

Part IB — Markov Chains

Based on lectures by Geoffrey Grimmet

Notes taken by Christopher Turnbull

Michaelmas 2017

0 Introduction

Probability, An Introduction, G. Grimmett¹, D. Welsh, Chapter 12

Andrey Markov was a fine soviet Russian Mathematician in the 20th century. A random variable is actually a function

$$X : \Omega \rightarrow S$$

It acts on outcomes. The mathematical formulation of an R.V. is a single function that varies randomly (eg. the number of heads in 5 flips of a coin) The passing of time plays an essential part in the world in which we inhabit, and consequently many applications of probability involve quantities which develop randomly as time passes. Such randomly evolving processes are called *random processes* or *stochastic processes*, and there are many different types of these (pollen count in Cambridge, number of people entering Cindies per night).

$$(X_0, X_1, X_2, \dots)$$

The interesting question is: are the separate events X_i iid, or do they depend² on each other? If the distribution of X_1 depends on X_0 ? This essentially means probabilities factorise as products.

Note that the use of ‘chain’ implies we are considering discrete time. For continuous time,

$$(X_t, t \in \mathbb{R})$$

these are fundamentally a much more complicated object.

The *general theory* of stochastic processes is much more challenging; there are too many to extract a general mathematical theory for now, so we must add a condition, such that we can extract a theory that has applications and is useful.

There are many types of conditions; there’s a lot of juice to get out of the ‘independence assumption’ but we’re throwing too much away with that. So we assume the Markov condition; loosely, all the information required to calculate the probabilities of the future is contained right now³.

0.1 Random Walks

In this section we consider the simplest type of random walk, where both space and time are *discrete*. We shall observe the particles position at each of the discrete time points $0, 1, 2, \dots$ and we assume that at each of these times the particle is located at one of the integer positions $\dots, -2, -1, 0, 1, 2, \dots$ of the real line. The particle moves in the following way: if S_N denotes the location of the particle at time n , then

$$S_n = \begin{cases} S_n + 1 & \text{with probability } p, \\ S_n - 1 & \text{with probability } q, \end{cases}$$

¹www.statstlab.cam.ac.uk/~grg/

²Dependence is a question about correlation. It is not a question about causality, that’s much harder.

³Eckhart Tolle, Power of Now

and we suppose that the random direction of each jump is independent of all earlier jumps. Therefore,

$$S_n = S_0 + X_1 + X_2 + \cdots + X_n \quad \text{for } n = 0, 1, 2, \dots$$

where S_0 is the starting position and X_1, X_2, \dots are independent random variables, with

$$X_i = \begin{cases} 1 & \text{with probability } p \\ -1 & \text{with probability } q \end{cases}$$

We call the process S_0, S_1, \dots a *simple random walk*.

Definition. If $p = q = \frac{1}{2}$ is it called a *symmetric random walk*, and *asymmetric* otherwise.

Theorem. Let $u_n = \mathbb{P}(S_n = S_0)$ be the probability that a simple random walk revisits its starting point at time n . Then $u_n = 0$ if n is odd, and

$$u_{2m} = \binom{2m}{m} p^m q^m$$

if $n = 2m$ is even

Proof. We may suppose without loss of generality that $S_0 = 0$. The first part is obvious, for the second, note that

$$S_{2m} = X_1 + X_2 + \cdots + X_{2m}$$

so that $S_{2m} = 0$ if and only if exactly m of the X_i equal $+1$ and exactly m equal -1 . There are exactly $\binom{2m}{m}$ ways of dividing the X_i into two sets of equal sizes, and the theorem follows. \square

For more general transition probabilities, the simplest way to proceed is to note that for $i \geq 1$, the random variable $\frac{1}{2}(X_i + 1)$ has the Bernoulli distribution with parameter p , giving that $B_n = \frac{1}{2}(S_n + n)$ has the binomial distribution with parameters n and p . Hence,

$$\begin{aligned} \mathbb{P}(S_n = k | S_0 = 0) &= \mathbb{P}(B_n = \frac{1}{2}(n+k)) \\ &= \binom{n}{\frac{1}{2}(n+k)} p^{\frac{1}{2}(n+k)} q^{\frac{1}{2}(n-k)} \end{aligned}$$

This is non-zero whenever k is such that $\frac{1}{2}(n+k)$ is an integer between 0 and n . The result of the previous theorem is obtained by setting $k = 0$.

1 Markov Chains

Let S be a countable set called the *state space*, and let $\mathbf{X} = (X_n : n \geq 0)$ be a sequence of random variables taking values in S .

Definition. (1.1) The sequence \mathbf{X} is called a *Markov Chain* if it satisfies the *Markov Property*:

$$\mathbb{P}(X_{n+1} = i_{n+1} | X_0 = i_0, X_1 = i_1, \dots, X_n = i_n) = \mathbb{P}(X_{n+1} = i_{n+1} | X_n = i_n)$$

$$\forall n \geq 0, i_0, \dots, i_{n+1} \in S$$

Definition. (1.2) The Markov Chain \mathbf{X} is called homogeneous⁴⁵ if, $\forall i, j \in S$, the conditional probability

$$\mathbb{P}(X_{n+1} = j | X_n = i)$$

does not depend on the value of n .

Example. Here are some examples of Markov chains

- (i) *Random walk*⁶. Let Z_1, Z_2, \dots be indep, $\mathbb{P}(Z_i = \pm 1) = p, 1 - p$. $X_n = Z_1 + \dots + Z_n$.
- (ii) *Branching process*. Let X_n be the size of the n th generation. Each member of the n th generate has a number of offspring that is independent of the past

In order to study the behaviour of a chain over a finite/infinite⁷ length of time, we need to know two quantities

- (i) the *initial distribution* $\lambda = (\lambda_i : i \in S)$ given by $\lambda_i = \mathbb{P}(X_0 = i)$
- (ii) the *transition matrix* $P = (p_{i,j} : i, j \in S)$ given by $p_{i,j} = \mathbb{P}(X_i = j | X_0 = i)$.

Proposition. (1.3)

- (i) The vector λ is a distribution in that $\lambda_i \geq 0$ for $i \in S$, and $\sum_{i \in S} \lambda_i = 1$.
- (ii) P is a stochastic matrix in that $p_{i,j} \geq 0$ for $i, j \in S$, and $\sum_{j \in S} p_{i,j} = 1$ for $i \in S$, so that P has row sums 1.

Proof. (i) Since λ_i is a probability, it is non-negative. Also, $\sum_{i \in S} \lambda_i = \sum_i \mathbb{P}(X_0 = i) = \mathbb{P}(X_0 \in S) = 1$

⁴Inhomogenous are more complicated, but not in a sense richer. So lets not obscure the mathematics.

⁵Henceforth (unless stated otherwise) all chains are assumed as homogenous.

⁶Often we will test new techniques on this simple example.

⁷this is where the beautiful maths lies!

(ii) Since $p_{i,j}$ is a probability, it is non-negative. Finally,

$$\begin{aligned}\sum_{j \in S} p_{i,j} &= \sum_{j \in S} \mathbb{P}(X_1 = j | X_0 = i) \\ &= \mathbb{P}(X_1 \in S | X_0 = i) = 1\end{aligned}$$

□

Every conditional probability can be expressed in terms of absolute probabilities, and to study stochastic processes in complete generality, we need the joint mass function. The following theorem will be useful later

Theorem. Let λ be a distribution (on S) and P a stochastic matrix. The sequence $X = (X_n : n \geq 0)$ is a Markov chain with initial distribution λ and transition matrix P if and only if:

$$\mathbb{P}(X_0 = i_0, X_1 = i_1, \dots, X_n = i_n) = \lambda_{i_0} p_{i_0, i_1} p_{i_1, i_2} \dots p_{i_{n-1}, i_n} (*)$$

for all $n \geq 0$ and $i_0, i_1, \dots, i_n \in S$.

Proof. ⁸ Write A_k for the event $\{X_k = i_k\}$. Then $(*)$ can be written as

$$\mathbb{P}(A_0 \cap A_1 \cap \dots \cap A_n) = \lambda_{i_0} p_{i_0, i_1} \dots p_{i_{n-1}, i_n}$$

Suppose X is a (λ, P) Markov chain. We prove by induction on n . When $n = 0$, $\mathbb{P}(X_0 = i_0) = \lambda_{i_0}$. Suppose $(*)$ holds for $n < N$.

$$\begin{aligned}\mathbb{P}(A_0 \cap \dots \cap A_N) &= \mathbb{P}(A_0 \cap \dots \cap A_N | A_0 \cap \dots \cap A_{N-1}) \mathbb{P}(A_0 \cap \dots \cap A_{N-1}) \\ &= \mathbb{P}(A_N | A_0 \cap \dots \cap A_{N-1}) \mathbb{P}(A_0 \cap \dots \cap A_{N-1}) \\ &= \mathbb{P}(A_N | A_{N-1}) \lambda_{i_0} p_{i_0, i_1} \dots p_{i_{N-2}, i_{N-1}}\end{aligned}$$

Conversely, suppose $(*)$ holds,

By $(*)$ with $n = 0$,

$$\mathbb{P}(X_0 = i_0) = \lambda_{i_0}$$

So X_0 has pmf λ .

$$\mathbb{P}(A_N | A_0 \cap \dots \cap A_N) = \frac{\mathbb{P}(A_0 \cap \dots \cap A_{n+1})}{\mathbb{P}(A_0 \cap \dots \cap A_N)} = p_{i_n, i_{n+1}}$$

Hence X is a Markov chain with transition matrix P

□

Theorem. (Extended Markov property)

Let X be an MC, and $n \geq 1$. Let H be a historic-event, ie. H is given in terms of $\{X_0, X_1, \dots, X_{n-1}\}$ and let F be a future event, ie. F given in terms of $\{X_{n+1}, X_{n+2}, \dots\}$. Then

$$\mathbb{P}(F | H, X_n = i) = \mathbb{P}(F | X_n = i)$$

⁸We're learning to drive here, to press the buttons and see what they do. There's nothing very deep in this proof, it's quite superficial.

Proof. Assume F depends on only finitely many of the future variables.⁹¹⁰

$$\begin{aligned}
 \mathbb{P}(F|H, X_n = i) &= \frac{\sum_{>N} \sum_{<N} \lambda_{i_0} p_{i_0, i_1}, \dots, p_{i_{N-1}, i} p_{i, i_{N+1}}}{\sum_{i < n} \lambda_{i_0} p_{i_0, i_1}, \dots, p_{i_{N-1}, i}} \\
 &= \sum_{>N} p_{i, i_{N+1}} \\
 &= \mathbb{P}(F|X_n = i)
 \end{aligned}$$

□

⁹We're going to prove it for F depending on a finite set of events, rather than infinitely many.

¹⁰ $\sum_{<n} :=$ or $(>)$ sum over all i_0, \dots, i_{N-1} contributing to H (or F)

2 Transition probabilities

One step transition probability: $p_{i,j} = \mathbb{P}(X_1 = j | X_0 = i)$

n -step $p_{i,j}(n) = \mathbb{P}(X_n = j | X_0 = i)$.

Example. How to compute n -step transition probabilities from the one-step prob?

$$P = (p_{i,j})_{i,j \in S}$$

$$P(n) = (p_{i,j}(n))_{i,j \in S}$$

Theorem.

$$P(n) = P^n$$

Proposition. (Chapman-Kolmogorov equations) ¹¹

$$p_{i,j}(n+m) = \sum_{k \in S} p_{i,k}(m) p_{k,j}(n)$$

Proof. ¹²

$$\begin{aligned} \mathbb{P}(X_{m+n} = j | X_0 = i) &= \sum_{k \in S} \mathbb{P}(X_{m+n} = j, X_m = k | X_0 = i) \\ &= \sum_k \mathbb{P}(X_{m+n} = j | X_m = k, X_0 = i) \mathbb{P}(X_m = k | X_0 = i) \\ &= \sum_k p_{i,k}(m) p_{k,j}(n) \end{aligned}$$

By CK, $P(m+n) = P(m)P(n)$.

Hence $P(n) = P(1)P(n-1) = \dots P(1)^n = P^n$

□

Example. $|S| = 2$, label them wlog $S = \{1, 2\}$

$$P = \begin{pmatrix} 1-\alpha & \alpha \\ \beta & 1-\beta \end{pmatrix}$$

where¹³ $0 < \alpha, \beta < 1$

(i) Diagonalize, so

$\det(P - \kappa I) = 0$, roots¹⁴ $\kappa_1 = 1$, $\kappa_2 = 1 - \alpha - \beta$.

$$P = U^{-1} \begin{pmatrix} 1 & 0 \\ 0 & 1 - \alpha - \beta \end{pmatrix} U$$

for some invertible U .

$$P^n = U^{-1} \begin{pmatrix} 1^n & 0 \\ 0 & (1 - \alpha - \beta)^n \end{pmatrix} U$$

¹¹breaking up trajectories according to the value of k

¹² $\mathbb{P}(A \cap B | C) = \mathbb{P}(A | B \cap C) \mathbb{P}(B | C)$

¹³avoid degenerate cases

¹⁴note that 1 is always an eigenvalue

So $p_{1,1}(n) = A + B(1 - \alpha - \beta)^n$, $P_{1,1}(0) = 1$, $P_{1,1}(1) = 1 - \alpha$. Thus

$$A = \frac{\beta}{\alpha + \beta}, \quad B = \frac{\alpha}{\alpha + \beta}$$

$$p_{1,2}(n) = 1 - P_{1,1}(n)$$

$p_{2,1}(n), P_{2,2}(n)$ obtained by interchanging α, β

(ii) by difference equations: Using CK,

$$\begin{aligned} p_{1,1}(n+1) &= \sum_k p_{1,k}(n)p_{k,1}(1) \\ &= p_{1,1}(n)(1 - \alpha) + p_{1,2}(n)\beta \\ &= p_{1,1}(n)(1 - \alpha) + (1 - p_{1,1}(n))\beta \end{aligned}$$

Thus

$$\pi_{n+1} = (1 - \alpha - \beta)\pi_n + \beta$$