

# Stochastic Processes (2021/22)

## Michaelmas term \*

January 11, 2022

### Contents

<b>1</b>	<b>Introduction</b>	<b>2</b>
1.1	Probability revision . . . . .	2
1.2	Generating functions: key properties . . . . .	7
<b>2</b>	<b>Branching processes</b>	<b>15</b>
2.1	Classification and extinction . . . . .	15
2.2	Critical case $m = 1$ . . . . .	20
2.3	Non-homogeneous case . . . . .	23
<b>3</b>	<b>Coupling</b>	<b>25</b>
3.1	Stochastic domination . . . . .	25
3.2	Total variation distance . . . . .	27
3.3	Applications to convergence . . . . .	29
3.3.1	Coupling of Markov chains . . . . .	29
3.3.2	The Law of rare events . . . . .	31
3.4	Additional problems . . . . .	32
<b>4</b>	<b>Martingales</b>	<b>33</b>
4.1	Definition and some examples . . . . .	33
4.2	A few remarks on conditional expectation . . . . .	34
4.3	Martingales and stopping times . . . . .	36
4.4	Optional stopping theorem . . . . .	38
4.5	Martingale convergence theorem . . . . .	42
4.6	Additional problems . . . . .	44

---

\*current version by Ellen Powell (ellen.g.powell@durham.ac.uk); based on previous notes by Nicholas Georgiou and Ostap Hryniv

# 1 Introduction

Here we shall briefly recall the most important facts from basic probability theory; for a more detailed recap, check your notes from the previous years.

## 1.1 Probability revision

### Probability triple

A probability triple is a collection  $(\Omega, \mathcal{F}, \mathbb{P})$ , where:

- $\Omega$  is a **sample space** (eg., the collection  $\{H, TH, TTH, \dots\}$  of all possible results in a coin tossing experiment until the first occurrence of H);
- the  $\sigma$ -field  $\mathcal{F}$  is the collection of all **events** under consideration (more precisely it is a set of subsets of  $\Omega$  satisfying (i)  $\emptyset \in \mathcal{F}$  (ii)  $A \in \mathcal{F} \Rightarrow \Omega \setminus A \in \mathcal{F}$  & (iii) if  $A_k \in \mathcal{F}$  for all  $k \geq 1$ , then  $\cup_{k \geq 1} A_k \in \mathcal{F}$ ); and
- the  $\sigma$ -additive<sup>1</sup> probability measure  $\mathbb{P} : \mathcal{F} \rightarrow [0, 1]$  assigns probabilities to all these events in a consistent way. That is,

$$\mathbb{P} : A \mapsto \mathbb{P}(A) \in [0, 1], \quad \mathbb{P}(\emptyset) = 0, \quad \mathbb{P}(\Omega) = 1, \text{ and}$$

$$(A_k)_{k \geq 1} \in \mathcal{F} \text{ and } A_j \cap A_k = \emptyset \text{ for all } j \neq k \Rightarrow \mathbb{P}(\cup_{k \geq 1} A_k) = \sum_{k \geq 1} \mathbb{P}(A_k).$$

One immediate consequence is **monotonicity**:

$$C, D \in \mathcal{F} \quad \text{and} \quad C \subseteq D \quad \implies \quad \mathbb{P}(C) \leq \mathbb{P}(D).$$

Another important property is the following **continuity** result:

*If the events  $A_k \in \mathcal{F}$ ,  $k \geq 1$  form a monotone increasing sequence, ie.,  $A_k \subseteq A_{k+1}$  for all  $k \geq 1$ , then we say that*

$$A_k \nearrow A \stackrel{\text{def}}{=} \bigcup_{j \geq 1} A_j \text{ as } k \rightarrow \infty.$$

*We then have that*

$$\mathbb{P}(A) = \lim_{k \rightarrow \infty} \mathbb{P}(A_k)$$

The proof is a straightforward, but instructive exercise.<sup>2</sup> By taking complements, one deduces an analogous result for **monotone decreasing** sequences  $B_k \in \mathcal{F}$ ,  $k \geq 1$  of events, ie., satisfying  $B_k \supseteq B_{k+1}$  for all  $k \geq 1$ .

**Example 1.1.** *In the standard coin flipping experiment, let  $B_k$  be the event {the first  $k$  results are T}; then  $B_k \searrow B \equiv \{\text{all results are T}\}$ . If the coin shows H with probability  $p > 0$ , then  $\mathbb{P}(B_k) = (1 - p)^k \searrow 0$  as  $k \rightarrow \infty$  and, by continuity,  $\mathbb{P}(B) = 0$ .*

---

<sup>1</sup>also called countably additive;

<sup>2</sup>See your 2nd year notes.

**Exercise 1.2.** In the setup of Example 1.1, let  $A_k$  be the event

$$\{\text{no H observed from } k^{\text{th}} \text{ flip onwards}\}.$$

Show that  $\mathbb{P}(A_k) = 0$  for all  $k \geq 1$ . Verify that  $(A_k)_{k \geq 1}$  is a monotone sequence of events with the limit

$$A \equiv \{\text{at most finitely many H observed}\}.$$

Use the continuity property of probability measures to deduce that  $\mathbb{P}(A) = 0$ .

### Conditional probability

If  $A, B$  are events (ie.,  $A, B \in \mathcal{F}$ ) with  $\mathbb{P}(B) > 0$ , then the conditional probability of  $A$  given  $B$  is

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

Notice that  $\mathbb{P}(\cdot | B) : \mathcal{F} \rightarrow [0, 1]$  is a probability measure.

### Formula of total probability (partition theorem)

Events  $B_1, B_2, \dots \in \mathcal{F}$  are said to form a partition of  $\Omega$ , if they are pairwise disjoint:

$$B_i \cap B_j = \emptyset \text{ for } i \neq j,$$

and cover the whole  $\Omega$ , ie.,

$$\cup_k B_k = \Omega.$$

If  $\{B_1, B_2, \dots\}$  form a partition of  $\Omega$ , then for every  $A \in \mathcal{F}$  the following formula of total probability holds (tacitly assuming that  $\mathbb{P}(A | B_k) \mathbb{P}(B_k) \equiv \mathbb{P}(A \cap B_k) = 0$  for  $\mathbb{P}(B_k) = 0$ ):

$$\mathbb{P}(A) = \sum_{k \geq 1} \mathbb{P}(A | B_k) \mathbb{P}(B_k).$$

**Example 1.3. (Discrete Renewals)** Consider a sequence of events that can only occur at discrete times  $k = 1, 2, \dots$  (eg., a light bulb burns out and is immediately replaced with a new one). Assume that the intervals  $X$  between consecutive events have a common distribution  $f_k = \mathbb{P}(X = k)$ ,  $k \geq 1$ . Let  $r_n$  denote the probability that an event occurs at time  $n$ ; ie.,  $r_n = \mathbb{P}(A_n)$  with  $A_n = \{\text{replacement at time } n\}$ ; we shall also assume that a replacement also occurs at time 0 so that  $r_0 = 1$ . Since the events  $B_k = \{\text{first bulb burns at time } k\}$  form a countable partition of  $\Omega$ , and  $\mathbb{P}(A_n | B_k) = r_{n-k}$  for all  $n \geq k$  (with  $r_1 = f_1$ ), the partition theorem implies

$$r_n = \sum_{k=1}^n r_{n-k} f_k. \tag{1.1}$$

As  $f_0 = 0$ , the RHS above is a convolution of the sequences  $(r_n)_{n \geq 0}$  and  $(f_k)_{k \geq 0}$ ; the large  $n$  behaviour of  $r_n$  can be analysed by taking various transforms (eg., generating functions) of these sequences.

### Random variables

Random variables are “nice” functions  $X : \Omega \rightarrow \mathbb{R}$  characterized by the fact that for every  $a \in \mathbb{R}$  we have  $\{\omega \in \Omega : X(\omega) \leq a\} \in \mathcal{F}$ . In other words, every inverse image  $X^{-1}((-\infty, a])$  is an event.

For discrete random variables, ie., those attaining at most countably many values, it is often convenient to replace the previous condition with its equivalent: for every  $a \in \mathbb{R}$  the set  $X^{-1}(\{a\})$  is an event.

### Expectation

If  $X$  is a discrete random variable taking values in  $\{x_1, x_2, \dots\}$  with probabilities

$$p_k = \mathbb{P}(X = x_k) \equiv \mathbb{P}(\{\omega \in \Omega : X(\omega) = x_k\}),$$

then if  $\sum_{k \geq 1} |x_k| p_k < \infty$ , we say that  $X$  is integrable and the expectation  $\mathbb{E}(X)$  of  $X$  is given by

$$\mathbb{E}(X) \stackrel{\text{def}}{=} \sum_{k \geq 1} x_k p_k \equiv \sum_{k \geq 1} x_k \mathbb{P}(X = x_k).$$

(If  $\sum |x_k| p_k = \infty$  we also write  $\mathbb{E}(|X|) = \infty$ ).

It is clear from this definition that if  $X, Y$  are two random variables defined on the same probability space with  $\mathbb{P}(X \leq Y) = 1$  then  $\mathbb{E}(X) \leq \mathbb{E}(Y)$ . In particular:

**Example 1.4. (Markov’s inequality)** *Suppose that  $X$  is a random variable taking real values, and  $a > 0$ . Then*

$$\mathbb{E}(|X|) \leq \mathbb{E}(|X| \mathbf{1}_{X \geq a}) \leq a \mathbb{P}(X \geq a)$$

which rearranges to give

$$\mathbb{P}(|X| \geq a) \leq \frac{\mathbb{E}(|X|)}{a}. \tag{1.2}$$

*This is a very useful tool for bounding the tail probabilities of  $X$  (i.e. the LHS above with  $a$  large).*

It is also straightforward to verify that expectation is linear: if  $X_1, \dots, X_n$  are integrable random variables and  $a_1, \dots, a_n \in \mathbb{R}$  then

$$\mathbb{E}\left(\sum_i a_i X_i\right) = \sum_i a_i \mathbb{E}(X_i).$$

The conditional expectation of  $X$  given  $B \in \mathcal{F}$  with  $P(B) > 0$  is computed similarly:

$$E(X | B) = \sum_{k \geq 1} x_k P(X = x_k | B).$$

We then have the following result.

### Partition theorem for expectations

If  $\{B_1, B_2, \dots\}$  form a partition of  $\Omega$ , then for every random variable  $X$

$$E(X) = \sum_{k \geq 1} E(X | B_k) P(B_k).$$

In some cases, the RHS above might not be well defined (i.e., the partial sums of the above series may not converge). However, everything is fine if

$$\sum_{k \geq 1} |E(X | B_k)| P(B_k) < \infty$$

(and we have equality in the sense that both sides are equal to  $\infty$  if  $P(X \geq 0) = 1$ ).

Recall also that for a random variable  $X$  with  $E(X^2) < \infty$ , we define its variance

$$\text{Var}(X) = E(X^2) - E(X)^2.$$

This measures, roughly speaking, how much the random variable deviates from its expected value. For example, if  $P(X = x) = 1$  for some fixed value  $x$ , then  $\text{Var}(X) = x^2 - (x)^2 = 0$ .

### Independence

Events  $A, B \in \mathcal{F}$  are independent if

$$P(A \cap B) = P(A)P(B).$$

Notice that if  $A, B \in \mathcal{F}$  are independent and if  $P(A | B)$  is well defined, then  $P(A | B) = P(A)$ . Also, if  $B \in \mathcal{F}$  satisfies  $P(B) \in \{0, 1\}$ , then for every  $A \in \mathcal{F}$  the events  $A$  and  $B$  are independent.

In general, a collection of events  $(A_\alpha)_{\alpha \in \mathcal{A}}$  is independent if every finite subcollection is independent, i.e., for every  $k \geq 1$  and all  $\alpha_1, \dots, \alpha_k \in \mathcal{A}$ ,

$$P(A_{\alpha_1} \cap A_{\alpha_2} \cap \dots \cap A_{\alpha_k}) = P(A_{\alpha_1})P(A_{\alpha_2}) \dots P(A_{\alpha_k}).$$

Two (discrete) random variables,  $X$  and  $Y$  are independent, if for all  $x, y$ ,

$$P(X = x, Y = y) = P(X = x)P(Y = y).$$



**Example 1.8.** Let  $(w_n^X)_{n \geq 0}$  be the random walk generated by independent copies of  $X \sim \text{Ber}_{\pm 1}(p_1)$ ;  $w_0^X = 0$ . Similarly, let  $(w_n^Y)_{n \geq 0}$  be the random walk generated by independent copies of  $Y \sim \text{Ber}_{\pm 1}(p_2)$ ;  $w_0^Y = 0$ . If  $p_1 < p_2$ , then the law of large numbers implies that  $\frac{1}{n} w_n^X$  grows linearly with slope  $2p_1 - 1$ , whereas  $\frac{1}{n} w_n^Y$  grows linearly with slope  $2p_2 - 1 > 2p_1 - 1$ . In other words, for times  $n$  large enough the trajectories of  $(w_n^X)_{n \geq 0}$  will lie below those of  $(w_n^Y)_{n \geq 0}$ . In fact, using the joint distribution from Example 1.7 (B), one can construct a joint distribution for the entire trajectories of these random walks such that the inequality  $w_n^X \leq w_n^Y$  holds for all times  $n \geq 0$ , and not only for  $n$  large enough. This is useful. For example, as a result, for every monotone increasing function  $f : \mathbb{R} \rightarrow \mathbb{R}$  one has

$$f(w_n^X) \leq f(w_n^Y) \text{ for all } n \geq 0, \text{ and therefore } \text{E}f(w_n^X) \leq \text{E}f(w_n^Y).$$

## 1.2 Generating functions: key properties

Lengthy calculations arising from even quite straightforward counting problems can be simplified by using generating functions. Recall that the generating function of a real sequence  $(a_k)_{k \geq 0}$  is

$$G(s) = G_a(s) \stackrel{\text{def}}{=} \sum_{k=0}^{\infty} a_k s^k \quad (1.3)$$

(defined whenever the sum on the RHS converges). Similarly, the (probability) generating function of a random variable  $X$  with values in

$$\mathbb{Z}^+ \stackrel{\text{def}}{=} \{0, 1, \dots\}$$

is just the generating function of its probability mass function:

$$G(s) \equiv G_X(s) \stackrel{\text{def}}{=} \text{E}(s^X) = \sum_{k=0}^{\infty} s^k \text{P}(X = k). \quad (1.4)$$

Notice that each probability generating function satisfies

$$|G_X(s)| \leq G_X(1) = \sum_{k=0}^{\infty} \text{P}(X = k) \leq 1,$$

i.e., is well defined and finite for all (complex)  $s$  with  $|s| \leq 1$ . In particular,  $G_X(s)$  can be differentiated term-by-term any number of times in the open unit disk  $|s| < 1$ .

Generating functions are very useful when studying sums of independent random variables. Indeed, Example 1.6 implies the following important fact:

**Example 1.9.** If  $X$  and  $Y$  are independent random variables with values in  $\mathbb{Z}^+$  and  $Z = X + Y$ , then their generating functions satisfy

$$G_Z(s) = G_{X+Y}(s) = G_X(s) G_Y(s)$$

for all  $s$  such that the RHS is well defined.

**Example 1.10.** Let  $X, X_1, \dots, X_n$  be independent identically distributed random variables<sup>3</sup> with values in  $\{0, 1, 2, \dots\}$  and let  $S_n = X_1 + \dots + X_n$ . Suppose that  $G_X(s)$  is well-defined. Then

$$G_{S_n}(s) = G_{X_1}(s) \dots G_{X_n}(s) \equiv [G_X(s)]^n.$$

**Example 1.11.** Let  $X, X_1, X_2, \dots$  be i.i.d. with values in  $\{0, 1, 2, \dots\}$  and let  $N \geq 0$  be an integer-valued random variable independent of  $\{X_k\}_{k \geq 1}$ . Then  $S_N = X_1 + \dots + X_N$  has generating function

$$G_{S_N} = G_N \circ G_X \tag{1.5}$$

This is a straightforward application of the partition theorem for expectations. Alternatively, the result follows from the standard properties of conditional expectation:

$$\mathbb{E}(s^{S_N}) = \mathbb{E}(\mathbb{E}(s^{S_N} | N)) = \mathbb{E}([G_X(s)]^N) = G_N(G_X(s)).$$

In general, we say a sequence  $\mathbf{c} = (c_n)_{n \geq 0}$  is the convolution of  $\mathbf{a} = (a_k)_{k \geq 0}$  and  $\mathbf{b} = (b_m)_{m \geq 0}$  (write  $\mathbf{c} = \mathbf{a} \star \mathbf{b}$ ), if<sup>4</sup>

$$c_n = \sum_{k=0}^n a_k b_{n-k}, \quad n \geq 0. \tag{1.6}$$

**Exercise 1.12.** If  $c = a \star b$ , show that the generating functions  $G_c, G_a$ , and  $G_b$  satisfy  $G_c = G_a \times G_b$ .

**Exercise 1.13.** In the setup of Example 1.3, let  $G_f$  and  $G_r$  be the generating functions of the sequences  $\mathbf{f} = (f_k)_{k \geq 1}$  and  $\mathbf{r} = (r_n)_{n \geq 0}$ . Show that  $G_r(s) = 1/(1 - G_f(s))$  for all  $|s| \leq 1$ .

Why do we care? If the generating function  $G_a$  of  $(a_n)_{n \geq 0}$  is analytic in a neighbourhood of the origin, then there is a one-to-one correspondence between  $G_a$  and  $(a_n)_{n \geq 0}$ . Namely,  $a_k$  can be recovered from  $G_a$  via<sup>5</sup>

$$a_k = \frac{1}{k!} \frac{d^k}{ds^k} G_a(s) \Big|_{s=0} \quad \text{or} \quad a_k = \frac{1}{2\pi i} \oint_{|z|=\rho} \frac{G_a(z)}{z^{k+1}} dz, \tag{1.7}$$

for suitable  $\rho > 0$ . This result is often referred to as the uniqueness property of generating functions.

<sup>3</sup>from now on we shall often abbreviate this to just i.i.d.

<sup>4</sup>If  $X$  and  $Y$  are independent variables in  $\mathbb{Z}_+$  and  $Z = X + Y$ , their p.m.f.s satisfy this equation.

<sup>5</sup>if a power series  $G_a(s)$  is finite for  $|s| < r$  with  $r > 0$ , then it can be differentiated in the disk  $|s| < r$ ; recall that each probability generating function is analytic in the unit disk  $|s| < 1$ .



**Example 1.14.** Let  $X \sim \text{Poi}(\lambda)$  and  $Y \sim \text{Poi}(\mu)$  be independent. A straightforward computation gives  $G_X(s) = e^{\lambda(s-1)}$  for all  $s$  so that if  $Z = X + Y$ , Example 1.9 implies that

$$G_Z(s) = G_X(s)G_Y(s) = e^{\lambda(s-1)}e^{\mu(s-1)} \equiv e^{(\lambda+\mu)(s-1)}.$$

This means that  $Z$  is  $\text{Poi}(\lambda + \mu)$  distributed.

A similar argument can be used in the following exercise.

**Exercise 1.15.** If  $X \sim \text{Bin}(n, p)$  and  $Y \sim \text{Bin}(m, p)$  are independent, show that  $X + Y \sim \text{Bin}(n + m, p)$ .

Another useful property of probability generating functions is that they can be used to compute moments:

**Theorem 1.16.** If  $X$  has generating function  $G_X$ , then

$$\mathbf{E}(X(X-1)\dots(X-k+1)) = G_X^{(k)}(1)$$

where  $G_X^{(k)}(1)$  is the shorthand for  $G_X^{(k)}(1_-) \equiv \lim_{s \uparrow 1} G_X^{(k)}(s)$ , the limiting value of the  $k$ th derivative of  $G(s)$  at  $s = 1$ . Since  $s^k G_X^{(k)}(s)$  is increasing in  $s$ , the RHS above is either  $+\infty$ , or finite. In the latter case,  $X(X-1)\dots(X-k+1)$  is integrable and the equality above holds. In the former,  $X(X-1)\dots(X-k+1)$  is not integrable.

**Exercise 1.17.** Prove Theorem 1.16.

**Remark 1.18.** The quantity  $\mathbf{E}(X(X-1)\dots(X-k+1))$  is called the  $k$ th factorial moment of  $X$ . Notice also that

$$\text{Var}(X) = G_X''(1) + G_X'(1) - (G_X'(1))^2. \quad (1.8)$$

**Remark 1.19.** Notice that  $\lim_{s \nearrow 1} G_X(s) \equiv \lim_{s \nearrow 1} \mathbf{E}(s^X) = \mathbf{P}(X < \infty)$ . This allows us to check whether the random variable  $X$  is finite, if we do not know this a priori. See Example 1.26 below.

**Remark 1.20.** The fact that a probability generating function is finite at  $u = 1$  (or has a finite left derivative there) does not, in general, imply any regularity beyond the unit disk. Indeed, let  $X$  be a random variable satisfying

$$\mathbf{P}(X = k) = \frac{1}{k(k+1)} \quad \text{for all } k \geq 1,$$

and let  $G_X$  be its generating function. It is easy to check that  $G_X(1) = 1$  while  $\mathbf{E}(X) = G_X'(1_-) = \infty$ , and thus  $|G_X(u)| \leq G_X(1) = 1$  if  $|u| \leq 1$  but  $G_X(u) = \infty$  for all  $|u| > 1$ .

**Exercise 1.21.** Let  $\mathbf{P}(X = k) = 4/(k(k+1)(k+2))$  for  $k \geq 1$ . Show that the generating function  $G_X(u) = \mathbf{E}(u^X)$  satisfies  $G_X(1) = 1$ ,  $G_X'(1_-) = 2 < \infty$ , but  $G_X''(1_-) = \infty$ . Notice that in this case  $|G_X'(u)| \leq 2$  uniformly in  $|u| < 1$  while  $G_X(u) = \infty$  for all  $|u| > 1$ .

**Exercise 1.22.** Following the approach of Exercise 1.21 or otherwise, for  $m \in \mathbb{N}$  find a generating function  $G$ , which is continuous and bounded on the closed unit disk  $|u| \leq 1$  together with derivatives up to order  $m$ , while  $G(u) = \infty$  for all  $|u| > 1$ .

**Exercise 1.23.** Let  $S_N = X_1 + \dots + X_N$  be a random sum of random variables, whose generating function is  $G_{S_N}(u) \equiv G_N(G_X(u))$ , recall Example 1.11. Use Theorem 1.16 to express  $E(S_N)$  and  $\text{Var}(S_N)$  in terms of  $E(X)$ ,  $E(N)$ ,  $\text{Var}(X)$  and  $\text{Var}(N)$ . Check your result for  $E(S_N)$  and  $\text{Var}(S_N)$  by directly applying the partition theorem for expectations.

**Exercise 1.24.** A bird lays  $N$  eggs, each being pink with probability  $p$  and blue otherwise. Assuming that  $N \sim \text{Poi}(\lambda)$ , find the distribution of the total number  $K$  of pink eggs.

**Exercise 1.25.** Suppose that in a population, each mature individual produces immature offspring according to a probability generating function  $F$ .

- (a) Assume that we start with a population of  $k$  immature individuals, each of which grows to maturity with probability  $p$  and then reproduces, independently of other individuals. Find the probability generating function of the number of immature individuals in the next generation.
- (b) Find the probability generating function of the number of mature individuals in the next generation, given that there are  $k$  mature individuals in the parent generation.
- (c) Show that the distributions in a) and b) above have the same mean, but not necessarily the same variance. You might prefer to first consider the case  $k = 1$ , and then generalise.

The next example is very important for applications.

**Example 1.26.** Let  $X_k$ ,  $k \geq 1$  be i.i.d. with the common distribution

$$P(X_k = 1) = p, \quad P(X_k = -1) = q = 1 - p.$$

Define the simple random walk  $(S_n)_{n \geq 0}$  via  $S_0 = 0$  and  $S_n = X_1 + \dots + X_n$  for  $n \geq 1$  and let

$$T \stackrel{\text{def}}{=} \inf\{n \geq 1 : S_n = 1\}$$

be the first time this random walk hits 1.

To calculate the generating function  $G_T$ , write  $p_k = P(T = k)$ , so that  $G_T(s) \equiv E(s^T) = \sum_{k \geq 0} s^k p_k$ . Conditioning on the outcome of the first step, and applying the partition theorem for expectations, we get

$$G_T(s) \equiv E(s^T) = E(s^T | X_1 = 1) p + E(s^T | X_1 = -1) q = ps + qs E(s^{T_2}),$$

where  $T_2$  is the time of the first visit to state 1 starting from  $S_0 = -1$ . By partitioning on the time  $T_1$  of the first visit to 0, it follows that

$$\mathbb{P}(T_2 = m) = \sum_{k=1}^{m-1} \mathbb{P}(T_1 = k, T_2 = m).$$

where of course, on the event  $\{S_0 = -1\}$ ,

$$\mathbb{P}(T_2 = m \mid S_0 = -1) \equiv \mathbb{P}(S_1 < 1, \dots, S_{m-1} < 1, S_m = 1 \mid S_0 = -1)$$

$$\mathbb{P}(T_1 = k \mid S_0 = -1) \equiv \mathbb{P}(S_1 < 0, \dots, S_{k-1} < 0, S_k = 0 \mid S_0 = -1).$$

Notice that by translation invariance the last probability is just  $\mathbb{P}(T = k \mid S_0 = 0) = p_k$ . We also observe that

$$\mathbb{P}(T_2 = m \mid T_1 = k) = \mathbb{P}(\text{first hit 1 from 0 after } m - k \text{ steps}) \equiv p_{m-k}.$$

The partition theorem now implies that

$$\mathbb{P}(T_2 = m) = \sum_{k=1}^{m-1} \mathbb{P}(T_2 = m \mid T_1 = k) \mathbb{P}(T_1 = k) \equiv \sum_{k=1}^{m-1} p_k p_{m-k} = \sum_{k=0}^m p_k p_{m-k},$$

ie.,  $G_{T_2}(s) = (G_T(s))^2$ . We deduce that  $G_T(s)$  solves the quadratic equation  $\varphi = ps + qs\varphi^2$ , so that<sup>6</sup>

$$G_T(s) = \frac{1 - \sqrt{1 - 4pqs^2}}{2qs} = \frac{2ps}{1 + \sqrt{1 - 4pqs^2}}.$$

Finally this allows us to deduce that

$$\mathbb{P}(T < \infty) \equiv G_T(1) = \frac{1 - |p - q|}{2q} = \begin{cases} 1, & p \geq q, \\ p/q, & p < q. \end{cases}$$

In particular,  $\mathbb{E}(T) = \infty$  for  $p < q$  (because  $\mathbb{P}(T = \infty) = (q - p)/q > 0$ ). For  $p \geq q$  we obtain

$$\mathbb{E}(T) \equiv G'_T(1) = \frac{1 - |p - q|}{2q|p - q|} = \begin{cases} \frac{1}{p - q}, & p > q, \\ \infty, & p = q, \end{cases}$$

ie.,  $\mathbb{E}(T) < \infty$  if  $p > q$  and  $\mathbb{E}(T) = \infty$  otherwise.

Notice that at criticality ( $p = q = 1/2$ ), the variable  $T$  is finite with probability 1, but has infinite expectation.

**Example 1.27.** In a sequence of independent Bernoulli experiments with success probability  $p \in (0, 1)$ , let  $D$  be the first time that two consecutive successful

<sup>6</sup>by recalling the fact that  $G_T(s) \rightarrow 0$  as  $s \rightarrow 0$ ;

outcomes have occurred (i.e., if two successes occurred in the first two experiments, then  $D$  would be equal to 2).

To find the generating function of  $D$ , there are two good methods, and both involve deriving a recursion relation.

**Method 1:** For  $n \geq 2$  let us write  $d_n = \mathbb{P}(D = n)$ , and consider the events

- $A := \{(n-2) \text{ failures followed by 2 successes}\}$ ,
- $A_k := \{\text{first failure immediately preceded by a success occurs at experiment } k\}$  for  $k = 2, \dots, n-2$ , and
- $B = (A \cup \bigcup_{k=2}^{n-2} A_k)^c$ .

These form a partition of the probability space, since they are disjoint by construction, and the definition of  $B$  means that they cover the whole space. Note that for  $\{D = n\}$  to occur, it must be that one of  $A$  or  $A_2, \dots, A_{n-2}$  occurs, so that  $\mathbb{P}(\{D = n\} \cap B) = 0$ . Also it is clear that  $A \subset \{D = n\}$  so  $\mathbb{P}(\{D = n\} \cap A) = \mathbb{P}(A) = q^{n-2}p^2$ . Thus we can write

$$d_n = \mathbb{P}(D = n) = q^{n-2}p^2 + \sum_{k=2}^{n-2} \mathbb{P}(\{D = n\} \cap A_k)$$

and we are left to calculate  $\mathbb{P}(\{D = n\} \cap A_k)$  for  $k = 2, \dots, n-2$ . For this, we observe that for  $\{D = n\} \cap A_k$  to occur, it must be that the first  $(k-2)$  experiments are failures, the  $(k-1)$ st is a success, the  $k$ th is a failure again, and then for the new sequence of experiments starting from the  $(k+1)$ st, the first time that 2 consecutive successes are seen is  $(n-k)$ . By independence of the experiments, the probability of this happening is just  $q^{k-2}pqd_{n-k}$ . Hence we obtain

$$d_n = q^{n-2}p^2 + \sum_{k=2}^{n-2} q^{k-2}pq d_{n-k},$$

and a standard method implies that

$$G_D(s) = \frac{p^2s^2}{1-qs} + \frac{pqs^2}{1-qs} G_D(s), \quad \text{or} \quad G_D(s) = \frac{p^2s^2}{1-qs-pqs^2}.$$

A straightforward computation gives  $G'_D(1) = \frac{1+p}{p^2}$ , so that on average it takes 42 tosses of a standard symmetric dice until the first two consecutive sixes appear.

**Method 2:** we derive a recursion relation directly for the generating function, by conditioning on the result of the first experiment. That is, we use the partition theorem for expectation to write

$$\begin{aligned} \mathbb{E}(s^D) &= \mathbb{E}(s^D \mid \text{failure})\mathbb{P}(\text{failure}) + \mathbb{E}(s^D \mid \text{success})\mathbb{P}(\text{success}) \\ &= s\mathbb{E}(s^{D-1} \mid \text{failure})q + s\mathbb{E}(s^{D-1} \mid \text{success})p. \end{aligned}$$

Now,  $E(s^{D-1} | \text{failure}) = E(s^D)$  since the new sequence of experiments starting from the 2nd have the same distribution as the whole sequence, and observing a failure for the first experiment means we are still just asking for the first time that two consecutive successes are observed in this new sequence. However, this is not the case for  $E(s^{D-1} | \text{success})$  since if we have already seen one success this could be the first one in a consecutive pair. So for this conditional expectation we will condition again, but now on the result of the 2nd experiment. If it is a success then  $s^{D-1}$  will be  $s$  with conditional probability one, and if it is a failure then we will be starting from scratch again. Thus we obtain that  $E(s^{D-1} | \text{success}) = ps + qsE(s^D)$ . Putting this all together and writing  $E(s^D) = G_D(s)$  we see that

$$G_D(s) = qsG_D(s) + p^2s^2 + pqs^2G_D(s)$$

and can rearrange to reach the same conclusion as for method 1.

**Exercise 1.28.** *In a sequence of independent Bernoulli experiments with success probability  $p \in (0, 1)$ , let  $M$  be the first time that  $m$  consecutive successful outcomes have occurred. Using the approach of Example 1.27 or otherwise, find the generating function of  $M$ .*

**Exercise 1.29.** *In the setup of Example 1.27, show that  $d_0 = d_1 = 0$ ,  $d_2 = p^2$ ,  $d_3 = qp^2$ , and, conditioning on the value of the first outcome, that  $d_n = qd_{n-1} + pqd_{n-2}$  for  $n \geq 3$ . Use these relations to re-derive the generating function  $G_D$ .*

**Exercise 1.30.** *Use the method of Exercise 1.29, to derive an alternative solution to Exercise 1.28. Compare the resulting expectation to that in Example 1.27.*

**Exercise 1.31.** *A biased coin showing ‘heads’ with probability  $p \in (0, 1)$  is flipped repeatedly. Let  $C_w$  be the first time that the word  $w$  appears in the observed sequence of results. Find the generating function of  $C_w$  and the expectation  $E[C_w]$  for each of the following words: HH, HT, TH and TT.*

**Example 1.32.** *If  $X_n \sim \text{Bin}(n, p)$  with  $p = p_n$  satisfying  $n \cdot p_n \rightarrow \lambda$  as  $n \rightarrow \infty$ , then  $G_{X_n}(s) \equiv (1 + p_n(s - 1))^n \rightarrow \exp\{\lambda(s - 1)\}$  for every fixed  $s \in [0, 1]$ , so that the distribution of  $X_n$  converges to that of  $X \sim \text{Poi}(\lambda)$ .*

**Exercise 1.33.** *For each  $n \geq 1$  let  $Y_n = \sum_{k=1}^n X_k^{(n)}$ , where  $X_k^{(n)}$  are independent Bernoulli random variables,*

$$p_k^{(n)} \stackrel{\text{def}}{=} P(X_k^{(n)} = 1) = 1 - P(X_k^{(n)} = 0).$$

*Assume that*

$$\delta^{(n)} \stackrel{\text{def}}{=} \max_{1 \leq k \leq n} p_k^{(n)} \rightarrow 0$$

as  $n \rightarrow \infty$  and that for a positive constant  $\lambda$  we have

$$\mathbf{E}(Y_n) \equiv \sum_{k=1}^n p_k^{(n)} \rightarrow \lambda.$$

Using generating functions or otherwise, show that the distribution of  $Y_n$  converges to that of a  $\text{Poi}(\lambda)$  random variable. This result is known as the law of rare events.

More generally, we have the following continuity result:

**Theorem 1.34.** For every fixed  $n$ , suppose that the sequence  $a_{0,n}, a_{1,n}, \dots$  is a probability distribution, ie.,  $a_{k,n} \geq 0$  and  $\sum_{k \geq 0} a_{k,n} = 1$ , and let  $G_n$  be the corresponding generating function,  $G_n(s) = \sum_{k \geq 0} a_{k,n} s^k$  for all  $s$  such that the RHS converges. In order that for every fixed  $k$

$$\lim_{n \rightarrow \infty} a_{k,n} = a_k$$

it is necessary and sufficient that  $\lim_{n \rightarrow \infty} G_n(s) = G_a(s)$  for every fixed  $s \in [0, 1)$ , where  $G_a(s) = \sum_{k \geq 0} a_k s^k$ , the generating function of the limiting sequence  $(a_k)$ .

**Remark 1.35.** In the probabilistic context, the convergence above:

$$a_{k,n} \equiv \mathbf{P}(X_n = k) \xrightarrow{n \rightarrow \infty} \mathbf{P}(X = k) = a_k \text{ for each } k,$$

is known as convergence in distribution.

Why do we care? In applications one often needs to describe the distribution of a random variable, which is obtained as a result of some limiting approach (or approximation). Then Theorem 1.34 can help to simplify the argument. This method is similar to proving the central limit theorem using moment generating functions  $\mathbf{E}(\exp\{tX_n\}) \equiv G_n(e^t)$ . Notice that  $G_n(e^t)$  exists for some  $t > 0$  only if the sequence  $a_{k,n} \equiv \mathbf{P}(X_n = k)$  decays sufficiently fast, recall Remark 1.20 above.

## 2 Branching processes

### 2.1 Classification and extinction

Informally, a branching process <sup>7</sup> is described as follows, where  $\{p_k\}_{k \geq 0}$  is a fixed probability mass function (pmf).

- A population starts with a single ancestor who forms generation number 0.
- This initial individual splits into  $k$  offspring with probability  $p_k$  for each  $k \geq 0$ ; the resulting offspring constitute the first generation.
- Each of the offspring in the first generation splits independently into a random number of offspring, again according to the pmf  $\{p_k\}_{k \geq 0}$ , and the resulting offspring constitute the second generation.
- This process continues until extinction, which occurs when all the members of a generation fail to produce offspring.

This model has a number of applications in biology (eg., it can be thought as a model of population growth), physics (chain reaction in nuclear fission), queueing theory, etc. Originally it arose from a study of the likelihood of survival of family names (“how fertile must a family be to ensure that in future generations the family name will not die out?”).

Formally, let  $\{Z_{n,k}\}$ ,  $n \geq 1$ ,  $k \geq 1$ , be a family of i.i.d. random variables in  $\mathbb{Z}^+$ , each having a common probability mass function  $\{p_k\}_{k \geq 0}$ . Then the branching process  $(Z_n)_{n \geq 0}$  (generated by  $\{p_k\}_{k \geq 0}$ ) is defined by setting  $Z_0 = 1$ , and, for  $n \geq 1$ ,

$$Z_n \stackrel{\text{def}}{=} Z_{n,1} + Z_{n,2} + \cdots + Z_{n,Z_{n-1}}, \quad (2.1)$$

where the empty sum is interpreted as zero. Notice that  $Z_n$  is a Markov chain in  $\mathbb{Z}^+$ . We shall use  $\mathbf{P}(\cdot) \equiv \mathbf{P}_1(\cdot)$  and  $\mathbf{E}(\cdot) \equiv \mathbf{E}_1(\cdot)$  to denote the corresponding probability measure and the expectation operator.<sup>8</sup> If  $\varphi_n(s) \equiv \mathbf{E}(s^{Z_n})$  is the generating function of  $Z_n$ , a straightforward induction based on (2.1) and (1.5) implies that

$$\begin{aligned} \varphi_0(s) &\equiv s, & \varphi(s) &\equiv \varphi_1(s) \equiv \mathbf{E}s^{Z_1}, \\ \varphi_k(s) &= \varphi_{k-1}(\varphi(s)) \equiv \varphi(\varphi_{k-1}(s)) = \underbrace{\varphi(\dots \varphi(s) \dots)}_{k \text{ times}} \dots \quad k > 1. \end{aligned} \quad (2.2)$$

Usually explicit calculations are hard, but at least in principle, equations (2.2) determine the distribution of  $Z_n$  for any  $n \geq 0$ .

**Example 2.1.** Let  $\varphi_1(s) \equiv \varphi(s) = q + ps$  for some  $0 < p = 1 - q < 1$ . Then

$$\varphi_n(s) \equiv q(1 + p + \cdots + p^{n-1}) + p^n s = 1 + p^n(s - 1).$$

Notice that here we have  $\varphi_n(s) \rightarrow 1$  as  $n \rightarrow \infty$ , for all  $s \in [0, 1]$ . In other words, the distribution of  $Z_n$  converges to that of  $Z_\infty \equiv 0$ , recall Theorem 1.34.

<sup>7</sup>sometimes called a Galton-Watson-Bienaymé process

<sup>8</sup>If  $Z_0 = k$ , we shall explicitly write  $\mathbf{P}_k(\cdot)$  and  $\mathbf{E}_k(\cdot)$ .

The following result is a straightforward corollary of (1.5).

**Lemma 2.2.** *In a branching process  $(Z_n)_{n \geq 0}$  with  $Z_0 = 1$ , let the offspring distribution have mean  $m$ . Then  $\mathbf{E}(Z_n) = m^n$  for all  $n \geq 1$ .*

*Proof.* Recall that a random variable  $X$  with values in  $\mathbb{Z}_{\geq 0}$  has finite mean equal to  $G'_X(1_-) := \lim_{s \uparrow 1} G'_X(s)$ , if and only if this limit exists and is finite. Since  $Z_1$  is assumed to have finite mean  $m$ , this implies that  $\varphi'(1_-) := \lim_{s \uparrow 1} \varphi'(s) = m$  (where  $\varphi$  is the generating function of  $Z_1$ ). We also know by (1.5) that the generating function of  $Z_n$  is given by  $\varphi_n$  which is just the composition of  $\varphi$  with itself  $n$  times. By the chain rule, and since  $\varphi_k(1) = 1$  for all  $k$ , we see that

$$\lim_{s \uparrow 1} \varphi'_n(1_-) = \varphi'(1_-)^n = m^n,$$

implying the result. This can alternatively be shown by induction, using a conditioning argument.  $\square$

**Exercise 2.3.** *In a branching process  $(Z_n)_{n \geq 0}$  with  $Z_0 = 1$ , let the offspring distribution have mean  $m$ , variance  $\sigma^2$ , and generating function  $\varphi$ . Write  $\varphi_n$  for the generating function of the  $n$ th generation size  $Z_n$ ,  $\varphi_n(s) \equiv \mathbf{E}(s^{Z_n})$ .*

- (a) *Using (2.2) or otherwise, show that  $\mathbf{Var}(Z_n) = \sigma^2 m^{n-1} (m^n - 1) / (m - 1)$  if  $m \neq 1$  and  $\mathbf{Var}(Z_n) = \sigma^2 n$  if  $m = 1$ .*
- (b) *Deduce that  $\mathbf{E}((Z_n/m^n)^2)$  is uniformly bounded for  $m \neq 1$ .*

This result suggests that if  $m \equiv \mathbf{E}(Z_1) \neq 1$ , the branching process might explode (for  $m > 1$ ) or die out (for  $m < 1$ ). One therefore classifies branching process as either **critical** (if  $m = 1$ ), **subcritical** ( $m < 1$ ), or **supercritical** ( $m > 1$ ).

**Example 2.4.** *It is straightforward to describe the case  $m < 1$ . Indeed, the Markov inequality (1.2) implies that*

$$\mathbf{P}(Z_n > 0) = \mathbf{P}(Z_n \geq 1) \leq \mathbf{E}(Z_n) = m^n,$$

so that  $\mathbf{P}(Z_n > 0) \rightarrow 0$  as  $n \rightarrow \infty$  (ie.,  $Z_n \rightarrow 0$  in probability). We also notice that the average total population in this case is finite,  $\mathbf{E}(\sum_{n \geq 0} Z_n) = \sum_{n \geq 0} m^n = (1 - m)^{-1} < \infty$ .

**Definition 2.5.** *The extinction event  $\mathcal{E}$  is the event  $\mathcal{E} = \cup_{n=1}^{\infty} \{Z_n = 0\}$ . Since  $\{Z_n = 0\} \subset \{Z_{n+1} = 0\}$  for all  $n \geq 0$ , the extinction probability  $\rho$  is defined as*

$$\rho = \mathbf{P}(\mathcal{E}) = \lim_{n \rightarrow \infty} \mathbf{P}(Z_n = 0),$$

where  $\mathbf{P}(Z_n = 0) \equiv \varphi_n(0)$  is the extinction probability before  $(n + 1)$ st generation.



The following result helps to derive the extinction probability  $\rho$  without needing to compute the iterates  $\varphi_n(\cdot)$  precisely. To avoid trivialities we shall assume that  $p_0 = \mathbf{P}(Z = 0)$  satisfies<sup>9</sup>  $0 < p_0 < 1$ ; notice that under this assumption  $\varphi(s)$  is a strictly increasing function of  $s \in [0, 1]$ .

**Theorem 2.6.** *If  $0 < p_0 < 1$ , then the extinction probability  $\rho$  is given by the smallest positive solution to the equation*

$$s = \varphi(s). \quad (2.3)$$

*In particular, if  $m = \mathbf{E}Z_1 \leq 1$ , then  $\rho = 1$ ; otherwise, we have  $0 < \rho < 1$ .*

In words, if the branching process is subcritical or critical then it eventually becomes extinct with probability one. However, if it is supercritical, the process has a positive probability to survive for all time.

**Remark 2.7.** *There is a clear probabilistic intuition behind the relation  $\rho = \varphi(\rho)$ . Indeed, if  $\rho = \mathbf{P}_1(\mathcal{E})$  is the extinction probability starting from a single individual,  $Z_0 = 1$ , then by independence we get  $\mathbf{P}_k(\mathcal{E}) \equiv \mathbf{P}(\mathcal{E} \mid Z_0 = k) = \rho^k$ , and thus the first step decomposition for the Markov chain  $Z_n$  gives*

$$\begin{aligned} \rho = \mathbf{P}(\mathcal{E}) &= \sum_{k \geq 0} \mathbf{P}(\mathcal{E}, Z_1 = k) = \sum_{k \geq 0} \mathbf{P}(\mathcal{E} \mid Z_1 = k) \mathbf{P}(Z_1 = k) \\ &= \sum_{k \geq 0} \rho^k \mathbf{P}(Z_1 = k) \equiv \mathbf{E}(\rho^{Z_1}) \equiv \varphi(\rho), \end{aligned}$$

*in agreement with (2.3).*

*Proof.* Let us now give the proof of Theorem 2.6. You may find it helpful to draw a picture!

Denote  $\rho_n = \mathbf{P}(Z_n = 0) \equiv \varphi_n(0)$ . By continuity and strict monotonicity of  $\varphi(\cdot)$  we have (recall (2.2))

$$0 < \rho_1 = \varphi(0) < \rho_2 = \varphi(\rho_1) < \cdots < 1,$$

so that the extinction probability  $\rho \in (0, 1]$  is the increasing limit of  $\rho_n$  as  $n \rightarrow \infty$ , and satisfies

$$\rho = \lim_n \varphi_n(0) = \lim_n \varphi(\varphi_{n-1}(0)) = \varphi(\lim_n \varphi_{n-1}(0)) = \varphi(\rho).$$

On the other hand, if  $\bar{\rho}$  is any other fixed point of  $\varphi(\cdot)$  in  $[0, 1]$ , i.e.,  $\bar{\rho} = \varphi(\bar{\rho})$ , then  $\bar{\rho} = \varphi_n(\bar{\rho}) \geq \varphi_n(0)$  for all  $n$ , meaning that  $\bar{\rho} \geq \lim_{n \rightarrow \infty} \varphi_n(0) = \rho$ . So,  $\rho$  is indeed the smallest positive solution to (2.3).

Next, we turn to the extinction criterion in terms of  $m$ . For this, observe that  $\varphi(\cdot)$  is convex on  $[0, 1]$ , since  $\varphi''(s) = \mathbf{E}(Z_1(Z_1 - 1)s^{Z_1 - 2}) \geq 0$  for  $s \in [0, 1]$  (actually unless  $p_1 = 1 - p_0$ , i.e., if there is some possibility of having more than

<sup>9</sup>otherwise the model is degenerate: if  $p_0 = 0$ , then  $Z_n \geq 1$  for all  $n \geq 0$  so that  $\rho = 0$ ; if  $p_0 = 1$ , then  $\mathbf{P}(Z_1 = 0) = \rho = 1$ .

1 child,  $\varphi''(s)$  is strictly positive and so  $\varphi$  is *strictly* convex on  $(0, 1)$ ). Hence if  $m = \varphi'(1_-) > 1$  we must have  $\varphi(s_0) < s_0$  for some  $s_0 < 1$  and therefore the curves  $y = s$  and  $y = \varphi(s)$  must cross at some point strictly in  $(0, 1)$  (recall that  $\varphi(0) = p_0 > 0$ ). A completely rigorous way to justify this is to define  $f(s) := \varphi(s) - s$ , which is continuous on  $[0, s_0]$  with  $f(s_0) < 0$  and  $f(0) > 0$ , so by the intermediate value theorem satisfies  $f(s) = 0$ , i.e.  $\varphi(s) = s$ , for some  $s \in (0, s_0)$ . Conversely, suppose that  $m \leq 1$  and  $p_1 \neq 1 - p_0$ . Then the condition  $m = \varphi'(1_-) \leq 1$  together with strict convexity implies that  $\varphi(s) - s$  is strictly decreasing on  $[0, 1]$ , from  $p_0$  at 0 to 0 at 1, and therefore cannot be 0 for any  $s < 1$ . The case  $p_1 = 1 - p_0$  give  $\varphi(s) = p_0 + p_1 s$  and it is immediate that the smallest solution of  $\varphi(s) = s$  in  $[0, 1]$  is at 1. This completes the proof.  $\square$

**Corollary 2.8.** *If  $s \in [0, 1)$ , we have  $\varphi_n(s) \equiv \mathbf{E}(s^{Z_n}) \rightarrow \rho \in (0, 1]$  as  $n \rightarrow \infty$ .*

**Remark 2.9.** *As a result, the distribution of  $Z_n$  converges to that of  $Z_\infty$ , where  $\mathbf{P}(Z_\infty = 0) = \rho$  and  $\mathbf{P}(Z_\infty = \infty) = 1 - \rho$ .*

**Exercise 2.10.** *For a branching process with generating function  $\varphi(s) = as^2 + bs + c$ , where  $a > 0$ ,  $b > 0$ ,  $c > 0$ ,  $\varphi(1) = 1$ , compute the extinction probability  $\rho$  and give the condition for sure extinction. Can you interpret your results?*

**Exercise 2.11.** *Let  $(Z_n)_{n \geq 0}$  be a branching process with generating function  $\varphi(s) \equiv \mathbf{E}s^{Z_1}$  satisfying  $0 < \varphi(0) < 1$ . Let*

$$\bar{\varphi}_n(u) \stackrel{\text{def}}{=} \mathbf{E}(u^{\bar{Z}_n})$$

*be the generating function of*

$$\bar{Z}_n = \sum_{k=0}^n Z_k,$$

*the total population size up to time  $n$ .*

(a) *Show that  $\bar{\varphi}_{n+1}(u) = u\varphi(\bar{\varphi}_n(u))$  for all  $n \geq 0$  and  $u \geq 0$ .*

(b) *If  $u \in (0, 1)$ , show that  $\bar{\varphi}_n(u) \rightarrow \bar{\varphi}(u) \stackrel{\text{def}}{=} \mathbf{E}(u^{\bar{Z}})$  as  $n \rightarrow \infty$ , where  $\bar{Z}$  is the total population size,  $\sum_{k \geq 0} Z_k$ , of  $(Z_n)_{n \geq 0}$ . Show that the limiting generating function  $\bar{\varphi}(u)$  is given by the smallest positive solution  $s$  to the equation  $s = u\varphi(s)$ .*

(c) *Let the process  $(Z_n)_{n \geq 0}$  be subcritical ( $\varphi'(1_-) < 1$ ) with offspring distribution having exponential tails ( $\varphi(s) < \infty$  for some  $s > 1$ ). Show that for some  $u > 1$  the equation  $s = u\varphi(s)$  has positive solutions  $s$ , the smallest of which coincides with  $\bar{\varphi}(u) = \mathbf{E}(u^{\bar{Z}})$ .*

(d) *Let the process  $(Z_n)_{n \geq 0}$  be supercritical ( $\mathbf{E}Z_1 > 1$ ) with  $0 < \mathbf{P}(Z_1 = 0) < 1$  and let  $u > 1$  be such that the equation  $s = u\varphi(s)$  has positive solutions  $s$ . Show that  $\bar{\varphi}_n(u) \rightarrow \infty$ , in agreement with the fact that with positive probability the process  $(Z_n)_{n \geq 0}$  survives forever,  $\mathbf{P}(\bar{Z} = \infty) > 0$ .*

(e) In the setting of part d), let  $\hat{\varphi}_n(u) \stackrel{\text{def}}{=} \mathbb{E}(u^{\bar{Z}_n} \mathbf{1}_{Z_n=0})$ , the generating function of the total population on the event that the process  $(Z_n)_{n \geq 0}$  dies out by time  $n$ . Show that  $\hat{\varphi}_{n+1}(u) = u\varphi(\hat{\varphi}_n(u))$  for all  $n \geq 0$  and  $u \geq 0$ . Deduce that for each  $u > 1$  such that the fixed point equation  $s = u\varphi(s)$  has positive solutions  $s$ , we have  $\hat{\varphi}_n(u) \rightarrow \mathbb{E}(u^{\bar{Z}} \mathbf{1}_{\bar{Z} < \infty})$ , where the latter coincides with the smallest positive  $s$  satisfying  $s = u\varphi(s)$ .

We now turn to classification of states for the Markov chain  $Z_n$  in  $\mathbb{Z}^+$ . Of course, since 0 is an absorbing state, it is recurrent.

**Lemma 2.12.** *If  $p_1 = \mathbb{P}(Z_1 = 1) \neq 1$ , then every state  $k \in \mathbb{N}$  is transient. As a result,*

$$\mathbb{P}(Z_n \rightarrow \infty) = 1 - \mathbb{P}(Z_n \rightarrow 0) = 1 - \rho.$$

*Proof.* We first show that every  $k \in \mathbb{N}$  is transient. If  $p_0 = 0$ , then  $Z_n$  is a non-decreasing Markov chain (ie.,  $Z_{n+1} \geq Z_n$ ), so that for every  $k \in \mathbb{N}$  the first passage probability  $f_{kk}$  satisfies

$$f_{kk} = \mathbb{P}(Z_{n+1} = k \mid Z_n = k) = \mathbb{P}_k(Z_1 = k) = (p_1)^k < 1.$$

On the other hand, for  $p_0 \in (0, 1]$  we have

$$f_{kk} \leq \mathbb{P}(Z_{n+1} \neq 0 \mid Z_n = k) = \mathbb{P}_k(Z_1 \neq 0) = 1 - \mathbb{P}_k(Z_1 = 0) = 1 - (p_0)^k < 1.$$

This means that

$$\mathbb{P}(Z_n = k \text{ i.o.}) = \lim_{m \rightarrow \infty} \mathbb{P}(Z_n \text{ returns to } k \text{ at least } m \text{ times}) \leq \lim_{m \rightarrow \infty} f_{kk}^{m-1} = 0$$

and the state  $k$  is transient.

Fix arbitrary  $K > 0$ . Since the states  $1, 2, \dots, K$  are transient, we see that  $\mathbb{P}(\{Z_n = 0\} \cup \{Z_n > K\}) \rightarrow 1$  as  $n \rightarrow \infty$  and therefore

$$\mathbb{P}(Z_n \rightarrow 0 \text{ or } Z_n \rightarrow \infty) = 1.$$

As the LHS above equals  $\mathbb{P}(Z_n \rightarrow 0) + \mathbb{P}(Z_n \rightarrow \infty)$ , the result follows from the observation that  $\mathbb{P}(Z_n \rightarrow 0) \equiv \mathbb{P}(\mathcal{E}) = \rho$ .  $\square$

**Exercise 2.13.** *For a supercritical branching process  $(Z_n)_{n \geq 0}$ , let  $T_0 = \min\{n \geq 0 : Z_n = 0\}$  be its extinction time and let  $\rho = \mathbb{P}(T_0 < \infty) > 0$  be its extinction probability. Define  $(\hat{Z}_n)_{n \geq 0}$  as  $Z_n$  conditioned on extinction, ie.,  $\hat{Z}_n = (Z_n \mid T_0 < \infty)$ .*

(a) *Show that the transition probabilities  $\hat{p}_{xy}$  of  $(\hat{Z}_n)_{n \geq 0}$  and the transition probabilities  $p_{xy}$  of the original process  $(Z_n)_{n \geq 0}$  are related via  $\hat{p}_{xy} = p_{xy}\rho^{y-x}$ ,  $x, y \geq 0$ .*

(b) *Deduce that the generating functions  $\hat{\varphi}(s) \equiv \mathbb{E}_1[s^{\hat{Z}_1}]$  and  $\varphi(s) \equiv \mathbb{E}_1[s^{Z_1}]$  are related via<sup>10</sup>  $\hat{\varphi}(s) = \frac{1}{\rho}\varphi(\rho s)$ ,  $0 \leq s \leq 1$ .*

<sup>10</sup> Geometrically, the graph of this generating function is a rescaled version of that of  $\varphi(\cdot)$ .

- (c) If the offspring distribution of  $(Z_n)_{n \geq 0}$  is  $\text{Poi}(\lambda)$  with  $\lambda > 1$ , use the fixed point equation  $\rho = e^{\lambda(\rho-1)}$  to show that  $\widehat{\varphi}(s) = e^{\lambda\rho(s-1)}$ , i.e., that the offspring distribution for  $(\widehat{Z}_n)_{n \geq 0}$  is just  $\text{Poi}(\lambda\rho)$ .

**Exercise 2.14.** Let  $(Z_n)_{n \geq 0}$  be a supercritical branching process with offspring distribution  $\{p_k\}_{k \geq 0}$ , offspring generating function  $\varphi(s)$  and extinction probability  $\rho \in [0, 1)$ .

- (a) If  $Z_0 = 1$ , let  $\tilde{p}_k$  be the probability that conditioned on survival the first generation has exactly  $k$  individuals with an infinite line of descent. Show that

$$\tilde{p}_k = \frac{1}{1-\rho} \sum_{n=k}^{\infty} p_n \binom{n}{k} (1-\rho)^k \rho^{n-k}.$$

- (b) Let  $(\tilde{Z}_n)_{n \geq 0}$  count only those individuals in  $(Z_n)_{n \geq 0}$ , who conditioned on survival have an infinite line of descent. Show that  $(\tilde{Z}_n)_{n \geq 0}$  is a branching process with offspring generating function<sup>10</sup>

$$\tilde{\varphi}(s) = \frac{1}{1-\rho} \left( \varphi((1-\rho)s + \rho) - \rho \right).$$

**Exercise 2.15.** Let  $(Z_n)_{n \geq 0}$  be a subcritical branching process whose generating function  $\varphi(s) = \mathbb{E}(s^{Z_1})$  is finite for some  $s > 1$ , i.e., the offspring distribution has finite exponential moments in a neighbourhood of the origin.

- (a) Using the result of Exercise 2.11 or otherwise, show that the total population size  $\bar{Z} = \sum_{k \geq 0} Z_k$  satisfies  $\mathbb{E}(u^{\bar{Z}}) < \infty$  for some  $u > 1$ .
- (b) Suppose that for each  $1 \leq i \leq \bar{Z}$ , individual  $i$  produces wealth of size  $W_i$ , where  $W_i$  are independent random variables with common distribution satisfying  $\mathbb{E}(s^W) < \infty$  for some  $s > 1$ . Show that for some  $u > 1$  we have  $\mathbb{E}(u^{\bar{W}}) < \infty$ , where  $\bar{W} = W_1 + \dots + W_{\bar{Z}}$  is the total wealth generated by  $(Z_n)_{n \geq 0}$ .

## 2.2 Critical case $m = 1$

The following example is one of very few for which the computation in the critical case  $m = \mathbb{E}(Z_1) = 1$  can be done explicitly.

**Example 2.16.** Consider the so-called linear-fractional case, where the offspring distribution is given by  $p_j = 2^{-(j+1)}$ ,  $j \geq 0$ . Then the offspring generating function is  $\varphi(s) = \sum_{j \geq 0} s^j / 2^{j+1} = (2-s)^{-1}$  and a straightforward induction gives (check this!)

$$\varphi_k(s) = \frac{k - (k-1)s}{(k+1) - ks} = \frac{k}{k+1} + \frac{1}{k(k+1)} \sum_{m \geq 1} \left( \frac{ks}{k+1} \right)^m,$$

so that  $\mathbb{P}(Z_k = 0) = \varphi_k(0) = k/(k+1)$ ,  $\mathbb{P}(Z_k > 0) = 1/(k+1)$ , and

$$\mathbb{P}(Z_k = m \mid Z_k > 0) = \frac{1}{k+1} \left( \frac{k}{k+1} \right)^{m-1}, \quad m \geq 1,$$

ie.,  $(Z_k \mid Z_k > 0)$  has geometric distribution with success probability  $1/(k+1)$ .

**Remark 2.17.** For each  $k \geq 0$ , by the partition theorem,

$$1 = \mathbb{E}(Z_k) = \mathbb{E}(Z_k \mid Z_k > 0) \mathbb{P}(Z_k > 0) + \mathbb{E}(Z_k \mid Z_k = 0) \mathbb{P}(Z_k = 0),$$

so that in the previous example we have

$$\mathbb{E}(Z_k \mid Z_k > 0) = \frac{1}{\mathbb{P}(Z_k > 0)} = k+1,$$

ie., conditional on survival, the average generation size grows linearly with time.

The following example is known as the general linear-fractional case:

**Exercise 2.18.** For fixed  $b > 0$  and  $p \in (0, 1)$ , consider a branching process with offspring distribution  $p_j = b p^{j-1}$ ,  $j \geq 1$ , and  $p_0 = 1 - \sum_{j \geq 1} p_j$ .

(a) Show that for  $b \in (0, 1-p)$  the distribution above is well defined; find the corresponding  $p_0$ , and show that

$$\varphi(s) = \frac{1-b-p}{1-p} + \frac{bs}{1-ps};$$

(b) Find  $b$  for which the branching process is critical and show that then

$$\varphi_k(s) = \mathbb{E}(s^{Z_k}) = \frac{kp - (kp + p - 1)s}{(1-p + kp) - kps};$$

(c) Deduce that  $(Z_k \mid Z_k > 0)$  is geometrically distributed with parameter  $\frac{1-p}{kp+1-p}$ .

Straightforward computer experiments show that a similar linear growth of  $\mathbb{E}(Z_k \mid Z_k > 0)$  takes place for other critical offspring distributions, eg., the one with  $\varphi(s) = (1+s^2)/2$ .

**Theorem 2.19.** If the offspring distribution of the branching process  $(Z_k)_{k \geq 0}$  has mean  $m = 1$  and finite variance  $\sigma^2 > 0$ , then  $k \mathbb{P}(Z_k > 0) \rightarrow \frac{2}{\sigma^2}$  as  $k \rightarrow \infty$ ; equivalently,

$$\frac{1}{k} \mathbb{E}(Z_k \mid Z_k > 0) \rightarrow \frac{\sigma^2}{2} \quad \text{as } k \rightarrow \infty. \quad (2.4)$$

**Remark 2.20.** This general result suggests that, conditional on survival, a general critical branching process exhibits linear intermittent behaviour;<sup>11</sup> namely, with small probability (of order  $2/(k\sigma^2)$ ) the values of  $Z_k$  are of order  $k$ .

<sup>11</sup>Intermittency follows from the criticality condition,  $1 = \mathbb{E}(Z_k \mid Z_k > 0) \mathbb{P}(Z_k > 0)$ ; it is the linearity which is surprising here!

Our argument is based on the following general fact:<sup>12</sup>

**Lemma 2.21.** *Let  $(y_n)_{n \geq 0}$  be a real-valued sequence. If for some constant  $a$  we have  $y_{n+1} - y_n \rightarrow a$  as  $n \rightarrow \infty$ , then  $n^{-1}y_n \rightarrow a$  as  $n \rightarrow \infty$ .*

*Proof.* By changing the variables  $y_n \mapsto y'_n = y_n - na$  if necessary, we can and shall assume that  $a = 0$ . Fix arbitrary  $\delta > 0$  and find  $K > 0$  such that for  $n \geq K$  we have  $|y_{n+1} - y_n| \leq \delta$ . Decomposing, for  $n > K$ ,

$y_n - y_K = \sum_{j=K}^{n-1} (y_{j+1} - y_j)$  we deduce that  $|y_n - y_K| \leq \delta(n - K)$  so that the claim follows from the estimate

$$\left| \frac{y_n}{n} \right| \leq \left| \frac{y_n - y_K}{n} \right| + \left| \frac{y_K}{n} \right| \leq \delta + \left| \frac{y_K}{n} \right| \leq 2\delta,$$

provided  $n$  is chosen sufficiently large.  $\square$

*Proof.* (of Theorem 2.19). We only derive the second claim of the theorem, (2.4). By assumptions and Taylor's theorem (here we are also using Theorem 1.16) the offspring generating function  $\varphi$  satisfies

$$1 - \varphi(s) = (1 - s) + \frac{\sigma^2}{2}(1 - s)^2 + R(s)(1 - s)^2 \text{ where } R(s) \rightarrow 0 \text{ as } s \uparrow 1.$$

Since  $\varphi_n(0) = \mathbb{P}(Z_n = 0) \rightarrow 1$  as  $n \rightarrow \infty$  this means in particular that  $1 - \varphi_{n+1}(0) = 1 - \varphi(\varphi_n(0)) = 1 - \varphi_n(0) + (1 - \varphi_n(0))^2(\frac{\sigma^2}{2} + r_n)$  with  $r_n := R(\varphi_n(0)) \rightarrow 0$  as  $n \rightarrow \infty$ . Setting

$$y_n = \frac{1}{1 - \varphi_n(0)} = \frac{\mathbb{E}(Z_n)}{\mathbb{P}(Z_n > 0)} = \mathbb{E}(Z_n | Z_n > 0)$$

we therefore have

$$y_{n+1} - y_n = \frac{1}{1 - \varphi_n(0)} \frac{1}{1 - (1 - \varphi_n(0))(\sigma^2/2 + r_n)} - 1$$

which we can rewrite as

$$\frac{(1 - \varphi_n(0))(\sigma^2/2 + r_n)}{1 - \varphi_n(0)} \frac{1}{1 - (1 - \varphi_n(0))(\sigma^2/2 + r_n)}$$

for each  $n$ . It therefore follows that  $y_{n+1} - y_n \rightarrow \sigma^2/2$  as  $n \rightarrow \infty$  and hence

$$\frac{y_n}{n} = \frac{\mathbb{E}(Z_n | Z_n > 0)}{n} \rightarrow \frac{\sigma^2}{2}$$

as well (by the Lemma). This completes the proof.  $\square$

<sup>12</sup>Compare the result to Cesàro limits of real sequences: if  $(a_k)_{k \geq 1}$  is a real-valued sequence, and  $s_n = a_1 + \dots + a_n$  is its  $n$ th partial sum, then  $\frac{1}{n}s_n$  are called the *Cesàro averages* for the sequence  $(a_k)_{k \geq 1}$ . Lemma 2.21 claims that if  $a_k \rightarrow a$  as  $k \rightarrow \infty$ , then the sequence of its Cesàro averages also converges to  $a$ . The converse is, of course, false. (Find a counterexample!)

**Remark 2.22.** With a bit of extra work <sup>13</sup> one can generalize the above proof to show that

$$\lim_{n \rightarrow \infty} n^{-1} \left( \frac{1}{1 - \varphi_n(s)} - \frac{1}{1 - s} \right) = \frac{\sigma^2}{2}$$

for any  $s \in [0, 1]$  and use this relation to derive the convergence in distribution:

**Theorem 2.23.** If  $\mathbb{E}Z_1 = 1$  and  $\text{Var}(Z_1) = \sigma^2 \in (0, \infty)$ , then for every  $z \geq 0$  we have

$$\lim_{k \rightarrow \infty} \mathbb{P} \left( \frac{Z_k}{k} > z \mid Z_k > 0 \right) = \exp \left\{ -\frac{2z}{\sigma^2} \right\},$$

ie., the distribution of  $(k^{-1}Z_k \mid Z_k > 0)$  is approximately exponential with parameter  $2/\sigma^2$ .

**Remark 2.24.** In the setup of Example 2.16, we have

$$\mathbb{P}(Z_k > m \mid Z_k > 0) = \left( \frac{k}{k+1} \right)^m = \left( 1 - \frac{1}{k+1} \right)^m,$$

so that  $\mathbb{P}(Z_k > kz \mid Z_k > 0) \rightarrow e^{-z}$  as  $k \rightarrow \infty$ ; in other words, for large  $k$  the distribution of  $(k^{-1}Z_k \mid Z_k > 0)$  is approximately  $\text{Exp}(1)$ .

**Exercise 2.25.** Let  $(Z_n)_{n \geq 0}$  be the critical branching process from Exercise 2.18, namely, the one whose offspring distribution is given by  $(p_j)_{j \geq 0}$ ,

$$p_j = bp^{j-1}, \quad j \geq 1, \quad p_0 = 1 - \sum_{j \geq 1} p_j,$$

where  $b > 0$  and  $p \in (0, 1)$  are fixed parameters. Show that the result of Theorem 2.23 holds: for every  $z \geq 0$

$$\lim_{k \rightarrow \infty} \mathbb{P} \left( \frac{Z_k}{k} > z \mid Z_k > 0 \right) = \exp \left\{ -\frac{2z}{\sigma^2} \right\},$$

where  $\text{Var}(Z_1) = \sigma^2 \in (0, \infty)$ .

### 2.3 Non-homogeneous case

If the offspring distribution changes with time, the previous approach must be modified. Let  $\psi_n(u)$  be the generating function of the offspring distribution of a single ancestor in the  $(n-1)$ st generation,

$$\psi_n(u) = \mathbb{E}(u^{Z_n} \mid Z_{n-1} = 1)$$

(so in the cases considered up to now,  $\psi_n = \varphi$  for every  $n$ ). Then the generating function  $\varphi_n(u) = \mathbb{E}(u^{Z_n} \mid Z_0 = 1)$  of the population size at time  $n$  given a single ancestor at time 0, can be defined recursively as follows:

$$\varphi_0(u) \equiv u, \quad \varphi_n(u) = \varphi_{n-1}(\psi_n(u)), \quad \forall n \geq 1.$$

<sup>13</sup>using the fact that every  $s \in (0, 1)$  satisfies  $0 < s < \varphi_k(0) < 1$  for some  $k \geq 1$ ;

If  $\mu_n = \mathbb{E}(Z_n \mid Z_{n-1} = 1) = \psi'_n(1)$  denotes the average offspring size in the  $n$ th generation given a single ancestor in the previous generation, then

$$m_n \equiv \mathbb{E}(Z_n \mid Z_0 = 1) = \mu_1 \mu_2 \dots \mu_{n-1} \mu_n.$$

It is natural to call the process  $(Z_n)_{n \geq 0}$  **supercritical** if  $m_n \rightarrow \infty$  and **subcritical** if  $m_n \rightarrow 0$  as  $n \rightarrow \infty$ .

**Exercise 2.26.** *A strain of phototrophic bacteria uses light as the main source of energy. As a result individual organisms reproduce with probability mass function  $p_0 = 1/4$ ,  $p_1 = 1/4$  and  $p_2 = 1/2$  per unit of time in light environment, and with probability mass function  $p_0 = 1-p$ ,  $p_1 = p$  (with some  $p > 0$ ) per unit of time in dark environment. A colony of such bacteria is grown in a laboratory, with alternating light and dark unit time intervals.*

*a) Model this experiment as a time non-homogeneous branching process  $(Z_n)_{n \geq 0}$  and describe the generating function of the population size at the end of the  $n$ th interval.*

*b) Characterise all values of  $p$  for which the branching process  $Z_n$  is subcritical and for which it is supercritical.*

*c) Let  $(D_k)_{k \geq 0}$  be the original process observed at the end of each even interval,  $D_k \stackrel{\text{def}}{=} Z_{2k}$ . Find the generating function of  $(D_k)_{k \geq 0}$  and derive the condition for sure extinction. Compare your result with that of part b).*



### 3 Coupling

Two random variables, say  $X$  and  $Y$ , are coupled if they are defined on the same probability space. To couple two given variables  $X$  and  $Y$  means to define a random vector  $(\tilde{X}, \tilde{Y})$  with joint probability  $\tilde{P}(\cdot, \cdot)$  on some probability space<sup>14</sup> so that the marginal distribution of  $\tilde{X}$  coincides with the distribution of  $X$  and the marginal distribution of  $\tilde{Y}$  coincides with the distribution of  $Y$ . Recall the following example:

**Example 3.1.** Fix  $p_1, p_2 \in [0, 1]$  such that  $p_1 \leq p_2$  and consider the following joint distributions (we write  $q_i = 1 - p_i$ ):

		0	1	$\tilde{X}$
(A)	0	$q_1 q_2$	$q_1 p_2$	$q_1$
	1	$p_1 q_2$	$p_1 p_2$	$p_1$
	$\tilde{Y}$	$q_2$	$p_2$	
		0	1	$\tilde{X}$
(B)	0	$q_2$	$p_2 - p_1$	$q_1$
	1	0	$p_1$	$p_1$
	$\tilde{Y}$	$q_2$	$p_2$	

It is easy to see that in both cases <sup>15</sup>

$$\tilde{X} \sim \text{Ber}(p_1), \tilde{Y} \sim \text{Ber}(p_2),$$

though in the first case  $\tilde{X}$  and  $\tilde{Y}$  are independent, whereas in the second case we have  $\tilde{P}(\tilde{X} \leq \tilde{Y}) = 1$ .

#### 3.1 Stochastic domination

If  $X \sim \text{Ber}(p)$ , its tail probabilities  $P(X > a)$  satisfy

$$P(X > a) = \begin{cases} 1, & a < 0, \\ p, & 0 \leq a < 1, \\ 0, & a \geq 1. \end{cases}$$

Consequently, in the setup of Example 3.1, for the variables  $X \sim \text{Ber}(p_1)$  and  $Y \sim \text{Ber}(p_2)$  with  $p_1 \leq p_2$  we have  $P(X > a) \leq P(Y > a)$  for all  $a \in \mathbb{R}$ . The last inequality is useful enough to deserve a name:

<sup>14</sup>A priori the original variables  $X$  and  $Y$  can be defined on arbitrary probability spaces, so that one has no reason to expect that these spaces can be “joined” in any way!

<sup>15</sup>and in fact, every convex linear combination of these two tables provides a joint distribution with the same marginals.

**Definition 3.2.** [Stochastic domination] A random variable  $X$  is stochastically smaller than a random variable  $Y$  (write  $X \preceq Y$ ) if the inequality

$$\mathbb{P}(X > x) \leq \mathbb{P}(Y > x) \quad (3.1)$$

holds for all  $x \in \mathbb{R}$ .

**Remark 3.3.** If  $X \preceq Y$  and  $g(\cdot) \geq 0$  is an arbitrary increasing function on  $\mathbb{R}$ , then  $g(X) \preceq g(Y)$ . If, in addition,  $X \geq 0$ ,  $Y \geq 0$ , and  $g(\cdot)$  is smooth with  $g(0) = 0$ , then

$$\mathbb{E}g(X) \equiv \int_0^\infty g'(z)\mathbb{P}(X > z) dz \leq \int_0^\infty g'(z)\mathbb{P}(Y > z) dz \equiv \mathbb{E}g(Y). \quad (3.2)$$

**Example 3.4.** If  $X$  is a random variable and  $a \geq 0$  is a fixed constant, it is immediate that  $Y = a + X$  stochastically dominates  $X$ , ie.,  $X \preceq Y$ : for each  $x \in \mathbb{R}$  we have  $\mathbb{P}(X > x) \leq \mathbb{P}(X > x - a) = \mathbb{P}(Y > x)$ . Similarly, if  $a \geq 1$  and  $Z = aX \geq 0$ , then  $X \preceq Z$ .

In the setup of Example 3.1, if  $X \sim \text{Ber}(p_1)$  and  $Y \sim \text{Ber}(p_2)$  then  $X$  is stochastically smaller than  $Y$  (ie.,  $X \preceq Y$ ) if and only if  $p_1 \leq p_2$ ; moreover, this is equivalent to existence of a coupling  $(\tilde{X}, \tilde{Y})$  of  $X$  and  $Y$  in which these variables are ordered with probability one,  $\tilde{\mathbb{P}}(\tilde{X} \leq \tilde{Y}) = 1$ . The next result shows that this is a rather generic situation.

**Lemma 3.5.** A random variable  $X$  is stochastically smaller than a random variable  $Y$  if and only if there exists a coupling  $(\tilde{X}, \tilde{Y})$  of  $X$  and  $Y$  such that  $\tilde{\mathbb{P}}(\tilde{X} \leq \tilde{Y}) = 1$ .

**Remark 3.6.** Notice that one claim of Lemma 3.5 is immediate from

$$\mathbb{P}(x < X) \equiv \tilde{\mathbb{P}}(x < \tilde{X}) = \tilde{\mathbb{P}}(x < \tilde{X} \leq \tilde{Y}) \leq \tilde{\mathbb{P}}(x < \tilde{Y}) \equiv \mathbb{P}(x < Y);$$

the other claim requires a more advanced argument (we shall not do it here!).

**Example 3.7.** If  $X \sim \text{Bin}(m, p)$  and  $Y \sim \text{Bin}(n, p)$  with  $m \leq n$ , then  $X \preceq Y$ . Indeed, Let  $Z_1 \sim \text{Bin}(m, p)$  and  $Z_2 \sim \text{Bin}(n - m, p)$  be independent variables defined on the same probability space. We then put  $\tilde{X} = Z_1$  and  $\tilde{Y} = Z_1 + Z_2$  so that  $\tilde{Y} - \tilde{X} = Z_2 \geq 0$  with probability one,  $\tilde{\mathbb{P}}(\tilde{X} \leq \tilde{Y}) = 1$ , and  $X \sim \tilde{X}$ ,  $Y \sim \tilde{Y}$ .

**Example 3.8.** If  $X \sim \text{Poi}(\lambda)$  and  $Y \sim \text{Poi}(\mu)$  with  $\lambda \leq \mu$ , then  $X \preceq Y$ . Indeed, Let  $Z_1 \sim \text{Poi}(\lambda)$  and  $Z_2 \sim \text{Poi}(\mu - \lambda)$  be independent variables defined on the same probability space.<sup>16</sup> We then put  $\tilde{X} = Z_1$  and  $\tilde{Y} = Z_1 + Z_2$  so that  $\tilde{Y} - \tilde{X} = Z_2 \geq 0$  with probability one,  $\tilde{\mathbb{P}}(\tilde{X} \leq \tilde{Y}) = 1$ , and  $X \sim \tilde{X}$ ,  $Y \sim \tilde{Y}$ .

**Example 3.9.** Let  $(X_n)_{n \geq 0}$  be a branching process with offspring distribution  $\{p_m\}_{m \geq 0}$  and  $X_0 = 1$ . Let  $(Y_n)_{n \geq 0}$  be a branching process with the same offspring distribution and  $Y_0 = 2$ . We can use stochastic domination to show that  $\mathbb{P}(X_n = 0) \geq \mathbb{P}(Y_n = 0)$  for all  $n$ .

<sup>16</sup>here and below we assume that  $Z \sim \text{Poi}(0)$  means that  $\mathbb{P}(Z = 0) = 1$ .

Indeed, it is enough to show that  $X_n \preceq Y_n$  for all  $n \geq 0$ . To this end consider two independent branching processes,  $(Z'_n)_{n \geq 0}$  and  $(Z''_n)_{n \geq 0}$ , having the same offspring distribution  $\{p_m\}_{m \geq 0}$  and satisfying  $Z'_0 = Z''_0 = 1$ . We then put  $\tilde{X}_n = Z'_n$  and  $\tilde{Y}_n = Z'_n + Z''_n$  for all  $n \geq 0$ , so that  $\tilde{Y}_n - \tilde{X}_n = Z''_n \geq 0$ , i.e.,  $\tilde{P}(\tilde{X}_n \leq \tilde{Y}_n) = 1$  and  $X_n \sim \tilde{X}_n$ ,  $Y_n \sim \tilde{Y}_n$  for all  $n \geq 0$ .

**Exercise 3.10.** For a given offspring distribution  $\{p_m\}_{m \geq 0}$ , let  $(X_n)_{n \geq 0}$  be the branching process with  $X_0 = k$  and let  $(Y_n)_{n \geq 0}$  be the branching process with  $Y_0 = l$ ,  $k < l$ . Show that  $X_n \preceq Y_n$  for all  $n \geq 0$ .

**Exercise 3.11.** Let  $X, Y, Z$  be random variables in  $\mathbb{Z}_+$ . If  $X \preceq Y$  and  $Z$  is independent of  $\{X, Y\}$ , show that  $X + Z \preceq Y + Z$ .

Let  $\{X_i, Y_i\}_{i=1, \dots, n}$  be random variables in  $\mathbb{Z}_+$ . If the pairs  $(X_i, Y_i)$  are mutually independent and  $X_i \preceq Y_i$  for each  $i$ , show that  $X_1 + \dots + X_n \preceq Y_1 + \dots + Y_n$ .

**Exercise 3.12.** Let  $(X_n)_{n \geq 0}$  and  $(Y_n)_{n \geq 0}$  be standard branching processes with  $X_0 = Y_0 = 1$ . Assume that the offspring distribution of  $X$  is stochastically smaller than that of  $Y$  ( $X_1 \preceq Y_1$ ), i.e., for all integer  $k \geq 0$ ,  $P(X_1 > k | X_0 = 1) \leq P(Y_1 > k | Y_0 = 1)$ . Show that  $X_n \preceq Y_n$  for all  $n \geq 0$ .

## 3.2 Total variation distance

**Definition 3.13.** [Total Variation Distance] Let  $\mu$  and  $\nu$  be two probability measures on the same probability space. The *total variation* distance between  $\mu$  and  $\nu$  is

$$d_{\text{TV}}(\mu, \nu) \stackrel{\text{def}}{=} \sup_A |\mu(A) - \nu(A)|. \quad (3.3)$$

If  $X, Y$  are random variables in  $\mathbb{Z}^+$  with respective p.m.f.s  $\{p\} = \{p_k\}_{k \geq 0}$  and  $\{q\} = \{q_\ell\}_{\ell \geq 0}$ , then

$$d_{\text{TV}}(X, Y) \stackrel{\text{def}}{=} d_{\text{TV}}(\{p\}, \{q\}) = \sup_{A \subseteq \mathbb{Z}^+} \left| \sum_{k \in A} p_k - \sum_{k \in A} q_k \right|.$$

In general if  $X, Y$  are two discrete random variables taking values in the same space  $\mathcal{X}$ , with  $P(X = x) = p_x$ ,  $P(Y = x) = q_x$  for  $x \in \mathcal{X}$ , then we define

$$d_{\text{TV}}(X, Y) \stackrel{\text{def}}{=} d_{\text{TV}}(\{p\}, \{q\}) = \sup_{A \subseteq \mathcal{X}} \left| \sum_{x \in A} p_x - \sum_{x \in A} q_x \right|.$$

**Example 3.14.** If  $X \sim \text{Ber}(p_1)$  and  $Y \sim \text{Ber}(p_2)$  we have

$$d_{\text{TV}}(X, Y) = \max\{|p_1 - p_2|, |q_1 - q_2|\} = |p_1 - p_2| = \frac{1}{2}(|p_1 - p_2| + |q_1 - q_2|).$$

**Exercise 3.15.** Let probability measures  $\mu$  and  $\nu$  have respective p.m.f.s  $\{p_x\}$  and  $\{q_y\}$ . Show that the total variation distance between  $\mu$  and  $\nu$  is

$$d_{\text{TV}}(\mu, \nu) \equiv d_{\text{TV}}(\{p\}, \{q\}) = \frac{1}{2} \sum_z |p_z - q_z|.$$

Deduce that  $d_{\text{TV}}(\cdot, \cdot)$  is a distance between probability measures<sup>17</sup> (ie., it is non-negative, symmetric, and satisfies the triangle inequality) such that  $d_{\text{TV}}(\cdot, \cdot) \leq 1$  for all probability measures  $\mu$  and  $\nu$ .

An important relation between coupling and the total variation distance is explained by the following fact.

**Example 3.16. [Maximal Coupling]** Let random variables  $X$  and  $Y$  be such that  $P(X = x) = p_x$  and  $P(Y = y) = q_y$ . Define the coupling  $(\tilde{X}, \tilde{Y})$  via<sup>18</sup>

$$\hat{P}(\tilde{X} = \tilde{Y} = z) = \min(p_z, q_z), \quad (3.4)$$

$$\hat{P}(\tilde{X} = x, \tilde{Y} = y) = \frac{(p_x - \min(p_x, q_x))(q_y - \min(p_y, q_y))}{d_{\text{TV}}(\{p\}, \{q\})}, \quad x \neq y. \quad (3.5)$$

**Exercise 3.17.** Show that

$$\sum_x (p_x - \min(p_x, q_x)) = \sum_y (q_y - \min(p_y, q_y)) = d_{\text{TV}}(\{p\}, \{q\}),$$

and deduce that (3.4) is indeed a coupling of  $X$  and  $Y$  (ie., that (3.4) defines a probability distribution with correct marginals).

**Example 3.18.** Consider  $X \sim \text{Ber}(p_1)$  and  $Y \sim \text{Ber}(p_2)$  with  $p_1 \leq p_2$ . It is a straightforward exercise to check that the second table in Example 3.1 provides the maximal coupling of  $X$  and  $Y$ . We notice also that in this case

$$\hat{P}(\tilde{X} \neq \tilde{Y}) = p_2 - p_1 = d_{\text{TV}}(X, Y).$$

**Lemma 3.19.** Let  $\hat{P}(\cdot, \cdot)$  be the maximal coupling (3.4) of  $X$  and  $Y$ . Then for every other coupling  $\tilde{P}(\cdot, \cdot)$  of  $X$  and  $Y$  we have

$$\tilde{P}(\tilde{X} \neq \tilde{Y}) \geq \hat{P}(\tilde{X} \neq \tilde{Y}) = d_{\text{TV}}(\tilde{X}, \tilde{Y}). \quad (3.6)$$

*Proof.* Summing the inequalities  $\tilde{P}(\tilde{X} = \tilde{Y} = z) \leq \min(p_z, q_z)$  we deduce

$$\tilde{P}(\tilde{X} \neq \tilde{Y}) \geq 1 - \sum_z \min(p_z, q_z) = \sum_z (p_z - \min(p_z, q_z)) = d_{\text{TV}}(\{p\}, \{q\}),$$

in view of Exercise 3.17 and Example 3.18.  $\square$

<sup>17</sup>so that all probability measures form a metric space for this distance!

<sup>18</sup>By Exercise 3.15, if  $d_{\text{TV}}(\{p\}, \{q\}) = 0$ , we have  $p_z = q_z$  for all  $z$ , and thus all off-diagonal terms in (3.4) vanish.

**Remark 3.20.** Notice that according to (3.6),

$$\tilde{\mathbb{P}}(\tilde{X} = \tilde{Y}) \leq \hat{\mathbb{P}}(\tilde{X} = \tilde{Y}) = 1 - d_{\text{TV}}(\tilde{X}, \tilde{Y}),$$

ie., the probability that  $\tilde{X} = \tilde{Y}$  is maximised under the optimal coupling  $\hat{\mathbb{P}}(\cdot, \cdot)$ .

**Example 3.21.** The maximal coupling of  $X \sim \text{Ber}(p)$  and  $Y \sim \text{Poi}(p)$  satisfies

$$\hat{\mathbb{P}}(\tilde{X} = \tilde{Y} = 0) = 1 - p, \quad \hat{\mathbb{P}}(\tilde{X} = \tilde{Y} = 1) = pe^{-p},$$

and  $\hat{\mathbb{P}}(\tilde{X} = \tilde{Y} = x) = 0$  for all  $x > 1$ . As a result,

$$d_{\text{TV}}(\tilde{X}, \tilde{Y}) \equiv \hat{\mathbb{P}}(\tilde{X} \neq \tilde{Y}) = 1 - \hat{\mathbb{P}}(\tilde{X} = \tilde{Y}) = p(1 - e^{-p}) \leq p^2. \quad (3.7)$$

Is either of the variables  $X$  and  $Y$  stochastically dominated by the other?

### 3.3 Applications to convergence

#### 3.3.1 Coupling of Markov chains

Coupling is a very useful tool for proving convergence towards equilibrium for various processes, including Markov chains and random walks. The following example uses the independent coupling<sup>19</sup> of two random walks on a complete graph.

**Example 3.22.** On the complete graph  $K_m$  on  $m$  vertices, consider the random walk  $(X_n)_{n \geq 0}$  with  $X_0 = x$ , such that at every step it jumps to any of the vertices of  $K_m$  uniformly at random (excluding the current one). To show that eventually  $(X_n)_{n \geq 0}$  forgets its initial distribution, one couples  $(X_n)_{n \geq 0}$  with another copy  $(Y_n)_{n \geq 0}$ ,  $Y_0 = y$ , of this random walk, and shows that for large  $n$  both processes coincide with large probability.

To do this, we treat the pair  $(\tilde{X}_n, \tilde{Y}_n)_{n \geq 0}$  as a two-component random walk on  $K_m \times K_m$  with the following jump probabilities: the transition  $(x, y) \mapsto (x', y')$  occurs with probability

$$\tilde{p}_{(x,y)(x',y')} = \begin{cases} \frac{1}{m^2}, & \text{if } x \neq y, \\ \frac{1}{m}, & \text{if } x = y \text{ and } x' = y', \end{cases}$$

and set  $\tilde{p}_{(x,y)(x',y')} = 0$  otherwise (ie., if  $x = y$  and  $x' \neq y'$ ). It is straightforward to check that this is a correct coupling.<sup>20</sup>

$$\sum_{y'} \tilde{p}_{(x,y)(x',y')} = m \cdot \frac{1}{m^2} = p_{x,x'}, \quad \sum_{y'} \tilde{p}_{(x,x)(x',y')} = \tilde{p}_{(x,x)(x',x')} = \frac{1}{m} = p_{x,x'}.$$

Notice that the walks  $(\tilde{X}_n)_{n \geq 0}$  and  $(\tilde{Y}_n)_{n \geq 0}$  run independently until they meet, and from that moment onwards they move together. For  $x \neq y$ , denote

$$T \stackrel{\text{def}}{=} \min\{n \geq 0 : \tilde{X}_n = \tilde{Y}_n\}. \quad (3.8)$$

<sup>19</sup>Recall the first table in Example 3.1.

<sup>20</sup>the argument for the other marginal is similar;

By a straightforward induction, it is easy to see that

$$\tilde{\mathbb{P}}(T > n) = \frac{1}{m} \left(1 - \frac{m-2}{(m-1)^2}\right)^n,$$

ie.,  $T$  is (basically) a geometric random variable. Since  $\tilde{X}_n$  and  $\tilde{Y}_n$  coincide after they meet, and  $\tilde{\mathbb{P}}, \tilde{X}_n, \tilde{Y}_n$  gives a coupling of  $(X_n, Y_n)$  for every  $n$  we have that

$$d_{\text{TV}}(X_n, Y_n) \leq \tilde{\mathbb{P}}(\tilde{X}_n \neq \tilde{Y}_n) = \tilde{\mathbb{P}}(T > n),$$

by (3.6). It therefore only remains to observe that the RHS above is exponentially small,  $\tilde{\mathbb{P}}(T > n) \leq \left(1 - \frac{1}{m}\right)^n \leq e^{-n/m}$ . In other words, the random walk  $(X_n)_{n \geq 0}$  forgets its initial state  $X_0 = x$  at least exponentially fast.

Notice that if  $(Y_n)_{n \geq 0}$  starts from the equilibrium (ie., its initial position is selected in  $K_m$  uniformly at random), our argument shows that for every initial state  $x$  and every state  $j$  we have  $|\mathbb{P}(X_n = j) - \frac{1}{m}| \leq e^{-n/m}$ , ie., convergence towards the equilibrium distribution is at least exponentially fast.<sup>21</sup>

The next example is important for computer science.

**Example 3.23.** [Random walk on a hypercube] An  $n$ -dimensional hypercube is just  $\mathcal{H}_n \equiv \{0, 1\}^n$ , a graph whose vertex set is the collection of all  $n$ -sequences made of 0 and 1 (there are  $2^n$  of them) and whose edges connect vertices which differ at a single position. The “lazy” random walk  $(W_k)_{k \geq 0}$  on  $\mathcal{H}_n$  is defined as follows: if  $W_k = v \in \mathcal{H}_n$ , select  $m$ ,  $1 \leq m \leq n$ , uniformly at random and flip a fair coin. If the coin shows heads, set the  $m$ th coordinate of  $v$  to 1, otherwise set it to 0. The walk  $(W_k)_{k \geq 0}$  is aperiodic and irreducible; as  $k$  increases, it tends to forget its initial state.

**Exercise 3.24.** Let  $(W_k)_{k \geq 0}$  be the random walk on the hypercube  $\mathcal{H}_n$ , recall Example 3.23. Use Example 3.22 to study its distribution for large  $k$ .

**Exercise 3.25.** Let  $(\xi_j)_{j \geq 1}$  be independent variables with  $\mathbb{P}(\xi = 1) = 1 - \mathbb{P}(\xi = -1) = p_x$  and let  $(\eta_j)_{j \geq 1}$  be independent variables with  $\mathbb{P}(\eta = 1) = 1 - \mathbb{P}(\eta = -1) = p_y$ . Define simple random walks  $(X_n)_{n \geq 0}$  and  $(Y_n)_{n \geq 0}$  on  $\mathbb{Z}$  via  $X_n = x + \sum_{j=1}^n \xi_j$  and  $Y_n = y + \sum_{j=1}^n \eta_j$ .

a) Show that the random walk  $(X_n)_{n \geq 0}$  on  $\mathbb{Z}$  forgets its initial state  $x$ ; namely, for  $p_x = p_y$  and  $y - x = 2k$ , construct a coupling of  $(X_n)_{n \geq 0}$  and  $(Y_n)_{n \geq 0}$  similar to that of Example 3.22.

b) Show that the random walk  $(X_n)_{n \geq 0}$  monotonously depends on  $p_x$ ; namely, for  $x \leq y$  and  $p_x \leq p_y$ , use the ideas from Example 3.1 to construct a monotone coupling of  $(X_n)_{n \geq 0}$  and  $(Y_n)_{n \geq 0}$ , ie., such that  $\tilde{\mathbb{P}}(\tilde{X}_n \leq \tilde{Y}_n) = 1$  for all  $n \geq 0$ .

---

<sup>21</sup>?? gives a more precise information about this convergence.

**Example 3.26.** [Random-to-top shuffling] For many practical purposes it is important to generate a random permutation of a collection of cards  $1, 2, \dots, m$ . One way is to define a Markov chain  $(X_n)_{n \geq 0}$  in the space of all possible permutations (so at any time  $n$ ,  $X_n$  will be a permutation of  $\{1, \dots, m\}$ ), and run it until the stationary distribution is reached. The stationary distribution is as usual the uniform distribution on all possible permutations.

One of the simplest algorithms—the random-to-top shuffling—is defined as follows: for a given state  $X_n = (\sigma(1), \dots, \sigma(m))$  of the chain, find the  $j$ th card by position (with  $j$  taken uniformly at random between 1 and  $m$ ) and put it at the top of the deck to get the new state  $X_{n+1} = (\sigma(j), \dots, \sigma(j-1), \sigma(j+1), \dots)$ . Note that we could take the  $j$ th card by value instead, and then the new state would be  $(j, \dots, \sigma(k-1), \sigma(k+1), \dots)$  for the unique  $k$  such that  $\sigma(k) = j$ . Since  $j$  is chosen uniformly at random, this gives the same algorithm (a Markov chain with the same transition probabilities).

To study the approach to stationarity, one couples two copies of the Markov chain,  $(X_n)_{n \geq 0}$  and  $(Y_n)_{n \geq 0}$ , one starting from a fixed state, and other starting from equilibrium, and shuffles until both configurations  $X_n$  and  $Y_n$  agree.

**Exercise 3.27.** Let  $(X_n)_{n \geq 0}$  and  $(Y_n)_{n \geq 0}$  be random-to-top shuffling (RTTS) Markov chains, recall Example 3.26. Suppose that a random card in  $X_n$  (by position) is chosen, say of value  $j$ , and then put at the top to get  $X_{n+1}$ . In  $Y_n$  find the (randomly positioned) card with number  $j$  and also move it to the top to get  $Y_{n+1}$ . Let  $T := \min\{k \geq 0 : Y_k = X_k\}$  be the coupling time for  $(X_n)_{n \geq 0}$  and  $(Y_n)_{n \geq 0}$ .

a) Show that the described algorithm provides a coupling<sup>22</sup> of the RTTS (by position) for  $(X_n)_{n \geq 0}$  and the RTTS (by value) for  $(Y_n)_{n \geq 0}$ .

b) Describe the distribution of  $T$ .

### 3.3.2 The Law of rare events

This subsection is optional and will not be examined.

The convergence result in Exercise 1.33 can also be derived using coupling:

**Theorem 3.28.** Let  $X = \sum_{k=1}^n X_k$ , where  $X_k \sim \text{Ber}(p_k)$  are independent random variables. Let, further,  $Y \sim \text{Poi}(\lambda)$ , where  $\lambda = \sum_{k=1}^n p_k$ . Then the maximal coupling of  $X$  and  $Y$  satisfies

$$d_{\text{TV}}(\tilde{X}, \tilde{Y}) \equiv \hat{\mathbb{P}}(\tilde{X} \neq \tilde{Y}) \leq \sum_{k=1}^n (p_k)^2.$$

*Proof.* Write  $Y = \sum_{k=1}^n Y_k$ , where  $Y_k \sim \text{Poi}(p_k)$  are independent rv's. Of course,  $\mathbb{P}(\sum_{k=1}^n X_k \neq \sum_{k=1}^n Y_k) \leq \sum_{k=1}^n \mathbb{P}(X_k \neq Y_k)$  for every joint distribution of  $(X_k)_{k=1}^n$  and  $(Y_k)_{k=1}^n$ . Let  $\hat{\mathbb{P}}_k$  be the maximal coupling for the pair  $\{X_k, Y_k\}$ ,

<sup>22</sup>try coupling RTTS (by value) and RTTS (by value) chains!

and let  $\widehat{P}_0$  be the maximal coupling for two sums. Notice that the LHS above is not smaller than  $d_{TV}(\widetilde{X}, \widetilde{Y}) \equiv \widehat{P}_0(\widetilde{X} \neq \widetilde{Y})$ ; on the other hand, using the (independent) product measure  $P = \widehat{P}_1 \times \cdots \times \widehat{P}_n$  on the right we deduce that then the RHS becomes just  $\sum_{k=1}^n \widehat{P}_k(\widetilde{X}_k \neq \widetilde{Y}_k)$ . The result now follows from (3.7).  $\square$

**Exercise 3.29.** Let  $X \sim \text{Bin}(n, \frac{\lambda}{n})$  and  $Y \sim \text{Poi}(\lambda)$  for some  $\lambda > 0$ . Show that

$$\frac{1}{2} |\mathbb{P}(X = k) - \mathbb{P}(Y = k)| \leq d_{TV}(\widetilde{X}, \widetilde{Y}) \leq \frac{\lambda^2}{n} \quad \text{for every } k \geq 0. \quad (3.9)$$

Deduce that for every fixed  $k \geq 0$ , we have  $\mathbb{P}(X = k) \rightarrow \frac{\lambda^k}{k!} e^{-\lambda}$  as  $n \rightarrow \infty$ .

**Remark 3.30.** Notice that (3.9) implies that if  $X \sim \text{Bin}(n, p)$  and  $Y \sim \text{Poi}(np)$  then for every  $k \geq 0$  the probabilities  $\mathbb{P}(X = k)$  and  $\mathbb{P}(Y = k)$  differ by at most  $2np^2$ . In particular, if  $n = 10$  and  $p = 0.01$  the discrepancy between any pair of such probabilities is bounded above by 0.002, i.e., they coincide in the first two decimal places.

### 3.4 Additional problems

**Exercise 3.31.** In the setting of Example 3.1 show that every convex linear combination of tables  $T_1$  and  $T_2$ , i.e., each table of the form  $T_\alpha = \alpha T_1 + (1-\alpha)T_2$  with  $\alpha \in [0, 1]$ , gives an example of a coupling of  $X \sim \text{Ber}(p_1)$  and  $Y \sim \text{Ber}(p_2)$ . Can you find all possible couplings for these variables?

**Exercise 3.32.** Generalise the inequality (3.2) to a broader class of functions  $g(\cdot)$  and verify that if  $X \preceq Y$ , then  $\mathbb{E}(X^{2k+1}) \leq \mathbb{E}(Y^{2k+1})$ ,  $\mathbb{E}e^{tX} \leq \mathbb{E}e^{tY}$  with  $t > 0$ ,  $\mathbb{E}s^X \leq \mathbb{E}s^Y$  with  $s > 1$  etc.

**Exercise 3.33.** Let  $\xi \sim \mathcal{U}(0, 1)$  be a standard uniform random variable. For fixed  $p \in (0, 1)$ , define  $X = \mathbf{1}_{\xi < p}$ . Show that  $X \sim \text{Ber}(p)$ , a Bernoulli random variable with parameter  $p$ . Now suppose that  $X = \mathbf{1}_{\xi < p_1}$  and  $Y = \mathbf{1}_{\xi < p_2}$  for some  $0 < p_1 \leq p_2 < 1$  and  $\xi$  as above. Show that  $X \preceq Y$  and that  $\mathbb{P}(X \leq Y) = 1$ . Compare your construction to the second table in Example 3.1.

**Exercise 3.34.** Let  $X \sim \text{Geom}(p)$  and  $Y \sim \text{Geom}(r)$  be two geometric random variables. If  $0 < p \leq r < 1$ , are  $X$  and  $Y$  stochastically ordered? Justify your answer.

**Exercise 3.35.** Let  $X \sim \Gamma(a, \lambda)$  and  $Y \sim \Gamma(b, \lambda)$  be two gamma random variables.<sup>23</sup> If  $0 < a \leq b < \infty$ , are  $X$  and  $Y$  stochastically ordered? Justify your answer.

<sup>23</sup>The density of  $Z \sim \Gamma(a, \lambda)$  is  $\frac{\lambda^a}{\Gamma(a)} x^{a-1} e^{-\lambda x}$  ( $x > 0$ ); notice that  $\Gamma(1, \lambda)$  is just  $\text{Exp}(\lambda)$ .



## 4 Martingales

### 4.1 Definition and some examples

A martingale is a generalized version of a “fair game”.

**Definition 4.1.** A process  $(M_n)_{n \geq 0}$  is a martingale if

a) for every  $n \geq 0$  the expectation  $\mathbf{E}M_n$  is finite, equivalently,  $\mathbf{E}|M_n| < \infty$ ;

b) for every  $n \geq 0$  and all  $m_n, m_{n-1}, \dots, m_0$  we have

$$\mathbf{E}(M_{n+1} \mid M_n = m_n, \dots, M_0 = m_0) = m_n. \quad (4.1)$$

For those familiar with the notion of conditioning on a random variable (see next section), (4.1) can just be written as  $\mathbf{E}(M_{n+1} \mid M_n, \dots, M_0) = M_n$ .

**Definition 4.2.** We say that  $(M_n)_{n \geq 0}$  is a supermartingale<sup>24</sup> if the equality in (4.1) holds with  $\leq$ , ie.,  $\mathbf{E}(M_{n+1} \mid M_n, \dots, M_0) \leq M_n$ ; and we say that  $(M_n)_{n \geq 0}$  is a submartingale, if (4.1) holds with  $\geq$ , ie.,  $\mathbf{E}(M_{n+1} \mid M_n, \dots, M_0) \geq M_n$ .

**Example 4.3.** Let  $(\xi_n)_{n \geq 1}$  be independent random variables<sup>25</sup> with

$$\mathbf{E}\xi_n = 0$$

for all  $n \geq 1$ . Then the process  $(M_n)_{n \geq 0}$  defined via

$$M_n \stackrel{\text{def}}{=} M_0 + \xi_1 + \dots + \xi_n$$

is a martingale as long as the random variable  $M_0$  is independent of  $(\xi_n)_{n \geq 1}$  and  $\mathbf{E}|M_0| < \infty$ . For example, we will often take  $M_0 = 0$  or some other deterministic constant.

Indeed, by the triangle inequality,  $\mathbf{E}|M_n| \leq \mathbf{E}|M_0| + \sum_{j=1}^n \mathbf{E}|\xi_j| < \infty$  for all  $n \geq 0$ ; on the other hand, the independence property implies

$$\mathbf{E}(M_{n+1} - M_n \mid M_n, \dots, M_0) \equiv \mathbf{E}(\xi_{n+1} \mid M_n, \dots, M_0) = \mathbf{E}\xi_{n+1} = 0.$$

**Remark 4.4.** Notice that if  $\mathbf{E}\xi_n \geq 0$  for all  $n \geq 1$ , then  $(M_n)_{n \geq 0}$  is a submartingale, whereas if  $\mathbf{E}\xi_n \leq 0$  for all  $n \geq 1$ , then  $(M_n)_{n \geq 0}$  is a supermartingale. More generally, if  $(\xi_n)_{n \geq 1}$  are independent random variables with  $\mathbf{E}|\xi_n| < \infty$  for all  $n \geq 1$ , then the process  $M_n = M_0 + (\xi_1 - \mathbf{E}\xi_1) + \dots + (\xi_n - \mathbf{E}\xi_n)$ ,  $n \geq 0$ , is a martingale.

**Exercise 4.5.** Given a sequence  $(\xi_k)_{k \geq 1}$  of independent Bernoulli variables with common distribution  $\mathbf{P}(\xi = 1) = p$  and  $\mathbf{P}(\xi = -1) = q = 1 - p$ , define the generated random walk via  $X_n = \sum_{k=1}^n \xi_k$ ,  $n \geq 0$ . Show that the process  $M_n = (q/p)^{X_n}$  is a martingale.

<sup>24</sup>If  $M_n$  traces your fortune, then “there is nothing super about a supermartingale”.

<sup>25</sup>Notice that we do not assume that all  $\xi_n$  have the same distribution!

**Example 4.6.** Let  $(Z_n)_{n \geq 0}$  be a branching process with  $\mathbb{E}Z_1 = m < \infty$ . We have  $\mathbb{E}|Z_n| < \infty$  and  $\mathbb{E}(Z_{n+1} | Z_n, \dots, Z_0) = mZ_n$  for all  $n \geq 0$ . In other words, the process  $(Z_n)_{n \geq 0}$  is a martingale, a submartingale or a supermartingale depending on whether  $m = 1$ ,  $m > 1$  or  $m < 1$ .

Notice also, that for every  $m \in (0, \infty)$  the process  $(Z_n/m^n)_{n \geq 0}$  is a martingale.

**Exercise 4.7.** Let  $\rho$  be the extinction probability for a branching process  $(Z_n)_{n \geq 0}$ ; show that  $M_n = \rho^{Z_n}$  is a martingale.

We can also construct many examples of submartingales and supermartingales if we have a martingale and apply certain functions to it. This is a consequence of:

**Lemma 4.8** (Conditional Jensen's inequality). Suppose that  $X$  is an integrable random variable on  $(\Omega, \mathcal{F}, \mathbb{P})$  and  $A \in \mathcal{F}$  with  $\mathbb{P}(A) > 0$ .

Suppose that  $\varphi$  is a convex function such that  $\varphi(X)$  is also integrable. Then

$$\mathbb{E}(\varphi(X)|A) \geq \varphi(\mathbb{E}(X|A)).$$

*Proof.* This follows from the non-conditional version of Jensen's inequality, and using the fact that  $\mathbb{P}(\cdot|A)$  is a valid probability measure.  $\square$

**Example 4.9.** Let  $(X_n)_{n \geq 0}$  be a martingale. If, for some convex function  $f(\cdot)$  we have  $\mathbb{E}(|f(X_n)|) < \infty$  for all  $n \geq 0$ , then the process  $f(X) := (f(X_n))_{n \geq 0}$  is a submartingale.

Similarly, if for some concave  $f(\cdot)$  we have  $\mathbb{E}(|f(X_n)|) < \infty$  for all  $n \geq 0$ , then the process  $f(X)$  is a supermartingale.

Both these claims follow immediately from the Jensen inequalities for conditional expectations.

## 4.2 A few remarks on conditional expectation

Recall the following basic definition:

**Example 4.10.** Let  $(\Omega, \mathcal{F}, \mathbb{P})$  be a probability triple and let  $X$  and  $Y$  be random variables taking values  $x_1, x_2, \dots, x_m$  and  $y_1, y_2, \dots, y_n$  respectively. On the event  $\{Y = y_j\}$  one defines the conditional probability

$$\mathbb{P}(X = x_i | Y = y_j) \stackrel{\text{def}}{=} \frac{\mathbb{P}(X = x_i, Y = y_j)}{\mathbb{P}(Y = y_j)}$$

and the conditional expectation:  $\mathbb{E}(X | Y = y_j) \equiv \sum_i x_i \mathbb{P}(X = x_i | Y = y_j)$ . Then the conditional expectation  $Z = \mathbb{E}(X | Y)$  of  $X$  given  $Y$  is defined as follows:

$$\text{if } Y(\omega) = y_j, \quad \text{then } Z(\omega) \stackrel{\text{def}}{=} z_j \equiv \mathbb{E}(X | Y = y_j).$$

Notice that the value  $z_j$  is completely determined by  $y_j$ ; in other words,  $Z$  is a function of  $Y$ , and as such, a random variable. Of course, if  $X$  and  $Y$  are independent, we have  $Z(\omega) \equiv \mathbb{E}X$ .

**Definition 4.11.** If  $\mathcal{D} = \{D_1, D_2, \dots, D_m\}$  is a finite partition<sup>26</sup> of  $\Omega$ , the collection  $\mathcal{G} = \sigma(\mathcal{D})$  of all  $2^m$  possible subsets of  $\Omega$  constructed from blocks  $D_i$  is called the  $\sigma$ -field generated by  $\mathcal{D}$ . A random variable  $Y$  is measurable with respect to  $\mathcal{G}$  if it is constant on every block  $D_i$  of the initial partition  $\mathcal{D}$ .

In Example 4.10 above we see that  $Z$  is constant on the events  $\{Y = y_j\}$ ; i.e.,  $Z$  is measurable w.r.t. the  $\sigma$ -field  $\sigma(Y) \equiv \sigma(\{Y = y_j\}_j)$ . Moreover, for every  $G_j \equiv \{Y = y_j\}$ , we have

$$\begin{aligned} \mathbb{E}(Z 1_{G_j}) &= z_j \mathbb{P}(Y = y_j) = \sum_i x_i \mathbb{P}(X = x_i | Y = y_j) \mathbb{P}(Y = y_j) \\ &= \sum_i x_i \mathbb{P}(X = x_i, Y = y_j) = \mathbb{E}(X 1_{G_j}), \end{aligned}$$

where the last equality follows from the observation that the random variable  $X 1_{G_j}$  equals  $x_i$  on every event  $G_j \cap \{X = x_i\} = \{Y = y_j\} \cap \{X = x_i\}$  and vanishes identically outside the set  $G_j$ .

**Remark 4.12.** The last two properties can be used to define conditional expectation in general: Let  $(\Omega, \mathcal{F}, \mathbb{P})$  be a probability triple, let  $\mathcal{G} \subseteq \mathcal{F}$  be a  $\sigma$ -field, and let  $X$  be a rv,  $X : \Omega \rightarrow \mathbb{R}$ . The conditional expectation of  $X$  w.r.t.  $\mathcal{G}$  is the unique random variable  $Z$  such that:  $Z$  is  $\mathcal{G}$  measurable and for every set  $G \in \mathcal{G}$  we have  $\int_G Z d\mathbb{P} \equiv \mathbb{E}(Z 1_G) = \mathbb{E}(X 1_G) \equiv \int_G X d\mathbb{P}$ .

Notice that when  $\mathcal{G} = \sigma(D)$ , the definition in the remark above coincides with Definition 4.11.

The following are the most important properties of conditional expectation:

**Lemma 4.13.** Let  $(\Omega, \mathcal{F}, \mathbb{P})$  be a probability triple, and let  $\mathcal{G} \subset \mathcal{F}$  a  $\sigma$ -field. Then for all random variables  $X, X_1, X_2$  and constants  $a_1, a_2$ , the following properties hold:

- a) If  $Z = \mathbb{E}(X | \mathcal{G})$  then  $\mathbb{E}Z = \mathbb{E}X$ ;
- b) If  $X$  is  $\mathcal{G}$ -measurable, then  $\mathbb{E}(X | \mathcal{G}) = X$ ;
- c)  $\mathbb{E}(a_1 X_1 + a_2 X_2 | \mathcal{G}) = a_1 \mathbb{E}(X_1 | \mathcal{G}) + a_2 \mathbb{E}(X_2 | \mathcal{G})$ ;
- d) If  $X \geq 0$ , then  $\mathbb{E}(X | \mathcal{G}) \geq 0$ ;
- e) If  $\mathcal{H}$  and  $\mathcal{G}$  are two  $\sigma$ -fields in  $\Omega$  such that  $\mathcal{H} \subseteq \mathcal{G}$ , then

$$\mathbb{E} \left[ \mathbb{E}(X | \mathcal{G}) | \mathcal{H} \right] = \mathbb{E} \left[ \mathbb{E}(X | \mathcal{H}) | \mathcal{G} \right] = \mathbb{E}(X | \mathcal{H}) ;$$

<sup>26</sup> In the general countable setting, if  $\mathcal{D} = \{D_1, D_2, \dots\}$  forms a denumerable (i.e., infinite countable) partition of  $\Omega$ , then the generated  $\sigma$ -field  $\mathcal{G} = \sigma(\mathcal{D})$  consists of all possible subsets of  $\Omega$  which are obtained by taking countable unions of blocks of  $\mathcal{D}$ . Similarly, a variable  $Y$  is measurable w.r.t.  $\mathcal{G}$ , if for every real  $y$  the event  $\{\omega : Y(\omega) \leq y\}$  belongs to  $\mathcal{G}$  (equivalently, can be expressed as a countable union of blocks of  $\mathcal{D}$ ).

f) If  $Z$  is  $\mathcal{G}$ -measurable, then  $\mathbf{E}[ZX | \mathcal{G}] = Z \mathbf{E}(X | \mathcal{G})$ .

g) If  $Z$  is independent of  $\mathcal{G}$ , then  $\mathbf{E}[Z | \mathcal{G}] = \mathbf{E}[Z]$ .

If  $(X_k)_{k \geq 1}$  is a sequence of random variables, we can define the generated  $\sigma$ -fields  $\mathcal{F}_1^X, \mathcal{F}_2^X, \dots$ , via

$$\mathcal{F}_k^X \stackrel{\text{def}}{=} \sigma(X_1, X_2, \dots, X_k); \quad (4.2)$$

here, the  $\sigma$ -field  $\mathcal{F}_k^X$  stores all information about the process  $(X_n)_{n \geq 1}$  up to time  $k$ . Observe that these  $\sigma$ -fields form a filtration  $(\mathcal{F}_n^X)_{n \geq 1}$  in the sense that

$$\mathcal{F}_1^X \subseteq \mathcal{F}_2^X \subseteq \dots \subseteq \mathcal{F}_k^X \subseteq \dots \quad (4.3)$$

The notion of filtration is very useful when working with martingales. Indeed, a process  $(M_n)_{n \geq 0}$  is a martingale if for all  $n \geq 0$ ,  $\mathbf{E}|M_n| < \infty$  and  $\mathbf{E}(M_{n+1} | \mathcal{F}_n^M) = M_n$ . We can also generalise the definition of a martingale:

**Definition 4.14.**  $(M_n)_{n \geq 0}$  is a martingale with respect to a filtration  $(\mathcal{F}_n)_{n \geq 0}$  if for all  $n \geq 0$ ,  $\mathbf{E}(|M_n|) < \infty$  and

$$\mathbf{E}(M_{n+1} | \mathcal{F}_n) = M_n.$$

The original definition of martingale is sometimes referred to as “being a martingale with respect to the natural filtration”.

We say that  $M$  is a martingale with respect to a sequence  $(X_n)_{n \geq 0}$  if it is a martingale with respect to the filtration  $(\mathcal{F}_n^X)_{n \geq 0}$ .

We also notice that by repeatedly applying the tower property in claim e) of Lemma 4.13 above to the sequence (4.3), we get the following result:

**Lemma 4.15.** If  $(M_n)_{n \geq 0}$  is a submartingale w.r.t.  $(X_n)_{n \geq 0}$ , then for all integer  $n \geq k \geq 0$ , we have  $\mathbf{E}M_n \geq \mathbf{E}M_k$ .

**Remark 4.16.** Changing signs, we get the inequality  $\mathbf{E}M_n \leq \mathbf{E}M_k$  for supermartingales, and therefore the equality  $\mathbf{E}M_n = \mathbf{E}M_k$  for martingales.

### 4.3 Martingales and stopping times

Martingales are extremely useful in studying stopping times:

**Example 4.17.** Let  $(X_k)_{k \geq 1}$  be a sequence of i.i.d. Bernoulli random variables with common distribution  $\mathbf{P}(X_1 = 1) = p$ ,  $\mathbf{P}(X_1 = -1) = q = 1 - p$ , where  $p \in (0, 1)$ ,  $p \neq 1/2$ . For fixed  $k \in \{0, 1, \dots, N\}$ , define the random walk  $(S_n)_{n \geq 0}$  defined via  $S_n = k + \sum_{j=1}^n X_j$ ,  $n \geq 0$ , and consider the process  $(Y_n)_{n \geq 0}$ , where  $Y_n = (q/p)^{S_n}$ . Clearly,  $Y_n$  is an  $(\mathcal{F}_n^X)_{n \geq 0}$ -martingale, so that  $\mathbf{E}(Y_n) = \mathbf{E}(Y_0) = (q/p)^k$  for all  $n \geq 0$ , recall Lemma 4.15.

Let  $T$  be the first time  $S_n$  hits 0 or  $N$ . If an analogue of the above equality,  $E(Y_T) = E(Y_0) = (q/p)^k$  were true for (random) time  $T$ , we could find the exit probability  $\mathbf{p}_k = P(S \text{ hits } 0 \text{ before } N \mid S_0 = k)$  from the expression

$$E(Y_T) = (q/p)^0 \mathbf{p}_k + (q/p)^N (1 - \mathbf{p}_k),$$

thus obtaining  $\mathbf{p}_k = ((q/p)^k - (q/p)^N) / (1 - (q/p)^N)$ .

**Remark 4.18.** The method used in the previous example relies on the assumption  $E(Y_T) = E(Y_0)$  and on the formula for  $E(Y_T)$  for a certain random variable  $T$ . An important part of the theory of martingales is to study random variables  $T$  for which the above statements are true.<sup>27</sup>

**Example 4.19.** Let  $S_n = \sum_{k=1}^n X_k$  be the simple symmetric random walk in  $\{-K, \dots, K\}$ , generated by a sequence of i.i.d. symmetric Bernoulli r.v.  $X_k$ , where  $P(X_1 = \pm 1) = 1/2$ . Similarly to Example 4.17 one can study the hitting time  $T$  of the boundary,  $T = \inf\{n \geq 0 : |S_n| = K\}$ : namely, since<sup>28</sup>

$$E((S_T)^2) = K^2 P(S_T = K) + K^2 P(S_T = -K) = K^2,$$

the same heuristics applied to the martingale  $(Y_n)_{n \geq 0}$ ,  $Y_n \stackrel{\text{def}}{=} (S_n)^2 - n$ , leads to  $0 = E(Y_0) = E(Y_T) = E((S_T)^2) - E(T)$ ; i.e., it suggests that  $E(T) = K^2$ .

One of our aims is to discuss general results that justify the above heuristics. To this end, we need to carefully define what we mean by a “stopping time”.

**Definition 4.20.** A variable  $T$  is a stopping time for a process  $(X_n)_{n \geq 0}$ , if the occurrence/non-occurrence of the event  $\{T = n\}$  = “we stop at time  $n$ ” can be determined by looking at the values  $X_0, X_1, \dots, X_n$  of the process up to time  $n$ . Equivalently, if we have  $\{T \leq n\} \in \mathcal{F}_n^X \equiv \sigma(X_0, \dots, X_n)$  for every  $n \geq 0$ .

**Example 4.21.** Let  $(X_n)_{n \geq 0}$  be a stochastic process with values in  $S$ , and let  $T$  be the hitting time of a set  $A \subset S$ , namely,  $T = \min\{n \geq 0 : X_n \in A\}$ . Then  $T$  is a stopping time for  $(X_n)_{n \geq 0}$ .

Indeed, for every fixed  $n \geq 0$ , we have  $\{T > n\} \equiv \{X_0 \notin A, X_1 \notin A, \dots, X_n \notin A\}$ ; therefore, the event  $\{T > n\}$  and its complement  $\{T \leq n\}$  both belong to  $\mathcal{F}_n^X$ .

By contrast, the last time that  $X$  visits  $A$ ,  $\tilde{T} = \max\{n \geq 0 : X_n \in A\}$  is not generally a stopping time. Because we generally cannot tell just by looking at  $X_0, \dots, X_n$ , whether the process will visit  $A$  after time  $n$  or not.

**Exercise 4.22.** Let  $k \in \mathbb{N}$  be fixed, and let  $S$  and  $T$  be stopping times for a process  $(X_n)_{n \geq 0}$ . Show that the following are stopping times:

(a)  $T \equiv k$ ,

<sup>27</sup>Notice that the gambler’s ruin problem can be solved by using the methods of finite Markov chains, so we indeed know that the result above is correct.

<sup>28</sup>the equality is correct, because  $P(T < \infty) = 1$  here!

(b)  $S \wedge T \equiv \min(S, T)$ ,

(c)  $S \vee T \equiv \max(S, T)$ .

Let  $A$  be a set of states, and let  $T = T_A$  be the moment of the first visit to  $A$ , i.e.,  $T = \min\{n \geq 0 : X_n \in A\}$ . Consider  $S = S_A = \min\{n > T_A : X_n \in A\}$ , the moment of the second visit to  $A$ . Show that  $S_A$  is a stopping time for  $(X_n)_{n \geq 0}$ . Is the variable  $L = \max\{n \geq 0 : X_n \in A\}$  a stopping time for  $(X_n)_{n \geq 0}$ ?

**Exercise 4.23.** Let  $(M_n)_{n \geq 0}$  be a submartingale w.r.t.  $(X_n)_{n \geq 0}$ , and let  $T$  be a stopping time for  $(X_n)_{n \geq 0}$ . Show that the process  $(L_n^T)_{n \geq 0}$ ,

$$L_n^T \stackrel{\text{def}}{=} M_{n \wedge T} \equiv \begin{cases} M_n, & n \leq T, \\ M_T, & n > T, \end{cases}$$

is a submartingale w.r.t.  $(X_n)_{n \geq 0}$ . Deduce that if  $(M_n)_{n \geq 0}$  is a martingale w.r.t.  $(X_n)_{n \geq 0}$ , then the stopped process  $(L_n^T)_{n \geq 0}$  is also a martingale w.r.t.  $(X_n)_{n \geq 0}$ .

#### 4.4 Optional stopping theorem

The optional stopping (or sampling) theorem (OST) tells us that, under quite general assumptions, whenever  $X_n$  is a martingale, then  $X_{T \wedge n}$  is a martingale for a stopping time  $T$ . Such results are very useful in deriving inequalities and probabilities of various events associated with such stochastic processes.

**Theorem 4.24** (Optional Stopping Theorem). Let  $(M_n)_{n \geq 0}$  be a martingale w.r.t.  $(X_n)_{n \geq 0}$ , and let  $T$  be a stopping time for  $(X_n)_{n \geq 0}$ . Then the equality

$$\mathbb{E}[M_T] = \mathbb{E}[M_0] \tag{4.4}$$

holds whenever **one** of the following conditions holds:

(OST-1) the stopping time  $T$  is bounded, i.e.,  $\mathbb{P}(T \leq N) = 1$  for some  $N < \infty$ ;

(OST-2)  $\mathbb{E}T < \infty$ , and the martingale  $(M_n)_{n \geq 0}$  has bounded increments, i.e.,  $|M_{n+1} - M_n| \leq K$  for all  $n$  and some constant  $K$ ;

(OST-3)  $\mathbb{P}(T < \infty) = 1$ , and the martingale  $(M_n)_{n \geq 0}$  is bounded, i.e.,  $|M_n| \leq K$  for all  $n$  and some constant  $K$ .

**Remark 4.25.** If  $M_n$  records gambler's fortune, by (OST-3), one cannot make money from a fair game, unless an unlimited amount of credit is available.

**Remark 4.26.** One can generalize (OST-3) by replacing the condition of boundedness,  $|M_n| \leq K$ , by that of uniform integrability for the martingale  $(M_n)_{n \geq 0}$ : a sequence of random variables  $(Y_n)_{n \geq 0}$  is uniformly integrable if

$$\lim_{K \rightarrow \infty} \sup_{n \geq 0} \mathbb{E}(|Y_n| \mathbf{1}_{\{|Y_n| > K\}}) = 0. \tag{4.5}$$

**Example 4.27.** Let the SSRW  $(S_n)_{n \geq 0}$  be generated as in Example 4.19. Put

$$H \stackrel{\text{def}}{=} \inf\{n \geq 0 : S_n = 1\}.$$

Since this RW is recurrent,<sup>29</sup> we deduce that  $\mathbb{P}(H < \infty) = 1$ . However, the (OST) does not apply, as  $0 = \mathbb{E}(S_0) \neq \mathbb{E}(S_H) \equiv 1$ . It is an instructive Exercise to check which conditions in each of the above (OST) are violated.

We now give the proof of Theorem 4.24.

*Proof.* Let us first assume that  $T$  satisfies (OST-1); that is, there is some  $N < \infty$  with  $0 \leq T \leq N$ . Then the decomposition

$$M_T = \sum_{n=0}^N M_T \mathbf{1}_{\{T=n\}} = \sum_{n=0}^N M_n \mathbf{1}_{\{T=n\}} = M_0 + \sum_{n=0}^{N-1} (M_{n+1} - M_n) \mathbf{1}_{\{T>n\}}$$

holds. Now for each  $0 \leq n \leq N-1$ , (a) of Lemma 4.13 allows us to rewrite  $\mathbb{E}((M_{n+1} - M_n) \mathbf{1}_{\{T>n\}})$  as

$$\mathbb{E}(\mathbb{E}((M_{n+1} - M_n) \mathbf{1}_{\{T>n\}} | \mathcal{F}_n^X)) = \mathbb{E}(\mathbf{1}_{\{T>n\}} \mathbb{E}(M_{n+1} - M_n | \mathcal{F}_n^X))$$

where we have used that  $\{T > n\}$  is  $\mathcal{F}_n^X$ -measurable (by definition of a stopping time). Since  $M$  is a martingale,  $\mathbb{E}(M_{n+1} - M_n | \mathcal{F}_n^X) = 0$  and hence  $\mathbb{E}((M_{n+1} - M_n) \mathbf{1}_{\{T>n\}}) = 0$  for each  $n$ . Linearity of expectation then implies that  $\mathbb{E}(M_T) = \mathbb{E}(M_0)$  as required.

Now let us assume that  $T$  satisfies (OST-2). For fixed  $n$ ,  $\min(T, n) = T \wedge n$  is clearly a bounded stopping time (it is less than  $n$  with probability one), so by (OST-1) we have that  $\mathbb{E}M_{T \wedge n} = \mathbb{E}M_0$  for all  $n \geq 0$ . We now want to take a limit as  $n \rightarrow \infty$ . For this we write

$$|M_T - M_{T \wedge n}| = \left| \sum_{k>n} (M_k - M_n) \mathbf{1}_{\{T=k\}} \right| = \left| \sum_{k>n} (M_k - M_{k-1}) \mathbf{1}_{\{T \geq k\}} \right|$$

(in particular, noting that this gives  $|M_T| \leq KT$  when  $n = 0$  and so  $M_T$  is integrable). Taking expectations then gives that  $\mathbb{E}(|M_T - M_{T \wedge n}|) \leq K \sum_{k>n} \mathbb{P}(T \geq k)$  which tends to 0 as  $n \rightarrow \infty$ . Hence  $\mathbb{E}(M_T) = \lim_{n \rightarrow \infty} \mathbb{E}(M_{T \wedge n}) = \lim_{n \rightarrow \infty} \mathbb{E}(M_0) = \mathbb{E}(M_0)$ .

Finally, we can assume (OST-3) and deduce the result in a similar way. The strategy is the same, but instead of writing  $M_T - M_{T \wedge n}$  as a telescoping sum, we note that  $|M_T - M_{T \wedge n}| \leq 2K \mathbf{1}_{\{T>n\}}$  so that

$$|\mathbb{E}M_T - \mathbb{E}M_{T \wedge n}| \leq 2K \mathbb{P}(T > n) \rightarrow 0 \quad \text{as } n \rightarrow \infty.$$

□

**Remark 4.28.** The proofs above will work exactly the same way if in (OST-2) one only assumes that  $|M_{T \wedge n} - M_{T \wedge (n+1)}| \leq K$  for all  $n$ , or if in (OST-3) if one only assumes that  $|M_{T \wedge n}| \leq K$  for all  $n$ .

<sup>29</sup>alternatively, recall the result in Example 1.26.

With suitable martingales (OST) gives very powerful results.  
 The following instructive example is due to D. Williams:

**Example 4.29. (ABRACADABRA)** *A monkey types symbols at random, one per unit time, on a typewriter having 26 keys. How long on average will it take him to produce the sequence ‘ABRACADABRA’?*

**Solution** *To compute the expected time, imagine a sequence of gamblers, each initially having £1, playing at a fair gambling casino. Gambler arriving just before time  $n$  ( $n \geq 1$ ) bets £1 on the event { $n$ th letter will be A}. If he loses, he leaves. If he wins, he receives £26 all of which he bets on the event { $n + 1$  st letter will be B}. If he loses, he leaves. If he wins, he receives £26<sup>2</sup> all of which he bets on the event { $n + 2$  nd letter will be R} and so on through the whole ABRACADABRA sequence.*

*Let  $X_n$  denote the total winnings of the casino after the  $n$ th day. Since all bets are fair the process  $(X_n)_{n \geq 0}$  is a martingale with mean zero. Let  $N$  denote the time until the sequence ABRACADABRA appears. At time  $N$ , gambler  $N - 10$  would have won £26<sup>11</sup> - 1, gambler  $N - 3$  would have won £26<sup>4</sup> - 1, gambler  $N$  would have won £26 - 1 and all other  $N - 3$  gamblers would have lost their initial fortune. Therefore,*

$$X_N = N - 3 - (26^{11} - 1) - (26^4 - 1) - (26 - 1) = N - 26^{11} - 26^4 - 26$$

*and since (OST-2) can be applied (check this!), we deduce the  $E(X_N) = E(X_0) = 0$ , that is*

$$E(N) = 26 + 26^4 + 26^{11}.$$

**Remark 4.30.** *Notice that the answer could also be obtained by considering a finite state Markov chain  $X_n$  on the state space of 12 strings representing the longest possible intersection of the tail of the typed text with the target word ABRACADABRA, ie., {ABRACADABRA, ABRACADABR, ..., ABRA, ABR, AB, A,  $\emptyset$ }, as there  $N$  is just the hitting time of the state ABRACADABRA from the initial condition  $X_0 = \emptyset$ .*

**Exercise 4.31.** *Use an appropriate (OST) to carefully derive the probability  $p_k$  in Example 4.17.*

**Exercise 4.32.** *Use an appropriate (OST) to carefully derive the expectation  $E(T)$  in Example 4.19.*

**Exercise 4.33.** *Consider the simple symmetric random walk  $(S_n)_{n \geq 0}$ , generated by a sequence of i.i.d. Bernoulli r.v.  $X_k$  with  $P(X_1 = \pm 1) = 1/2$ , ie.,  $S_n = \sum_{k=1}^n X_k$ . For integer  $a < 0 < b$ , let  $T$  be the stopping time  $T = \inf\{n \geq 0 : S_n \notin (a, b)\} \equiv \inf\{n \geq 0 : S_n \in \{a, b\}\}$ .*

*(a) Show that  $(S_n)_{n \geq 0}$  is a martingale and use an appropriate (OST) to find  $P(S_T = a)$  and  $P(S_T = b)$ .*



- (b) Show that  $(M_n)_{n \geq 0}$  defined via  $M_n = (S_n)^2 - n$  is a martingale w.r.t. the process  $(S_n)_{n \geq 0}$ .
- (c) For fixed integer  $K > 0$ , carefully apply an appropriate (OST) to  $M_n$  and prove that  $\mathbf{E}(T \wedge K) = \mathbf{E}[(S_{T \wedge K})^2]$ .
- (d) Deduce that  $\mathbf{E}(T) = -ab$ .

**Exercise 4.34.** A coin showing heads with probability  $p$  is tossed repeatedly. Let  $w$  be a fixed sequence of outcomes such as 'HTH', and let  $N$  denote the number of (independent) tosses until the word  $w$  is observed. Using an appropriate martingale, find the expectation  $\mathbf{E}N$  for each of the following sequences: 'HH', 'HTH', 'HHTTHH'.

**Lemma 4.35.** Let  $(Y_n)_{n \geq 0}$  be a supermartingale w.r.t. a sequence  $(X_n)_{n \geq 0}$  and let  $H_n \in \mathcal{F}_{n-1}^X = \sigma(X_0, \dots, X_{n-1})$  satisfy  $0 \leq H_n \leq c_n$ , where the constant  $c_n$  might depend on  $n$ . Then the process  $W_n = W_0 + \sum_{m=1}^n H_m(Y_m - Y_{m-1})$ ,  $n \geq 0$ , is a supermartingale w.r.t.  $(X_n)_{n \geq 0}$ .

*Proof.* Following the proof of the optional stopping theorem, we observe that since  $(Y_n)_{n \geq 0}$  is a supermartingale w.r.t.  $(X_n)_{n \geq 0}$ ,

$$\mathbf{E}(H_m(Y_m - Y_{m-1})) = \mathbf{E}[H_m \mathbf{E}(Y_m - Y_{m-1} | \mathcal{F}_{m-1})] \leq 0.$$

□

**Example 4.36.** If  $(Y_n)_{n \geq 0}$  describes the stock price process, and  $H_m$  is the number of stocks held during the time  $(m-1, m]$  (decided when the price  $Y_{m-1}$  is known), then  $W_n$  describes the fortune of an investor at time  $n$ . As  $(W_n)_{n \geq 0}$  is a supermartingale w.r.t.  $(X_n)_{n \geq 0}$ , we have  $\mathbf{E}W_n \leq \mathbf{E}W_0$  for all  $n \geq 0$ .

**Remark 4.37.** The famous "doubling martingale" corresponds to doubling the bet size until one wins, i.e., to taking  $H_m = 2^{m-1} \mathbf{1}_{\{T > m\}}$ , where  $T$  is the first moment when the price goes up, i.e.,  $T = \min\{m > 0 : Y_m - Y_{m-1} = 1\}$ . Since the stopped process  $(W_{n \wedge T})_{n \geq 0}$  is a supermartingale, for all  $n \geq 0$  we have  $\mathbf{E}(W_{n \wedge T}) \leq \mathbf{E}(W_0)$ , i.e., on average, the doubling strategy does not produce money if one bets against a (super)martingale.

**Example 4.38.** [Wald's equation] Let  $(S_n)_{n \geq 0}$  be a random walk generated by a sequence  $(X_n)_{n \geq 0}$  of i.i.d. steps with  $\mathbf{E}|X| < \infty$  and  $\mathbf{E}(X) = m$ . If  $T$  is a stopping time for  $(X_n)_{n \geq 0}$  with  $\mathbf{E}(T) < \infty$ , then the optional stopping theorem implies that  $S_T$  is integrable and

$$\mathbf{E}(S_T - S_0) = m \mathbf{E}(T).$$

To show this, first notice that  $S_n - nEX = S_n - mn$  is a martingale and for every  $n \geq 0$  the variable  $T \wedge n$  is a bounded stopping time. By (OST-1), we have

$$\mathbf{E}(S_0) = \mathbf{E}(S_{n \wedge T} - m(n \wedge T)). \quad (4.6)$$

This rearranges to

$$E(S_{n \wedge T} - S_0) = mE(n \wedge T)$$

for every  $n$ . Now, the RHS converges to  $E(T)$  as  $n \rightarrow \infty$  since  $|E(T) - E(n \wedge T)| \leq E(T1_{\{T > n\}}) = \sum_{k > n} kP(T = k)$ , where the tail sums on the right go to zero as  $n \rightarrow \infty$  by the assumption that  $T$  is integrable. Next, by writing  $|S_T| = \sum_{k \geq 0} |X_k|1_{\{T \geq k\}}$  as a telescoping sum, where  $E(|X_k|1_{\{T \geq k\}}) = P(T \geq k)E(|X_1|)$  and  $\sum_k P(T \geq k) < \infty$ , we see that  $|S_T|$  is integrable.<sup>30</sup> Similarly, we can bound  $E(|S_T - S_{T \wedge n}|) \leq E(|X_1|) \sum_{k > n} P(T \geq k)$  which tends to 0 as  $n \rightarrow \infty$ . This implies that  $E(S_{T \wedge n}) \rightarrow E(S_T)$  as  $n \rightarrow \infty$ , and combining the above completes the argument.

## 4.5 Martingale convergence theorem

This subsection is optional and will not be examined.

The following example has a number of important applications.

**Example 4.39** (Pólya's urn). *An urn contains one green and one red ball. At every step a ball is selected at random, and then replaced together with another ball of the same colour. Let  $X_n$  be the number of green balls after  $n$ th draw,  $X_0 = 1$ . Then the fraction  $M_n = X_n/(n+2)$  of green balls is a martingale w.r.t. the filtration  $(\mathcal{F}_n^X)_{n \geq 0}$ .*

*Indeed, as  $|M_n| \leq 1$  we have  $E|M_n| \leq 1$  for all  $n \geq 0$ , and since*

$$P(X_{n+1} = k+1 | X_n = k) = \frac{k}{n+2}, \quad P(X_{n+1} = k | X_n = k) = 1 - \frac{k}{n+2},$$

*we get  $E(X_{n+1} | \mathcal{F}_n^X) = \frac{n+3}{n+2} X_n$ , equivalently,  $E(M_{n+1} | \mathcal{F}_n^X) = M_n$ .*

**Exercise 4.40.** *Show that  $P(M_n = \frac{k}{n+2}) = \frac{1}{n+1}$  for  $1 \leq k \leq n+1$ , ie.,  $M_n$  is uniformly distributed in  $\{\frac{1}{n+2}, \frac{2}{n+2}, \dots, \frac{n+1}{n+2}\}$ .*

Exercise 4.40 suggests that in the limit  $n \rightarrow \infty$  the distribution of  $M_n$  becomes uniform in  $(0, 1)$ :

**Exercise 4.41.** *Show that  $\lim_{n \rightarrow \infty} P(M_n < x) = x$  for every  $x \in (0, 1)$ .*

In view of Exercise 4.40, a natural question is: does the proportion  $M_n$  of green balls fluctuate between 0 and 1 infinitely often or does it eventually settle down to a particular value? The following example shows that the latter is true. Our argument is based upon the following observation: if a real sequence  $y_n$  does not converge, for some real  $a, b$  with  $-\infty < a < b < \infty$  the sequence  $y_n$  must go from the region below  $a$  to the region above  $b$  (and back) infinitely often.

<sup>30</sup>Here we are really using the monotone convergence theorem: if  $0 \leq Z_n \leq Z$  for every  $n$  and  $Z_n \uparrow Z$  as  $n \rightarrow \infty$ , then  $E(Z_n) \rightarrow E(Z)$  as  $n \rightarrow \infty$ .

**Example 4.42.** For fixed  $n \geq 0$  let  $M_n < a \in (0, 1)$ , and let  $N = \min\{k > n : M_n > b\}$  for some  $b \in (a, 1)$ . Since  $N_m = N \wedge m$  is a bounded stopping time, by (OST-1) we have  $\mathbb{E}M_{N_m} = \mathbb{E}M_n < a$  if only  $m > n$ . On the other hand,

$$\mathbb{E}M_{N_m} \geq \mathbb{E}(M_{N_m} \mathbf{1}_{N \leq m}) \equiv \mathbb{E}(M_N \mathbf{1}_{N \leq m}) > b \mathbb{P}(N \leq m).$$

In other words,  $\mathbb{P}(N \leq m) < \frac{a}{b}$  and consequently  $\mathbb{P}(N < \infty) \leq \frac{a}{b} < 1$ , i.e., the fraction  $M_n$  of green balls ever gets above level  $b$  with probability at most  $\frac{a}{b}$ . Suppose that at certain moment  $N \in (n, \infty)$  the fraction of green balls became bigger than  $b$ . Then a similar argument shows that with probability at most  $(1-b)/(1-a)$  the value  $M_n$  becomes smaller than  $a$  at a later moment.

Put  $S_0 = \min\{n \geq 0 : M_n < a\}$ , and then, inductively, for  $k \geq 0$ ,

$$T_k = \min\{n > S_k : M_n > b\}, \quad S_{k+1} = \min\{n > T_k : M_n < a\}. \quad (4.7)$$

The argument above implies that

$$\mathbb{P}(S_k < \infty) \leq \prod_{j=1}^k \left( \mathbb{P}(T_{j-1} < \infty \mid S_{j-1} < \infty) \mathbb{P}(S_j < \infty \mid T_{j-1} < \infty) \right) \quad (4.8)$$

with the RHS bounded above by  $\left(\frac{a}{b}\right)^k \left(\frac{1-b}{1-a}\right)^k \rightarrow 0$  as  $k \rightarrow \infty$ . As a result, the probability of infinitely many crossing (i.e.,  $S_k < \infty$  for all  $k \geq 0$ ) vanishes.

Clearly, the argument above applies to all strips  $(a, b) \subset (0, 1)$  with rational endpoints. Thus, with probability one,<sup>31</sup> trajectories of  $M_n$  eventually converge to a particular value.<sup>32</sup>

**Exercise 4.43.** Write  $U_{(a,b)}$  for the total number of upcrossings of the strip  $(a, b)$  by the process  $(M_n)_{n \geq 0}$ . By using the approach of Example 4.42 and noticing that  $\{U_{(a,b)} \geq m\} \subset \{S_{m-1} < \infty\}$  or otherwise, show that  $\mathbb{E}U_{(a,b)} < \infty$ .

The argument in Example 4.42 also works in general. Let  $(M_n)_{n \geq 0}$  be a martingale w.r.t. filtration  $(\mathcal{F}_n^X)_{n \geq 0}$ . For real  $a, b$  with  $-\infty < a < b < \infty$  let  $U_{(a,b)}$  be the total number of upcrossings of the strip  $(a, b)$ . The following result (or some of its variants) is often referred to as Doob's Upcrossing Lemma:

**Lemma 4.44.** Let the martingale  $(M_n)_{n \geq 0}$  have uniformly bounded expectations, i.e., for some constant  $K$  and all  $n \geq 0$ ,  $\mathbb{E}|M_n| < K < \infty$ . If  $U_{(a,b)}$  is the number of upcrossings of a strip  $(a, b)$ , then  $\mathbb{E}U_{(a,b)} < \infty$ .

*Proof.* With stopping times as in (4.7), put  $H_n = 1$  if  $S_m < n \leq T_m$  and put  $H_n = 0$  otherwise. Then the process  $W_n = \sum_{k=1}^n H_k (M_k - M_{k-1})$  is a martingale w.r.t.  $(M_n)_{n \geq 0}$ , cf. Lemma 4.35. It is easy to check that  $W_n \geq (b-a)U_n - |M_n - a|$  (draw the picture!), where  $U_n = \max\{m \geq 0 : T_m \leq n\}$  is the number of upcrossings of the strip  $(a, b)$  up to time  $n$ . As a result

$$0 = \mathbb{E}W_0 = \mathbb{E}W_n \geq (b-a)\mathbb{E}U_n - \mathbb{E}|M_n - a| \geq (b-a)\mathbb{E}U_n - (K + |a|),$$

so that  $\mathbb{E}U_n \leq (K + |a|)/(b-a)$  for all  $n \geq 0$ , and thus  $\mathbb{E}U_{(a,b)} < \infty$ .  $\square$

<sup>31</sup>If  $M_n$  does not converge, it must cross at least one of countably many strips  $(a, b)$  with rational points infinitely many times.

<sup>32</sup>which is random and depends on the trajectory

**Theorem 4.45.** Let  $(M_n)_{n \geq 0}$  be a martingale as in Lemma 4.44. Then there exists a random variable  $M_\infty$  such that  $M_n \rightarrow M_\infty$  with probability one.

*Proof.* If  $M_n$  does not converge, for some rational  $a, b$  with  $-\infty < a < b < \infty$  we must have  $U_{(a,b)} = \infty$ . However, by Lemma 4.44,  $\mathbb{E}U_{(a,b)} < \infty$  implying that  $\mathbb{P}(U_{(a,b)} = \infty) = 0$ . As the number of such pairs  $(a, b)$  is countable, the result follows.  $\square$

**Exercise 4.46.** Let  $(X_k)_{k \geq 1}$  be independent variables with

$$\mathbb{P}\left(X = \frac{3}{2}\right) = \mathbb{P}\left(X = \frac{1}{2}\right) = \frac{1}{2}.$$

Put  $M_n = X_1 \cdots X_n$  with  $M_0 = 1$ . Show that  $M_n$  is an  $(\mathcal{F}_n^X)_{n \geq 0}$  martingale. Deduce that  $M_n \rightarrow M_\infty$  with probability one. Can you compute  $\mathbb{E}(M_\infty)$ ?

## 4.6 Additional problems

**Exercise 4.47.** Let  $(\eta_n)_{n \geq 1}$  be independent positive random variables with  $\mathbb{E}\eta_n = 1$  for all  $n \geq 1$ . If a random variable  $M_0 > 0$  is independent of  $(\eta_n)_{n \geq 0}$  and  $\mathbb{E}M_0 < \infty$ , then the process  $(M_n)_{n \geq 0}$  defined via  $M_n = M_0 \prod_{j=1}^n \eta_j$  is a martingale w.r.t.  $(\eta_n)_{n \geq 1}$ .

Interpreting  $\eta_n - 1$  as the (fractional) change in the value of a stock during the  $n$ th time interval, the martingale  $(M_n)_{n \geq 0}$  can be used to model stock prices. Two often used examples are:

Discrete Black-Sholes model: take  $\eta_j = e^{\zeta_j}$ , where  $\zeta_j$  is Gaussian,  $\zeta_j \sim \mathcal{N}(\mu, \sigma^2)$ ;

Binomial model: take  $\eta_j = (1+a)e^{-r}$  and  $\eta_j = (1+a)^{-1}e^{-r}$  with probabilities  $p$  and  $1-p$  respectively.

**Exercise 4.48.** Let  $(S_n)_{n \geq 0}$  be the random walk from Example 4.19. Find constants  $a, b, c$  such that the process  $(S_n)^4 + an(S_n)^2 + bn^2 + cn$  is an  $(\mathcal{F}_n)_{n \geq 0}$ -martingale. Use the heuristic in Example 4.19 to predict the value of the second moment  $\mathbb{E}(T^2)$  of the exit time  $T$ .

**Exercise 4.49.** A standard symmetric dice is tossed repeatedly. Let  $N$  be the number of (independent) tosses until a fixed pattern is observed. Using an appropriate martingale, find  $\mathbb{E}N$  for the sequences '123456' and '123321'.

**Exercise 4.50.** Suppose that the process in Example 4.39 is modified as follows: for a fixed integer  $c > 1$ , every time a random ball is selected, it is replaced together with other  $c$  balls of the same colour. If, as before,  $X_n$  denotes the total number of green balls after  $n$  draws, show that the fraction  $M_n = \frac{X_n}{2+nc}$  of green balls forms a martingale w.r.t.  $(\mathcal{F}_n^X)_{n \geq 0}$ .

**Exercise 4.51.** Find the large- $n$  limit of the distribution of the martingale  $(M_n)_{n \geq 0}$  from Exercise 4.50.

**Exercise 4.52.** Let  $(X_n)_{n \geq 0}$  be a birth-and-death process in  $\mathcal{S} = \{0, 1, \dots\}$ , ie., a Markov chain in  $\mathcal{S}$  with transition probabilities  $p_{00} = r_0$ ,  $p_{01} = p_0$ , and  $p_{m,m-1} = q_m$ ,  $p_{m,m} = r_m$ ,  $p_{m,m+1} = p_m$  for  $m > 0$ , while  $p_{m,k} = 0$  for all other pairs  $(m, k) \in \mathcal{S}^2$ . Let  $X_0 = x$ , and for  $y \geq 0$  denote  $T_y \stackrel{\text{def}}{=} \min\{n \geq 0 : X_n = y\}$ .

(a) Show that the process  $(\varphi(X_n))_{n \geq 0}$  with  $\varphi(z) \stackrel{\text{def}}{=} \sum_{y=1}^z \prod_{x=1}^{y-1} \frac{q_x}{p_x}$  is a martingale.

(b) Show that for all  $0 \leq a < X_0 = x < b$  we have

$$\mathbb{P}(T_b < T_a) = (\varphi(x) - \varphi(a)) / (\varphi(b) - \varphi(a)).$$

Deduce that state 0 is recurrent iff  $\varphi(b) \rightarrow \infty$  as  $b \rightarrow \infty$ .

(c) Now suppose that  $p_m \equiv p$ ,  $q_m \equiv q = 1 - p$ , and  $r_m = 0$  for  $m > 0$ , whereas  $p_0 = p$  and  $r_0 = q$ . Show that in this case the result in part b) above becomes

$$\mathbb{P}(T_b < T_a) = ((q/p)^a - (q/p)^x) / ((q/p)^a - (q/p)^b).$$

(d) Find  $\mathbb{P}(T_b < T_a)$  if in the setup of part c) one has  $p = q = 1/2$ .

**Exercise 4.53.** Let  $(\xi_k)_{k \geq 1}$  be i.i.d. random variables with  $\mathbb{P}(\xi = 1) = p < \frac{1}{2}$ ,  $\mathbb{P}(\xi = -1) = q = 1 - p$ , and  $\mathbb{E}\xi > 0$ . Let  $(S_n)_{n \geq 0}$  be the generated random walk,  $S_n = x + \xi_1 + \dots + \xi_n$ , and let  $T_0 = \min\{n \geq 0 : S_n = 0\}$  be the hitting time of 0. Deduce that for all  $x > 0$ ,  $\mathbb{P}(T_0 < \infty) = (q/p)^x$ . Compare this to the result of Example 1.26.

**Exercise 4.54.** Let  $(Z_n)_{n \geq 0}$  be a homogeneous branching process with  $Z_0 = 1$ ,  $m = \mathbb{E}Z_1 > 0$  and finite variance  $\sigma^2 = \text{Var}(Z_1)$ . Show that  $M_n = Z_n/m^n$  is a martingale.

(a) Let  $m > 1$ . By using Exercise 2.3 or otherwise show that  $\mathbb{E}(M_n)$  is uniformly bounded. Deduce that  $M_n \rightarrow M_\infty$  almost surely. What can you say about  $\mathbb{E}M_\infty$ ?

(b) What happens if  $m \leq 1$ ? Compute  $\mathbb{E}(M_\infty)$ .

Hint: Recall Exercise 4.56.

**Exercise 4.55.** Let  $(X_n)_{n \geq 0}$  be a sequence of i.i.d. Bernoulli random variables with  $\mathbb{P}(X = 1) = p$  and  $\mathbb{P}(X = -1) = 1 - p = q$ . Let  $(S_n)_{n \geq 0}$  be the generated random walk with  $S_0 = x > 0$ , and let  $T = \min\{n \geq 0 : S_n = 0\}$  be the hitting time of the origin. Example 1.26 suggests that  $\mathbb{E}(T) = x/(q - p) < \infty$  for  $q > p$ .

1. Use (4.6) to deduce that  $(q - p)\mathbf{E}(n \wedge T) = \mathbf{E}(S_0 - S_{n \wedge T}) \leq \mathbf{E}(S_0) = x$ ; then take  $n \rightarrow \infty$  to show that  $\mathbf{E}(T) < \infty$ ;
2. Use the Wald equation to deduce that indeed  $\mathbf{E}(T) = \frac{x}{q-p}$ . Can you give a heuristic explanation of this result?
3. Argue, without using the Wald equation, that  $\mathbf{E}(T) = cx$  for some constant  $c$ .
4. Use the Wald equation and an argument by contradiction to show that if  $p \geq q$ , then  $\mathbf{E}(T) = \infty$  for all  $x > 0$ .

**Exercise 4.56.** Let a variable  $Y$  satisfy  $\mathbf{E}(Y^2) < \infty$ . Show that  $\mathbf{E}|Y| < \infty$ .  
Hint Notice that  $\text{Var}(|Y|) \geq 0$ .

**Exercise 4.57.** Let  $(X_n)_{n \geq 0}$  be an irreducible Markov chain in  $\mathcal{S} = \{0, 1, \dots\}$  with bounded jumps, and let a function  $\varphi : \mathcal{S} \rightarrow \mathbb{R}^+$  satisfy  $\varphi(x) \rightarrow \infty$  as  $x \rightarrow \infty$ . Let  $K \geq 0$  be such that

$$\mathbf{E}_x \varphi(X_1) \stackrel{\text{def}}{=} \mathbf{E}[\varphi(X_1) | X_0 = x] \leq \varphi(x)$$

for all  $x \geq K$ .

- (a) If the function  $\varphi(x)$  is monotone, show that the set of states  $\{0, 1, \dots, K\}$ , and thus the whole space  $\mathcal{S}$  is recurrent for  $(X_n)_{n \geq 0}$ .  
Hint If  $H_K = \min\{n \geq 0 : 0 \leq X_n \leq K\}$ , show that  $\varphi(X_{n \wedge H_K})$  is a supermartingale. Deduce that if  $T_M = \min\{n \geq 0 : X_n \geq M\}$ , then  $\varphi(x) \geq \varphi(M)\mathbf{P}(T_M < H_K)$ .
- (b) Argue that the result holds for  $\varphi(x) \geq 0$  not necessarily monotone, but only satisfying  $\varphi(x) \rightarrow \infty$  as  $x \rightarrow \infty$ .  
Hint With  $T_M$  as above, show that  $\varphi_M^* \stackrel{\text{def}}{=} \min\{\varphi(x) : x \geq M\} \rightarrow \infty$  as  $M \rightarrow \infty$ .

**Exercise 4.58.** Let  $(X_k)_{k \geq 1}$  be independent variables with  $\mathbf{P}(X = \pm 1) = \frac{1}{2}$ . Show that the process

$$M_n = \sum_{k=1}^n \frac{1}{k} X_k$$

is a martingale w.r.t.  $(\mathcal{F}_n^X)_{n \geq 0}$  and that  $\mathbf{E}[(M_n)^2] < K < \infty$  for some constant  $K$  and all  $n \geq 0$ . By using Exercise 4.56 or otherwise, deduce that with probability one,  $M_n \rightarrow M_\infty$  for some random variable  $M_\infty$ . In other words, the random sign harmonic series converges with probability one.

**Exercise 4.59.** [Wright-Fischer model] Thinking of a population of  $N$  haploid individuals who have one copy of each of their chromosomes, consider a fixed population of  $N$  genes that can be of two types  $A$  or  $a$ . In the simplest version of

this model the population at time  $n+1$  is obtained by sampling with replacement from the population at time  $n$ . If we let  $X_n$  to be the number of  $A$  alleles at time  $n$ , then  $X_n$  is a Markov chain with transition probability

$$p_{ij} = \binom{N}{j} \left(\frac{i}{N}\right)^j \left(1 - \frac{i}{N}\right)^{N-j}.$$

Starting from  $i$  of the  $A$  alleles and  $N - i$  of  $a$  alleles, what is the probability that the population fixates in the all  $A$  state? **Hint** You can use the heuristics of Example 4.17 but need to justify your computation!

**Exercise 4.60.** Let  $(X_n)_{n \geq 0}$  be a Markov chain with a (countable) state space  $\mathcal{S}$  and the transition matrix  $\mathbf{P}$ , and let  $h(x, n)$  be a function of the state  $x$  and time  $n$  such that<sup>33</sup>

$$h(x, n) = \sum_y p_{xy} h(y, n+1).$$

Show that  $(M_n)_{n \geq 0}$  with  $M_n = h(X_n, n)$  is a martingale w.r.t.  $(X_n)_{n \geq 0}$ .

**Exercise 4.61.** Let  $(X_n)_{n \geq 0}$  be a Markov chain with a (countable) state space  $\mathcal{S}$  and the transition matrix  $\mathbf{P}$ . If  $\psi$  is a right eigenvector of  $\mathbf{P}$  corresponding to the eigenvalue  $\lambda > 0$ , i.e.,  $\mathbf{P}\psi = \lambda\psi$ , show that the process  $M_n = \lambda^{-n}\psi(X_n)$  is a martingale w.r.t.  $(X_n)_{n \geq 0}$ .

---

<sup>33</sup>This result is useful, eg., if  $h(x, n) = x^2 - cn$  or  $h(x, n) = \exp\{x - cn\}$  for a suitable  $c$ .