3. DISCRETE RANDOM VARIABLES

3.1 Introduction

When an experiment is conducted there may be a number of quantities associated with the outcome $\omega \in \Omega$ that may be of interest. Suppose that the experiment is choosing a male student at random from the audience of the IA Probability lecture – there are many different measurements, or attributes, of the person chosen that may be of interest: his height, his weight, his IQ, the colour of his eyes, etc. Rather than think of each of these as the outcome of a separate experiment it is more useful to view them as functions of the outcome ω. This leads to the following definition which is a central notion of probability.

Definition A **random variable** X, taking values in a set S, is a function $X : \Omega \to S$.

Typically, S may be a subset of the real numbers, $\mathbb{R}$, as would be the case if the height of the student was of interest; or it could be a subset of $\mathbb{R}^k$, if more than one measurement is made on the subject as would be the case, with $k = 2$, if height and weight are measured; or, S could be some arbitrary set such as $S = \{\text{Blue}, \text{Green}, \text{Brown}\}$, say, if it is the colour of the subject's eyes that are to be recorded. The most frequent situation that we will encounter is the case when $S \subseteq \mathbb{R}$, and X is then said to be a real-valued random variable. Denote by Ω_X the range of X, so that $\Omega_X = \{X(\omega) : \omega \in \Omega\}$. In this chapter we will assume that the sample space Ω is either a finite or a countable set, so that Ω_X is finite or countable.

For $T \subseteq S$, we denote the event $\{\omega : X(\omega) \in T\}$ as $\{X \in T\}$, so that the dependence of X on ω is suppressed in the notation. Suppose that we enumerate the points in Ω_X (equivalently the values taken on by X), so that $\Omega_X = \{x_j : j \in J\}$, then we write the event $\{\omega : X(\omega) = x_j\} = \{X = x_j\}$. If we let $p_j = \mathbb{P}(X = x_j)$, $j \in J$, then $\{p_j : j \in J\}$ is a probability distribution on the space Ω_X, and is referred to as the **probability distribution of the random variable** X. Note that it is a probability distribution on the set Ω_X, not on the underlying sample space Ω.

Example 3.1 Suppose that two standard dice are rolled so that the sample space is

$\Omega = \{(i,j) : 1 \leqslant i,j \leqslant 6\}$, and we are interested in the sum of the numbers shown so that the random variable $X : \Omega \to \mathbb{R}$ is given by $X(i,j) = i + j$. The probability of each point in Ω is $\frac{1}{36}$ with the set of possible values taken on by X being $\Omega_X = \{2, 3, \ldots, 12\}$ and, for example,

$$\mathbb{P}(X = 6) = \mathbb{P}(\{(1,5),(2,4),(3,3),(4,2),(5,1)\}) = \frac{5}{36}.$$

If we set $p_j = \mathbb{P}(X = j)$, for $j = 2, \ldots, 12$, then the table

j	2	3	4	5	6	7	8	9	10	11	12
p_j	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$

gives the full probability distribution of the random variable X. □

Terminology If the probability distribution of X is a standard distribution such as the binomial distribution (or Poisson, or geometric), we say that X is a binomial (respectively, Poisson, or geometric) random variable. We often write $X \sim \text{Bin}\,(n,p)$, for example, for the statement that X is binomial distribution where the parameters are n and p, or $X \sim \text{Poiss}(\lambda)$ for a Poisson random variable with parameter λ..

Example 3.2 Suppose that a coin is tossed n times and a 1 is recorded whenever a head occurs and a 0 is recorded for each tail. Then $\Omega = \{(i_1, i_2, \ldots, i_n) : i_j = 1 \text{ or } 0\}$. If p is the probability of a head and tosses are independent then the probability on Ω is specified by

$$\mathbb{P}(i_1, i_2, \ldots, i_n) = p^{i_1 + \cdots + i_n}(1-p)^{n - i_1 - \cdots - i_n}.$$

Let X denote the number of heads obtained, so that $X(i_1, \ldots, i_n) = i_1 + \cdots + i_n$, then X is a binomial random variable since the distribution of X is given by

$$\mathbb{P}(X = k) = \binom{n}{k} p^k (1-p)^{n-k}, \quad \text{for} \quad 0 \leqslant k \leqslant n.$$ □

For a function $g : S \to T$, mapping from the set S to the set T, then if X is a random variable taking values in S, $g(X)$ is the random variable taking values in T, with $g(X) : \Omega \to T$ specified by $g(X)(\omega) = g(X(\omega))$. For subsets $C \subseteq T$ we have

$\mathbb{P}(g(X) \in C) = \mathbb{P}\left(X \in g^{-1}(C)\right)$ and the distribution of $g(X)$ may be obtained from that of X by observing that

$$\mathbb{P}(g(X) = y) = \mathbb{P}\left(X \in g^{-1}(y)\right) = \sum_{x \in g^{-1}(y)} \mathbb{P}(X = x).$$

A real-valued random variable which takes on just the two values 0 and 1 is known as an **indicator random variable**; suppose that the event on which it takes the value 1 is $A \subseteq \Omega$ then the random variable is denoted by I_A, so that

$$I_A(\omega) = \begin{cases} 1 & \text{for } \omega \in A, \\ 0 & \text{for } \omega \notin A, \end{cases}$$

and I_A is 1 or 0 according as the event A occurs or does not occur. The following properties of indicator random variables should be noted for events A and B:

1. $I_{A^c} = 1 - I_A$.
2. $I_{A \cap B} = I_A I_B$.
3. $I_{A \cup B} = 1 - (1 - I_A)(1 - I_B)$.

and, for events $A_1, A_2, \ldots, A_n$, Properties 2 and 3 generalize to $I_{A_1 \cap A_2 \cap \cdots \cap A_n} = \prod_1^n I_{A_i}$, and

$$\begin{aligned} I_{A_1 \cup A_2 \cup \cdots \cup A_n} &= 1 - \prod_1^n (1 - I_{A_i}) \\ &= \sum_i I_{A_i} - \sum_{i_1 < i_2} I_{A_{i_1}} I_{A_{i_2}} + \sum_{i_1 < i_2 < i_3} I_{A_{i_1}} I_{A_{i_2}} I_{A_{i_3}} - \cdots + (-1)^{n-1} I_{A_1} \cdots I_{A_n} \\ &= \sum_i I_{A_i} - \sum_{i_1 < i_2} I_{A_{i_1} \cap A_{i_2}} + \sum_{i_1 < i_2 < i_3} I_{A_{i_1} \cap A_{i_2} \cap A_{i_3}} - \cdots + (-1)^{n-1} I_{A_1 \cap \cdots \cap A_n}. \end{aligned}$$

In the next section we see how this last relation provides an alternate proof of the inclusion-exclusion formula.

3.2 Expectation, variance and covariance

From now on, unless we indicate to the contrary, the random variables we will consider will take real values. For a non-negative random variable X, that is one for which $X(\omega) \geqslant 0$

for all $\omega \in \Omega$, (usually just written as $X \geqslant 0$),we define the **expectation** (or **expected value** or **mean value**) of X to be

$$\mathbb{E}\,X = \sum_{\omega\in\Omega} X(\omega)\mathbb{P}(\{\omega\});$$

since all the terms in the sum are non-negative the sum is well defined (although it may take the value $+\infty$). Note that, since $\Omega = \bigcup_{x\in\Omega_X} \{X = x\}$, we have a more useful form for the expectation given by

$$\mathbb{E}\,X = \sum_{\omega\in\Omega} X(\omega)\mathbb{P}(\{\omega\}) = \sum_{x\in\Omega_X}\sum_{\omega\in\{X=x\}} X(\omega)\mathbb{P}(\{\omega\}) = \sum_{x\in\Omega_X} x\mathbb{P}\left(X = x\right).$$

Thus the expectation is the average of the values taken on by the random variable, averaged with weights corresponding to the probabilities of the values.

Example 3.3 Suppose that $X \sim \mathrm{Bin}(n,p)$, so that $\mathbb{P}\left(X = k\right) = \binom{n}{k} p^k(1-p)^{n-k}$, for $0 \leqslant k \leqslant n$, then

$$\begin{aligned}\mathbb{E}\,X &= \sum_{k=0}^{n} k\binom{n}{k}p^k(1-p)^{n-k} = p\sum_{k=1}^{n}\frac{n!}{(n-k)!(k-1)!}p^{k-1}(1-p)^{n-k}\\ &= np\sum_{k=1}^{n}\binom{n-1}{k-1}p^{k-1}(1-p)^{n-k} = np\left[p+(1-p)\right]^{n-1} = np.\end{aligned}$$ □

Example 3.4 Suppose that $X \sim \mathrm{Poiss}(\lambda)$, so that $\mathbb{P}(X = k) = e^{-\lambda}\frac{\lambda^k}{k!}$, for $k = 0, 1, 2, \ldots$, then

$$\mathbb{E}\,X = \sum_{k=0}^{\infty} ke^{-\lambda}\frac{\lambda^k}{k!} = \lambda e^{-\lambda}\sum_{k=1}^{\infty}\frac{\lambda^{k-1}}{(k-1)!} = \lambda e^{-\lambda}e^{\lambda} = \lambda.$$ □

For any random variable X denote by $X_+ = \max(X, 0)$, the positive part of X, and $X_- = \max(-X, 0)$, the negative part of X are non-negative random variables, for which so that $X = X_+ - X_-$ and $|X| = X_+ + X_-$. Provided not both $\mathbb{E}\,X_+ = \infty$ and $\mathbb{E}\,X_- = \infty$, we define the expectation of X to be

$$\mathbb{E}\,X = \mathbb{E}\,X_+ - \mathbb{E}\,X_- = \sum_{x\in\Omega_X} x\mathbb{P}\left(X = x\right);$$

if both $\mathbb{E}\,X_+$ and $\mathbb{E}\,X_-$ are infinite then the expectation of X is not defined. In the following, when we write $\mathbb{E}\,X$ for a random variable X, it may be assumed that the expectation of X is well defined.

Properties of $\mathbb{E}\,X$

1. If $X \geqslant 0$, then $\mathbb{E}\,X \geqslant 0$, and $\mathbb{E}\,X = 0$ implies that $\mathbb{P}(X = 0) = 1$.
2. If c is a constant then $\mathbb{E}\,(cX) = c\mathbb{E}\,X$, and $\mathbb{E}\,c = c$.
3. For random variables X and Y, $\mathbb{E}\,(X + Y) = \mathbb{E}\,X + \mathbb{E}\,Y$.

Properties 2 and 3 show the important property that the operator $\mathbb{E}\,(\cdot)$ is a linear operator and they generalize, by induction, to the case of random variables $X_1, \ldots, X_n$ and constants $c_1, \ldots, c_n$ so that $\mathbb{E}\left(\sum_{i=1}^{n} c_i X_i\right) = \sum_{i=1}^{n} c_i \mathbb{E}\,X_i$.

4. $\mathbb{E}\,g(X) = \sum_{x \in \Omega_X} g(x)\mathbb{P}(X = x)$.

To see this, let $Y = g(X)$, then

$$\begin{aligned}
\mathbb{E}\,g(X) = \mathbb{E}\,Y &= \sum_{y \in \Omega_Y} y\mathbb{P}\,(Y = y) = \sum_{y \in \Omega_Y} y \left(\sum_{x \in g^{-1}(y)} \mathbb{P}\,(X = x) \right) \\
&= \sum_{y \in \Omega_Y} \sum_{x \in g^{-1}(y)} y\mathbb{P}\,(X = x) = \sum_{y \in \Omega_Y} \sum_{x \in g^{-1}(y)} g(x)\mathbb{P}\,(X = x) \\
&= \sum_{x \in \Omega_X} g(x)\mathbb{P}\,(X = x)\,.
\end{aligned}$$

5. For the indicator of any event $A \subseteq \Omega$ we have $\mathbb{E}\,I_A = \mathbb{P}(A)$.
6. If $X \geqslant 0$ and X takes integer values, then $\mathbb{E}\,X = \sum_{n=1}^{\infty} \mathbb{P}\,(X \geqslant n)$.

Proof. We have

$$\mathbb{E}\,X = \sum_{k=1}^{\infty} k\mathbb{P}\,(X = k) = \sum_{k=1}^{\infty} \sum_{n=1}^{k} \mathbb{P}\,(X = k) = \sum_{n=1}^{\infty} \sum_{k=n}^{\infty} \mathbb{P}\,(X = k) = \sum_{n=1}^{\infty} \mathbb{P}\,(X \geqslant n)\,,$$

after interchanging the order of the summations. □

Terminology For a random variable X, the expected values of powers of X are known as **moments** of X; thus $\mathbb{E}\,(X^r)$ (assuming it is well defined) is the rth moment of X and $\mathbb{E}\,(|X|^r)$ is the rth absolute moment of X.

Example 3.5 *Another proof of inclusion-exclusion.* For events $A_1, \ldots, A_n$, use the previous expression for the product of indicators to calculate

$$\mathbb{P}(A_1 \cup \cdots \cup A_n) = \mathbb{E}\,(I_{A_1 \cup A_2 \cup \cdots \cup A_n}) = \mathbb{E}\left(1 - \prod_1^n (1 - I_{A_i})\right)$$

$$= \mathbb{E}\left(\sum_i I_{A_i} - \sum_{i_1 < i_2} I_{A_{i_1} \cap A_{i_2}} + \sum_{i_1 < i_2 < i_3} I_{A_{i_1} \cap A_{i_2} \cap A_{i_3}} - \cdots + (-1)^{n-1} I_{A_1 \cap \cdots \cap A_n}\right)$$

then using the linearity of the expectation, this

$$= \sum_i \mathbb{E}\,(I_{A_i}) - \sum_{i_1 < i_2} \mathbb{E}\,\left(I_{A_{i_1} \cap A_{i_2}}\right) + \cdots + (-1)^{n-1} \mathbb{E}\,(I_{A_1 \cap \cdots \cap A_n})$$

$$= \sum_i \mathbb{P}(A_i) - \sum_{i_1 < i_2} \mathbb{P}\left(A_{i_1} \cap A_{i_j}\right) + \cdots + (-1)^{n-1} \mathbb{P}\,(A_1 \cap \cdots \cap A_n)\,,$$

which is the required expression for the inclusion-exclusion formula. □

For any random variable, X with finite mean, the **variance** is defined to be

$$\mathbb{V}\mathrm{ar}\,(X) = \mathbb{E}\,(X - \mathbb{E}\,X)^2\,,$$

and it is a measure of how much the distribution of X is spread out around the mean; the smaller the distribution the more the distribution of X is concentrated close to $\mathbb{E}\,X$. The quantity $\sqrt{\mathbb{V}\mathrm{ar}\,(X)}$ is known as the **standard deviation** of X. When we use the notation $\mathbb{V}\mathrm{ar}\,(X)$ we will assume implicitly that it is a finite quantity.

Properties of $\mathbb{V}\mathrm{ar}\,(X)$

1. $\mathbb{V}\mathrm{ar}\,(X) = \mathbb{E}\,X^2 - (\mathbb{E}\,X)^2$.

Proof. We have, using Properties 2 and 3 of the expectation,

$$\mathbb{E}\,(X - \mathbb{E}\,X)^2 = \mathbb{E}\,\left(X^2 - 2X\mathbb{E}\,X + (\mathbb{E}\,X)^2\right)$$
$$= \mathbb{E}\,X^2 - 2\mathbb{E}\,X\mathbb{E}\,X + (\mathbb{E}\,X)^2 = \mathbb{E}\,X^2 - (\mathbb{E}\,X)^2.$$
□

2. If c is a constant, $\mathbb{V}\mathrm{ar}\,(cX) = c^2\mathbb{V}\mathrm{ar}\,(X)$.
3. If c is a constant, $\mathbb{V}\mathrm{ar}\,(X + c) = \mathbb{V}\mathrm{ar}\,(X)$.
4. $\mathbb{V}\mathrm{ar}\,(X) \geqslant 0$, and $\mathbb{V}\mathrm{ar}\,(X) = 0$ if and only if $\mathbb{P}(X = c) = 1$, for some constant c.
5. The expression $\mathbb{E}\,(X - c)^2$ is minimized over constants c when $c = \mathbb{E}\,X$, so that $\mathbb{E}\,(X - c)^2 \geqslant \mathbb{V}\mathrm{ar}\,(X)$, for all c, with equality when $c = \mathbb{E}\,X$.

Proof. Expand out the expression

$$\mathbb{E}\,(X-c)^2 = \mathbb{E}\,\left(X^2 - 2cX + c^2\right) = \mathbb{E}\,X^2 - 2c\mathbb{E}\,X + c^2,$$

and minimize the right-hand side in c to see that the minimum occurs at $c = \mathbb{E}\,X$. □

Example 3.6 For $X \sim \text{Bin}(n,p)$, we have

$$\begin{aligned}\mathbb{E}\,(X(X-1)) &= \sum_{k=0}^{n} k(k-1)\binom{n}{k}p^k(1-p)^{n-k} = \sum_{k=2}^{n} \frac{n!}{(k-2)!(n-k)!}p^k(1-p)^{n-k} \\ &= n(n-1)p^2 \sum_{r=0}^{n-2}\binom{n-2}{r}p^r(1-p)^{n-2-r} = n(n-1)p^2;\end{aligned}$$

then it follows that

$$\mathbb{E}\,X^2 = \mathbb{E}\,(X(X-1)) + \mathbb{E}\,X = n(n-1)p^2 + np,$$

since we had seen that $\mathbb{E}\,X = np$; hence $\mathbb{V}\text{ar}\,(X) = \mathbb{E}\,X^2 - (\mathbb{E}\,X)^2 = np(1-p)$. □

Example 3.7 Suppose that $X \sim \text{Poiss}(\lambda)$, then a similar calculation to that in the previous Example gives

$$\mathbb{E}\,(X(X-1)) = \sum_{k=0}^{\infty} k(k-1)e^{-\lambda}\frac{\lambda^k}{k!} = \sum_{k=2}^{\infty} e^{-\lambda}\frac{\lambda^k}{(k-2)!} = \lambda^2 e^{-\lambda}\sum_{r=0}^{\infty}\frac{\lambda^r}{r!} = \lambda^2;$$

recalling that in this case $\mathbb{E}\,X = \lambda$, we have $\mathbb{E}\,X^2 = \lambda^2 + \lambda$, so that $\mathbb{V}\text{ar}\,(X) = \lambda$, showing that for a Poisson random variable the mean is the same as the variance. □

Example 3.8 *Use of indicators* Return to the situation, considered in Chapter 2, where n students leave their n coats outside the lecture room and when they leave they pick up their coats at random. Let N be the number of students who get their own coat, then $N = \sum_{i=1}^{n} I_{A_i}$, where A_i is the event that student i obtains his own coat. It follows that

$$\begin{aligned}\mathbb{E}\,N &= \mathbb{E}\left(\sum_{i=1}^{n} I_{A_i}\right) = \sum_{i=1}^{n}\mathbb{E}\,(I_{A_i}) = \sum_{i=1}^{n}\mathbb{P}\,(A_i) = \sum_{i=1}^{n}\frac{1}{n} = 1, \quad \text{and} \\ \mathbb{E}\,N^2 &= \mathbb{E}\left(\sum_{i=1}^{n} I_{A_i}\right)^2 = \mathbb{E}\left(\sum_{i=1}^{n}(I_{A_i})^2 + \sum_{i}\sum_{j\neq i} I_{A_i}I_{A_j}\right);\end{aligned}$$

since $I_{A_i}I_{A_j} = I_{A_i\cap A_j}$ we have $(I_{A_i})^2 = I_{A_i}$, and we see that

$$\mathbb{E}\, N^2 = \mathbb{E}\left(\sum_{i=1}^{n} I_{A_i} + \sum_i \sum_{j\neq i} I_{A_i \cap A_j}\right) = \sum_{i=1}^{n} \mathbb{P}(A_i) + \sum_i \sum_{j\neq i} \mathbb{P}(A_i \cap A_j)$$
$$= \sum_{i=1} \frac{1}{n} + \sum_i \sum_{j\neq i} \frac{1}{n(n-1)} = n \times \frac{1}{n} + n(n-1) \times \frac{1}{n(n-1)} = 1 + 1 = 2.$$

That gives $\mathbb{V}\mathrm{ar}\,(N) = \mathbb{E}\, N^2 - (\mathbb{E}\, N)^2 = 1$. The fact that the mean and the variance are both the same might suggest that the distribution of the random variable N is close to being Poisson (with mean $\lambda = 1$) as is indeed the case when n is large. If we let $p_n = \mathbb{P}(N = 0)$, the probability that when there are n students none of them gets his own coat, then we have seen previously (using inclusion-exclusion) that

$$p_n = 1 - \frac{1}{1!} + \frac{1}{2!} - \frac{1}{3!} + \cdots + (-1)^n \frac{1}{n!} \to e^{-1}, \quad \text{as} \quad n \to \infty;$$

take $p_0 = 1$. The probability that exactly k students get their own coats is

$$\mathbb{P}(N = k) = \binom{n}{k} \frac{1}{n!}((n-k)!)p_{n-k} = \frac{1}{k!} p_{n-k} \to \frac{1}{k!} e^{-1}, \quad \text{as} \quad n \to \infty,$$

showing that the distribution of N is approximately Poisson. $\square$

Theorem 3.9 (Cauchy–Schwarz inequality) *For any random variables X and Y,*

$$(\mathbb{E}\,(XY))^2 \leqslant \mathbb{E}\left(X^2\right) \mathbb{E}\left(Y^2\right);$$

if $\mathbb{E}\left(Y^2\right) > 0$, equality occurs if and only if $X = aY$ for some constant $a \in \mathbb{R}$.

Proof. For any $a \in \mathbb{R}$, observe that $\mathbb{E}\,(X - aY)^2 \geqslant 0$, so that

$$0 \leqslant \mathbb{E}\left(X^2 - 2aXY + a^2Y^2\right) = \mathbb{E}\left(X^2\right) - 2a\mathbb{E}\,(XY) + a^2\mathbb{E}\left(Y^2\right),$$

showing that the quadratic in a on the right-hand side has at most one real root, whence the discriminant $4\left((\mathbb{E}\,(XY))^2 - \mathbb{E}\left(X^2\right)\mathbb{E}\left(Y^2\right)\right) \leqslant 0$, giving the inequality. There is clearly equality if $X = aY$ for some $a \in \mathbb{R}$, whereas if $\mathbb{E}\left(Y^2\right) > 0$ and the discriminant is 0 then the quadratic is 0 for $a = \mathbb{E}\,(XY)/\mathbb{E}\left(Y^2\right)$, and for that value of a, $\mathbb{E}\,(X - aY)^2 = 0$ and so $X = aY$. Of course, if $\mathbb{E}\left(Y^2\right) = 0$ then $Y = 0$ and equality occurs. $\square$

For two random variable X and Y, we define the **covariance** between X and Y as

$$\mathbb{C}\text{ov}\,(X,Y) = \mathbb{E}\,((X - \mathbb{E}\,X)\,(Y - \mathbb{E}\,Y))\,.$$

We shall see that this is a measure of the dependence between the random variables X and Y.

Properties of $\mathbb{C}\text{ov}\,(X,Y)$

1. $\mathbb{C}\text{ov}\,(X,Y) = \mathbb{C}\text{ov}\,(Y,X)$.
2. $\mathbb{C}\text{ov}\,(X,Y) = \mathbb{E}\,(XY) - (\mathbb{E}\,X)\,(\mathbb{E}\,Y)$.

Proof. We have

$$\begin{aligned}\mathbb{C}\text{ov}\,(X,Y) &= \mathbb{E}\,(XY - X(\mathbb{E}\,Y) - Y(\mathbb{E}\,X) + (\mathbb{E}\,X)(\mathbb{E}\,Y)) \\ &= \mathbb{E}\,(XY) - (\mathbb{E}\,X)(\mathbb{E}\,Y) - (\mathbb{E}\,X)(\mathbb{E}\,Y) + (\mathbb{E}\,X)(\mathbb{E}\,Y) \\ &= \mathbb{E}\,(XY) - (\mathbb{E}\,X)\,(\mathbb{E}\,Y)\,.\end{aligned}$$ □

3. $\mathbb{C}\text{ov}\,(X,X) = \mathbb{V}\text{ar}\,(X)$.
4. $\mathbb{V}\text{ar}\,(X+Y) = \mathbb{V}\text{ar}\,(X) + \mathbb{V}\text{ar}\,(Y) + 2\mathbb{C}\text{ov}\,(X,Y)$.

Proof. We have

$$\begin{aligned}\mathbb{V}\text{ar}\,(X+Y) &= \mathbb{E}\,(X + Y - \mathbb{E}\,X - \mathbb{E}\,Y)^2 = \mathbb{E}\,((X - \mathbb{E}\,X) + (Y - \mathbb{E}\,Y))^2 \\ &= \mathbb{E}\,\Big((X - \mathbb{E}\,X)^2 + (Y - \mathbb{E}\,Y)^2 + 2\,(X - \mathbb{E}\,X)\,(Y - \mathbb{E}\,Y)\Big) \\ &= \mathbb{E}\,(X - \mathbb{E}\,X)^2 + \mathbb{E}\,(Y - \mathbb{E}\,Y)^2 + 2\mathbb{E}\,(X - \mathbb{E}\,X)\,(Y - \mathbb{E}\,Y)\,.\end{aligned}$$ □

5. If c is a constant, $\mathbb{C}\text{ov}\,(X,c) = 0$.
6. If c is a constant, $\mathbb{C}\text{ov}\,(X+c,Y) = \mathbb{C}\text{ov}\,(X,Y)$.
7. If c is a constant, $\mathbb{C}\text{ov}\,(cX,Y) = c\,\mathbb{C}\text{ov}\,(X,Y)$.
8. $\mathbb{C}\text{ov}\,(X+Z,Y) = \mathbb{C}\text{ov}\,(X,Y) + \mathbb{C}\text{ov}\,(Z,Y)$.

These last two generalize to the case of random variables $X_1,\ldots,X_n$ and $Y_1,\ldots,Y_n$ and constants $c_1,\ldots,c_n$ and $d_1,\ldots d_n$ to give, by induction,

$$\mathbb{C}\text{ov}\,\left(\sum_{i=1}^{n} c_i X_i, \sum_{j=1}^{n} d_j Y_j\right) = \sum_{i=1}^{n}\sum_{j=1}^{n} c_i d_j \mathbb{C}\text{ov}\,(X_i, Y_j)\ . \tag{3.10}$$

Using the fact that $\mathbb{V}\text{ar}\,(X) = \mathbb{C}\text{ov}\,(X, X)$, we see that a special case of this is

$$\mathbb{V}\text{ar}\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} \mathbb{V}\text{ar}\;(X_i) + \sum_{i=1}^{n} \sum_{j \neq i} \mathbb{C}\text{ov}\;(X_i, X_j)\,, \tag{3.11}$$

for any random variables $X_1, \ldots, X_n$.

The **correlation coefficient** (or just the **correlation**) between random variables X and Y with $\mathbb{V}\text{ar}\,(X) > 0$ and $\mathbb{V}\text{ar}\,(Y) > 0$ is

$$\mathbb{C}\text{orr}\,(X, Y) = \frac{\mathbb{C}\text{ov}\,(X, Y)}{\sqrt{\mathbb{V}\text{ar}\,(X)\mathbb{V}\text{ar}\,(Y)}}.$$

Notice that by the Cauchy-Schwarz inequality

$$|\mathbb{C}\text{orr}\,(X, Y)| \leqslant 1, \quad \text{for all } X \text{ and } Y;$$

this follows by applying the inequality to the random variables $\overline{X} = X - \mathbb{E}\,X$ and $\overline{Y} = Y - \mathbb{E}\,Y$. It may be further seen that $|\mathbb{C}\text{orr}\,(X, Y)| = 1$ if and only if $X = aY + b$ for some constants a and b. One property of correlation that we should note is that for constants a, b, c and d with $ac \neq 0$, we have

$$\mathbb{C}\text{orr}\;(aX + b, cY + d) = \begin{cases} \mathbb{C}\text{orr}\,(X, Y) & \text{when } \; ac > 0, \\ -\mathbb{C}\text{orr}\,(X, Y) & \text{when } \; ac < 0. \end{cases}$$

This follows easily from the definition of correlation and the properties of the covariance and variance; notice that when $ac = 0$, $\mathbb{C}\text{ov}\;(aX + b, cY + d) = 0$, and the correlation is not defined because at least one of $\mathbb{V}\text{ar}\;(aX + b) = 0$ or $\mathbb{V}\text{ar}\;(cY + d) = 0$.

Notice that one consequence of this fact is that the correlation between two random variables is scale invariant – if we multiply the observation of X and Y by positive constants we do not alter the correlation.

3.3 Independence

Discrete random variables $X_1, X_2, \ldots, X_n$ are **independent** if, for all choices of $x_i \in \Omega_{X_i}$, $1 \leqslant i \leqslant n$, we have

$$\mathbb{P}\,(X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n) = \prod_{i=1}^{n} \mathbb{P}\,(X_i = x_i)\,. \tag{3.12}$$

Notice that $X_1, X_2, \ldots, X_n$ are independent if and only if, for all choices of subsets $S_i \subseteq \Omega_{X_i}$, $1 \leqslant i \leqslant n$, we have

$$\mathbb{P}\left(X_1 \in S_1, X_2 \in S_2, \ldots, X_n \in S_n\right) = \prod_{i=1}^{n} \mathbb{P}\left(X_i \in S_i\right). \tag{3.13}$$

To see this, if (3.13) holds, take $S_i = \{x_i\}$ for each i and we see that (3.12) is true; conversely, the left hand side of (3.13) is

$$\sum_{x_1 \in S_1} \sum_{x_2 \in S_2} \cdots \sum_{x_n \in S_n} \mathbb{P}\left(X_1 = x_1, X_2 = x_x, \ldots, X_n = x_n\right)$$

and we see that, if (3.12) holds, then this is expression is

$$\sum_{x_1 \in S_1} \sum_{x_2 \in S_2} \cdots \sum_{x_n \in S_n} \prod_{i=1}^{n} \mathbb{P}\left(X_i = x_i\right) = \prod_{i=1}^{n} \left(\sum_{x_i \in S_i} \mathbb{P}\left(X_i = x_i\right) \right) = \prod_{i=1}^{n} \mathbb{P}\left(X_i \in S_i\right),$$

which gives (3.13).

Notice that events $A_1, \ldots, A_n$ are independent, as defined in the previous chapter, if and only if their indicator random variables $I_{A_1}, \ldots, I_{A_n}$ are independent random variables. Observe also that if random variables are independent then they are independent in pairs (this follows by taking $S_i = \Omega_{X_i}$ for all but two of the subsets S_i in (3.13)) – they are said to be pairwise independent; a similar argument shows that if any collection of random variables is independent then any sub-collection of them is independent. By considering indicators, the example from the last chapter shows that pairwise independence of random variables does not imply independence in general.

Properties of independent random variables

1. If $X_1, \ldots, X_n$ are independent random variables and $g_i : \mathbb{R} \to \mathbb{R}$, $1 \leqslant i \leqslant n$, are functions then $g_1(X_1), \ldots, g_n(X_n)$ are independent random variables.

Proof. For $y_i \in \Omega_{g_i(X_i)}$, $1 \leqslant i \leqslant n$, we have

$$\begin{aligned}\mathbb{P}\left(g_1(X_1) = y_1, \ldots, g_n(X_n) = y_n\right) &= \mathbb{P}\left(X_1 \in g_1^{-1}(y_1), \ldots, X_n \in g_n^{-1}(y_n)\right) \\ &= \prod_{i=1}^{n} \mathbb{P}\left(X_i \in g_i^{-1}(y_i)\right) = \prod_{i=1}^{n} \mathbb{P}\left(g_i(X_i) = y_i\right)\end{aligned}$$

after using (3.13), showing that the random variables $g_1(X_1), \dots, g_n(X_n)$ are independent. $\square$

2. If $X_1, \dots, X_n$ are independent random variables, then

$$\mathbb{E}\left(\prod_{i=1}^{n} X_i\right) = \prod_{i=1}^{n} \mathbb{E}\,(X_i)\,;$$

that is, the expectation of the product of independent random variables is the product of their expectations.

Proof. In a similar way to the previous proof, we may represent the event

$$\left(\prod_{i=1}^{n} X_i = y\right) = \bigcup (X_1 = x_1, X_2 = x_2, \dots, X_n = x_n)$$

as a disjoint union of events over values of $x_1, \dots, x_n$ with $\prod_i x_i = y$. Then

$$\begin{aligned}
\mathbb{E}\left(\prod_{i=1}^{n} X_i\right) &= \sum_y y\mathbb{P}\left(\prod_{i=1}^{n} X_i = y\right) = \sum_y y \sum_{x_i:\prod_i x_i = y} \mathbb{P}\,(X_1 = x_1, \dots, X_n = x_n) \\
&= \sum_y \sum_{x_i:\prod_i x_i = y} y \prod_{i=1}^{n} \mathbb{P}\,(X_i = x_i)\,, \quad \text{by independence} \\
&= \sum_{x_1,\dots,x_n} \prod_{i=1}^{n} (x_i \mathbb{P}\,(X_i = x_i)) = \prod_{i=1}^{n}\left(\sum_{x_i} x_i \mathbb{P}\,(X_i = x_i)\right) = \prod_{i=1}^{n} \mathbb{E}\,(X_i)\,,
\end{aligned}$$

as required. $\square$

3. If X and Y are independent random variables then $\mathbb{C}\text{ov}\,(X, Y) = 0$ (and hence $\mathbb{C}\text{orr}\,(X, Y) = 0$). The converse is not true in general (see Example 3.14 below): that is, $\mathbb{C}\text{ov}\,(X, Y) = 0$ does not imply that X and Y are independent.

Proof. Property 1 shows that $X - \mathbb{E}\,X$ and $Y - \mathbb{E}\,Y$ are independent random variables and then by Property 2,

$$\mathbb{C}\text{ov}\,(X, Y) = \mathbb{E}\,((X - \mathbb{E}\,X)\,(Y - \mathbb{E}\,Y)) = \mathbb{E}\,(X - \mathbb{E}\,X)\,\mathbb{E}\,(Y - \mathbb{E}\,Y) = 0$$

since $\mathbb{E}\,(X - \mathbb{E}\,X) = \mathbb{E}\,(X) - \mathbb{E}\,(X) = 0$ (and similarly $\mathbb{E}\,(Y - \mathbb{E}\,Y) = 0$). $\square$

4. If $X_1, \dots, X_n$ are independent random variables then

$$\mathbb{V}\text{ar}\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} \mathbb{V}\text{ar}\,(X_i)\,;$$

that is, the variance of the sum of independent random variables is the sum of their variances.

Proof. Use Property 3 to see that for $j \neq i$, $\mathbb{C}\text{ov}\,(X_i, X_j) = 0$ and the result follows from the relation (3.11). □

5. If $X_1, \dots, X_n$ are independent random variables then the conditional probability

$$\mathbb{P}\left(X_1, = x_1, \dots, X_{n-1} = x_{n-1} \mid X_n = x_n\right) = \mathbb{P}\left(X_1, = x_1, \dots, X_{n-1} = x_{n-1}\right),$$

for all choices of $x_i \in \Omega_{X_i}$, $1 \leqslant i \leqslant n$.

Proof. We have the conditional probability on the left-hand side is

$$\frac{\mathbb{P}\left(X_1, = x_1, \dots, X_n = x_n\right)}{\mathbb{P}\left(X_n = x_n\right)} = \frac{\prod_{i=1}^{n} \mathbb{P}\left(X_i = x_i\right)}{\mathbb{P}\left(X_n = x_n\right)} = \prod_{i=1}^{n-1} \mathbb{P}\left(X_i = x_i\right),$$

which equals the right-hand side, again by independence. □

Terminology Random variables with the same distribution are usually said to be **identically distributed**, and if they are also independent they are **i.i.d.** (independent and identically distributed). If $X_1, \dots, X_n$ are i.i.d. then, from Property 4,

$$\mathbb{V}\text{ar}\left(\frac{X_1 + \cdots + X_n}{n}\right) = \frac{\mathbb{V}\text{ar}\left(X_1\right)}{n}.$$

Example 3.14 *Covariance equal to* 0 *does not imply independence* Suppose that X is a random variable with distribution determined by

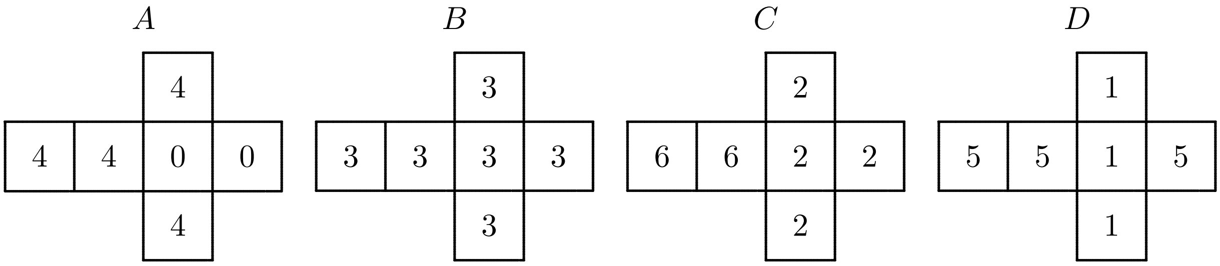
x	2	1	-1	-2
$\mathbb{P}\left(X = x\right)$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$

and let $Y = X^2$. Then $\mathbb{E}\,X = 0$ and $\mathbb{E}\left(X^3\right) = 0$ so that $\mathbb{C}\text{ov}\left(X, Y\right) = \mathbb{E}\left(X^3\right) = 0$, but

$$\mathbb{P}\left(X = 2, Y = 4\right) = \tfrac{1}{4} \neq \mathbb{P}\left(X = 2\right)\mathbb{P}\left(Y = 4\right) = \tfrac{1}{4} \times \tfrac{1}{2},$$

so that X and Y are not independent. $\square$

Example 3.15 *Efron's dice* An interesting example showing that odds are not transitive is given by a set of 4 dice with the following faces:



If each of the dice is rolled with respective outcomes A, B, C and D then

$$\mathbb{P}(A > B) = \mathbb{P}(B > C) = \mathbb{P}(C > D) = \mathbb{P}(D > A) = \frac{2}{3}.$$ $\square$

3.4 Probability generating functions

Consider a random variable, X, taking values in the non-negative integers $0, 1, 2, \ldots$ with distribution determined by $p_r = \mathbb{P}(X = r)$, $r = 0, 1, 2, \ldots$. The **probability generating function** (p.g.f.) of X is defined to be

$$p(z) = \mathbb{E}\left(z^X\right) = \sum_{r=0}^{\infty} p_r z^r, \quad \text{for} \quad 0 \leqslant z \leqslant 1.$$

Since the terms in the sum are all non-negative and $0 \leqslant \sum_r p_r z^r \leqslant \sum_r p_r = 1$, the probability generating function is well defined and takes values in $[0, 1]$. Its importance stems from the following result.

Theorem 3.16 *The probability generating function of X, $p(z)$, $0 \leqslant z \leqslant 1$, determines the probability distribution of X uniquely.*

Proof. Suppose that $p(z) = \sum_{r=0}^{\infty} p_r z^r = \sum_{r=0}^{\infty} q_r z^r$, for all $0 \leqslant z \leqslant 1$, where $p_r \geqslant 0$, and $q_r \geqslant 0$ for each r, and $\sum_{r=0}^{\infty} p_r = 1 = \sum_{r=0}^{\infty} q_r$. We will show by induction on n that $p_n = q_n$ for all n. First see, by setting $z = 0$, that $p_0 = q_0$. Now assume that $p_i = q_i$ for $0 \leqslant i \leqslant n$, then for $0 < z \leqslant 1$

$$\sum_{r=n+1}^{\infty} p_r z^r = \sum_{r=n+1}^{\infty} q_r z^r.$$

Divide through both sides by z^{n+1} and let $z \downarrow 0$ to see that $p_{n+1} = q_{n+1}$ to complete the induction. □

In addition to determining the distribution uniquely, the probability generating function may be used to compute moments of the random variable by evaluating derivatives of the function.

Theorem 3.17 *Let X be a random variable with probability generating function $p(z)$, then the mean of X is*

$$\mathbb{E}\, X = \lim_{z \uparrow 1} p'(z) = p'(1-).$$

Proof. First assume that $\mathbb{E}\, X < \infty$. For $0 < z < 1$,

$$p'(z) = \sum_{r=1}^{\infty} r p_r z^{r-1} \leqslant \sum_{r=1}^{\infty} r p_r = \mathbb{E}\, X.$$

We see that $p'(z)$ is non-decreasing in z so that $\lim_{z \uparrow 1} p'(z) \leqslant \mathbb{E}\, X$. Take $\epsilon > 0$, and choose N so that $\sum_{r=1}^{N} r p_r \geqslant \mathbb{E}\, X - \epsilon$. Then

$$\lim_{z \uparrow 1} p'(z) \geqslant \lim_{z \uparrow 1} \sum_{r=1}^{N} r p_r z^{r-1} = \sum_{r=1}^{N} r p_r \geqslant \mathbb{E}\, X - \epsilon;$$

this is true for each $\epsilon > 0$, whence $\lim_{z \uparrow 1} p'(z) \geqslant \mathbb{E}\, X$ and it follows that $\lim_{z \uparrow 1} p'(z) = \mathbb{E}\, X$. If $\mathbb{E}\, X = \infty$, then for any $M > 0$ choose N so that $\sum_{r=1}^{N} r p_r \geqslant M$, and, as above, see that

$$\lim_{z \uparrow 1} p'(z) \geqslant \lim_{z \uparrow 1} \sum_{r=1}^{N} r p_r z^{r-1} = \sum_{r=1}^{N} r p_r \geqslant M;$$

this is true for any M, whence $\lim_{z \uparrow 1} p'(z) = \infty$. □

Note By considering the second derivative of $p(z)$, a similar argument to that of Theorem 3.17 may be used to show that

$$p''(1-) = \lim_{z \uparrow 1} p''(z) = \lim_{z \uparrow 1} \sum_{r=1}^{\infty} r(r-1) p_r z^{r-2} = \mathbb{E}\, \left(X(X-1) \right),$$

and by considering the kth derivative, $k \geqslant 1$, we have

$$
\begin{aligned}
p^{(k)}(1-) = \lim_{z \uparrow 1} p^{(k)}(z) &= \lim_{z \uparrow 1} \sum_{r=1}^{\infty} r(r-1) \cdots (r-k+1) p_r z^{r-2} \\
&= \mathbb{E}\ (X(X-1) \cdots (X-k+1)) .
\end{aligned}
$$

In particular, $\mathbb{V}\text{ar}\ (X) = p''(1-) + p'(1-) - (p'(1-))^2$.

Example 3.18 *Geometric distribution* Let X be a random variable with probability distribution given by $\mathbb{P}\,(X = r) = p(1-p)^r = pq^r$, $r = 0, 1, 2, \ldots$, where $0 < p = 1 - q < 1$. Then X may be thought of as the number of tails obtained before getting the first head when successively tossing a coin with probability p of heads on each toss. The probability generating function of X is

$$
p(z) = \mathbb{E}\ \left(z^X\right) = \sum_{r=0}^{\infty} p_r z^r = \sum_{r=0}^{\infty} pq^r z^r = \frac{p}{1 - qz} \ .
$$

We have $p'(z) = pq/\,(1-qz)^2$, so that $\mathbb{E}\,X = p'(1) = q/p$. Also, $p''(z) = 2pq^2/\,(1-qz)^3$, so that $\mathbb{E}\ (X(X-1)) = 2q^2/p^2$, from which we deduce that

$$
\mathbb{E}\,(X^2) = \frac{2q^2}{p^2} + \frac{q}{p} \quad \text{and} \quad \mathbb{V}\text{ar}\,(X) = \mathbb{E}\,(X^2) - (\mathbb{E}\,X)^2 = \frac{q}{p^2}. \qquad \square
$$

Note The term geometric distribution is often also given to the situation where $\mathbb{P}\,(X = r) = pq^{r-1}$, $r = 1, 2, \ldots$ for $0 < p = 1 - q < 1$. Here, X would be the number of tosses required to achieve the first head where the probability of heads is p. This just corresponds to replacing X in Example 3.18 by $X+1$, so the probability generating function becomes $pz/(1 - qz)$, the mean is $1/p$ and the variance is unchanged at q/p^2. $\square$

Another use for probability generating functions is that they provide an easy way of dealing with sums of independent random variables. Suppose that $X_1, \ldots, X_n$ are independent random variables with probability generating functions $p_1(z), \ldots, p_n(z)$ respectively. Then, since $z^{X_1}, \ldots, z^{X_n}$ are independent, we have that the probability generating function of $X_1 + \cdots + X_n$ is

$$
\mathbb{E}\ \left(z^{X_1 + \cdots + X_n}\right) = \prod_{i=1}^{n} \mathbb{E}\ \left(z^{X_i}\right) = \prod_{i=1}^{n} p_i(z).
$$

In the special case when $X_1, \ldots, X_n$ are i.i.d. with common probability generating function $p(z)$ we have

$$\mathbb{E}\left(z^{X_1+\cdots+X_n}\right) = (p(z))^n .$$

Example 3.19 *Sums of Binomial random variables* Consider independent random variables $X \sim \text{Bin}\,(n,p)$ and $Y \sim \text{Bin}\,(m,p)$, where $0 < p = 1-q < 1$. The probability generating function of X is

$$\mathbb{E}\left(z^X\right) = \sum_{r=0}^{n} \binom{n}{r} p^r q^{n-r} z^r = (pz+q)^n ,$$

so that the probability generating function of Y is $(pz+q)^m$. It follows that the probability generating function of $X+Y$ is the product of the two generating functions and is therefore $(pz+q)^{m+n}$. From Theorem 3.16 we conclude that $X+Y \sim \text{Bin}(n+m,p)$. The probabilistic interpretation is immediate, of course; X is the number of heads in n tosses of a coin with probability p of heads and Y is the number of tosses in m (independent) tosses of the coin, so that $X+Y$ is the number of heads in $n+m$ tosses. This of course generalizes, by induction, to the case of independent random variables $X_1, \ldots, X_k$ with $X_i \sim \text{Bin}\,(n_i,p)$, to give $X_1+\cdots+X_k \sim \text{Bin}\,\left(\sum_1^k n_i, p\right)$. □

Example 3.20 *Sums of Poisson random variables* Consider independent random variables $X \sim \text{Poiss}\,(\lambda)$ and $Y \sim \text{Poiss}\,(\mu)$, where $\lambda > 0$ and $\mu > 0$. The probability generating function of X is

$$\mathbb{E}\left(z^X\right) = \sum_{r=0}^{\infty} z^r e^{-\lambda} \frac{\lambda^r}{r!} = e^{-\lambda(1-z)}.$$

The probability generating function is the same expression with μ replacing λ and the probability generating function of $X+Y$ is

$$e^{-\lambda(1-z)} e^{-\mu(1-z)} = e^{-(\lambda+\mu)(1-z)};$$

from Theorem 3.16 we conclude that $X+Y \sim \text{Poiss}\,(\lambda+\mu)$; for an alternative argument see Example 3.22 below. □

Example 3.21 *Negative binomial distribution* Consider a random variable X which has distribution given by

$$\mathbb{P}(X = r) = \binom{r-1}{n-1} p^n (1-p)^{r-n}, \quad \text{for} \quad r = n, n+1, \ldots,$$

where $0 < p = 1 - q < 1$, and $n \geqslant 1$. Here, X represents the number of tosses of a coin to get n heads for the first time, where the probability of heads is p. The probability generating function of X is

$$\mathbb{E}\left(z^X\right) = \sum_{r=n}^{\infty} z^r \binom{r-1}{n-1} p^n q^{r-n} = (pz)^n \sum_{r=n}^{\infty} \binom{r-1}{n-1} (qz)^{r-n} = \left(pz/(1-qz)\right)^n .$$

From the note following Example 3.18 we see that X may be represented as the sum $X_1 + \cdots + X_n$ of n i.i.d. random variables each with the same geometric distribution

$$\mathbb{P}(X_1 = r) = pq^{r-1}, \quad \text{for} \quad r = 1, 2, \ldots.$$

The distribution of X is usually referred to as the negative binomial distribution. □

3.5 Conditional distributions

The **joint distribution** of random variables $X_1, \ldots, X_n$, is given by

$$\mathbb{P}(X_1 = x_1, \ldots, X_n = x_n) \quad \text{for} \quad x_1 \in \Omega_{X_1}, \ldots, x_n \in \Omega_{X_n},$$

and it is a probability distribution on $\Omega_{X_1} \times \cdots \times \Omega_{X_n}$, and the **marginal distribution** of X_i is

$$\mathbb{P}(X_i = x_i) = \sum \mathbb{P}(X_1 = x_1, \ldots, X_n = x_n),$$

where the summation is over $x_1, \ldots, x_{i-1}, x_{i+1} \ldots, x_n$; this identity is a consequence of the law of total probability. Now consider the case $n = 2$ and (to avoid unnecessary subscripts) consider the random variables X and Y. The **conditional distribution** of X, given $Y = y$, is a probability distribution on Ω_X given by

$$\mathbb{P}(X = x \mid Y = y) \quad \text{for} \quad x \in \Omega_X,$$

where, of course, $\mathbb{P}(X = x \mid Y = y) = \mathbb{P}(X = x, Y = y)/\mathbb{P}(Y = y)$. Again, by the law of total probability

$$\mathbb{P}(X = x) = \sum_{y \in \Omega_Y} \mathbb{P}(X = x, Y = y) = \sum_{y \in \Omega_Y} \mathbb{P}(X = x \mid Y = y)\, \mathbb{P}(Y = y).$$

Example 3.22 *Sum of two independent random variables* Suppose that X and Y are independent random variables, then we may express the distribution of their sum as follows

$$\begin{aligned}\mathbb{P}(X + Y = z) &= \sum_{y \in \Omega_Y} \mathbb{P}(X + Y = z \mid Y = y)\, \mathbb{P}(Y = y) = \sum_{y \in \Omega_Y} \mathbb{P}(X = z - y)\, \mathbb{P}(Y = y) \\ &= \sum_{x \in \Omega_X} \mathbb{P}(X = x)\, \mathbb{P}(Y = z - x), \qquad (3.23)\end{aligned}$$

where the last expression is obtained if we condition on X initially instead of Y. This procedure gives the **convolution** of the distributions of X and Y. For example, if $X \sim \text{Poiss}(\lambda)$ and $Y \sim \text{Poiss}(\mu)$,

$$\begin{aligned}\mathbb{P}(X + Y = n) &= \sum_{r=0}^{\infty} \mathbb{P}(X = n - r)\, \mathbb{P}(Y = r) \\ &= \sum_{r=0}^{n} e^{-\lambda} \frac{\lambda^{n-r}}{r!}\, e^{-\mu} \frac{\mu^r}{(n-r)!}, \quad \text{since } \mathbb{P}(X = k) = 0, \text{ for } k < 0, \\ &= \frac{e^{-(\lambda+\mu)}}{n!} \sum_{r=0}^{n} \binom{n}{r} \lambda^{n-r} \mu^r = e^{-(\lambda+\mu)} \frac{(\lambda + \mu)^n}{n!},\end{aligned}$$

so that $X + Y \sim \text{Poiss}(\lambda + \mu)$, as seen in Example 3.20 previously using generating functions. □

The **conditional expectation of X given $Y = y$** is

$$\mathbb{E}(X \mid Y = y) = \sum_{x \in \Omega_X} x \mathbb{P}(X = x \mid Y = y) = \sum_{\omega : Y(\omega) = y} X(\omega) \mathbb{P}(\{\omega\}) / \mathbb{P}(Y = y).$$

Note that $\mathbb{E}(X \mid Y = y)$ is a function of y, $g(y)$ say, then the random variable $g(Y)$ is known as the **conditional expectation of X given Y** and is written $\mathbb{E}(X \mid Y)$. It is important to emphasize that $\mathbb{E}(X \mid Y)$ is a random variable and it is a function of Y, in contrast to $\mathbb{E}(X \mid Y = y)$, which is a real number.

Example 3.24 Consider tossing a coin n times where the probability of a head is p, $0 < p = 1 - q < 1$, and let $X_i = 1$ if the ith toss produces a head and $X_i = 0$, otherwise. Let $Y = X_1 + \cdots + X_n$ denote the total number of heads so that $Y \sim \text{Bin}\,(n, p)$. Then, for $r \geqslant 1$,

$$\mathbb{P}\,(X_1 = 1 \mid Y = r) = \frac{\mathbb{P}\,(X_1 = 1, Y = r)}{\mathbb{P}\,(Y = r)} = \frac{\mathbb{P}\,(X_1 = 1, X_1 + \cdots + X_n = r)}{\mathbb{P}\,(Y = r)}$$
$$= \frac{\mathbb{P}\,(X_1 = 1, X_2 + \cdots + X_n = r - 1)}{\mathbb{P}\,(Y = r)},$$

then by independence and the fact that $X_2 + \cdots + X_n \sim \text{Bin}\,(n-1, p)$ this

$$= \frac{\mathbb{P}\,(X_1 = 1)\,\mathbb{P}\,(X_2 + \cdots + X_n = r - 1)}{\mathbb{P}\,(Y = r)}$$
$$= \frac{p\binom{n-1}{r-1}p^{r-1}q^{n-r}}{\binom{n}{r}p^r q^{n-r}} = \frac{r}{n};$$

we may see also that $\mathbb{P}\,(X_1 = 1 \mid Y = 0) = 0$. Then

$$\mathbb{E}\,(X_1 \mid Y = r) = 1 \times \mathbb{P}\,(X_1 = 1 \mid Y = r) + 0 \times \mathbb{P}\,(X_1 = 0 \mid Y = r) = \frac{r}{n}, \quad 0 \leqslant r \leqslant n.$$

In this case we have $\mathbb{E}\,(X_1 \mid Y) = Y/n$. □

Properties of conditional expectation

1. For c, a constant, then $\mathbb{E}\,(cX \mid Y) = c\mathbb{E}\,(X \mid Y)$ and $\mathbb{E}\,(\,c \mid Y) = c$.
2. For random variables $X_1, \ldots, X_n$, $\mathbb{E}\,\left(\sum_i X_i \mid Y\right) = \sum_i \mathbb{E}\,(X_i \mid Y)$.
3. $\mathbb{E}\,(\mathbb{E}\,(X \mid Y)) = \mathbb{E}\,(X)$.

Proof. We have

$$\mathbb{E}\,(\mathbb{E}\,(X \mid Y)) = \sum_{y \in \Omega_Y} \left(\sum_{x \in \Omega_X} x\mathbb{P}\,(X = x \mid Y = y) \right) \mathbb{P}\,(Y = y)$$
$$= \sum_{x \in \Omega_X} x \sum_{y \in \Omega_Y} \mathbb{P}\,(X = x, Y = y) = \sum_{x \in \Omega_X} x\mathbb{P}\,(X = x) = \mathbb{E}\,(X)\,.$$ □

4. When X and Y are independent, $\mathbb{E}\,(X \mid Y) = \mathbb{E}\,(X)$.

Proof. For $y \in \Omega_Y$,

$$\mathbb{E}\,(X \mid Y = y) = \sum_{x \in \Omega_X} x\mathbb{P}\,(X = x \mid Y = y) = \sum_{x \in \Omega_X} x\mathbb{P}\,(X = x) = \mathbb{E}\,(X)\,.$$ □

5. When Y and Z are independent, $\mathbb{E}\left(\mathbb{E}\left(X \mid Y\right) \mid Z\right) = \mathbb{E}\left(X\right)$.

Proof. Since $\mathbb{E}\left(X \mid Y\right)$ is a function of Y it is independent of Z, so using Property 4 and then Property 3, we have

$$\mathbb{E}\left(\mathbb{E}\left(X \mid Y\right) \mid Z\right) = \mathbb{E}\left(\mathbb{E}\left(X \mid Y\right)\right) = \mathbb{E}\left(X\right). \qquad \square$$

6. For any function $h : \mathbb{R} \to \mathbb{R}$, we have $\mathbb{E}\left(h(Y)X \mid Y\right) = h(Y)\mathbb{E}\left(X \mid Y\right)$.

Proof. We have, for $y \in \Omega_Y$,

$$\mathbb{E}\left(h(Y)X \mid Y = y\right) = \sum_{\omega : Y(\omega) = y} h(Y(\omega))X(\omega)\mathbb{P}\left(\{\omega\}\right) / \mathbb{P}\left(Y = y\right) = h(y)\mathbb{E}\left(X \mid Y = y\right).$$

A particular consequence of this and Property 1 is that $\mathbb{E}\left(\mathbb{E}\left(X \mid Y\right) \mid Y\right) = \mathbb{E}\left(X \mid Y\right)$. $\square$

7. The conditional expectation $\mathbb{E}\left(X \mid Y\right)$ is that function $h(Y)$ of Y which minimizes $\mathbb{E}\left(X - h(Y)\right)^2$ over all functions h.

Proof. Write

$$\mathbb{E}\left(X - h(Y)\right)^2 = \mathbb{E}\left[X - \mathbb{E}\left(X \mid Y\right) + \mathbb{E}\left(X \mid Y\right) - h(Y)\right]^2,$$

which may be expanded to

$$\mathbb{E}\left[X - \mathbb{E}\left(X \mid Y\right)\right]^2 + \mathbb{E}\left[\mathbb{E}\left(X \mid Y\right) - h(Y)\right]^2 + 2\mathbb{E}\left[\big(X - \mathbb{E}\left(X \mid Y\right)\big)\big(\mathbb{E}\left(X \mid Y\right) - h(Y)\big)\right].$$

Now consider half the cross-product term,

$$\mathbb{E}\left[\big(X - \mathbb{E}\left(X \mid Y\right)\big)\big(\mathbb{E}\left(X \mid Y\right) - h(Y)\big)\right] = \mathbb{E}\left(\mathbb{E}\left[\big(X - \mathbb{E}\left(X \mid Y\right)\big)\big(\mathbb{E}\left(X \mid Y\right) - h(Y)\big) \;\big|\; Y\right]\right)$$

by using Property 3, and then, using Property 5, this

$$= \mathbb{E}\left(\big(\mathbb{E}\left(X \mid Y\right) - h(Y)\big)\mathbb{E}\left[\big(X - \mathbb{E}\left(X \mid Y\right)\big) \;\big|\; Y\right]\right);$$

but $\mathbb{E}\left[\big(X - \mathbb{E}\left(X \mid Y\right)\big) \;\big|\; Y\right] = \mathbb{E}\left(X \mid Y\right) - \mathbb{E}\left(X \mid Y\right) = 0$, so that

$$\mathbb{E}\left(X - h(Y)\right)^2 = \mathbb{E}\left[X - \mathbb{E}\left(X \mid Y\right)\right]^2 + \mathbb{E}\left[\mathbb{E}\left(X \mid Y\right) - h(Y)\right]^2$$

from which the result follows, since the first term in this expression does not involve h and the second term is minimized by $h(Y) = \mathbb{E}\left(X \mid Y\right)$. $\square$

Example 3.25 *Sum of a random number of random variables* Let $X_1, X_2, \ldots$ be independent and identically distributed random variables with common probability generating function $p(z)$. Let N be a non-negative integer valued random variable independent of the $\{X_i\}$ and having probability generating function $q(z)$. We consider the p.g.f. of the random variable $X_1 + \cdots + X_N$; (here the sum is 0 if $N = 0$).

$$\begin{aligned} r(z) = \mathbb{E}\left(z^{X_1+\cdots+X_N}\right) = \mathbb{E}\left(\mathbb{E}\left(z^{X_1+\cdots+X_N} \mid N\right)\right) = \mathbb{E}\left(\left(\mathbb{E}\, z^{X_1}\right)^N\right) \\ = \mathbb{E}\left((p(z))^N\right) = q(p(z)). \end{aligned}$$

If at a first reading you find the second equality too cryptic, you might wish to spell out the argument as

$$\begin{aligned} \mathbb{E}\left(z^{X_1+\cdots+X_N}\right) &= \sum_{n=0}^{\infty} \mathbb{E}\left(z^{X_1+\cdots+X_N} \mid N = n\right) \mathbb{P}(N = n) \\ &= \sum_{n=0}^{\infty} \mathbb{E}\left(z^{X_1+\cdots+X_n} \mid N = n\right) \mathbb{P}(N = n) \\ &= \sum_{n=0}^{\infty} \mathbb{E}\left(z^{X_1+\cdots+X_n}\right) \mathbb{P}(N = n) = \sum_{n=0}^{\infty} (p(z))^n\, \mathbb{P}(N = n) = q(p(z)). \end{aligned}$$

After some practice you should find the conditional expectation shorthand notation given first more helpful. It follows from the expression for $r(z)$ that $r'(z) = q'(p(z))p'(z)$, so that

$$\mathbb{E}\left(X_1 + \cdots + X_N\right) = q'(p(1-))p'(1-) = (\mathbb{E}\, N)\,(\mathbb{E}\, X_1)\,,$$

since $p(1-) = 1$. Furthermore, since $r''(z) = q''(p(z)\,(p'(z))^2 + q'(p(z))p''(z)$, and the fact that $q''(1-) = \mathbb{E}\,(N)^2 - \mathbb{E}\, N$ and $p''(1-) = \mathbb{E}\,(X_1)^2 - \mathbb{E}\, X_1$, we may calculate that

$$\begin{aligned} \mathbb{V}\mathrm{ar}\,(X_1 + \cdots + X_N) &= r''(1-) + r'(1-) - (r'(1-))^2 \\ &= (\mathbb{E}\, N)\,\mathbb{V}\mathrm{ar}\,(X_1) + (\mathbb{E}\, X_1)^2\,\mathbb{V}\mathrm{ar}\,(N)\,. \end{aligned}$$

Notice that the variance of $X_1+\cdots+X_N$ is increased over what it would be if N is constant, $N \equiv \mathbb{E}\, N = n$, say, by the amount $(\mathbb{E}\, X_1)^2\, \mathbb{V}\mathrm{ar}\,(N)$; if $\mathbb{V}\mathrm{ar}\,(N) = 0$ and N is constant we get the usual expression for the variance of a sum of n i.i.d. random variables. □

3.6 Branching processes

As an example of conditional expectations and of generating functions we will consider a model of population growth and extinction known as the Bienaymé-Galton-Watson process. Consider a sequence of random variables $X_0, X_1, \ldots$, where X_n represents the number of individuals in the nth generation. We will assume that the population is initiated by one individual, take $X_0 \equiv 1$, and when he dies he is replaced by k individuals with probability g_k, $k = 0, 1, 2, \ldots$. These individuals behave independently and identically to the parent individual, as do those in subsequent generations. The number in the $(n+1)$st generation, X_{n+1}, depends on the number in the nth generation and is given by

$$X_{n+1} = \begin{cases} Y_1^n + Y_2^n + \cdots + Y_{X_n}^n & \text{when } X_n \geqslant 1, \\ 0 & \text{when } X_n = 0. \end{cases}$$

Here $\left\{Y_j^n : n \geqslant 1, j \geqslant 1\right\}$ are independent, identically distributed random variables with $\mathbb{P}\left(Y_j^n = k\right) = g_k$, for $k \geqslant 0$ and Y_j^n represents the number of offspring of the jth individual in the nth generation, $j \leqslant X_n$.

Assumptions (i) $g_0 > 0$; and (ii) $g_0 + g_1 < 1$.

Assumption (i) means that the population can die out (extinction) since in each generation there is positive probability that all individuals have no offspring; assumption (ii) means that the population may grow, there is positive probability that the next generation has more individuals than the present one. Now let $G(z) = \sum_{k=0}^{\infty} g_k z^k = \mathbb{E}\left(z^{X_1}\right)$ and set $G_n(z) = \mathbb{E}\left(z^{X_n}\right)$, for $n \geqslant 1$, so that $G_1 = G$.

Theorem 3.26 *For all $n \geqslant 1$, $G_{n+1}(z) = G_n(G(z)) = G(\cdots(G(z))\cdots) = G(G_n(z))$.*

Proof. Note that $Y_1^n, Y_2^n, \ldots$ are independent of X_n, so that

$$\begin{aligned} G_{n+1}(z) = \mathbb{E}\left(z^{X_{n+1}}\right) &= \sum_{k=0}^{\infty} \mathbb{E}\left(z^{X_{n+1}} \mid X_n = k\right) \mathbb{P}(X_n = k) \\ &= \sum_{k=0}^{\infty} \mathbb{E}\left(z^{Y_1^n + \cdots + Y_k^n}\right) \mathbb{P}(X_n = k) = \sum_{k=0}^{\infty} (G(z))^k \, \mathbb{P}(X_n = k) \\ &= \mathbb{E}\left((G(z))^{X_n}\right) = G_n(G(z)). \end{aligned}$$

□

Corollary 3.27 *For* $m = \mathbb{E}\,(X_1) = \sum_{k=1}^{\infty} k g_k$ *and* $\sigma^2 = \mathbb{V}\text{ar}\,(X_1) = \sum_{k=0}^{\infty} (k-m)^2 g_k$*, then for* $n \geqslant 1$*, we have*

$$\mathbb{E}\,(X_n) = m^n, \quad \mathbb{V}\text{ar}\,(X_n) = \begin{cases} \dfrac{\sigma^2 m^{n-1}(m^n - 1)}{m-1} & \text{when } m \neq 1, \\ n\sigma^2 & \text{when } m=1. \end{cases}$$

Proof. Differentiating $G_n(z) = G_{n-1}(G(z))$ to obtain $G_n'(z) = G_{n-1}'(G(z))G'(z)$ and letting $z \uparrow 1$, it follows that

$$\mathbb{E}\,(X_n) = m\mathbb{E}\,(X_{n-1}) = \cdots = m^n \mathbb{E}\,(X_0) = m^n,$$

since $X_0 = 1$. Differentiating $G_n(z)$ a second time gives

$$G_n''(z) = G_{n-1}''(G(z))(G'(z))^2 + G_{n-1}'(G(z))\,G''(z),$$

and letting $z \uparrow 1$ again we have

$$\mathbb{E}\,(X_n(X_n - 1)) = m^2\mathbb{E}\,(X_{n-1}(X_{n-1}-1)) + \left(\sigma^2 + m^2 - m\right)\mathbb{E}\,(X_{n-1}).$$

We then have, using the fact that $\mathbb{E}\,X_n = m^n$,

$$\begin{aligned} \mathbb{V}\text{ar}\,(X_n) &= \mathbb{E}\,(X_n(X_n-1)) + \mathbb{E}\,(X_n) - (\mathbb{E}\,X_n)^2 \\ &= m^2\mathbb{E}\,(X_{n-1}(X_{n-1}-1)) + \left(\sigma^2 + m^2 - m\right)\mathbb{E}\,(X_{n-1}) + m^n - m^{2n} \\ &= m^2\left[\mathbb{V}\text{ar}\,(X_{n-1}) - \mathbb{E}\,(X_{n-1}) + (\mathbb{E}\,X_{n-1})^2\right] + \left(\sigma^2 + m^2\right)m^{n-1} - m^{2n} \\ &= m^2\mathbb{V}\text{ar}\,(X_{n-1}) + \sigma^2 m^{n-1}. \end{aligned}$$

Iterating this, we see that

$$\begin{aligned} \mathbb{V}\text{ar}\,(X_n) &= m^2\mathbb{V}\text{ar}\,(X_{n-1}) + \sigma^2 m^{n-1} = m^4\mathbb{V}\text{ar}\,(X_{n-2}) + \sigma^2\left(m^{n-1} + m^n\right) = \cdots \\ &= m^{2n}\mathbb{V}\text{ar}\,(X_0) + \sigma^2\left(m^{n-1} + \cdots + m^{2n-2}\right) \\ &= \sigma^2\left(m^{n-1} + \cdots + m^{2n-2}\right), \text{ since } \mathbb{V}\text{ar}\,(X_0) = 0 \text{ because } X_0 = 1, \end{aligned}$$

and then the result may be obtained immediately. □

Probability of extinction Notice that $X_n = 0$ implies that $X_{n+1} = 0$ so that if we let $A_n = (X_n = 0)$, the event that the population is extinct at or before generation n, we have $A_n \subseteq A_{n+1}$ and $A = \bigcup_{n=1}^{\infty} A_n$ represents the event that extinction ever occurs. Notice that $\mathbb{P}(A_n) = G_n(0)$ and by the continuity property of probabilities on increasing events we see that the extinction probability, q, say, is

$$q = \mathbb{P}(A) = \lim_{n\to\infty} \mathbb{P}(A_n) = \lim_{n\to\infty} G_n(0) = \lim_{n\to\infty} \mathbb{P}(X_n = 0).$$

Theorem 3.28 *The extinction probability q is the smallest positive root of the equation $G(z) = z$. When m, the mean number of offspring per individual, satisfies $m \leqslant 1$ then $q = 1$; when $m > 1$ then $q < 1$.*

Proof. The fact that the extinction probability q is well defined follows from the above and since G is continuous and $q = \lim_{n\to\infty} G_n(0)$ we have $G\left(\lim_{n\to\infty} G_n(0)\right) = \lim_{n\to\infty} G_{n+1}(0)$, so that $G(q) = q$, that is q is a root of $G(z) = z$; note that 1 is always a root since $G(1) = \sum_{r=0}^{\infty} g_r = 1$. Now let $H(z) = G(z) - z$, then $H'' = \sum_{r=0}^{\infty} r(r-1) g_r z^{r-2} > 0$ for $0 < z < 1$ provided $g_0 + g_1 < 1$, so the derivative of H is strictly increasing in the range $0 < z < 1$, hence H can have at most one root different from 1 in $[0, 1]$ (Rolle's Theorem).

Firstly, suppose that H has no root in $[0, 1)$ then, since $H(0) = g_0 > 0$ we must have $H(z) > 0$ for all $0 < z < 1$, so $H(1) - H(z) < H(1) = 0$ and so

$$H'(1-) = \lim_{z\uparrow 1} \frac{H(1) - H(z)}{1 - z} \leqslant 0, \quad \text{whence } m = G'(1-) \leqslant 1.$$

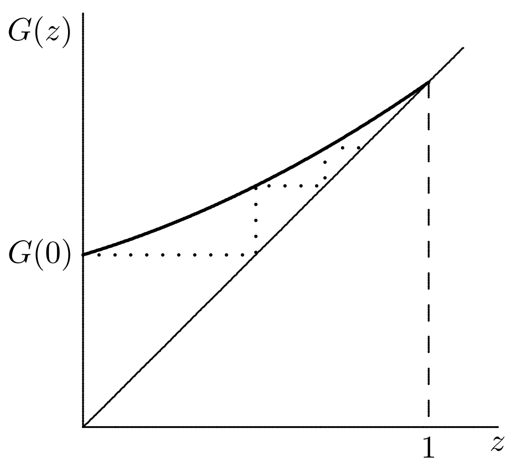


Fig. 1: $m \leqslant 1$, $q = 1$

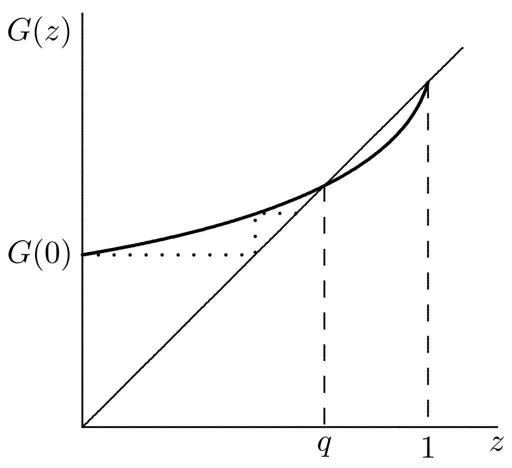


Fig. 2: $m > 1$, $q < 1$

Next, suppose that H has a unique root r in $[0, 1)$, then H' must have a root in $[r, 1)$; that is $H'(z) = G'(z) - 1 = 0$ for some z, $r \leqslant z < 1$. The function G' is strictly increasing (since $g_0 + g_1 < 1$) so that $m = G'(1-) > G'(z) = 1$. Thus we see that $m \leqslant 1$, if and only if, $q = 1$. Furthermore, if $m > 1$, let $\alpha > 0$ be the smallest positive root of $G(z) = z$, so that because G is increasing, $\alpha = G(\alpha) \geqslant G(0)$, and repeating n times we have $\alpha \geqslant G_n(0)$, whence $\alpha \geqslant \lim_{n\to\infty} G_n(0) = q$, so that we must have $\alpha = q$, completing the proof. □

Note Figures 1 and 2 illustrate the two situations $m \leqslant 1$ and $m > 1$; the dotted lines illustrate the iteration $G_{n+1}(0) = G(G_n(0))$ tending to the smallest positive root, q.

3.7 Random walks

Let $X_1, X_2, \dots$ be i.i.d. random variables and set $S_n = S_0 + X_1 + \cdots + X_k$ for $k \geqslant 1$ where S_0 is a constant then $\{S_k, k \geqslant 0\}$ is known as a (one-dimensional) **random walk**. When each X_i just takes the two values $+1$ and -1 with probabilities p and $q = 1 - p$, respectively, it is a **simple random walk** and further when $p = q = \frac{1}{2}$ it is a **simple, symmetric random walk**. We will consider simple random walks.

Recurrence relations The problems we will look at for the simple random walk often reduce to the solution of recurrence relations (or difference equations). We consider the general solution of such equations in the simplest situations which have constant coefficients.

1. First-order equations: The general first-order equation is $x_{n+1} = ax_n + b$, for $n \geqslant 0$,where a and b are constants; the case $b = 0$ gives the general first-order homogeneous equation $x_{n+1} = ax_n$, which trivially may be solved as $x_n = a^n x_0$; if y_n is any solution of the inhomogeneous equation, then the general solution of the inhomogeneous equation is of the form $x_n = Ca^n + y_n$ for some constant C (because $x_n - y_n$ must be a solution of the homogeneous equation). The constant is determined by a boundary condition.
2. Second-order equations: $x_{n+1} = ax_n + bx_{n-1} + c$, for $n \geqslant 1$, where a, b and c are constants. First consider the homogeneous case where $c = 0$. Then write the

relation in matrix form as follows:

$$\begin{pmatrix} x_{n+1} \\ x_n \end{pmatrix} = \begin{pmatrix} a & b \\ 1 & 0 \end{pmatrix} \begin{pmatrix} x_n \\ x_{n-1} \end{pmatrix} = A \begin{pmatrix} x_n \\ x_{n-1} \end{pmatrix}, \quad \text{where } A = \begin{pmatrix} a & b \\ 1 & 0 \end{pmatrix}.$$

It follows that

$$\begin{pmatrix} x_{n+1} \\ x_n \end{pmatrix} = A^n \begin{pmatrix} x_1 \\ x_0 \end{pmatrix};$$

find the eigenvalues of A, by solving

$$\begin{vmatrix} a-\lambda & b \\ 1 & -\lambda \end{vmatrix} = 0, \text{ to give the equation } \lambda^2 - a\lambda - b = 0,$$

with roots λ_1 and λ_2, say. This equation is known as the **auxiliary equation** of the recurrence relation; it corresponds to seeking a solution of the form $x_n = \lambda^n$. If λ_1 and λ_2 are distinct then for some matrix Λ we may write

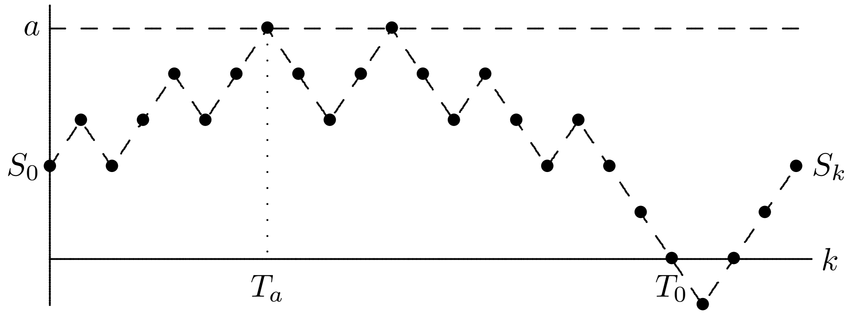
$$A = \Lambda^{-1} \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix} \Lambda \quad \text{and then} \quad A^n = \Lambda^{-1} \begin{pmatrix} \lambda_1^n & 0 \\ 0 & \lambda_2^n \end{pmatrix} \Lambda,$$

so that the general solution of the homogeneous equation may be seen to be of the form $x_n = C\lambda_1^n + D\lambda_2^n$ for some constants C and D. If the eigenvalues are not distinct, $\lambda_1 = \lambda_2 = \lambda$, then

$$A = \Lambda^{-1} \begin{pmatrix} \lambda & 1 \\ 0 & \lambda \end{pmatrix} \Lambda \quad \text{and then} \quad A^n = \Lambda^{-1} \begin{pmatrix} \lambda^n & n\lambda^{n-1} \\ 0 & \lambda^n \end{pmatrix} \Lambda,$$

and then the general solution of the homogeneous equation may be seen to be of the form $x_n = \lambda^n(C + Dn)$ for some constants C and D. As before, if y_n is any particular solution of the inhomogeneous equation, the general solution is of the form $x_n + y_n$ where x_n is the general solution of the homogeneous equation.

Example 3.29 *Gambler's ruin* For the simple random walk, $\{S_k\}$ may represent the fortune of a gambler after k plays of a game where on each play he either wins £1, with probability p, or loses £1 with probability $q = 1-p$; his initial fortune is £S_0 and a classical problem is to calculate the probability that his fortune achieves the level a, $a > S_0$, before the time of ruin, that is the time that he goes bankrupt (his fortune hits the level 0). If T_a denotes the first time that the random walk hits the level a and T_0 the time the random



walk first hits the level 0, we would wish to calculate $\mathbb{P}(T_a < T_0)$, given that his fortune starts at $S_0 = r$, $0 < r < a$.

The figure illustrates a path of the random walk–although, in the case of the game, it finishes at the instant T_0, the time of bankruptcy! Let $x_r = \mathbb{P}(T_a < T_0)$ when $S_0 = r$, for $0 \leqslant r \leqslant a$, so that we have the boundary conditions $x_a = 1$ and $x_0 = 0$. A general rule in problems of this type in probability may be summed up as 'condition on the first thing that happens', which here would be a shorthand for using the Law of Total Probability to express the probability conditional on the outcome of the first play of the game, that is, whether $X_1 = 1$ or $X_1 = -1$, or equivalently, $S_1 = r+1$ or $S_1 = r-1$. Thus, for $0 < r < a$,

$$\begin{aligned} x_r &= \mathbb{P}(T_a < T_0 \mid S_1 = r+1)\,\mathbb{P}(X_1 = 1) + \mathbb{P}(T_a < T_0 \mid S_1 = r-1)\,\mathbb{P}(X_1 = -1) \\ &= p x_{r+1} + q x_{r-1}. \end{aligned}$$

The auxiliary equation for this recurrence relation is $p\lambda^2 - \lambda + q = 0$, and since $p + q = 1$, this may be factored as $(\lambda - 1)(p\lambda - q) = 0$ to give roots $\lambda = 1$ and $\lambda = q/p$.

Case $p \neq q$: the roots are distinct and the general solution is of the form $x_r = A + B\,(q/p)^r$ for some constants A and B; the boundary conditions at $r = a$ and $r = 0$, fix A and B and we conclude that

$$x_r = \mathbb{P}(T_a < T_0) = \frac{1 - (q/p)^r}{1 - (q/p)^a}, \quad \text{for} \quad 0 \leqslant r \leqslant a.$$

Case $p = q = \frac{1}{2}$: here $\lambda = 1$ is a repeated root of the auxiliary equation so that the general solution of the recurrence relation is $x_r = A + Br$, which, after using the boundary conditions, leads to the solution $x_r = r/a$, $0 \leqslant r \leqslant a$.

We do not know necessarily that at least one of T_0 and T_a must be finite, but if we interchange p and q and replace r by $a - r$, (or just calculate directly as above) we may obtain, for $S_0 = r$, $0 \leqslant r \leqslant a$, that

$$\mathbb{P}(T_0 < T_a) = \begin{cases} \dfrac{(q/p)^r - (q/p)^a}{1 - (q/p)^a} & \text{when } p \neq q, \\ 1 - r/a & \text{when } p = q = \frac{1}{2}. \end{cases}$$

It follows, in both cases, that $P(T_a < T_0) + \mathbb{P}(T_0 < T_a) = 1$, so that at least one of the the two barriers, 0 or a, must be reached with certainty. □

Example 3.30 *Probability of ruin* From the previous calculation we may derive an expression for $\mathbb{P}(T_0 < \infty)$ given $S_0 = r > 0$, which is the probability that ruin ever happens. We see that the event that ruin occurs may be written as

$$(T_0 < \infty) = \bigcup_{a=r+1}^{\infty} (T_0 < T_a);$$

the events in the union are expanding as a increases, so by the continuity of the probability on expanding events, we have

$$\mathbb{P}(T_0 < \infty) = \lim_{a \to \infty} \mathbb{P}(T_0 < T_a) = \begin{cases} (q/p)^r & \text{when } p > q, \\ 1 & \text{when } p \leqslant q, \end{cases}$$

so that ruin is certain except in the case when the probability of winning a play is strictly larger than $\frac{1}{2}$. □

Example 3.31 *Expected duration of the game* Suppose that the gambler plays either until his fortune reaches a or until he goes bankrupt, whichever is sooner. That is the number of plays is $\min(T_0, T_a) = T_0 \wedge T_a$. We will derive the expected length of the game, $\mathbb{E}(T_0 \wedge T_a)$, given that $S_0 = r$, $0 \leqslant r \leqslant a$, which we will denote by m_r. We do not know whether m_r is finite. Consider blocks of jumps of the random walk of length a, that is

$$\begin{array}{cccc} X_1 & X_2 & \cdots & X_a \\ X_{a+1} & X_{a+2} & \cdots & X_{2a} \\ X_{2a+1} & X_{2a+2} & \cdots & X_{3a} \\ \vdots & \vdots & \vdots & \vdots \end{array}$$

and for $i \geqslant 1$ set $Y_i = 1$ if either $X_{(i-1)a+1} = X_{(i-1)a+2} = \cdots = X_{ia} = 1$ or $X_{(i-1)a+1} = X_{(i-1)a+2} = \cdots = X_{ia} = -1$, otherwise $Y_i = 0$. Thus $Y_i = 1$ if and only if the ith block of plays is a run of all wins or all losses, and $\mathbb{P}(Y_i = 1) = 1 - \mathbb{P}(Y_i = 0) = p^a + q^a = \theta$, say. If we let Z be the first i such that $Y_i = 1$, then Z has a geometric distribution $\mathbb{P}(Z = j) = (1-\theta)^{j-1}\theta$, $j \geqslant 1$, and so $\mathbb{E}(Z) = 1/\theta < \infty$. But it is clear that $T_0 \wedge T_a \leqslant aZ$, hence we see that $\mathbb{E}(T_0 \wedge T_a) \leqslant a\mathbb{E}(Z) < \infty$. To compute m_r, we again condition on the first thing to happen, that is whether the first play is a win or loss, to see that for $0 < r < a$,

$$m_r = p\left(1 + m_{r+1}\right) + q\left(1 + m_{r-1}\right) = 1 + pm_{r+1} + qm_{r-1}, \quad \text{with} \quad m_0 = m_a = 0;$$

here the 1 in the recurrence relation counts the initial play of the game. The solution of the homogeneous equation is again $m_r = A + B\,(q/p)^r$ when $p \neq q$ and $m_r = A + Br$ for the case $p = q = \frac{1}{2}$.

Case $p \neq q$: look for a particular solution of the inhomogeneous equation with $m_r = cr$, then

$$cr = 1 + pc(r+1) + qc(r-1), \quad \text{so that} \quad c = 1/(q-p),$$

so that the general solution is $m_r = r/(q-p) + A + B\,(q/p)^r$, and after using the boundary conditions we have

$$m_r = \frac{r}{q-p} - \left(\frac{a}{q-p}\right)\frac{1-(q/p)^r}{1-(q/p)^a}\ .$$

Case $p = q = \frac{1}{2}$: a particular solution of the inhomogeneous equation is $-r^2$, so the general solution is $m_r = A + Br - r^2$ and after using the boundary conditions we have $m_r = r(a-r)$. □