University of Warwick

Stochastic Processes

Written by Khallil Benyattou
Lectured by Nikos Zygouras

October 9, 2019

Contents

1 Review of Probability A & B

The sample space Ω of an experiment is the set of all possible outcomes of said experiment. An event A is a subset of Ω.

A random variable is a map $X: \Omega \to \mathbb{R}$. The range of X is defined as $\text{range}(X) := \{X(\omega): \omega \in \Omega\}$.

A sequence of random variables $X_1, X_2, \dots$ are called independent if

$$\mathbb{P}(X_1 \in A_1, X_2 \in A_2, \dots) = \prod_{i \in \mathbb{N}} \mathbb{P}(X_i \in A_i).$$

For a discrete random variable X, its expectation is given by

$$\mathbb{E}[X] := \sum_{x \in \Omega_x} x f(x).$$

As a small but important corollary, for any functions $(f_i)_{i \geqslant 1}$:

$$\mathbb{E}\left[\prod_{i \geqslant 1} f_i(X_i)\right] = \prod_{i \geqslant 1} \mathbb{E}[f_i(X_i)].$$

A moment generating function is a function on $z \in \mathbb{C}$ that is defined by

$$\mathbb{E}[z^X] = \sum_a z^a \ \mathbb{P}(X = a).$$

In a probabilistic setting, the expectation of X is

$$\mathbb{E}[X] = \sum_a a\mathbb{P}(X = a) = \frac{\mathrm{d}\mathbb{E}}{\mathrm{d}\lambda}[e^{\lambda X}]\Big|_{\lambda=0}.$$

The variance of X is given by

$$\text{Var}(X) = \mathbb{E}[X^2] - (\mathbb{E}[X])^2 = \frac{\mathrm{d}^2 \log(\mathbb{E})}{\mathrm{d}\lambda^2}[e^{\lambda X}]\Big|_{\lambda=0}.$$

An example of a basic distribution is the Bernoulli $(a, b; p)$ distribution which is defined by $\mathbb{P}(X = a) = p$, $\mathbb{P}(X = b) = 1 - p$. Another distribution is the Binomial distribution which is defined, for identically and independently distributed random variables $X_1, X_2, \dots$ with Bernoulli $(0, 1; p)$ distribution, by

$$\mathbb{P}(S_n = k) = \binom{n}{k} p^k (1-p)^{n-k}$$

where $n \geqslant 1$ and $k = 0, \dots, n$.

The expectation of S_n can be computed by noting that

$$\mathbb{E}\left[\sum_{i=1}^n X_i\right] = \sum_{i=1}^n \mathbb{E}[X_i] = np.$$

It can also be computed directly by manipulating the original definition

$$\mathbb{E}[S_n] = \sum_{k=0}^n k \binom{n}{k} p^k (1-p)^{n-k}.$$

Similarly, the independence of the X_i implies that

$$\text{Var}\left(\sum_{i=1}^n X_i\right) = \sum_{i=1}^n \text{Var}(X_i)) = np(1-p).$$

2 Stochastic Processes

2.1 Basic Definitions

Let Ω be a sample space, $\mathcal{A}$ be a σ-algebra on Ω and $\mathbb{P}$ be a probability measure on Ω.

Definition 2.1 A **stochastic** (or **random**) **process** is formally defined to be a collection of random variables defined on a common probability space $(\Omega, \mathcal{A}, \mathbb{P})$ and the random variables, indexed by some set T, all take values in the same range-space I; this may be $\mathbb{R}^n$ (a **vector-valued** process) or some other measurable space.

- The set T will generally be $\mathbb{R}, \mathbb{R}^+ = [0, \infty), \mathbb{Z} = \{\dots, -1, 0, 1, 2, \dots\}$ or $\mathbb{Z}^+ = \{0, 1, 2, \dots\}$. In all of these cases, the parameter $t \in T$ may be thought of as time.
  - If $T = \mathbb{Z}$ or $\mathbb{Z}^+$, one sometimes speaks of a **random sequence**.
  - If $T = \mathbb{R}^n$ with $n > 1$, the process is often called a **random field**.
- The range I of the random variables is called the **state space**.

In describing a stochastic process as we have done, there is a certain psychological bias: one tends to regard the process primarily as a function on T whose values for each $t \in T$ are random variables. Of course, we're really dealing with one function of two variables, say $X = X(t, \omega)$, where $t \in T$, $\omega \in \Omega$, and where for each fixed t the function $X(t, \cdot)$ is measurable with respect to $\mathcal{A}$. If instead of t we fix an $\omega \in \Omega$, we obtain a function $X(\cdot, \omega)\colon T \to \mathbb{R}^1$ (or into whichever state space space I may be) which is called a **trajectory** or a **path/sample-function** of the process.

2.2 Random Walks

Definition 2.2 A **random walk** is the process $(S_n)_{n \geqslant 1}$ where $S_n = X_1 + \cdots + X_n$ and $(X_i)_{i \geqslant 1}$ is a sequence of independent and identically distributed random variables.

The special case in which each X_i possesses a Bernoulli $(\pm 1, 1/2)$ distribution is called a **simple random walk**.

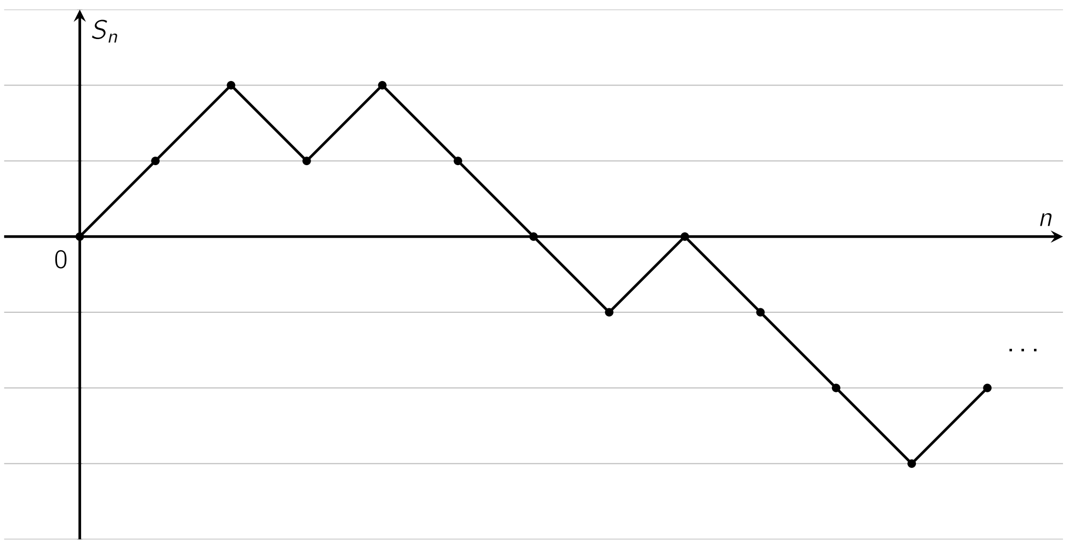


Figure 1: An example of a random walk $(S_n)_{n \geqslant 1}$.

We can ask a few questions about S_n:

1) What is the probability $\mathbb{P}(S_n = k)$?
2) What is the probability that S_n will visit k by time n?
3) Does S_n always return to its starting point?
4) How long do we expect it to take for S_n to return to its starting point?

We'll denote the starting position of a process with a subscript e.g. $\mathbb{P}_x(A)$ refers to the probability of some event A occurring given that the initial state of the process is x.

e.g. Let $(S_n)_{n\geqslant 0}$ be a simple random walk with $S_0 = 0$. What is $\mathbb{P}_0(S_n = k)$?

Denote the number of steps up and down by m and k respectively. Then $m + l = n$ and $m - l = k$. This implies that $m = (n + k)/2$ and so

$$\mathbb{P}_0(S_n = k) = \mathbb{P}(\{\text{no. of steps up}\}) = \mathbb{P}\left(\frac{n+k}{2}\right) = \binom{n}{\frac{n+k}{2}} p^{\frac{n+k}{2}}(1-p)^{\frac{n-k}{2}}.$$

However, the remaining 3 questions don't have such an easy answer. We'll need to develop a systematic method to find their solutions.

Definition 2.3 (Geometric Variables) Let $X_1, X_2, \dots$ be i.i.d. Bernoulli $(1, 0; p)$ and define $\tau := \min\{n \colon X_n = 1\}$. Then $\mathbb{P}(\tau = n) = (1-p)^{n-1}p$ because the first $(n-1)$ values of $X_1, \dots, X_{n-1}$ need to be equal to 0 and $X_n = 1$. τ can be thought of as the first time a random walk makes an upwards step.

Definition 2.4 (Conditional Probabilities) If X and Y are discrete random variables, then

$$\begin{aligned}\mathbb{P}(X = a|Y = b) &:= \frac{\mathbb{P}(X = a, Y = b)}{\mathbb{P}(X = b)} \\ &= \frac{\mathbb{P}(Y = b \mid X = a) \cdot \mathbb{P}(X = a)}{\mathbb{P}(Y = b)}.\end{aligned}$$

Conditional probabilities are important because they can be used to define the statistics of a stochastic process through transition probabilities $\pi_{x,y} := \mathbb{P}(S_{n+1} = x \mid S_n = y)$. We'll see this later on.

Definition 2.5 The **conditional expectation** of X given Y is defined by

$$E[X \mid Y] := \sum_a \mathbb{P}(X = a \mid Y).$$

Conditional expectation is just an expectation but it's computed with respect to a conditional probability. Informally, it is what we expect X to be knowing (the value of) Y. Since Y is a random variable, $\mathbb{E}[X \mid Y]$ is also a random variable.

2.3 Simple Random Walks

In this section, we'll familiarise ourselves with some basic techniques used in stochastic processes to compute things. We'll explore these through the example of a simple random walk. We already defined a simple random walk as $S_n = X_1 + \cdots + X_n$ where $(X_i)_{i\geqslant 1}$ are Bernoulli $(\pm 1; 1/2)$. However, there is an alternate formulation which is more general and can be extended to define general stochastic processes. This definition relies on specifying the conditional probabilities:

Definition 2.6 A simple random walk starting at a is a sequence of random variables $(S_n)_{n\geqslant 1}$ such that

- $S_0 = a$ with probability 1
- $\mathbb{P}(S_n = x \mid S_{n-1} = y, S_{n-2}, \cdots, S_1, S_0) = \mathbb{P}(S_n = x \mid S_{n-1} = y) = 1/2$ if $x = y \pm 1$.

Remark In general, we'll be making use of joint probabilities $\mathbb{P}(S_{n_1} = a_1, \cdots, S_{n_k} = a_k)$ with $n_1 < \cdots < n_k$. Using the above conditional law it turns out to be equal to

$$\prod_{i=1}^{k} \mathbb{P}(S_{n_i} = a_i \mid S_{n_{i-1}} = a_{i-1}).$$

2.3.1 A First Computation: Reflection Principle

e.g. Let (S_n) be a simple random walk starting at 0. Compute $\mathbb{P}_0\left(\max_{k\leqslant n} S_k \geqslant b\right)$ for $b \in \mathbb{Z}^+$.

Definition 2.7 The hitting time of a point b will be denoted by $\tau_b := \min\{k \geqslant 1 \colon S_k = b\}$. We can think of this hitting time as the first time that the random walk attains the value b.

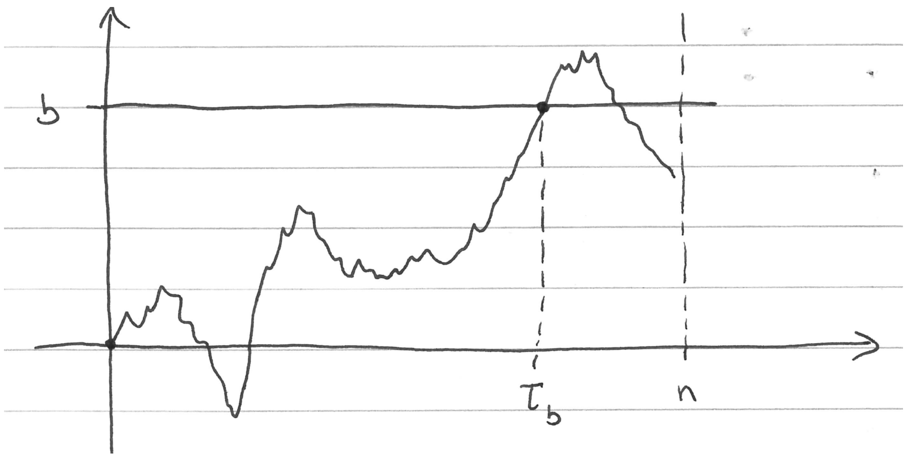


Figure 2: Pictorial representation of τ_b.

Note that $\{\max_{k\leqslant n} S_k \geqslant b\} = \{\tau_b \leqslant n\}$. First of all $\mathbb{P}_0(S_n \geqslant b) = \mathbb{P}_0(S_n \geqslant b, \tau_b \leqslant n) = \displaystyle\sum_{k=1}^{n} \mathbb{P}_0(S_n \geqslant b, \tau_b = k)$.

This implies, by the correspondence of events above, that: $\{\tau_b \leqslant n\} = \bigcup_{k=1}^{n} \{\tau_b = k\}$.

$$
\begin{aligned}
\mathbb{P}_0(S_n \geqslant b) &= \mathbb{P}_0(S_n \geqslant b, \tau_b \leqslant n) \\
&= \sum_{k=1}^{n} \mathbb{P}_0(S_n \geqslant b, \tau_b = k) \\
&= \sum_{k=1}^{n} \mathbb{P}_0(S_n \geqslant b, S_k = b, \max\{S_1, \dots, S_k\} < b) \\
&= \sum_{k=1}^{n} \mathbb{P}_0(S_n \geqslant b \mid S_k = b, \max\{S_1, \dots, S_k\} < b) \cdot \mathbb{P}_0(S_k = b, \max\{S_1, \dots, S_k\} < b) \text{ by conditioning} \\
&= \sum_{k=1}^{n} \mathbb{P}_0(S_n \geqslant b \mid S_k = b) \cdot \mathbb{P}_0(S_k = b, \max\{S_1, \dots, S_k\} < b) \text{ by the Markov property} \\
&= \sum_{k=1}^{n} \mathbb{P}_b(S_{n-k} \geqslant b) \cdot \mathbb{P}_0(S_k = b, \max\{S_1, \dots, S_k\} < b) \\
&= \sum_{k=1}^{n} \left\{\mathbb{P}_b(S_{n-k} > b) + \mathbb{P}_b(S_{n-k} = b)\right\} \cdot \mathbb{P}_0(S_k = b, \max\{S_1, \dots, S_k\} < b) \\
&\overset{(1)}{=} \sum_{k=1}^{n} \left\{\frac{1}{2} + \frac{1}{2}\mathbb{P}_b(S_{n-k} = b)\right\} \cdot \mathbb{P}_0(\tau_b = k) \\
&= \frac{1}{2}\sum_{k=1}^{n} \mathbb{P}_0(\tau_b = k) + \frac{1}{2}\sum_{k=1}^{n} \mathbb{P}_b(\tau_b = k) \cdot \mathbb{P}_b(S_{n-k} = b) \\
&= \frac{1}{2}\mathbb{P}_0(\tau_b \leqslant n) + \frac{1}{2}\mathbb{P}_0(S_n = b)
\end{aligned}
$$

Where we used in (1) the fact that via symmetry:

$$
\begin{aligned}
1 &= \mathbb{P}_b(S_{n-k} = b) + \mathbb{P}_b(S_{n-k} > b) + \mathbb{P}_b(S_{n-k} < b) \\
&= \mathbb{P}_b(S_{n-k} = b) + 2\mathbb{P}_b(S_{n-k} > b)
\end{aligned}
$$

which implies that

$$
\mathbb{P}_b(S_{n-k} > b) = \frac{1}{2} - \frac{1}{2}\mathbb{P}_b(S_{n-k} = b). \tag{1}
$$

By rearranging, we arrive at the equation

$$
\mathbb{P}_0(\tau_b \leqslant n) = 2\mathbb{P}_0(S_n \geqslant b) - \mathbb{P}_0(S_n = b).
$$

Definition 2.8 Let $p \neq 1/2$. We call $S_n = X_1 + \cdots + X_n$ an **asymmetric simple random walk** if $(X_i)_{i\geqslant 1}$ are independent, identically distributed with $\mathbb{P}(X_i = 1) = p = 1 - \mathbb{P}(X_i = -1)$.

e.g. Let $(S_n)_{n\geqslant 1}$ be an asymmetric simple random walk starting from 0 with probability of step-up being equal to p. Let $a < 0 < b$. Compute $\mathbb{P}_0(\tau_a < \tau_b)$.

Let's define $u(x) := \mathbb{P}_x(\tau_a < \tau_b)$ i.e. $u(x)$ represents the probability starting from x that S_n will hit a before it hits b. We'll set up an equation and this will be a prototype example that we'll develop into a method later.

The idea is to decompose according to the first step. S_n can either take a step up or down from x:

$$
\begin{aligned}
u(x) &:= \mathbb{P}_x(\tau_a < \tau_b) \\
&= \mathbb{P}_x(\tau_a < \tau_b, S_1 = x+1) + \mathbb{P}_x(\tau_a < \tau_b, S_1 = x-1) \\
&= \mathbb{P}_x(S_1 = x+1) \cdot \mathbb{P}_x(\tau_a < \tau_b \mid S_1 = x+1) + \mathbb{P}_x(S_1 = x-1) \cdot \mathbb{P}_x(\tau_a < \tau_b \mid S_1 = x-1) \text{ by conditioning} \\
&= p\mathbb{P}_{x+1}(\tau_a < \tau_b) + (1-p)\mathbb{P}_{x-1}(\tau_a < \tau_b) \text{ by the Markov property} \\
&= pu(x+1) + (1-p)u(x-1)
\end{aligned}
$$

This is a 2-term recursive relation i.e. a difference equation. To solve it, we also need boundary conditions: $u(a) = 1$ i.e. the probability of starting from a and hitting a is certain and $u(b) = 0$ i.e. the probability of starting at a and hitting b first is impossible.

Thus, we need to solve the boundary value problem:

$$
\begin{cases}
u(x) = pu(x+1) + (1-p)u(x-1) \\
u(a) = 1 \\
u(b) = 2
\end{cases}
$$

The general method to solve such a problem involves guessing a solution of the form t^x where g is a constant parameter to be determined. Inserting this into the difference equation and dividing through by t^{x-1} (for $t \neq 0$) gives
$pt^2 - t + (1-p) = 0$. This has solutions

$$
t_{1,2} = \frac{1 \pm \sqrt{1 - 4(1-p)p}}{2p}.
$$

This means that the recurrence relation of order 2 is satisfied by any linear combination of $(t_1)^x$ and $(t_2)^x$ i.e.
$u(x) = At_1^x + Bt_2^x$.

In the case that $p = 1/2$, $t_1 = t_2$ and so $u(x) = A + Bx$. Let's assume that $p \neq 1-p$ as the simple random walk is asymmetric. The constants A, B can be determined by the boundary conditions

$$
\begin{cases}
1 = u(a) = At_1^a + Bt_2^a \\
0 = u(b) = At_1^b + Bt_2^b
\end{cases}
$$

We know how to solve this system of two equations with in unknowns:

$$
A = \frac{\begin{vmatrix} 1 & t_2^a \\ 0 & t_2^b \end{vmatrix}}{\begin{vmatrix} t_1^a & t_2^b \\ t_1^b & t_2^b \end{vmatrix}}, \quad B = \frac{\begin{vmatrix} t_1^a & 1 \\ t_1^b & 0 \end{vmatrix}}{\begin{vmatrix} t_1^a & t_2^b \\ t_1^b & t_2^b \end{vmatrix}}
$$

So with these constants, the desired probability is $\mathbb{P}_0(\tau_a < \tau_b) = u(0) = A + B$.

Simplifying in the case that $p = 1/2$, the system of equations becomes

$$
\begin{cases}
1 = u(a) = Aa + B \\
0 = u(b) = Ab + B
\end{cases}
\implies \quad A = \frac{1}{a-b}, \ B = \frac{b}{b-a}
$$

so the general solution is $u(x) = \dfrac{x}{a-b} + \dfrac{b}{b-a}$ and the desired probability is

$$
\mathbb{P}_0(\tau_a < \tau_b) = u(0) = \frac{b}{b-a}.
$$

As a sanity check, we can interpret $\lim_{b\to\infty} \mathbb{P}_0(\tau_a < \tau_b)$ as $\mathbb{P}_0(\tau_a < \infty)$ and we can verify that $\lim_{b\to\infty} \mathbb{P}_0(\tau_a < \tau_b) = 1$ i.e. the probability that you will ever hit a is 1. This property is called recurrence i.e. for a symmetric simple random walk, the probability that you will always come back to a certain point is certain.

However, if we consider an asymmetric simple random walk e.g. $p > 1/2$

$$\lim_{b\to\infty} \mathbb{P}_0(\tau_a < \tau_b) < 1.$$

We call this property transience i.e. there's a non-trivial probability that the process will never return to a state from which it started.

2.4 Generating Functions

Let's recall the definition of a generating function of a discrete probability distribution. Let $X\colon \Omega \to A \subseteq \mathbb{R}$ be a discrete random variable defined on a sample space Ω. The **probability distribution** (or **mass**) **function** $p_X\colon A \to [0,1]$ for X is defined $\forall x \in A$ by $p_X(a) = \mathbb{P}(X = a) = \mathbb{P}(\{\omega \in \Omega\colon X(\omega) = a\})$ and satisfies

$$\sum_{a\in A} p_X(a) = 1.$$

Definition 2.9 The **probability generating function** of a discrete, non-negative random variable X is the map $\hat{p}_X$ defined for $z \in \mathbb{C}$ by

$$\hat{p}_X(z) := \mathbb{E}[z^X] = \sum_{a=0}^{\infty} z^a p_X(a).$$

If we let $z = e^\lambda$, we obtain the Laplace transform.

Now consider a random walk $S_n = X_1 + \cdots + X_n$ with $(X_i)_{i\geqslant 1}$ i.i.d. variables. The generating function, which we denote by $\hat{p}_{S_n}(t)$ will be given by

$$\begin{aligned}\hat{p}_{S_n}(t) \;=\; \mathbb{E}[t^{S_n}] \;=\; \mathbb{E}[t^{X_1+\cdots+X_n}] = \mathbb{E}\left[\prod_{i=1}^n t^{X_i}\right] &= \prod_{i=1}^n \mathbb{E}[t^{X_i}] \;\text{ by independence}\\ &= \mathbb{E}[t^{X_1}]^n \;\text{ as the } X_i \text{ are identically distributed}\\ &=: (\hat{p}_X(t))^n\end{aligned}$$

where $\hat{p}_X$ denotes the generating function of the random variable X.

2.4.1 Computations involving generating functions

Let $(S_n)_{n\geqslant 1}$ be a simple random walk and define

- $p_0(n) := \mathbb{P}_0(S_n = 0)$
- $\tau_0 = \min\{n \geqslant 1\colon S_n = 0\}$
- $f_0(n) := \mathbb{P}_0(\tau_0 = n) = \mathbb{P}_0(S_1 \neq 0, \ldots, S_{n-1} \neq 0, S_n = 0)$.

We can compute

$$p_0(n) = \mathbb{P}_0(S_n = 0) = \binom{n}{n/2} p^{n/2}(1-p)^{n/2}\mathbb{1}_{\{n \text{ even}\}}$$

because in order to hit 0 at time n, $\#\{\text{steps up}\} = \#\{\text{steps down}\} = n/2$. If n is odd, then $p_0(n) = 0$. However, $f_0(\cdot)$ is less easy to compute. We'll do this by setting up an equation:

Lemma 1 It holds that $p_0(n) = \sum_{k=1}^{n} f_0(k) p_0(n-k)$.

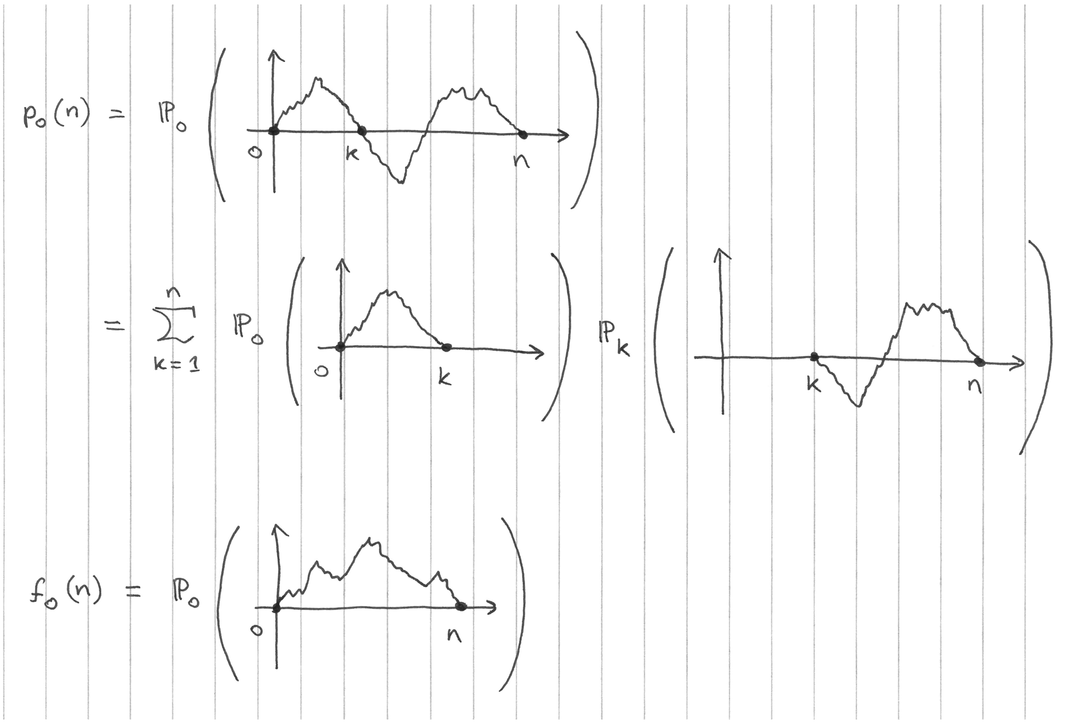


Figure 3: Pictorial representations of p_0 and f_0.

This is difficult to solve for $f_0(k)$ so we'll transform it by using generating functions. To do so, multiply both sides by s^n and sum over n. For the series to converge, we have to choose $|s| < 1$.

$$
\begin{aligned}
\hat{p}_0(s) &:= \sum_{n=0}^{\infty} s^n p_0(n) \\
&= p_0(0) + \sum_{n=1}^{\infty} s^n p_0(n) \\
&= 1 + \sum_{n=1}^{\infty} s^n \sum_{k=1}^{n} f_0(k) p_0(n-k) \\
&= 1 + \sum_{n=1}^{\infty} \sum_{k=1}^{n} s^k f_0(k) \cdot s^{n-k} p_0(n-k) \\
&= 1 + \sum_{k=1}^{\infty} \sum_{n=k}^{\infty} s^k f_0(k) \cdot s^{n-k} p_0(n-k) \\
&= 1 + \sum_{k=1}^{\infty} s^k f_0(k) \sum_{n=k}^{\infty} s^{n-k} p_0(n-k) \\
&= 1 + \sum_{k=1}^{\infty} s^k f_0(k) \sum_{n=0}^{\infty} s^n p_0(n) \\
&=: 1 + \hat{f}_0(s) \hat{p}_0(s).
\end{aligned}
$$

Thus, we've derived the equation $\hat{p}_0(s) = 1 + \hat{f}_0(s)\hat{p}_0(s)$ which is trivial to solve:

$$\hat{f}_0(s) = \frac{\hat{p}_0(s) - 1}{\hat{p}_0(s)}. \tag{2}$$

Inverting $\hat{f}_0(s)$ to get $f_0(n)$, though possible, is not trivial. Nevertheless, we obtain useful information from our solution for $\hat{f}_0(s)$. For example, we can compute $\hat{f}_0(1)$ by taking limits:

$$\begin{aligned}
\hat{f}_0(s) &:= \lim_{s\uparrow 1} \sum_{n=0}^{\infty} s^n f_0(n) \\
&= \sum_{n=0}^{\infty} \lim_{s\uparrow 1} s^n f_0(n) \\
&= \sum_{n=0}^{\infty} f_0(n) \\
&= \sum_{n=0}^{\infty} \mathbb{P}_0(\tau_0 = n) = \mathbb{P}_0(\tau_0 < \infty)
\end{aligned}$$

Thus, $\hat{f}_0(1)$ gives the probability that the random walk will return to 0 in finite time. Furthermore, we have from (2) that

$$\hat{f}_0(1) = \lim_{s\uparrow 1} \frac{\hat{p}_0(s) - 1}{\hat{p}_0(s)} \tag{3}$$

Now we focus on computing $\hat{p}_0(s)$:

$$\begin{aligned}
\hat{p}_0(s) &:= \sum_{n=0}^{\infty} s^n p_0(n) \\
&= \sum_{n=0}^{\infty} s^n \binom{n}{n/2} p^{n/2}(1-p)^{n/2} \mathbb{1}_{n \text{ even}} \\
&\overset{n=2k}{=} \sum_{k=0}^{\infty} (s^2 p(1-p))^k \binom{2k}{k} \\
&= \frac{1}{\sqrt{1 - 4s^2(1-p)p}}
\end{aligned}$$

Therefore, $\hat{p}_0(1) = \dfrac{1}{\sqrt{1-4(1-p)p}}$. If $p \neq 1/2$, $\hat{p}_0(1) < \infty$. However, if $p = 1/2$ then $\hat{p}_0(1) = \infty$.

Substituting back into equation (3), we have that

(i) $p = 1/2 \implies \mathbb{P}_0(\tau_0 < \infty) =: \hat{f}_0(1) = \dfrac{\hat{p}_0(1) - 1}{\hat{p}_0(1)} = 1$

(ii) $p \neq 1/2 \implies \mathbb{P}_0(\tau_0 < \infty) = 1 - |2p - 1| < 1$.

Remark

- In case (i), we'll say that "0 is recurrent" i.e. 0 will be revisited an infinite number of times.
- In case (ii), we'll say that "0 is invariant" i.e. 0 will be visited only finitely many times.

It's important to know that although we know we'll return to 0 at some point, it may be an "infinite" amount of time/number of steps before we do.

We can also compute the expected return time $\mathbb{E}_0[\tau_0]$:

$$\begin{aligned}\frac{\mathrm{d}\hat{f}(s)}{\mathrm{d}s}\Big|_{s=1} &= \frac{\mathrm{d}}{\mathrm{d}s}\left(\sum_{n=1}^{\infty} s^n f_0(n)\right)\Big|_{s=1} \\ &= \sum_{n=1}^{\infty} n f_0(n) \\ &= \sum_{n=1}^{\infty} n\ \mathbb{P}_0(\tau_0 = n) \\ &= \mathbb{E}[\tau_0 \mathbb{1}_{\tau_0<\infty}]\end{aligned}$$

Using our expression for $\hat{f}_0(s)$, we compute its derivative at $s = 1$ as:

$$\hat{f}_0'(s)\Big|_{s=1} = \frac{4p(1-p)}{\sqrt{1-4s^2p(1-p)}}\Bigg|_{s=1} = \frac{4p(1-p)}{\sqrt{1-4p(1-p)}}$$

We already know that $p = 1/2$ means that $\mathbb{P}_0(\tau_0 < \infty) = 1$. Therefore,

$$\begin{aligned}\mathbb{E}_0[\tau_0] &= \mathbb{E}_0[\tau_0\left(\mathbb{1}_{\tau_0<\infty} + \mathbb{1}_{\tau_0=\infty}\right)] \\ &= \mathbb{E}_0[\tau_0 \mathbb{1}_{\tau_0<\infty}] + \mathbb{E}_0[\mathbb{1}_{\tau_0=\infty}] \\ &= \mathbb{E}_0[\tau_0 \mathbb{1}_{\tau_0<\infty}] + \infty\cdot\underbrace{\mathbb{P}_0(\tau_0=\infty)}_{=0} \\ &= \mathbb{E}_0[\tau_0 \mathbb{1}_{\tau_0<\infty}] \\ &= \hat{f}'(1) = \infty\end{aligned}$$

- $\mathbb{E}_0[\tau_0] = \infty$ will be referred to as the state 0 being **null recurrent**.
- $\mathbb{E}_0[\tau_0] < \infty$ will be referred to as the state 0 being **positive recurrent**".

In the case that $p \neq 1/2$, we have that $\mathbb{E}_0[\tau_0 \mathbb{1}_{\tau_0<\infty}] = \dfrac{4p(1-p)}{\sqrt{1-4p(1-p)}}$.

Furthermore,

$$\begin{aligned}\mathbb{E}_0[\tau_0] &= \mathbb{E}_0[\tau_0 \mathbb{1}_{\tau_0<\infty}] + \mathbb{E}_0[\mathbb{1}_{\tau_0=\infty}] \\ &\geqslant \mathbb{E}_0[\tau_0 \mathbb{1}_{\tau_0=\infty}] \\ &= \infty\cdot\underbrace{\mathbb{P}_0(\tau_0=\infty)}_{>0} = \infty.\end{aligned}$$

If the simple random walk is asymmetric, we may never return back to 0 i.e. we expect that it may take an infinite amount of time. It's also important to note that if a state is transient, it's also null recurrent.

e.g. Let $|s| < 1$. We wish to compute the generating function $\mathbb{E}_1[s^{\tau_0}]$ of $\tau_0 := \min\{\mathbf{n} \geqslant \mathbf{0} \colon S_n = 0\}$.

Notice that we've redefined the starting time for τ_0 but the starting location is 1 instead of 0. This change of state makes the method developed earlier not applicable. We'll compute the generating function by deriving a difference equation starting at a general state $x > 0$.

$$u(x) := \mathbb{E}_x[s^{\tau_0}]$$

The intuition is the same as before. You can either go up or down a step.

We'll also use that $(*)$ τ_0 should be thought of as a function of the random walk i.e. $\tau_0(S_0, S_1, S_2, \dots)$.

$$\begin{aligned}
u(x) &:= \mathbb{E}_x[s^{\tau_0}] \\
&= \mathbb{E}_x[s^{\tau_0} : S_1 = x+1] + \mathbb{E}_x[s^{\tau_0} : S_1 = x-1] \\
&= \mathbb{E}_x[s^{\tau_0} \mid S_1 = x+1] \cdot \mathbb{P}_x(S_1 = x+1) + \mathbb{E}_x[s^{\tau_0} \mid S_1 = x-1] \cdot \mathbb{P}_x(S_1 = x-1) \text{ by conditioning} \\
&= \mathbb{E}_x[s^{\tau_0(S_0,S_1,\dots)} \mid S_1 = x+1] \cdot p + \mathbb{E}_x[s^{\tau_0(S_0,S_1,\dots)} \mid S_1 = x-1] \cdot (1-p) \\
&= \mathbb{E}_x[s^{1+\tau_0(S_1,S_2,\dots)} \mid S_1 = x+1] \cdot p + \mathbb{E}_x[s^{1+\tau_0(S_1,S_2,\dots)} \mid S_1 = x-1] \cdot (1-p) \text{ by the Markov property}
\end{aligned}$$

Going back to the equation, we have that

$$\begin{aligned}
u(x) &= s\mathbb{E}_x[s^{\tau_0(S_1,S_2,\dots)} \mid S_1 = x+1] \cdot p + s\mathbb{E}_x[s^{\tau_0(S_1,S_2,\dots)} \mid S_1 = x-1] \cdot (1-p) \\
&= s\mathbb{E}_x[s^{\tau_0(S_0,S_1,\dots)} \mid S_1 = x+1] \cdot p + s\mathbb{E}_x[s^{\tau_0(S_0,S_1,\dots)} \mid S_1 = x-1] \cdot (1-p) \text{ by the Markov property} \\
&= ps \cdot \mathrm{u}(x+1) + (1-p)s \cdot \mathrm{u}(x-1)
\end{aligned}$$

As before, we need boundary conditions

- $u(0) = \mathbb{E}_0[s^{\tau_0}] = \mathbb{E}_0[s^0] = 1$
- $u(\infty) = \mathbb{E}_\infty[s^{\tau_0}] = \mathbb{E}_\infty[s^\infty] = \mathbb{E}_\infty[0] = 0 \quad \because |s| < 1$

Thus, we need to solve the boundary value problem:

$$\begin{cases} u(x) = ps \cdot \mathrm{u}(x+1) + (1-p)s \cdot \mathrm{u}(x-1), & x > 0 \\ u(0) = 1 \\ u(\infty) = 0 \end{cases}$$

The solutions will again be of the form t^x and upon substitution, we obtain the equation $pst^2 - t + (1-p)s = 0$ which has solutions

$$t_{1,2} = \frac{1 \pm \sqrt{1 - 4s^2p(1-p)}}{2ps}.$$

The solution has the form $u(x) = At_1^x + Bt_2^x$ and with the boundary conditions, $A = 0$ and $B = 1$ so

$$u(x) = t_2^x = \left(\frac{1 - \sqrt{1 - 4s^2p(1-p)}}{2ps}\right)^x$$

As an example, we can use the above formula to find that

$$\begin{aligned}
\frac{1 - \sqrt{1 - 4p(1-p)}}{2p} \; = \; \lim_{s \uparrow 1} \mathbb{E}_x[s^{\tau_0}] &= \lim_{s \uparrow 1} \left\{ \mathbb{E}_x[s^{\tau_0} : \tau_0 < \infty] + \cancelto{0}{\mathbb{E}_x[s^{\tau_0} : \tau_0 = \infty]} \right\} \\
&= \lim_{s \uparrow 1} \mathbb{E}_x[s^{\tau_0} : \tau_0 < \infty] \\
&= \mathbb{E}_x[\lim_{s \uparrow 1} s^{\tau_0} : \tau_0 < \infty] \text{ by the DCT since all the } s^{\tau_0} \text{ are bounded} \\
&= \mathbb{E}_x[\mathbb{1}_{\tau_0 < \infty}] \\
&= \mathbb{P}_x(\tau_0 < \infty).
\end{aligned}$$

2.5 Branching Processes

Let X be the integer-valued ($\mathbb{Z}_{\geqslant 0}$) random variable denoting the number of offspring of an individual. Let X_k^j denote the number of offspring of the k^{th} person in the j^{th} generation. The collection $(X_k^j)_{k=1,2,3,\dots}^{j=0,1,2,\dots}$ is independent and identically distributed to X.

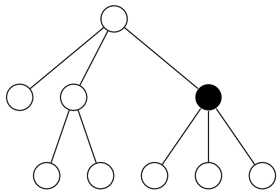


Figure 4: An example of a branching process where the highlighted node is represented by $X_3^1 = 3$.

Let Z_n be the random variable describing the number of individuals in generation n. This random variable has a recursive nature described by

$$Z_{n+1} = \sum_{i=1}^{Z_n} X_i^n.$$

We can now ask the question of whether the genealogy will become extinct or survive ad infinitum. By introducing the random variable Z_n that describes the number of individuals in generation n, we can reformulate the question to finding out what $\mathbb{P}(\{Z_n = 0 \text{ eventually}\})$ is:

$$\begin{aligned}
\mathbb{P}(\{Z_n = 0 \text{ eventually}\}) &= \mathbb{P}\left(\bigcup_n \{Z_n = 0\}\right) \\
&= \mathbb{P}\left(\bigcup_n \bigcap_{m \geqslant n} \{Z_m = 0\}\right) \\
&= \lim_{n\to\infty} \mathbb{P}(\{Z_n = 0\}) \text{ by monotonicity.}
\end{aligned}$$

So we're interested in computing $\eta := \lim_{n\to\infty} \mathbb{P}(\{Z_n = 0\})$. We can do this by using moment generating functions. Define

$$\begin{aligned}
\hat{p}_{Z_n}(t) &:= \mathbb{E}[t^{Z_n}] \\
&= \mathbb{E}[t^{Z_n}, Z_n = 0] + \mathbb{E}[t^{Z_n}, Z_n \neq 0] \\
&= \mathbb{P}(Z_n = 0) + \mathbb{E}[t^{Z_n}, Z_n \neq 0]
\end{aligned}$$

and taking the limit as $t \downarrow 0$ gives

$$\begin{aligned}
\hat{p}_{Z_n}(0) &= \mathbb{P}(Z_n = 0) + \lim_{t\downarrow 0} \mathbb{E}[t^{Z_n}, Z_n \neq 0] \\
&= \mathbb{P}(Z_n = 0) + \mathbb{E}[\lim_{t\downarrow 0} t^{Z_n}, Z_n \neq 0] \\
&= \mathbb{P}(Z_n = 0) + \mathbb{E}[0, Z_n \neq 0] \\
&= \mathbb{P}(Z_n = 0).
\end{aligned}$$

We'll use the recursive nature of Z_n to find a recursion for $\hat{p}_{Z_n}(t)$.

First of all, $\hat{p}_{Z_1}(t) := \mathbb{E}[t^{Z_1}] = \mathbb{E}[t^X] = \hat{p}_X(t)$ and:

$$
\begin{aligned}
\hat{p}_{Z_{n+1}}(t) &:= \mathbb{E}[t^{Z_{n+1}}] = \mathbb{E}[t^{X_1^n + \cdots + X_{Z_n}^n}] \\
&= \mathbb{E}[\mathbb{E}[t^{X_1^n + \cdots + X_{Z_n}^n} \mid Z_n]] \\
&= \sum_k \mathbb{E}[t^{X_1^n + \cdots + X_{Z_n}^n} \mid Z_n = k]\mathbb{P}(Z_n = k) \\
&= \sum_k \mathbb{E}[t^{X_1^n + \cdots + X_k^n}]\mathbb{P}(Z_n = k) \\
&= \sum_k \mathbb{E}[t^X]^k \mathbb{P}(Z_n = k) \quad \text{by independence and identical distribution} \\
&= \sum_k \hat{p}_X(t)^k \mathbb{P}(Z_n = k) \\
&= \mathbb{E}[\hat{p}_X(t)^{Z_n}] \\
&= \hat{p}_{Z_n}(\hat{p}_X(t))
\end{aligned}
$$

So we conclude that $\hat{p}_{Z_{n+1}}(t) = \hat{p}_{Z_n}(\hat{p}_X(t))$.

We can iterate this relation to obtain $\hat{p}_{Z_{n+1}}(t) = \underbrace{(\hat{p}_X \circ \cdots \circ \hat{p}_X)}_{(n+1) \text{ times}}(t) = \hat{p}_X(\hat{p}_{Z_n}(t))$.

Setting $t = 0$ to obtain $\hat{p}_{Z_{n+1}}(0) = \hat{p}_X(\hat{p}_{Z_n}(0))$ and letting $n \to \infty$ gives $\eta = \hat{p}_X(\eta)$. Since $\hat{p}_X$ is an expectation, we must justify passing a limit inside to the argument as $n \to \infty$. This can be done with the Dominated Convergence Theorem. Thus, $\eta := \mathbb{P}(Z_n = 0 \text{ eventually})$ is a fixed point of $\hat{p}_X$. We cannot solve it exactly but we can make some progress via numerical methods.

Remark

- It's important to note that in general, if $(A_n)_{n \geqslant 1}$ is a sequence of events, then $\mathbb{P}(A_k \text{ happens eventually}) \neq \lim_{n\to\infty} \mathbb{P}(A_n)$ but we know that if $z_n = 0$, then for all $k \geqslant n$, $z_k = 0$. This is actually a property of measures (of which $\mathbb{P}$ is an example) called upward monotone convergence/continuity from below.
- We couldn't have computed the moment generating function of Z_{n+1} by regular means i.e. as

$$
\hat{p}_{Z_{n+1}}(t) := \mathbb{E}[t^{Z_{n+1}}] = \mathbb{E}\left[t^{\sum_{i=1}^{Z_n} X_i^n}\right] \overset{\text{ind.}}{=} \prod_{i=1}^{Z_n} \mathbb{E}[t^{X_i^n}]
$$

because Z_n is a random variable and not a fixed number. In Nikos' words - "If something is random but you wish for it to be a fixed number, then condition it." This is the reason for the conditional expectation calculation.

2.5.1 Numerical Solutions

We can use information about $\hat{p}_X$ to figure out what it looks like graphically. This will guide us to locating any fixed point solutions η. Note that $\hat{p}_X(1) = \mathbb{E}[t^X]\Big|_{t=1} = 1$ and $t \longmapsto \hat{p}_X(t)$ is convex because $\hat{p}_X''(t) > 0$.

Solutions to the fixed point equation will lie on the curve $f(t) = t$. Thus, how many solutions we have depends on the number of intersections between $\hat{p}_X(t)$ and t. The aforementioned convexity means that we have two cases to distinguish depending on the slope of $\hat{p}_X(t)$ at $t = 1$.

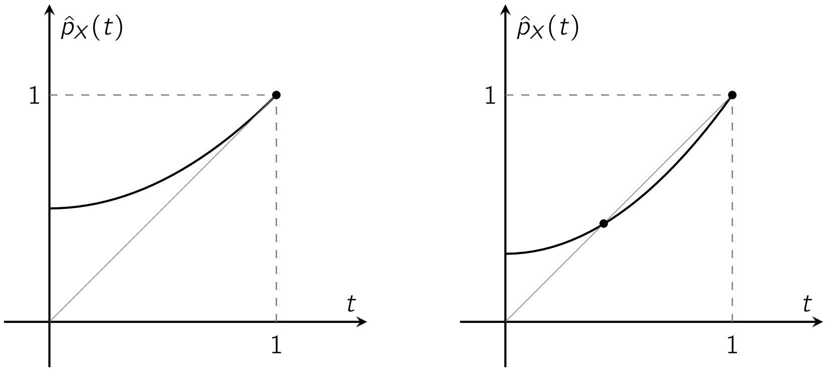


Figure 5: Two sketches of fixed point solutions of $\hat{p}_X(t)$ depending on the slope at 1: the left is $\leqslant 1$ and the right is > 1.

The slope can be computed at 1 as

$$\hat{p}_X'(1) = \frac{\mathrm{d}}{\mathrm{d}t}\mathbb{E}[t^X]\,\Big|_{t=1} = \mathbb{E}\left[\frac{\mathrm{d}}{\mathrm{d}t}t^X\right]\Big|_{t=1} = \mathbb{E}[Xt^{X-1}]\,\Big|_{t=1} = \mathbb{E}[X].$$

- If $\hat{p}_X'(1) = \mathbb{E}[X] \leqslant 1$, then the only solution of $\hat{p}_X(\eta) = \eta$ is $\eta = 1$.
- If $\hat{p}_X'(1) = \mathbb{E}[X] > 1$, then there is another[1] solution in $[0, 1]$ beside $\eta = 1$.

Numerically, the idea is to start with some initial point $\eta_0 \in (0, 1)$ and iterate the equation $\eta_{i+1} = \hat{p}_X(\eta_i)$ for $i = 0, 1, \ldots$ in order to obtain a sequence $(\eta_i)_{i\in\mathbb{N}}$ converging to some value. A natural question to ask in this case is if the sequence actually converges. Let $n, m \geqslant 0$.

$$\begin{aligned}
|\eta_{m+1} - \eta_m| = |\hat{p}_X(\eta_m) - \hat{p}_X(\eta_{m-1})| \overset{\text{MVT}}{=} & \ |\hat{p}_X'(\eta_{m-1}) \cdot (\eta_m - \eta_{m-1})| \\
& < \alpha|\eta_m - \eta_{m-1}| \\
& < \alpha^2|\eta_{m-1} - \eta_{m-2}| \\
& < \ \ldots \\
& < \alpha^m \longrightarrow 0 \ \text{ because } \ \alpha \in (0, 1).
\end{aligned}$$

Thus, $(\eta_i)_{i\in\mathbb{N}} \subset \mathbb{R}$ is Cauchy and therefore convergent because $(\mathbb{R}, |\cdot|)$ is complete.

[1]We can disregard any solutions greater than 1 because $\eta = \lim_{n\to\infty} \mathbb{P}(Z_n = 0) \leqslant 1$.

3 Markov Processes

3.1 Formalisms

Let $(X_n)_{n\geqslant 1}$ be a stochastic process. Let I be a countable set.

To define a Markov process and describe its behaviour, we intuitively need know only two pieces of information:

(1) The nature of the start of the process: We may start from a random location so we may not know the value of X_0 but we often know its distribution, the initial distribution of the process.

(2) How we move from one state to the next: We'll define this through objects called transition probabilities.

Definition 3.1

- Each $i \in I$ is called a **state** and I is called the **state-space**.
- We say that $\lambda = (\lambda_i : i \in I)$ is a **measure** on I if $0 \leqslant \lambda_i < \infty$ for all $i \in I$. If, in addition, the total mass $\sum_{i\in I} \lambda_i = 1$, then we call λ a **distribution**.
- For a random variable $X : \Omega \to I$, suppose that we set

$$\lambda_i = \mathbb{P}(X = i) = \mathbb{P}(\{\omega : X(\omega) = i\}).$$

Then λ defines a distribution, the **distribution of** X. We think of X as modelling a random state which takes the value i with probability λ_i.

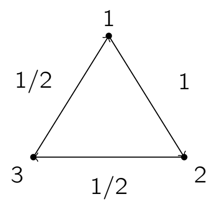
Definition 3.2 A matrix $P = (P_{ij} : i, j \in I)$ is called **stochastic** if every row is a distribution i.e. for all $i, j \in I$:

- $\displaystyle\sum_{j\in I} P_{ij} = 1$
- $P_{i,j} \in [0, 1]$.

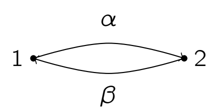
There is a one-to-one correspondence between stochastic matrices and state diagrams like those we'll see below. We realise stochastic matrices in terms of transition probabilities. For example, the probability to move from state 3 at $t = 0$ to state 1 at $t = 1$ in the first diagram below is equal to $1/2$. We'll formalise this properly after some more exposition.

e.g. Consider the following state diagrams and their corresponding stochastic matrices:

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



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



Thus, we can formalise the rules for a Markov chain with a definition involving the corresponding matrices P:

Definition 3.3 A process $(X_n)_{n\geqslant 0}$ is called a **Markov chain** with **initial distribution** λ and **transition matrix** P if

(i) X_0 has distribution λ;

(ii) for $n \geqslant 0$, conditional on $X_n = i$, X_{n+1} has distribution $(P_{ij} : j \in I)$ and is independent of $X_0, \cdots, X_{n-1}$.

More explicitly, these conditions state that for $n \geqslant 0$ and $i_0, \cdots, i_{n+1} \in I$:

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

(ii) $\mathbb{P}(X_{n+1} = i_{n+1} \mid X_0 = i_0, \cdots, X_n = i_n) = P_{i_n i_{n+1}}$.

For short, we say that $(X_n)_{n\geqslant 0}$ is Markov (λ, P).

Theorem 1 A discrete-time random process $(X_n)_{n\geqslant 0}$ is Markov (λ, P) iff $\forall i_0, \ldots, i_n \in I$:

$$\mathbb{P}(X_0 = i_0, \ldots, X_n = i_n) = \lambda_{i_0} \prod_{i=1}^{n} P_{i-1,i}$$

where λ is the initial distribution and P is the probability matrix.

Note that each entry in $\lambda = (\lambda_i)$ with $\lambda_i \geqslant 0$ and $\sum \lambda_i = 1$ is the probability that X is at position i.

Proof. The forward implication begins with supposing that X_i is a Markov (λ, P) chain.

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

$$\begin{aligned}\mathbb{P}(X_0 = i_0, X_1 = i_1) &= \mathbb{P}(X_1 = i_1 \mid X_0 = i_0) \cdot \mathbb{P}(X_0 = i_0)\\ &= P_{i_0,i_1}\lambda_{i_0}\end{aligned}$$

$$\begin{aligned}\mathbb{P}(X_0 = i_0, X_1 = i_1, X_2 = i_2) &= \mathbb{P}(X_2 = i_2 \mid X_0 = i_0, X_1 = i_1) \cdot \mathbb{P}(X_0 = i_0, X_1 = i_1)\\ &= \mathbb{P}(X_2 = i_2 \mid X_0 = i_0, X_1 = i_1) P_{i_0,i_1}\lambda_{i_0}\\ &= \mathbb{P}(X_2 = i_2 \mid X_1 = i_1) P_{i_0,i_1}\lambda_{i_0} \quad \text{by the Markov property}\\ &= \lambda_{i_0} P_{i_0,i_1} P_{i_1,i_2}\end{aligned}$$

The general case is given by

$$\begin{aligned}\mathbb{P}(X_0 = i_0, X_1 = i_1, \ldots, X_n = i_n) &= \mathbb{P}(X_n = i_n \mid X_0 = i_0, \ldots, X_{n-1} = i_{n-1}) \cdot \mathbb{P}(X_0 = i_0, \ldots, X_{n-1} = i_{n-1})\\ &= P_{i_{n-1},i_n} \cdot \ldots \cdot P_{i_0,i_1}\lambda_{i_0} \quad \text{by the Markov property}\\ &= \lambda_{i_0} \prod_{j=1}^{n} P_{i_{j-1},i_j}\end{aligned}$$

The reverse implication is as follows:

$$\begin{aligned}\mathbb{P}(X_n = i_n \mid X_{n-1} = i_{n-1}, \ldots, X_0 = i_0) &= \frac{\mathbb{P}(X_n = i_n, \ldots, X_0 = i_0)}{\mathbb{P}(X_{n-1} = i_{n-1}, \ldots, X_0 = i_0)}\\ &= \frac{\lambda_{i_0} \prod_{j=1}^{n} P_{i_{j-1},i_j}}{\lambda_{i_0} \prod_{j=1}^{n-1} P_{i_{j-1},i_j}} = P_{i_{n-1},i_n}\end{aligned}$$

Thus, X_i is a Markov (λ, P) chain. $\square$

The next result reinforces the idea that a Markov chain has no memory. Write $\delta_i = (\delta_{ij} : j \in I)$ for the unit mass at i, where

$$\delta_{ij} = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{otherwise.} \end{cases}$$

3.2 Markov Property

Theorem 2 Let $(X_n)_{n\geqslant 0}$ be Markov (λ, P). Then, conditional on $X_m = i$, $(X_{m+n})_{n\geqslant 0}$ is Markov (δ_i, P) and is independent of the random variables $X_0, \ldots, X_m$.

Proof. The goal is to show that for any event A determined by $X_0, \ldots, X_m$, we have that

$$\mathbb{P}(\{X_m = i_m, \ldots, X_{m+n} = i_{m+n}\} \cap A \mid X_m = i) = \delta_{i,i_m} P_{i_m,i_{m+1}} \cdot \ldots \cdot P_{i_{m+n-1},i_{m+n}} \mathbb{P}(A \mid X_m = i).$$

The result will thusly follow from the prior theorem. We'll begin by considering the case of elementary events $A = \{X_0 = i_0, \ldots, X_m = i_m\}$. By the prior theorem, we have that

$$\frac{\mathbb{P}(\{X_0 = i_0, \ldots, X_{m+n} = i_{m+n} \text{ and } i = i_m\})}{\mathbb{P}(X_m = i)} = \frac{\delta_{i,i_m} P_{i_m,i_{m+1}} \cdot \ldots \cdot P_{i_{m+n-1},i_{m+n}} \times \mathbb{P}(X_0 = i_0, \ldots, X_m = i_m \text{ and } i = i_m)}{\mathbb{P}(X_m = i)}.$$

Since any event A determined by $X_0, \ldots, X_m$ can be written as a countable disjoint union of elementary events $A = \bigsqcup_{k=1}^{\infty} A_k$, the desired identity for A holds by summing up the corresponding identities for the A_k. $\square$

The rest of this section concerns the following question: What is the probability that after n steps, our Markov chain is in a given state? In other words, what is the value of $\mathbb{P}(X_n = i \mid X_0 = j)$?

Notation

- We regard distributions and measures λ as row vectors whose components are indexed by I, just as P is a matrix whose entries are indexed by $I \times I$.
- Thus, we can define a new measure λP by straight-forward matrix multiplication. This works for infinite matrices and infinite row vectors as well.
- We'll write $P_{i,j}^{(n)} = (P^n)_{i,j}$ for the $(i,j)^{\text{th}}$ entry of P^n.
- In the case where $\lambda_i > 0$, we'll write $\mathbb{P}_i(A)$ for the conditional probability $\mathbb{P}(A \mid X_0 = i)$.

By the Markov property at time $m = 0$, under $\mathbb{P}_i$, $(X_n)_{n\geqslant 0}$ is Markov (δ_i, P) so the behaviour of $(X_n)_{n\geqslant 0}$ under $\mathbb{P}_i$ doesn't depend on λ.

Theorem 3 Let $(X_n)_{n\geqslant 0}$ be Markov (λ, P). Then, for all $n, m \geqslant 0$:

(i) $\mathbb{P}(X_n = j) = (\lambda P^n)_j$;

(ii) $\mathbb{P}_i(X_n = j) = \mathbb{P}(X_{n+m} = j \mid X_m = i) = P_{i,j}^{(n)}$.

Proof.

(i) By theorem 1, we have that

$$\begin{aligned}\mathbb{P}(X_n = j) &= \sum_{i_0 \in I} \dots \sum_{i_{n-1} \in I} \mathbb{P}(X_0 = i_0, \dots, X_{n-1} = i_{n-1}, X_n = j) \\ &= \sum_{i_0 \in I} \dots \sum_{i_{n-1} \in I} \lambda_{i_0} P_{i_0,i_1} \cdot \dots \cdot P_{i_{n-1},j} \\ &= (\lambda P^n)_j\end{aligned}$$

(ii) By the Markov property, conditional on $X_m = i$, $(X_{m+n})_{n \geqslant 0}$ is Markov (δ_i, P) so we just take $\lambda = \delta_i$ in (i).

□

In light of this theorem, we call $P_{i,j}^{(n)}$ the **n-step transition probability from state i to state j**.

Consider the second 2-state diagram from the beginning of this chapter. Note that

$$\mathbb{P}_1(X_n = 1) = \begin{cases} 1 & \text{if } n = 0 \\ 1 - \alpha & \text{if } n = 1 \\ (1-\alpha)^2 + \alpha\beta & \text{if } n = 2 \\ ? & \text{for } n > 2. \end{cases}$$

How can one compute $\mathbb{P}_1(X_n = 1)$ for $n > 2$? Note that:

$$P^2 = \begin{pmatrix} (1-\alpha)^2 + \alpha\beta & \alpha(1-\alpha) + \alpha(1-\beta) \\ \beta(1-\alpha) + (1-\beta)\beta & \alpha\beta + (1-\beta)^2 \end{pmatrix}$$

In general, we can see that $\mathbb{P}_1(X_n = 1) = P_{1,1}^n$ so we need to compute P^n:

$$\begin{aligned}P^{n+1} = P^n P = \begin{pmatrix} P_{1,1}^{(n+1)} & P_{1,2}^{(n+1)} \\ P_{2,1}^{(n+1)} & P_{2,2}^{(n+1)} \end{pmatrix} &= \begin{pmatrix} P_{1,1}^{(n)} & P_{1,2}^{(n)} \\ P_{2,1}^{(n)} & P_{2,2}^{(n)} \end{pmatrix} \begin{pmatrix} 1-\alpha & \alpha \\ \beta & 1-\beta \end{pmatrix} \\ &= \begin{pmatrix} (1-\alpha)P_{1,1}^{(n)} + \beta P_{1,2}^{(n)} & \alpha P_{1,1}^{(n)} + (1-\beta)P_{1,2}^{(n)} \\ (1-\alpha)P_{2,1}^{(n)} + \beta P_{2,2}^{(n)} & \alpha P_{2,1}^{(n)} + (1-\beta)P_{2,2}^{(n)} \end{pmatrix}\end{aligned}$$

This tells us that the $(1,1)$ entry in the P^{n+1} matrix is given by the recursive formula

$$P_{1,1}^{(n+1)} = (1-\alpha)P_{1,1}^{(n)} + \beta P_{1,1}^{(n)}.$$

This is a non-closed equation so, in principle, it cannot be solved on its own. However, in this case we can close it because $P_{1,2}^{(n)} = 1 - P_{1,1}^{(n)}$ which implies that

$$P_{1,1}^{(n+1)} = (1-\alpha-\beta)P_{1,1}^{(n)} + \beta \quad \text{where } P_{1,1}^{(0)} = 1.$$

This is an inhomogeneous recursive equation of order 1.

We can solve equations like these by:

(1) Find the general solution to the homogeneous equation $P_{1,1}^{(n+1)} = (1-\alpha-\beta)P_{1,1}^{(n)}$.

(2) Find a special solution to the inhomogeneous equation (by guessing).

(3) Finally, form a linear combination of the two and use initial/boundary conditions to determine the constants.

When we do this, the general solution of the equation that describes our process is given by

$$P_{1,1}^{(n)} = 1 \cdot \frac{\alpha}{\alpha+\beta} + \frac{\alpha}{\alpha+\beta} \cdot (1-\alpha-\beta)^n.$$

$$\therefore \ \mathbb{P}_1(X_n = 1) = (P^n)_{1,1} =: P_{1,1}^{(n)} = \begin{cases} 1 & \text{if } n = 0 \\ 1-\alpha & \text{if } n = 1 \\ (1-\alpha)^2 + \alpha\beta & \text{if } n = 2 \\ \dfrac{\alpha}{\alpha+\beta} + \dfrac{\alpha}{\alpha+\beta}(1-\alpha-\beta)^n & \text{if } n > 2. \end{cases}$$