$ such that*

$$\mathbb{P}_p(0 \leftrightarrow \partial\Lambda_n) \leq e^{-n\psi(p)}.$$

Proof. Fix p and drop the subscript from $\mathbb{P}_p$. Let $m, k \in \mathbb{N}$ and suppose $0 \leftrightarrow \partial\Lambda_{m+k}$.

Thus there is an open self avoiding walk from 0 to some $y \in \partial\Lambda_{m+k}$.

Let $x \in \partial\Lambda_m$ be the last site visited by the SAW in Λ_m.

The remainder of the SAW from x to y is an open path disjoint from the path 0 to x. Further, the final part of the path must hit a site in the translated box $\partial\Lambda_k + x$.

I.e.

$$\{0 \leftrightarrow \partial\Lambda_{m+k}\} \subset \bigcup_{x\in\partial\Lambda_m} \{0 \leftrightarrow x\}\Box\{x \leftrightarrow \partial\Lambda_k + x\}.$$

But the events $\{0 \leftrightarrow x\}$,$\{x \leftrightarrow \partial\Lambda_k + x\}$ are both increasing and depend only on finitely many edges, so we can use the BK-inequality for their disjoint occurrence probability–

$$\mathbb{P}(0 \leftrightarrow \partial\Lambda_{m+k}) \leq \sum_{x\in\partial\Lambda_m} \mathbb{P}(0 \leftrightarrow x)\mathbb{P}(x \leftrightarrow \partial\Lambda_k + x\} = \mathbb{P}(0 \leftrightarrow \partial\Lambda_k) \sum_{x\in\partial\Lambda_m} \mathbb{P}(0 \leftrightarrow x).$$

Now, if $\chi(p) < \infty$ then

$$\infty > \mathbb{E}[|C_0|] = \sum_{m=0}^{\infty} \sum_{x\in\partial\Lambda_m} \mathbb{P}(0 \leftrightarrow x),$$

so $\sum_{x\in\partial\Lambda_m} \mathbb{P}(0 \leftrightarrow x) \to 0$ as $m \to \infty$.

Assume m is big enough so that

$$\eta = \sum_{x\in\partial\Lambda_m} \mathbb{P}(0 \leftrightarrow x) < 1.$$

Suppose $n = mj + r$ where $0 \leq r < m$. Then using the inequality above repeatedly,

$$\mathbb{P}(0 \leftrightarrow \partial\Lambda_n) \leq \mathbb{P}(0 \leftrightarrow \Lambda_{mj}) \leq \mathbb{P}(0 \leftrightarrow \partial\Lambda_{m(j-1)}) \sum_{x\in\partial\Lambda_m} \mathbb{P}(0 \leftrightarrow x) \leq \ldots \leq \eta^j.$$

But $j = (n - r)/m \geq m/m - 1$. So $\eta^j \leq \eta^{n/m-1}$ and for any $\eta^{1/m} < \tilde{\eta} < 1$ we get

$$\mathbb{P}(0 \leftrightarrow \partial\Lambda_n) \leq \eta^{-1}(\eta^{1/m})^n < \tilde{\eta}^n$$

for n big enough, say $n > N$.

For any $n \geq 1$ note that

$$\mathbb{P}(0 \leftrightarrow \partial\Lambda_n) \leq \mathbb{P}(0 \leftrightarrow \partial\Lambda_1) \leq 1 - (1 - p)^{2d} =: b < 1.$$

So $0 < \psi(p) < \min\{-\ln(b)/N, -\ln(\tilde{\eta})\}$ gives

$$\mathbb{P}(0 \leftrightarrow \partial\Lambda_n) \leq \mathbb{P}(0 \leftrightarrow \partial\Lambda_1) \leq b \leq \exp(-N\psi(p)) \leq \exp(-n\psi(p))$$

for $1 \leq n \leq N$ and $\mathbb{P}(0 \leftrightarrow \partial\Lambda_n) \leq \tilde{\eta}^n \leq \exp(-\psi(p)n)$ for $n > N$.

$\Box$