3

Discrete random variables

Summary. The distribution of a discrete random variable may be specified via its probability mass function. The key notion of independence for discrete random variables is introduced. The concept of expectation, or mean value, is defined for discrete variables, leading to a definition of the variance and the moments of a discrete random variable. Joint distributions, conditional distributions, and conditional expectation are introduced, together with the ideas of covariance and correlation. The Cauchy–Schwarz inequality is presented. The analysis of sums of random variables leads to the convolution formula for mass functions. Random walks are studied in some depth, including the reflection principle, the ballot theorem, the hitting time theorem, and the arc sine laws for visits to the origin and for sojourn times.

3.1 Probability mass functions

Recall that a random variable X is *discrete* if it takes values only in some countable set $\{x_1, x_2, \dots\}$. Its distribution function $F(x) = \mathbb{P}(X \le x)$ is a jump function; just as important as its distribution function is its mass function.

(1) Definition. The **(probability) mass function**† of a discrete random variable X is the function $f : \mathbb{R} \to [0, 1]$ given by $f(x) = \mathbb{P}(X = x)$.

The distribution and mass functions are related by

$$F(x) = \sum_{i:x_i \le x} f(x_i), \qquad f(x) = F(x) - \lim_{y \uparrow x} F(y).$$

(2) Lemma. *The probability mass function* $f : \mathbb{R} \to [0, 1]$ *satisfies*:

(a) *the set of x such that $f(x) \ne 0$ is countable,*

(b) $\sum_i f(x_i) = 1$, *where $x_1, x_2, \dots$ are the values of x such that $f(x) \ne 0$.*

Proof. The proof is obvious. ■

This lemma characterizes probability mass functions.

†Some refer loosely to the mass function of X as its distribution.

(3) Example. Binomial distribution. A coin is tossed n times, and a head turns up each time with probability p $(= 1 - q)$. Then $\Omega = \{\mathrm{H}, \mathrm{T}\}^n$. The total number X of heads takes values in the set $\{0, 1, 2, \dots, n\}$ and is a discrete random variable. Its probability mass function $f(x) = \mathbb{P}(X = x)$ satisfies

$$f(x) = 0 \quad \text{if} \quad x \notin \{0, 1, 2, \dots, n\}.$$

Let $0 \le k \le n$, and consider $f(k)$. Exactly $\binom{n}{k}$ points in Ω give a total of k heads; each of these points occurs with probability $p^k q^{n-k}$, and so

$$f(k) = \binom{n}{k} p^k q^{n-k} \quad \text{if} \quad 0 \le k \le n.$$

The random variable X is said to have the *binomial distribution* with parameters n and p, written bin(n, p). It is the sum $X = Y_1 + Y_2 + \cdots + Y_n$ of n Bernoulli variables (see Example (2.1.8)). ●

(4) Example. Poisson distribution. If a random variable X takes values in the set $\{0, 1, 2, \dots\}$ with mass function

$$f(k) = \frac{\lambda^k}{k!} e^{-\lambda}, \qquad k = 0, 1, 2, \dots,$$

where $\lambda > 0$, then X is said to have the *Poisson distribution* with parameter λ. ●

Exercises for Section 3.1

1. For what values of the constant C do the following define mass functions on the positive integers $1, 2, \dots$?
(a) Geometric: $f(x) = C2^{-x}$.
(b) Logarithmic: $f(x) = C2^{-x}/x$.
(c) Inverse square: $f(x) = Cx^{-2}$.
(d) 'Modified' Poisson: $f(x) = C2^x/x!$.

2. For a random variable X having (in turn) each of the four mass functions of Exercise (1), find:
(i) $\mathbb{P}(X > 1)$,
(ii) the most probable value of X,
(iii) the probability that X is even.

3. We toss n coins, and each one shows heads with probability p, independently of each of the others. Each coin which shows heads is tossed again. What is the mass function of the number of heads resulting from the second round of tosses?

4. Let S_k be the set of positive integers whose base-10 expansion contains exactly k elements (so that, for example, $1024 \in S_4$). A fair coin is tossed until the first head appears, and we write T for the number of tosses required. We pick a random element, N say, from S_T, each such element having equal probability. What is the mass function of N?

5. Log-convexity. (a) Show that, if X is a binomial or Poisson random variable, then the mass function $f(k) = \mathbb{P}(X = k)$ has the property that $f(k-1)f(k+1) \le f(k)^2$.
(b) Show that, if $f(k) = 90/(\pi k)^4$, $k \ge 1$, then $f(k-1)f(k+1) \ge f(k)^2$.
(c) Find a mass function f such that $f(k)^2 = f(k-1)f(k+1)$, $k \ge 1$.

3.2 Independence

Remember that events A and B are called 'independent' if the occurrence of A does not change the subsequent probability of B occurring. More rigorously, A and B are independent if and only if $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$. Similarly, we say that discrete variables X and Y are 'independent' if the numerical value of X does not affect the distribution of Y. With this in mind we make the following definition.

(1) Definition. Discrete variables X and Y are **independent** if the events $\{X = x\}$ and $\{Y = y\}$ are independent for all x and y.

Suppose X takes values in the set $\{x_1, x_2, \dots\}$ and Y takes values in the set $\{y_1, y_2, \dots\}$. Let

$$A_i = \{X = x_i\}, \qquad B_j = \{Y = y_j\}.$$

Notice (see Problem (2.7.2)) that X and Y are linear combinations of the indicator variables I_{A_i}, I_{B_j}, in that

$$X = \sum_i x_i I_{A_i} \qquad \text{and} \qquad Y = \sum_j y_j I_{B_j}.$$

The random variables X and Y are independent if and only if A_i and B_j are independent for all pairs i, j. A similar definition holds for collections $\{X_1, X_2, \dots, X_n\}$ of discrete variables.

(2) Example. Poisson flips. A coin is tossed once and heads turns up with probability $p = 1 - q$. Let X and Y be the numbers of heads and tails respectively. It is no surprise that X and Y are not independent. After all,

$$\mathbb{P}(X = Y = 1) = 0, \qquad \mathbb{P}(X = 1)\mathbb{P}(Y = 1) = p(1 - p).$$

Suppose now that the coin is tossed a random number N of times, where N has the Poisson distribution with parameter λ. It is a remarkable fact that the resulting numbers X and Y of heads and tails *are* independent, since

$$\begin{aligned}\mathbb{P}(X = x,\ Y = y) &= \mathbb{P}\big(X = x, Y = y \,\big|\, N = x + y\big)\mathbb{P}(N = x + y)\\ &= \binom{x+y}{x} p^x q^y \frac{\lambda^{x+y}}{(x+y)!} e^{-\lambda} = \frac{(\lambda p)^x (\lambda p)^y}{x!\, y!} e^{-\lambda}.\end{aligned}$$

However, by Lemma (1.4.4),

$$\begin{aligned}\mathbb{P}(X = x) &= \sum_{n \geq x} \mathbb{P}(X = x \mid N = n)\mathbb{P}(N = n)\\ &= \sum_{n \geq x} \binom{n}{x} p^x q^{n-x} \frac{\lambda^n}{n!} e^{-\lambda} = \frac{(\lambda p)^x}{x!} e^{-\lambda p};\end{aligned}$$

a similar result holds for Y, and so

$$\mathbb{P}(X = x,\ Y = y) = \mathbb{P}(X = x)\mathbb{P}(Y = y).$$ ●

If X is a random variable and $g : \mathbb{R} \to \mathbb{R}$, then $Z = g(X)$, defined by $Z(\omega) = g(X(\omega))$, is a random variable also. We shall need the following.

(3) Theorem. *If X and Y are independent and $g, h : \mathbb{R} \to \mathbb{R}$, then $g(X)$ and $h(Y)$ are independent also.*

Proof. *Exercise.* See Problem (3.11.1). ■

More generally, we say that a family $\{X_i : i \in I\}$ of (discrete) random variables is *independent* if the events $\{X_i = x_i\}$, $i \in I$, are independent for all possible choices of the set $\{x_i : i \in I\}$ of the values of the X_i. That is to say, $\{X_i : i \in I\}$ is an independent family if and only if

$$\mathbb{P}(X_i = x_i \text{ for all } i \in J) = \prod_{i \in J} \mathbb{P}(X_i = x_i)$$

for all sets $\{x_i : i \in I\}$ and for all finite subsets J of I. The conditional independence of a family of random variables, given an event C, is defined similarly to the conditional independence of events; see equation (1.5.5).

Independent families of random variables are very much easier to study than dependent families, as we shall see soon. Note that pairwise-independent families are not necessarily independent.

Exercises for Section 3.2

1. Let X and Y be independent random variables, each taking the values -1 or 1 with probability $\frac{1}{2}$, and let $Z = XY$. Show that X, Y, and Z are pairwise independent. Are they independent?

2. Let X and Y be independent random variables taking values in the positive integers and having the same mass function $f(x) = 2^{-x}$ for $x = 1, 2, \dots$. Find:
(a) $\mathbb{P}(\min\{X, Y\} \le x)$, (b) $\mathbb{P}(Y > X)$,
(c) $\mathbb{P}(X = Y)$, (d) $\mathbb{P}(X \ge kY)$, for a given positive integer k,
(e) $\mathbb{P}(X \text{ divides } Y)$, (f) $\mathbb{P}(X = rY)$, for a given positive rational r.

3. Let X_1, X_2, X_3 be independent random variables taking values in the positive integers and having mass functions given by $\mathbb{P}(X_i = x) = (1 - p_i)p_i^{x-1}$ for $x = 1, 2, \dots$, and $i = 1, 2, 3$.
(a) Show that

$$\mathbb{P}(X_1 < X_2 < X_3) = \frac{(1 - p_1)(1 - p_2)p_2 p_3^2}{(1 - p_2 p_3)(1 - p_1 p_2 p_3)}.$$

(b) Find $\mathbb{P}(X_1 \le X_2 \le X_3)$.

4. Three players, A, B, and C, take turns to roll a die; they do this in the order ABCABCA....
(a) Show that the probability that, of the three players, A is the first to throw a 6, B the second, and C the third, is 216/1001.
(b) Show that the probability that the first 6 to appear is thrown by A, the second 6 to appear is thrown by B, and the third 6 to appear is thrown by C, is 46656/753571.

5. Let X_r, $1 \le r \le n$, be independent random variables which are symmetric about 0; that is, X_r and $-X_r$ have the same distributions. Show that, for all x, $\mathbb{P}(S_n \ge x) = \mathbb{P}(S_n \le -x)$ where $S_n = \sum_{r=1}^{n} X_r$.
Is the conclusion necessarily true without the assumption of independence?

3.3 Expectation

Let $x_1, x_2, \dots, x_N$ be the numerical outcomes of N repetitions of some experiment. The average of these outcomes is

$$m = \frac{1}{N}\sum_i x_i.$$

In advance of performing these experiments we can represent their outcomes by a sequence $X_1, X_2, \dots, X_N$ of random variables, and we shall suppose that these variables are discrete with a common mass function f. Then, roughly speaking (see the beginning of Section 1.3), for each possible value x, about $Nf(x)$ of the X_i will take that value x. So the average m is about

$$m \simeq \frac{1}{N}\sum_x xNf(x) = \sum_x xf(x)$$

where the summation here is over all possible values of the X_i. This average is called the 'expectation' or 'mean value' of the underlying distribution with mass function f.

(1) Definition. The mean value, or expectation, or expected value of the random variable X with mass function f is defined to be

$$\mathbb{E}(X) = \sum_{x:f(x)>0} xf(x)$$

whenever this sum is absolutely convergent.

We require *absolute* convergence in order that $\mathbb{E}(X)$ be unchanged by reordering the x_i. We can, for notational convenience, write $\mathbb{E}(X) = \sum_x xf(x)$. This appears to be an uncountable sum; however, all but countably many of its contributions are zero. If the numbers $f(x)$ are regarded as masses $f(x)$ at points x then $\mathbb{E}(X)$ is just the position of the centre of gravity; we can speak of X as having an 'atom' or 'point mass' of size $f(x)$ at x. We sometimes omit the parentheses and simply write $\mathbb{E}X$.

(2) Example (2.1.5) revisited. The random variables X and W of this example have mean values

$$\mathbb{E}(X) = \sum_x x\mathbb{P}(X = x) = 0 \cdot \tfrac{1}{4} + 1 \cdot \tfrac{1}{2} + 2 \cdot \tfrac{1}{4} = 1,$$

$$\mathbb{E}(W) = \sum_x x\mathbb{P}(W = x) = 0 \cdot \tfrac{3}{4} + 4 \cdot \tfrac{1}{4} = 1.$$

●

If X is a random variable and $g : \mathbb{R} \to \mathbb{R}$, then $Y = g(X)$, given formally by $Y(\omega) = g(X(\omega))$, is a random variable also. To calculate its expectation we need first to find its probability mass function f_Y. This process can be complicated, and it is avoided by the following lemma (called by some the 'law of the unconscious statistician'!).

(3) Lemma. *If X has mass function f and $g : \mathbb{R} \to \mathbb{R}$, then*

$$\mathbb{E}(g(X)) = \sum_x g(x) f(x)$$

whenever this sum is absolutely convergent.

Proof. This is Problem (3.11.3). ■

(4) Example. Suppose that X takes values $-2, -1, 1, 3$ with probabilities $\frac{1}{4}, \frac{1}{8}, \frac{1}{4}, \frac{3}{8}$ respectively. The random variable $Y = X^2$ takes values $1, 4, 9$ with probabilities $\frac{3}{8}, \frac{1}{4}, \frac{3}{8}$ respectively, and so

$$\mathbb{E}(Y) = \sum_x x\mathbb{P}(Y = x) = 1 \cdot \tfrac{3}{8} + 4 \cdot \tfrac{1}{4} + 9 \cdot \tfrac{3}{8} = \tfrac{19}{4}.$$

Alternatively, use the law of the unconscious statistician to find that

$$\mathbb{E}(Y) = \mathbb{E}(X^2) = \sum_x x^2 \mathbb{P}(X = x) = 4 \cdot \tfrac{1}{4} + 1 \cdot \tfrac{1}{8} + 1 \cdot \tfrac{1}{4} + 9 \cdot \tfrac{3}{8} = \tfrac{19}{4}.$$ ●

Lemma (3) provides a method for calculating the 'moments' of a distribution; these are defined as follows.

(5) Definition. If k is a positive integer, the kth **moment** m_k of X is defined to be $m_k = \mathbb{E}(X^k)$. The kth **central moment** σ_k is $\sigma_k = \mathbb{E}((X - m_1)^k)$.

The two moments of most use are $m_1 = \mathbb{E}(X)$ and $\sigma_2 = \mathbb{E}((X - \mathbb{E}X)^2)$, called the *mean* (or *expectation*) and *variance* of X. These two quantities are measures of the mean and dispersion of X; that is, m_1 is the average value of X, and σ_2 measures the amount by which X tends to deviate from this average. The mean m_1 is often denoted μ, and the variance of X is often denoted $\mathrm{var}(X)$. The positive square root $\sigma = \sqrt{\mathrm{var}(X)}$ is called the *standard deviation*, and in this notation $\sigma_2 = \sigma^2$. The central moments $\{\sigma_i\}$ can be expressed in terms of the ordinary moments $\{m_i\}$. For example, $\sigma_1 = 0$ and

$$\begin{aligned}
\sigma_2 &= \sum_x (x - m_1)^2 f(x) \\
&= \sum_x x^2 f(x) - 2m_1 \sum_x x f(x) + m_1^2 \sum_x f(x) \\
&= m_2 - m_1^2,
\end{aligned}$$

which may be written as

$$\mathrm{var}(X) = \mathbb{E}\big((X - \mathbb{E}X)^2\big) = \mathbb{E}(X^2) - (\mathbb{E}X)^2.$$

Remark. Experience with student calculations of variances causes us to stress the following elementary fact: *variances cannot be negative*. We sometimes omit the parentheses and write simply $\mathrm{var}\, X$. The expression $\mathbb{E}(X)^2$ means $(\mathbb{E}(X))^2$ and must not be confused with $\mathbb{E}(X^2)$.

(6) Example. Bernoulli variables. Let X be a Bernoulli variable, taking the value 1 with probability p $(= 1 - q)$. Then

$$\mathbb{E}(X) = \sum_x x f(x) = 0 \cdot q + 1 \cdot p = p,$$

$$\mathbb{E}(X^2) = \sum_x x^2 f(x) = 0 \cdot q + 1 \cdot p = p,$$

$$\operatorname{var}(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2 = pq.$$

Thus the indicator variable I_A has expectation $\mathbb{P}(A)$ and variance $\mathbb{P}(A)\mathbb{P}(A^c)$. ●

(7) Example. Binomial variables. Let X be $\operatorname{bin}(n, p)$. Then

$$\mathbb{E}(X) = \sum_{k=0}^{n} k f(k) = \sum_{k=0}^{n} k \binom{n}{k} p^k q^{n-k}.$$

To calculate this, differentiate the identity

$$\sum_{k=0}^{n} \binom{n}{k} x^k = (1 + x)^n,$$

multiply by x to obtain

$$\sum_{k=0}^{n} k \binom{n}{k} x^k = nx(1 + x)^{n-1},$$

and substitute $x = p/q$ to obtain $\mathbb{E}(X) = np$. A similar argument shows that the variance of X is given by $\operatorname{var}(X) = npq$. ●

We can think of the process of calculating expectations as a linear operator on the space of random variables.

(8) Theorem. *The expectation operator* $\mathbb{E}$ *has the following properties:*

(a) *if* $X \geq 0$ *then* $\mathbb{E}(X) \geq 0$,

(b) *if* $a, b \in \mathbb{R}$ *then* $\mathbb{E}(aX + bY) = a\mathbb{E}(X) + b\mathbb{E}(Y)$,

(c) *the random variable* 1, *taking the value* 1 *always, has expectation* $\mathbb{E}(1) = 1$.

Proof. (a) and (c) are obvious.

(b) Let $A_x = \{X = x\}$, $B_y = \{Y = y\}$. Then

$$aX + bY = \sum_{x,y} (ax + by) I_{A_x \cap B_y}$$

and the solution of the first part of Problem (3.11.3) shows that

$$\mathbb{E}(aX + bY) = \sum_{x,y} (ax + by) \mathbb{P}(A_x \cap B_y).$$

However,

$$\sum_y \mathbb{P}(A_x \cap B_y) = \mathbb{P}\left(A_x \cap \left(\bigcup_y B_y\right)\right) = \mathbb{P}(A_x \cap \Omega) = \mathbb{P}(A_x)$$

and similarly $\sum_x \mathbb{P}(A_x \cap B_y) = \mathbb{P}(B_y)$, which gives

$$\begin{aligned}\mathbb{E}(aX + bY) &= \sum_x ax \sum_y \mathbb{P}(A_x \cap B_y) + \sum_y by \sum_x \mathbb{P}(A_x \cap B_y)\\ &= a\sum_x x\mathbb{P}(A_x) + b\sum_y y\mathbb{P}(B_y)\\ &= a\mathbb{E}(X) + b\mathbb{E}(Y).\end{aligned}$$

■

Remark. It is not in general true that $\mathbb{E}(XY)$ is the same as $\mathbb{E}(X)\mathbb{E}(Y)$.

(9) Lemma. *If X and Y are independent then* $\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y)$.

Proof. Let A_x and B_y be as in the proof of (8). Then

$$XY = \sum_{x,y} xyI_{A_x \cap B_y}$$

and so

$$\begin{aligned}\mathbb{E}(XY) &= \sum_{x,y} xy\mathbb{P}(A_x)\mathbb{P}(B_y) \qquad \text{by independence}\\ &= \sum_x x\mathbb{P}(A_x)\sum_y y\mathbb{P}(B_y) = \mathbb{E}(X)\mathbb{E}(Y).\end{aligned}$$

■

(10) Definition. X and Y are called **uncorrelated** if $\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y)$.

Lemma (9) asserts that independent variables are uncorrelated. The converse is not true, as Problem (3.11.16) indicates.

(11) Theorem. *For random variables X and Y,*

(a) $\operatorname{var}(aX) = a^2 \operatorname{var}(X)$ *for* $a \in \mathbb{R}$,

(b) $\operatorname{var}(X + Y) = \operatorname{var}(X) + \operatorname{var}(Y)$ *if X and Y are uncorrelated.*

Proof. (a) Using the linearity of $\mathbb{E}$,

$$\begin{aligned}\operatorname{var}(aX) &= \mathbb{E}\{(aX - \mathbb{E}(aX))^2\} = \mathbb{E}\{a^2(X - \mathbb{E}X)^2\}\\ &= a^2\mathbb{E}\{(X - \mathbb{E}X)^2\} = a^2 \operatorname{var}(X).\end{aligned}$$

(b) We have when X and Y are uncorrelated that

$$\begin{aligned}\operatorname{var}(X + Y) &= \mathbb{E}\{\big(X + Y - \mathbb{E}(X + Y)\big)^2\}\\ &= \mathbb{E}\Big[(X - \mathbb{E}X)^2 + 2\big(XY - \mathbb{E}(X)\mathbb{E}(Y)\big) + (Y - \mathbb{E}Y)^2\Big]\\ &= \operatorname{var}(X) + 2\big[\mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y)\big] + \operatorname{var}(Y)\\ &= \operatorname{var}(X) + \operatorname{var}(Y).\end{aligned}$$

■

Theorem (11a) shows that the variance operator 'var' is *not* a linear operator, even when it is applied only to uncorrelated variables.

Sometimes the sum $S = \sum x f(x)$ does not converge absolutely, and the mean of the distribution does not exist. If $S = -\infty$ or $S = +\infty$, then we can sometimes speak of the mean as taking these values also. Of course, there exist distributions which do not have a mean value.

(12) Example. A distribution without a mean. Let X have mass function

$$f(k) = Ak^{-2} \quad \text{for} \quad k = \pm 1, \pm 2, \ldots$$

where A is chosen so that $\sum f(k) = 1$. The sum $\sum_k kf(k) = A\sum_{k\neq 0} k^{-1}$ does not converge absolutely, because both the positive and the negative parts diverge. ●

This is a suitable opportunity to point out that we can base probability theory upon the expectation operator $\mathbb{E}$ rather than upon the probability measure $\mathbb{P}$. After all, our intuitions about the notion of 'average' are probably just as well developed as those about quantitative chance. Roughly speaking, the way we proceed is to postulate axioms, such as (a), (b), and (c) of Theorem (8), for a so-called 'expectation operator' $\mathbb{E}$ acting on a space of 'random variables'. The probability of an event can then be recaptured by defining $\mathbb{P}(A) = \mathbb{E}(I_A)$. Whittle (2000) is an able advocate of this approach.

This method can be easily and naturally adapted to deal with probabilistic questions in quantum theory. In this major branch of theoretical physics, questions arise which cannot be formulated entirely within the usual framework of probability theory. However, there still exists an expectation operator $\mathbb{E}$, which is applied to linear operators known as observables (such as square matrices) rather than to random variables. There does not exist a sample space Ω, and nor therefore are there any indicator functions, but nevertheless there exist analogues of other concepts in probability theory. For example, the *variance* of an operator X is defined by $\mathrm{var}(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2$. Furthermore, it can be shown that $\mathbb{E}(X) = \mathrm{tr}(UX)$ where tr denotes *trace* and U is a non-negative definite operator with unit trace.

(13) Example. Wagers. Historically, there has been confusion amongst probabilists between the price that an individual may be willing to pay in order to play a game, and her expected return from this game. For example, I conceal £2 in one hand and nothing in the other, and then invite a friend to pay a fee which entitles her to choose a hand at random and keep the contents. Other things being equal (my friend is neither a compulsive gambler, nor particularly busy), it would seem that £1 would be a 'fair' fee to ask, since £1 is the expected return to the player. That is to say, faced with a modest (but random) gain, then a fair 'entrance fee' would seem to be the expected value of the gain. However, suppose that I conceal $£2^{10}$ in one hand and nothing in the other; what now is a 'fair' fee? Few persons of modest means can be expected to offer $£2^9$ for the privilege of playing. There is confusion here between fairness and reasonableness: we do not generally treat large payoffs or penalties in the same way as small ones, even though the relative odds may be unquestionable. The customary resolution of this paradox is to introduce the notion of 'utility'. Writing $u(x)$ for the 'utility' to an individual of $£x$, it would be fairer to charge a fee of $\frac{1}{2}(u(0) + u(2^{10}))$ for the above prospect. Of course, different individuals have different utility functions, although such functions have presumably various features in common: $u(0) = 0$, u is non-decreasing, $u(x)$ is near to x for small positive x, and u is concave, so that in particular $u(x) \leq xu(1)$ when $x \geq 1$.

The use of expectation to assess a 'fair fee' may be convenient but is sometimes inappropriate. For example, a more suitable criterion in the finance market would be absence of arbitrage; see Exercise (3.3.7) and Section 13.10. And, in a rather general model of financial markets, there is a criterion commonly expressed as 'no free lunch with vanishing risk'. ●

Exercises for Section 3.3

1. Is it generally true that $\mathbb{E}(1/X) = 1/\mathbb{E}(X)$? Is it ever true that $\mathbb{E}(1/X) = 1/\mathbb{E}(X)$?

2. Coupons. Every package of some intrinsically dull commodity includes a small and exciting plastic object. There are c different types of object, and each package is equally likely to contain any given type. You buy one package each day.
(a) Find the mean number of days which elapse between the acquisitions of the jth new type of object and the $(j+1)$th new type.
(b) Find the mean number of days which elapse before you have a full set of objects.

3. Each member of a group of n players rolls a die.
(a) For any pair of players who throw the same number, the group scores 1 point. Find the mean and variance of the total score of the group.
(b) Find the mean and variance of the total score if any pair of players who throw the same number scores that number.

4. St Petersburg paradox†. A fair coin is tossed repeatedly. Let T be the number of tosses until the first head. You are offered the following prospect, which you may accept on payment of a fee. If $T = k$, say, then you will receive £2^k. What would be a 'fair' fee to ask of you?

5. Let X have mass function

$$f(x) = \begin{cases} \{x(x+1)\}^{-1} & \text{if } x = 1, 2, \dots, \\ 0 & \text{otherwise,} \end{cases}$$

and let $\alpha \in \mathbb{R}$. For what values of α is it the case‡ that $\mathbb{E}(X^\alpha) < \infty$?

6. Show that $\mathrm{var}(a + X) = \mathrm{var}(X)$ for any random variable X and constant a.

7. Arbitrage. Suppose you find a warm-hearted bookmaker offering payoff odds of $\pi(k)$ against the kth horse in an n-horse race where $\sum_{k=1}^{n}\{\pi(k)+1\}^{-1} < 1$. Show that you can distribute your bets in such a way as to ensure you win.

8. You roll a conventional fair die repeatedly. If it shows 1, you must stop, but you may choose to stop at any prior time. Your score is the number shown by the die on the final roll. What stopping strategy yields the greatest expected score? What strategy would you use if your score were the square of the final roll?

9. Continuing with Exercise (8), suppose now that you lose c points from your score each time you roll the die. What strategy maximizes the expected final score if $c = \frac{1}{3}$? What is the best strategy if $c = 1$?

---

†This problem was mentioned by Nicholas Bernoulli in 1713, and Daniel Bernoulli wrote about the question for the Academy of St Petersburg.

‡If α is not integral, than $\mathbb{E}(X^\alpha)$ is called the *fractional moment of order* α of X. A point concerning notation: for real α and complex $x = re^{i\theta}$, x^α should be interpreted as $r^\alpha e^{i\theta\alpha}$, so that $|x^\alpha| = r^\alpha$. In particular, $\mathbb{E}(|X^\alpha|) = \mathbb{E}(|X|^\alpha)$.

3.4 Indicators and matching

This section contains light entertainment, in the guise of some illustrations of the uses of indicator functions. These were defined in Example (2.1.9) and have appeared occasionally since. Recall that

$$I_A(\omega) = \begin{cases} 1 & \text{if } \omega \in A, \\ 0 & \text{if } \omega \in A^c, \end{cases}$$

and $\mathbb{E}I_A = \mathbb{P}(A)$.

(1) Example. Proofs of Lemma (1.3.4c, d). Note that

$$I_A + I_{A^c} = I_{A \cup A^c} = I_\Omega = 1$$

and that $I_{A \cap B} = I_A I_B$. Thus

$$\begin{aligned} I_{A \cup B} &= 1 - I_{(A \cup B)^c} = 1 - I_{A^c \cap B^c} \\ &= 1 - I_{A^c} I_{B^c} = 1 - (1 - I_A)(1 - I_B) \\ &= I_A + I_B - I_A I_B. \end{aligned}$$

Take expectations to obtain

$$\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B).$$

More generally, if $B = \bigcup_{i=1}^n A_i$ then

$$I_B = 1 - \prod_{i=1}^n (1 - I_{A_i});$$

multiply this out and take expectations to obtain

(2) $$\mathbb{P}\left(\bigcup_{i=1}^n A_i\right) = \sum_i \mathbb{P}(A_i) - \sum_{i<j} \mathbb{P}(A_i \cap A_j) + \cdots + (-1)^{n+1} \mathbb{P}(A_1 \cap \cdots \cap A_n).$$

This very useful identity is known as the *inclusion–exclusion formula*. ●

(3) Example. Matching problem. A number of melodramatic applications of (2) are available, of which the following is typical. A secretary types n different letters together with matching envelopes, drops the pile down the stairs, and then places the letters randomly in the envelopes. Each arrangement is equally likely, and we ask for the probability that exactly r are in their correct envelopes. Rather than using (2), we shall proceed directly by way of indicator functions. (Another approach is presented in Exercise (3.4.9).)

Solution. Let $L_1, L_2, \ldots, L_n$ denote the letters. Call a letter *good* if it is correctly addressed and *bad* otherwise; write X for the number of good letters. Let A_i be the event that L_i is good, and let I_i be the indicator function of A_i. Let $j_1, \ldots, j_r, k_{r+1}, \ldots, k_n$ be a permutation of the numbers $1, 2, \ldots, n$, and define

(4) $$S = \sum_\pi I_{j_1} \cdots I_{j_r} (1 - I_{k_{r+1}}) \cdots (1 - I_{k_n})$$

where the sum is taken over all such permutations π. Then

$$S = \begin{cases} 0 & \text{if } X \neq r, \\ r!\,(n-r)! & \text{if } X = r. \end{cases}$$

To see this, let $L_{i_1}, \dots, L_{i_m}$ be the good letters. If $m \neq r$ then each summand in (4) equals 0. If $m = r$ then the summand in (4) equals 1 if and only if $j_1, \dots, j_r$ is a permutation of $i_1, \dots, i_r$ and $k_{r+1}, \dots, k_n$ is a permutation of the remaining numbers; there are $r!\,(n-r)!$ such pairs of permutations. It follows that I, given by

$$(5) \qquad I = \frac{1}{r!\,(n-r)!} S,$$

is the indicator function of the event $\{X = r\}$ that exactly r letters are good. We take expectations of (4) and multiply out to obtain

$$\mathbb{E}(S) = \sum_{\pi} \sum_{s=0}^{n-r} (-1)^s \binom{n-r}{s} \mathbb{E}(I_{j_1} \cdots I_{j_r} I_{k_{r+1}} \cdots I_{k_{r+s}})$$

by a symmetry argument. However,

$$(6) \qquad \mathbb{E}(I_{j_1} \cdots I_{j_r} I_{k_{r+1}} \cdots I_{k_{r+s}}) = \frac{(n-r-s)!}{n!}$$

since there are $n!$ possible permutations, only $(n-r-s)!$ of which allocate $L_{i_1}, \dots, L_{j_r}$, $L_{k_{r+1}}, \dots, L_{k_{r+s}}$ to their correct envelopes. We combine (4), (5), and (6) to obtain

$$\begin{aligned}
\mathbb{P}(X = r) = \mathbb{E}(I) &= \frac{1}{r!\,(n-r)!} \mathbb{E}(S) \\
&= \frac{1}{r!\,(n-r)!} \sum_{s=0}^{n-r} (-1)^s \binom{n-r}{s} n! \frac{(n-r-s)!}{n!} \\
&= \frac{1}{r!} \sum_{s=0}^{n-r} (-1)^s \frac{1}{s!} \\
&= \frac{1}{r!} \left(\frac{1}{2!} - \frac{1}{3!} + \cdots + \frac{(-1)^{n-r}}{(n-r)!} \right) \qquad \text{for } r \leq n-2 \text{ and } n \geq 2.
\end{aligned}$$

In particular, as the number n of letters tends to infinity, we obtain the possibly surprising result that the probability that no letter is put into its correct envelope approaches e^{-1}. It is left as an *exercise* to prove this without using indicators. ●

(7) Example. Reliability. When you telephone your friend in Cambridge, your call is routed through the telephone network in a way which depends on the current state of the traffic. For example, if all lines into the Ascot switchboard are in use, then your call may go through the switchboard at Newmarket. Sometimes you may fail to get through at all, owing to a combination of faulty and occupied equipment in the system. We may think of the network as comprising nodes joined by edges, drawn as 'graphs' in the manner of the examples of

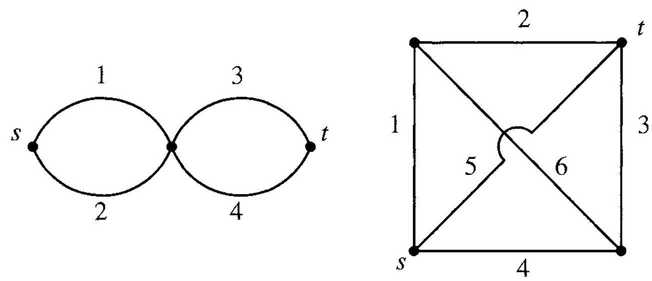

Figure 3.1. Two networks with source s and sink t.

Figure 3.1. In each of these examples there is a designated 'source' s and 'sink' t, and we wish to find a path through the network from s to t which uses available channels. As a simple model for such a system in the presence of uncertainty, we suppose that each edge e is 'working' with probability p_e, independently of all other edges. We write $\mathbf{p}$ for the vector of edge probabilities p_e, and define the *reliability* $R(\mathbf{p})$ of the network to be the probability that there is a path from s to t using only edges which are working. Denoting the network by G, we write $R_G(\mathbf{p})$ for $R(\mathbf{p})$ when we wish to emphasize the role of G.

We have encountered questions of reliability already. In Example (1.7.2) we were asked for the reliability of the first network in Figure 3.1 and in Problem (1.8.19) of the second, assuming on each occasion that the value of p_e does not depend on the choice of e.

Let us write

$$X_e = \begin{cases} 1 & \text{if edge } e \text{ is working,} \\ 0 & \text{otherwise,} \end{cases}$$

the indicator function of the event that e is working, so that X_e takes the values 0 and 1 with probabilities $1 - p_e$ and p_e respectively. Each realization X of the X_e either includes a working connection from s to t or does not. Thus, there exists a *structure function* ζ taking values 0 and 1 such that

(8)
$$\zeta(X) = \begin{cases} 1 & \text{if such a working connection exists,} \\ 0 & \text{otherwise;} \end{cases}$$

thus $\zeta(X)$ is the indicator function of the event that a working connection exists. It is immediately seen that $R(\mathbf{p}) = \mathbb{E}(\zeta(X))$. The function ζ may be expressed as

(9)
$$\zeta(X) = 1 - \prod_{\pi} I_{\{\pi \text{ not working}\}} = 1 - \prod_{\pi}\left(1 - \prod_{e \in \pi} X_e\right)$$

where π is a typical path in G from s to t, and we say that π is working if and only if every edge in π is working.

For instance, in the case of the first example of Figure 3.1, there are four different paths from s to t. Numbering the edges as indicated, we have that the structure function is given by

(10)
$$\zeta(X) = 1 - (1 - X_1X_3)(1 - X_1X_4)(1 - X_2X_3)(1 - X_2X_4).$$

As an *exercise*, expand this and take expectations to calculate the reliability of the network when $p_e = p$ for all edges e. ●

(11) Example. The probabilistic method†. Probability may be used to derive non-trivial results not involving probability. Here is an example. There are 17 fenceposts around the perimeter of a field, exactly 5 of which are rotten. Show that, irrespective of which these 5 are, there necessarily exists a run of 7 consecutive posts at least 3 of which are rotten.

Our solution involves probability. We label the posts $1, 2, \dots, 17$, and let I_k be the indicator function that post k is rotten. Let R_k be the number of rotten posts amongst those labelled $k+1, k+2, \dots, k+7$, all taken modulo 17. We now pick a random post labelled K, each being equally likely. We have that

$$\mathbb{E}(R_K) = \sum_{k=1}^{17} \frac{1}{17}(I_{k+1} + I_{k+2} + \cdots + I_{k+7}) = \sum_{j=1}^{17} \frac{7}{17} I_j = \frac{7}{17} \cdot 5.$$

Now $\frac{35}{17} > 2$, implying that $\mathbb{P}(R_K > 2) > 0$. Since R_K is integer valued, it must be the case that $\mathbb{P}(R_K \geq 3) > 0$, implying that $R_k \geq 3$ for some k. ●

Exercises for Section 3.4

1. A biased coin is tossed n times, and heads shows with probability p on each toss. A *run* is a sequence of throws which result in the same outcome, so that, for example, the sequence HHTHTTH contains five runs. Show that the expected number of runs is $1 + 2(n-1)p(1-p)$. Find the variance of the number of runs.

2. An urn contains n balls numbered $1, 2, \dots, n$. We remove k balls at random (without replacement) and add up their numbers. Find the mean and variance of the total.

3. Of the $2n$ people in a given collection of n couples, exactly m die. Assuming that the m have been picked at random, find the mean number of surviving couples. This problem was formulated by Daniel Bernoulli in 1768.

4. Urn R contains n red balls and urn B contains n blue balls. At each stage, a ball is selected at random from each urn, and they are swapped. Show that the mean number of red balls in urn R after stage k is $\frac{1}{2}n\{1 + (1 - 2/n)^k\}$. This 'diffusion model' was described by Daniel Bernoulli in 1769.

5. Consider a square with diagonals, with distinct source and sink. Each edge represents a component which is working correctly with probability p, independently of all other components. Write down an expression for the Boolean function which equals 1 if and only if there is a working path from source to sink, in terms of the indicator functions X_i of the events {edge i is working} as i runs over the set of edges. Hence calculate the reliability of the network.

6. A system is called a 'k out of n' system if it contains n components and it works whenever k or more of these components are working. Suppose that each component is working with probability p, independently of the other components, and let X_c be the indicator function of the event that component c is working. Find, in terms of the X_c, the indicator function of the event that the system works, and deduce the reliability of the system.

7. **The probabilistic method.** Let $G = (V, E)$ be a finite graph. For any set W of vertices and any edge $e \in E$, define the indicator function

$$I_W(e) = \begin{cases} 1 & \text{if } e \text{ connects } W \text{ and } W^c, \\ 0 & \text{otherwise.} \end{cases}$$

Set $N_W = \sum_{e \in E} I_W(e)$. Show that there exists $W \subseteq V$ such that $N_W \geq \frac{1}{2}|E|$.

†Generally credited to Erdős.

8. A total of n bar magnets are placed end to end in a line with random independent orientations. Adjacent like poles repel, ends with opposite polarities join to form blocks. Let X be the number of blocks of joined magnets. Find $\mathbb{E}(X)$ and $\mathrm{var}(X)$.

9. Matching. (a) Use the inclusion–exclusion formula (3.4.2) to derive the result of Example (3.4.3), namely: in a random permutation of the first n integers, the probability that exactly r retain their original positions is

$$\frac{1}{r!}\left(\frac{1}{2!}-\frac{1}{3!}+\cdots+\frac{(-1)^{n-r}}{(n-r)!}\right).$$

(b) Let d_n be the number of derangements of the first n integers (that is, rearrangements with no integers in their original positions). Show that $d_{n+1} = nd_n + nd_{n-1}$ for $n \geq 2$. Deduce the result of part (a).

3.5 Examples of discrete variables

(1) Bernoulli trials. A random variable X takes values 1 and 0 with probabilities p and q $(= 1 - p)$, respectively. Sometimes we think of these values as representing the 'success' or the 'failure' of a trial. The mass function is

$$f(0) = 1 - p, \qquad f(1) = p,$$

and it follows that $\mathbb{E}X = p$ and $\mathrm{var}(X) = p(1-p)$. ●

(2) Binomial distribution. We perform n independent Bernoulli trials $X_1, X_2, \ldots, X_n$ and count the total number of successes $Y = X_1 + X_2 + \cdots + X_n$. As in Example (3.1.3), the mass function of Y is

$$f(k) = \binom{n}{k} p^k (1-p)^{n-k}, \qquad k = 0, 1, \ldots, n.$$

Application of Theorems (3.3.8) and (3.3.11) yields immediately

$$\mathbb{E}Y = np, \qquad \mathrm{var}(Y) = np(1-p);$$

the method of Example (3.3.7) provides a more lengthy derivation of this. ●

(3) Trinomial distribution. More generally, suppose we conduct n trials, each of which results in one of three outcomes (red, white, or blue, say), where red occurs with probability p, white with probability q, and blue with probability $1 - p - q$. The probability of r reds, w whites, and $n - r - w$ blues is

$$\frac{n!}{r!\,w!\,(n-r-w)!} p^r q^w (1-p-q)^{n-r-w}.$$

This is the *trinomial distribution*, with parameters n, p, and q. The 'multinomial distribution' is the obvious generalization of this distribution to the case of some number, say t, of possible outcomes. ●

(4) Poisson distribution. A *Poisson* variable is a random variable with the Poisson mass function

$$f(k) = \frac{\lambda^k}{k!}e^{-\lambda}, \qquad k = 0, 1, 2, \dots$$

for some $\lambda > 0$. It can be obtained in practice in the following way. Let Y be a $\text{bin}(n, p)$ variable, and suppose that n is very large and p is very small (an example might be the number Y of misprints on the front page of the *Grauniad*, where n is the total number of characters and p is the probability for each character that the typesetter has made an error). Now, let $n \to \infty$ and $p \to 0$ in such a way that $\mathbb{E}(Y) = np$ approaches a non-zero constant λ. Then, for $k = 0, 1, 2, \dots$,

$$\mathbb{P}(Y = k) = \binom{n}{k} p^k (1-p)^{n-k} \sim \frac{1}{k!}\left(\frac{np}{1-p}\right)^k (1-p)^n \to \frac{\lambda^k}{k!}e^{-\lambda}.$$

Check that both the mean and the variance of this distribution are equal to λ. Now do Problem (2.7.7) again (*exercise*). ●

(5) Geometric distribution. A *geometric* variable is a random variable with the geometric mass function

$$f(k) = p(1-p)^{k-1}, \qquad k = 1, 2, \dots$$

for some number p in $(0, 1)$. This distribution arises in the following way. Suppose that independent Bernoulli trials (parameter p) are performed at times $1, 2, \dots$. Let W be the time which elapses before the first success; W is called a *waiting time*. Then $\mathbb{P}(W > k) = (1-p)^k$ and thus

$$\mathbb{P}(W = k) = \mathbb{P}(W > k-1) - \mathbb{P}(W > k) = p(1-p)^{k-1}.$$

The reader should check, preferably at this point, that the mean and variance are p^{-1} and $(1-p)p^{-2}$ respectively. ●

(6) Negative binomial distribution. More generally, in the previous example, let W_r be the waiting time for the rth success. Check that W_r has mass function

$$\mathbb{P}(W_r = k) = \binom{k-1}{r-1} p^r (1-p)^{k-r}, \qquad k = r, r+1, \dots;$$

it is said to have the *negative binomial distribution* with parameters r and p. The random variable W_r is the sum of r independent geometric variables. To see this, let X_1 be the waiting time for the first success, X_2 the *further* waiting time for the second success, X_3 the *further* waiting time for the third success, and so on. Then $X_1, X_2, \dots$ are independent and geometric, and

$$W_r = X_1 + X_2 + \cdots + X_r.$$

Apply Theorems (3.3.8) and (3.3.11) to find the mean and the variance of W_r. ●

Exercises for Section 3.5

1. **De Moivre trials.** Each trial may result in any of t given outcomes, the ith outcome having probability p_i. Let N_i be the number of occurrences of the ith outcome in n independent trials. Show that

$$\mathbb{P}(N_i = n_i \text{ for } 1 \le i \le t) = \frac{n!}{n_1!\, n_2! \cdots n_t!} p_1^{n_1} p_2^{n_2} \cdots p_t^{n_t}$$

for any collection $n_1, n_2, \ldots, n_t$ of non-negative integers with sum n. The vector N is said to have the *multinomial distribution.*

2. In your pocket is a random number N of coins, where N has the Poisson distribution with parameter λ. You toss each coin once, with heads showing with probability p each time. Show that the total number of heads has the Poisson distribution with parameter λp.

3. Let X be Poisson distributed where $\mathbb{P}(X = n) = p_n(\lambda) = \lambda^n e^{-\lambda}/n!$ for $n \ge 0$. Show that $\mathbb{P}(X \le n) = 1 - \int_0^\lambda p_n(x)\, dx$.

4. **Capture–recapture.** A population of b animals has had a number a of its members captured, marked, and released. Let X be the number of animals it is necessary to recapture (without re-release) in order to obtain m marked animals. Show that

$$\mathbb{P}(X = n) = \frac{a}{b}\binom{a-1}{m-1}\binom{b-a}{n-m} \bigg/ \binom{b-1}{n-1},$$

and find $\mathbb{E}X$. This distribution has been called *negative hypergeometric.*

3.6 Dependence

Probability theory is largely concerned with families of random variables; these families will not in general consist entirely of independent variables.

(1) Example. Suppose that we back three horses to win as an accumulator. If our stake is £1 and the starting prices are α, β, and γ, then our total profit is

$$W = (\alpha + 1)(\beta + 1)(\gamma + 1) I_1 I_2 I_3 - 1$$

where I_i denotes the indicator of a win in the ith race by our horse. (In checking this expression remember that a bet of £B on a horse with starting price α brings a return of £$B(\alpha + 1)$, should this horse win.) We lose £1 if some backed horse fails to win. It seems clear that the random variables W and I_1 are *not* independent. If the races are run independently, then

$$\mathbb{P}(W = -1) = \mathbb{P}(I_1 I_2 I_3 = 0),$$

but

$$\mathbb{P}(W = -1 \mid I_1 = 1) = \mathbb{P}(I_2 I_3 = 0)$$

which are different from each other unless the first backed horse is guaranteed victory. ●

We require a tool for studying collections of dependent variables. Knowledge of their individual mass functions is little help by itself. Just as the main tools for studying a random variable is its distribution function, so the study of, say, a pair of random variables is based on its 'joint' distribution function and mass function.

(2) Definition. The **joint distribution function** $F : \mathbb{R}^2 \to [0, 1]$ of X and Y, where X and Y are discrete variables, is given by

$$F(x, y) = \mathbb{P}(X \le x \text{ and } Y \le y).$$

Their **joint mass function** $f : \mathbb{R}^2 \to [0, 1]$ is given by

$$f(x, y) = \mathbb{P}(X = x \text{ and } Y = y).$$

Joint distribution functions and joint mass functions of larger collections of variables are defined similarly. The functions F and f can be characterized in much the same way (Lemmas (2.1.6) and (3.1.2)) as the corresponding functions of a single variable.We omit the details. We write $F_{X,Y}$ and $f_{X,Y}$ when we need to stress the role of X and Y. You may think of the joint mass function in the following way. If $A_x = \{X = x\}$ and $B_y = \{Y = y\}$, then

$$f(x, y) = \mathbb{P}(A_x \cap B_y).$$

The definition of independence can now be reformulated in a lemma.

(3) Lemma. *The discrete random variables X and Y are independent if and only if*

$$f_{X,Y}(x, y) = f_X(x) f_Y(y) \quad \textit{for all } x, y \in \mathbb{R}. \tag{4}$$

More generally, X and Y are independent if and only if $f_{X,Y}(x, y)$ can be factorized as the product $g(x)h(y)$ of a function of x alone and a function of y alone.

Proof. This is Problem (3.11.1). ∎

Suppose that X and Y have joint mass function $f_{X,Y}$ and we wish to check whether or not (4) holds. First we need to calculate the *marginal mass functions* f_X and f_Y from our knowledge of $f_{X,Y}$. These are found in the following way:

$$\begin{aligned} f_X(x) = \mathbb{P}(X = x) &= \mathbb{P}\Big(\bigcup_y (\{X = x\} \cap \{Y = y\})\Big) \\ &= \sum_y \mathbb{P}(X = x, Y = y) = \sum_y f_{X,Y}(x, y), \end{aligned}$$

and similarly $f_Y(y) = \sum_x f_{X,Y}(x, y)$. Having found the marginals, it is a trivial matter to see whether (4) holds or not.

Remark. We stress that the factorization (4) must hold for *all* x and y in order that X and Y be independent.

(5) Example. Calculation of marginals. In Example (3.2.2) we encountered a pair X, Y of variables with a joint mass function

$$f(x, y) = \frac{\alpha^x \beta^y}{x!\, y!} e^{-\alpha-\beta} \quad \text{for} \quad x, y = 0, 1, 2, \ldots$$

where $\alpha, \beta > 0$. The marginal mass function of X is

$$f_X(x) = \sum_y f(x, y) = \frac{\alpha^x}{x!}e^{-\alpha}\sum_{y=0}^{\infty}\frac{\beta^y}{y!}e^{-\beta} = \frac{\alpha^x}{x!}e^{-\alpha}$$

and so X has the Poisson distribution with parameter α. Similarly Y has the Poisson distribution with parameter β. It is easy to check that (4) holds, whence X and Y are independent. ●

For any discrete pair X, Y, a real function $g(X, Y)$ is a random variable. We shall often need to find its expectation. To avoid explicit calculation of its mass function, we shall use the following more general form of the law of the unconscious statistician, Lemma (3.3.3).

(6) Lemma. $\mathbb{E}(g(X, Y)) = \sum_{x,y} g(x, y) f_{X,Y}(x, y)$.

Proof. As for Lemma (3.3.3). ■

For example, $\mathbb{E}(XY) = \sum_{x,y} xy f_{X,Y}(x, y)$. This formula is particularly useful to statisticians who may need to find simple ways of explaining dependence to laymen. For instance, suppose that the government wishes to announce that the dependence between defence spending and the cost of living is very small. It should *not* publish an estimate of the joint mass function unless its object is obfuscation alone. Most members of the public would prefer to find that this dependence can be represented in terms of a single number on a prescribed scale. Towards this end we make the following definition†.

(7) Definition. The **covariance** of X and Y is

$$\text{cov}(X, Y) = \mathbb{E}\big[(X - \mathbb{E}X)(Y - \mathbb{E}Y)\big].$$

The **correlation (coefficient)** of X and Y is

$$\rho(X, Y) = \frac{\text{cov}(X, Y)}{\sqrt{\text{var}(X) \cdot \text{var}(Y)}}$$

as long as the variances are non-zero.

Note that the concept of covariance generalizes that of variance in that $\text{cov}(X, X) = \text{var}(X)$. Expanding the covariance gives

$$\text{cov}(X, Y) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y).$$

Remember, Definition (3.3.10), that X and Y are called *uncorrelated* if $\text{cov}(X, Y) = 0$. Also, independent variables are always uncorrelated, although the converse is not true. Covariance itself is not a satisfactory measure of dependence because the scale of values which $\text{cov}(X, Y)$ may take contains no points which are clearly interpretable in terms of the relationship between X and Y. The following lemma shows that this is not the case for correlations.

(8) Lemma. *The correlation coefficient* ρ *satisfies* $|\rho(X, Y)| \leq 1$ *with equality if and only if* $\mathbb{P}(aX + bY = c) = 1$ *for some* $a, b, c \in \mathbb{R}$.

---

†The concepts and terminology in this definition were formulated by Francis Galton in the late 1880s.

The proof is an application of the following important inequality.

(9) Theorem. Cauchy–Schwarz inequality. *For random variables X and Y,*

$$\{\mathbb{E}(XY)\}^2 \le \mathbb{E}(X^2)\mathbb{E}(Y^2)$$

with equality if and only if $\mathbb{P}(aX = bY) = 1$ for some real a and b, at least one of which is non-zero.

Proof. We can assume that $\mathbb{E}(X^2)$ and $\mathbb{E}(Y^2)$ are strictly positive, since otherwise the result follows immediately from Problem (3.11.2). For $a, b \in \mathbb{R}$, let $Z = aX - bY$. Then

$$0 \le \mathbb{E}(Z^2) = a^2\mathbb{E}(X^2) - 2ab\mathbb{E}(XY) + b^2\mathbb{E}(Y^2).$$

Thus the right-hand side is a quadratic in the variable a with at most one real root. Its discriminant must be non-positive. That is to say, if $b \neq 0$,

$$\mathbb{E}(XY)^2 - \mathbb{E}(X^2)\mathbb{E}(Y^2) \le 0.$$

The discriminant is zero if and only if the quadratic has a real root. This occurs if and only if

$$\mathbb{E}\big((aX - bY)^2\big) = 0 \quad \text{for some } a \text{ and } b,$$

which, by Problem (3.11.2), completes the proof. ■

Proof of (8). Apply (9) to the variables $X - \mathbb{E}X$ and $Y - \mathbb{E}Y$. ■

A more careful treatment than this proof shows that $\rho = +1$ if and only if Y *increases* linearly with X and $\rho = -1$ if and only if Y *decreases* linearly as X increases.

(10) Example. Here is a tedious numerical example of the use of joint mass functions. Let X and Y take values in $\{1, 2, 3\}$ and $\{-1, 0, 2\}$ respectively, with joint mass function f where $f(x, y)$ is the appropriate entry in Table 3.1.

	$y=-1$	$y=0$	$y=2$	f_X
$x=1$	$\frac{1}{18}$	$\frac{3}{18}$	$\frac{2}{18}$	$\frac{6}{18}$
$x=2$	$\frac{2}{18}$	0	$\frac{3}{18}$	$\frac{5}{18}$
$x=3$	0	$\frac{4}{18}$	$\frac{3}{18}$	$\frac{7}{18}$
f_Y	$\frac{3}{18}$	$\frac{7}{18}$	$\frac{8}{18}$	

Table 3.1. The joint mass function of the random variables X and Y. The indicated row and column sums are the marginal mass functions f_X and f_Y.

A quick calculation gives

$$\mathbb{E}(XY) = \sum_{x,y} xyf(x, y) = 29/18,$$

$$\mathbb{E}(X) = \sum_{x} xf_X(x) = 37/18, \quad \mathbb{E}(Y) = 13/18,$$

$$\text{var}(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2 = 233/324, \quad \text{var}(Y) = 461/324,$$

$$\text{cov}(X, Y) = 41/324, \quad \rho(X, Y) = 41/\sqrt{107413}.$$ ●

Exercises for Section 3.6

1. Show that the collection of random variables on a given probability space and having finite variance forms a vector space over the reals.

2. Find the marginal mass functions of the multinomial distribution of Exercise (3.5.1).

3. Let X and Y be discrete random variables with joint mass function

$$f(x, y) = \frac{C}{(x + y - 1)(x + y)(x + y + 1)}, \qquad x, y = 1, 2, 3, \dots.$$

Find the marginal mass functions of X and Y, calculate C, and also the covariance of X and Y.

4. Let X and Y be discrete random variables with mean 0, variance 1, and covariance ρ. Show that $\mathbb{E}(\max\{X^2, Y^2\}) \leq 1 + \sqrt{1 - \rho^2}$.

5. Mutual information. Let X and Y be discrete random variables with joint mass function f.

(a) Show that $\mathbb{E}(\log f_X(X)) \geq \mathbb{E}(\log f_Y(X))$.

(b) Show that the *mutual information*

$$I = \mathbb{E}\left(\log\left\{\frac{f(X, Y)}{f_X(X) f_Y(Y)}\right\}\right)$$

satisfies $I \geq 0$, with equality if and only if X and Y are independent.

6. Voter paradox. Let X, Y, Z be discrete random variables with the property that their values are distinct with probability 1. Let $a = \mathbb{P}(X > Y)$, $b = \mathbb{P}(Y > Z)$, $c = \mathbb{P}(Z > X)$.

(a) Show that $\min\{a, b, c\} \leq \frac{2}{3}$, and give an example where this bound is attained.

(b) Show that, if X, Y, Z are independent and identically distributed, then $a = b = c = \frac{1}{2}$.

(c) Find $\min\{a, b, c\}$ and $\sup_p \min\{a, b, c\}$ when $\mathbb{P}(X = 0) = 1$, and Y, Z are independent with $\mathbb{P}(Z = 1) = \mathbb{P}(Y = -1) = p$, $\mathbb{P}(Z = -2) = \mathbb{P}(Y = 2) = 1 - p$. Here, $\sup_p$ denotes the supremum as p varies over $[0, 1]$.

[Part (a) is related to the observation that, in an election, it is possible for more than half of the voters to prefer candidate A to candidate B, more than half B to C, and more than half C to A.]

7. Benford's distribution, or the law of anomalous numbers. If one picks a numerical entry at random from an almanac, or the annual accounts of a corporation, the first two significant digits, X, Y, are found to have approximately the joint mass function

$$f(x, y) = \log_{10}\left(1 + \frac{1}{10x + y}\right), \qquad 1 \leq x \leq 9,\ 0 \leq y \leq 9.$$

Find the mass function of X and an approximation to its mean. [A heuristic explanation for this phenomenon may be found in the second of Feller's volumes (1971).]

8. Let X and Y have joint mass function

$$f(j, k) = \frac{c(j + k)a^{j+k}}{j!\,k!}, \qquad j, k \geq 0,$$

where a is a constant. Find c, $\mathbb{P}(X = j)$, $\mathbb{P}(X + Y = r)$, and $\mathbb{E}(X)$.

3.7 Conditional distributions and conditional expectation

In Section 1.4 we discussed the conditional probability $\mathbb{P}(B \mid A)$. This may be set in the more general context of the conditional distribution of one variable Y given the value of another variable X; this reduces to the definition of the conditional probabilities of events A and B if $X = I_A$ and $Y = I_B$.

Let X and Y be two discrete variables on $(\Omega, \mathcal{F}, \mathbb{P})$.

(1) Definition. The **conditional distribution function** of Y given $X = x$, written $F_{Y|X}(\cdot \mid x)$, is defined by

$$F_{Y|X}(y \mid x) = \mathbb{P}(Y \le y \mid X = x)$$

for any x such that $\mathbb{P}(X = x) > 0$. The **conditional (probability) mass function** of Y given $X = x$, written $f_{Y|X}(\cdot \mid x)$, is defined by

(2) $$f_{Y|X}(y \mid x) = \mathbb{P}(Y = y \mid X = x)$$

for any x such that $\mathbb{P}(X = x) > 0$.

Formula (2) is easy to remember as $f_{Y|X} = f_{X,Y}/f_X$. Conditional distributions and mass functions are undefined at values of x for which $\mathbb{P}(X = x) = 0$. Clearly X and Y are independent if and only if $f_{Y|X} = f_Y$.

Suppose we are told that $X = x$. Conditional upon this, the new distribution of Y has mass function $f_{Y|X}(y \mid x)$, which we think of as a function of y. The expected value of this distribution, $\sum_y y f_{Y|X}(y \mid x)$, is called the *conditional expectation* of Y given $X = x$ and is written $\psi(x) = \mathbb{E}(Y \mid X = x)$. Now, we observe that the conditional expectation depends on the value x taken by X, and can be thought of as a function $\psi(X)$ of X itself.

(3) Definition. Let $\psi(x) = \mathbb{E}(Y \mid X = x)$. Then $\psi(X)$ is called the **conditional expectation** of Y given X, written as $\mathbb{E}(Y \mid X)$.

Although 'conditional expectation' sounds like a number, it is actually a random variable. It has the following important property.

(4) Theorem. *The conditional expectation $\psi(X) = \mathbb{E}(Y \mid X)$ satisfies*

$$\mathbb{E}(\psi(X)) = \mathbb{E}(Y).$$

Proof. By Lemma (3.3.3),

$$\begin{aligned}\mathbb{E}(\psi(X)) &= \sum_x \psi(x) f_X(x) = \sum_{x,y} y f_{Y|X}(y \mid x) f_X(x)\\ &= \sum_{x,y} y f_{X,Y}(x, y) = \sum_y y f_Y(y) = \mathbb{E}(Y).\end{aligned}$$ ■

This is an extremely useful theorem, to which we shall make repeated reference. It often provides a useful method for calculating $\mathbb{E}(Y)$, since it asserts that

$$\mathbb{E}(Y) = \sum_x \mathbb{E}(Y \mid X = x)\mathbb{P}(X = x).$$

(5) Example. A hen lays N eggs, where N has the Poisson distribution with parameter λ. Each egg hatches with probability p $(= 1 - q)$ independently of the other eggs. Let K be the number of chicks. Find $\mathbb{E}(K \mid N)$, $\mathbb{E}(K)$, and $\mathbb{E}(N \mid K)$.

Solution. We are given that

$$f_N(n) = \frac{\lambda^n}{n!}e^{-\lambda}, \qquad f_{K|N}(k \mid n) = \binom{n}{k} p^k (1-p)^{n-k}.$$

Therefore

$$\psi(n) = \mathbb{E}(K \mid N = n) = \sum_k k f_{K|N}(k \mid n) = pn.$$

Thus $\mathbb{E}(K \mid N) = \psi(N) = pN$ and

$$\mathbb{E}(K) = \mathbb{E}(\psi(N)) = p\mathbb{E}(N) = p\lambda.$$

To find $\mathbb{E}(N \mid K)$ we need to know the conditional mass function $f_{N|K}$ of N given K. However,

$$\begin{aligned} f_{N|K}(n \mid k) &= \mathbb{P}(N = n \mid K = k) \\ &= \frac{\mathbb{P}(K = k \mid N = n)\mathbb{P}(N = n)}{\mathbb{P}(K = k)} \\ &= \frac{\binom{n}{k} p^k (1-p)^{n-k} (\lambda^n / n!) e^{-\lambda}}{\sum_{m \geq k} \binom{m}{k} p^k (1-p)^{m-k} (\lambda^m / m!) e^{-\lambda}} \qquad \text{if } n \geq k \\ &= \frac{(q\lambda)^{n-k}}{(n-k)!} e^{-q\lambda}. \end{aligned}$$

Hence

$$\mathbb{E}(N \mid K = k) = \sum_{n \geq k} n \frac{(q\lambda)^{n-k}}{(n-k)!} e^{-q\lambda} = k + q\lambda,$$

giving $\mathbb{E}(N \mid K) = K + q\lambda$. ●

There is a more general version of Theorem (4), and this will be of interest later.

(6) Theorem. *The conditional expectation* $\psi(X) = \mathbb{E}(Y \mid X)$ *satisfies*

$$\mathbb{E}\big(\psi(X)g(X)\big) = \mathbb{E}\big(Yg(X)\big) \tag{7}$$

for any function g for which both expectations exist.

Setting $g(x) = 1$ for all x, we obtain the result of (4). Whilst Theorem (6) is useful in its own right, we shall see later that its principal interest lies elsewhere. The conclusion of the theorem may be taken as a *definition* of conditional expectation—as a function $\psi(X)$ of X such that (7) holds for all suitable functions g. Such a definition is convenient when working with a notion of conditional expectation more general than that dealt with here.

Proof. As in the proof of (4),

$$\begin{aligned} \mathbb{E}\big(\psi(X)g(X)\big) &= \sum_x \psi(x)g(x)f_X(x) = \sum_{x,y} yg(x)f_{Y|X}(y \mid x)f_X(x) \\ &= \sum_{x,y} yg(x)f_{X,Y}(x,y) = \mathbb{E}\big(Yg(X)\big). \end{aligned}$$ ■

Exercises for Section 3.7

1. Show the following:
(a) $\mathbb{E}(aY + bZ \mid X) = a\mathbb{E}(Y \mid X) + b\mathbb{E}(Z \mid X)$ for $a, b \in \mathbb{R}$,
(b) $\mathbb{E}(Y \mid X) \geq 0$ if $Y \geq 0$,
(c) $\mathbb{E}(1 \mid X) = 1$,
(d) if X and Y are independent then $\mathbb{E}(Y \mid X) = \mathbb{E}(Y)$,
(e) ('pull-through property') $\mathbb{E}(Yg(X) \mid X) = g(X)\mathbb{E}(Y \mid X)$ for any suitable function g,
(f) ('tower property') $\mathbb{E}\{\mathbb{E}(Y \mid X, Z) \mid X\} = \mathbb{E}(Y \mid X) = \mathbb{E}\{\mathbb{E}(Y \mid X) \mid X, Z\}$.

2. Uniqueness of conditional expectation. Suppose that X and Y are discrete random variables, and that $\phi(X)$ and $\psi(X)$ are two functions of X satisfying

$$\mathbb{E}\big(\phi(X)g(X)\big) = \mathbb{E}\big(\psi(X)g(X)\big) = \mathbb{E}\big(Yg(X)\big)$$

for any function g for which all the expectations exist. Show that $\phi(X)$ and $\psi(X)$ are almost surely equal, in that $\mathbb{P}(\phi(X) = \psi(X)) = 1$.

3. Suppose that the conditional expectation of Y given X is defined as the (almost surely) unique function $\psi(X)$ such that $\mathbb{E}(\psi(X)g(X)) = \mathbb{E}(Yg(X))$ for all functions g for which the expectations exist. Show (a)–(f) of Exercise (1) above (with the occasional addition of the expression 'with probability 1').

4. How should we define $\text{var}(Y \mid X)$, the conditional variance of Y given X? Show that $\text{var}(Y) = \mathbb{E}(\text{var}(Y \mid X)) + \text{var}(\mathbb{E}(Y \mid X))$.

5. The lifetime of a machine (in days) is a random variable T with mass function f. Given that the machine is working after t days, what is the mean subsequent lifetime of the machine when:
(a) $f(x) = (N+1)^{-1}$ for $x \in \{0, 1, \dots, N\}$,
(b) $f(x) = 2^{-x}$ for $x = 1, 2, \dots$.
(The first part of Problem (3.11.13) may be useful.)

6. Let $X_1, X_2, \dots$ be identically distributed random variables with mean μ, and let N be a random variable taking values in the non-negative integers and independent of the X_i. Let $S = X_1 + X_2 + \cdots + X_N$. Show that $\mathbb{E}(S \mid N) = \mu N$, and deduce that $\mathbb{E}(S) = \mu\mathbb{E}(N)$.

7. A factory has produced n robots, each of which is faulty with probability ϕ. To each robot a test is applied which detects the fault (if present) with probability δ. Let X be the number of faulty robots, and Y the number detected as faulty. Assuming the usual independence, show that

$$\mathbb{E}(X \mid Y) = \big\{n\phi(1-\delta) + (1-\phi)Y\big\}/(1-\phi\delta).$$

8. Families. Each child is equally likely to be male or female, independently of all other children.
(a) Show that, in a family of predetermined size, the expected number of boys equals the expected number of girls. Was the assumption of independence necessary?
(b) A randomly selected child is male; does the expected number of his brothers equal the expected number of his sisters? What happens if you do not require independence?

9. Let X and Y be independent with mean μ. Explain the error in the following equation:

$$\text{`}\mathbb{E}(X \mid X + Y = z) = \mathbb{E}(X \mid X = z - Y) = \mathbb{E}(z - Y) = z - \mu\text{'}.$$

10. A coin shows heads with probability p. Let X_n be the number of flips required to obtain a run of n consecutive heads. Show that $\mathbb{E}(X_n) = \sum_{k=1}^{n} p^{-k}$.

3.8 Sums of random variables

Much of the classical theory of probability concerns sums of random variables. We have seen already many such sums; the number of heads in n tosses of a coin is one of the simplest such examples, but we shall encounter many situations which are more complicated than this. One particular complication is when the summands are dependent. The first stage in developing a systematic technique is to find a formula for describing the mass function of the sum $Z = X + Y$ of two variables having joint mass function $f(x, y)$.

(1) Theorem. *We have that* $\mathbb{P}(X + Y = z) = \sum_x f(x, z - x)$.

Proof. The union

$$\{X + Y = z\} = \bigcup_x \big(\{X = x\} \cap \{Y = z - x\}\big)$$

is disjoint, and at most countably many of its contributions have non-zero probability. Therefore

$$\mathbb{P}(X + Y = z) = \sum_x \mathbb{P}(X = x,\ Y = z - x) = \sum_x f(x, z - x). \qquad \blacksquare$$

If X and Y are independent, then

$$\mathbb{P}(X + Y = z) = f_{X+Y}(z) = \sum_x f_X(x) f_Y(z - x) = \sum_y f_X(z - y) f_Y(y).$$

The mass function of $X + Y$ is called the *convolution* of the mass functions of X and Y, and is written

$$f_{X+Y} = f_X * f_Y. \tag{2}$$

(3) Example (3.5.6) revisited. Let X_1 and X_2 be independent geometric variables with common mass function

$$f(k) = p(1 - p)^{k-1}, \qquad k = 1, 2, \dots.$$

By (2), $Z = X_1 + X_2$ has mass function

$$\begin{aligned}
\mathbb{P}(Z = z) &= \sum_k \mathbb{P}(X_1 = k)\mathbb{P}(X_2 = z - k) \\
&= \sum_{k=1}^{z-1} p(1 - p)^{k-1} p(1 - p)^{z-k-1} \\
&= (z - 1)p^2(1 - p)^{z-2}, \qquad z = 2, 3, \dots
\end{aligned}$$

in agreement with Example (3.5.6). The general formula for the sum of a number, r say, of geometric variables can easily be verified by induction. ●

Exercises for Section 3.8

1. Let X and Y be independent variables, X being equally likely to take any value in $\{0, 1, \dots, m\}$, and Y similarly in $\{0, 1, \dots, n\}$. Find the mass function of $Z = X + Y$. The random variable Z is said to have the *trapezoidal distribution*.

2. Let X and Y have the joint mass function

$$f(x, y) = \frac{C}{(x + y - 1)(x + y)(x + y + 1)}, \qquad x, y = 1, 2, 3, \dots.$$

Find the mass functions of $U = X + Y$ and $V = X - Y$.

3. Let X and Y be independent geometric random variables with respective parameters α and β. Show that

$$\mathbb{P}(X + Y = z) = \frac{\alpha\beta}{\alpha - \beta}\big\{(1 - \beta)^{z-1} - (1 - \alpha)^{z-1}\big\}.$$

4. Let $\{X_r : 1 \le r \le n\}$ be independent geometric random variables with parameter p. Show that $Z = \sum_{r=1}^{n} X_r$ has a negative binomial distribution. [Hint: No calculations are necessary.]

5. Pepys's problem†. Sam rolls $6n$ dice once; he needs at least n sixes. Isaac rolls $6(n + 1)$ dice; he needs at least $n + 1$ sixes. Who is more likely to obtain the number of sixes he needs?

6. Let N be Poisson distributed with parameter λ. Show that, for any function g such that the expectations exist, $\mathbb{E}(Ng(N)) = \lambda\mathbb{E}g(N + 1)$. More generally, if $S = \sum_{r=1}^{N} X_r$, where $\{X_r : r \ge 0\}$ are independent identically distributed non-negative integer-valued random variables, show that

$$\mathbb{E}\big(Sg(S)\big) = \lambda\mathbb{E}\big(g(S + X_0)X_0\big).$$

3.9 Simple random walk

Until now we have dealt largely with general theory; the final two sections of this chapter may provide some lighter relief. One of the simplest random processes is so-called 'simple random walk'‡; this process arises in many ways, of which the following is traditional. A gambler G plays the following game at the casino. The croupier tosses a (possibly biased) coin repeatedly; each time heads appears, he gives G one franc, and each time tails appears he takes one franc from G. Writing S_n for G's fortune after n tosses of the coin, we have that $S_{n+1} = S_n + X_{n+1}$ where X_{n+1} is a random variable taking the value 1 with some fixed probability p and -1 otherwise; furthermore, X_{n+1} is assumed independent of the results of all previous tosses. Thus

$$S_n = S_0 + \sum_{i=1}^{n} X_i, \tag{1}$$

†Pepys put a simple version of this problem to Newton in 1693, but was reluctant to accept the correct reply he received.

‡Karl Pearson coined the term 'random walk' in 1906, and (using a result of Rayleigh) demonstrated the theorem that "the most likely place to find a drunken walker is somewhere near his starting point", empirical verification of which is not hard to find.

so that S_n is obtained from the initial fortune S_0 by the addition of n independent random variables. We are assuming here that there are no constraints on G's fortune imposed externally, such as that the game is terminated if his fortune is reduced to zero.

An alternative picture of 'simple random walk' involves the motion of a particle—a particle which inhabits the set of integers and which moves at each step either one step to the right, with probability p, or one step to the left, the directions of different steps being independent of each other. More complicated random walks arise when the steps of the particle are allowed to have some general distribution on the integers, or the reals, so that the position S_n at time n is given by (1) where the X_i are independent and identically distributed random variables having some specified distribution function. Even greater generality is obtained by assuming that the X_i take values in $\mathbb{R}^d$ for some $d \geq 1$, or even some vector space over the real numbers. Random walks may be used with some success in modelling various practical situations, such as the numbers of cars in a toll queue at 5 minute intervals, the position of a pollen grain suspended in fluid at 1 second intervals, or the value of the Dow–Jones index each Monday morning. In each case, it may not be too bad a guess that the $(n+1)$th reading differs from the nth by a random quantity which is independent of previous jumps but has the same probability distribution. The theory of random walks is a basic tool in the probabilist's kit, and we shall concern ourselves here with 'simple random walk' only.

At any instant of time a particle inhabits one of the integer points of the real line. At time 0 it starts from some specified point, and at each subsequent epoch of time $1, 2, \ldots$ it moves from its current position to a new position according to the following law. With probability p it moves one step to the right, and with probability $q = 1 - p$ it moves one step to the left; moves are independent of each other. The walk is called *symmetric* if $p = q = \frac{1}{2}$. Example (1.7.4) concerned a symmetric random walk with 'absorbing' barriers at the points 0 and N. In general, let S_n denote the position of the particle after n moves, and set $S_0 = a$. Then

(2)
$$S_n = a + \sum_{i=1}^{n} X_i$$

where $X_1, X_2, \ldots$ is a sequence of independent Bernoulli variables taking values $+1$ and -1 (rather than $+1$ and 0 as before) with probabilities p and q.

We record the motion of the particle as the sequence $\{(n, S_n) : n \geq 0\}$ of Cartesian coordinates of points in the plane. This collection of points, joined by solid lines between neighbours, is called the *path* of the particle. In the example shown in Figure 3.2, the particle has visited the points $0, 1, 0, -1, 0, 1, 2$ in succession. This representation has a confusing aspect in that the direction of the particle's steps is parallel to the y-axis, whereas we have previously been specifying the movement in the traditional way as to the right or to the left. In future, any reference to the x-axis or the y-axis will pertain to a diagram of its path as exemplified by Figure 3.2.

The sequence (2) of partial sums has three important properties.

(3) Lemma. *The simple random walk is* spatially homogeneous; *that is*

$$\mathbb{P}(S_n = j \mid S_0 = a) = \mathbb{P}(S_n = j + b \mid S_0 = a + b).$$

Proof. Both sides equal $\mathbb{P}\big(\sum_1^n X_i = j - a\big)$. ■

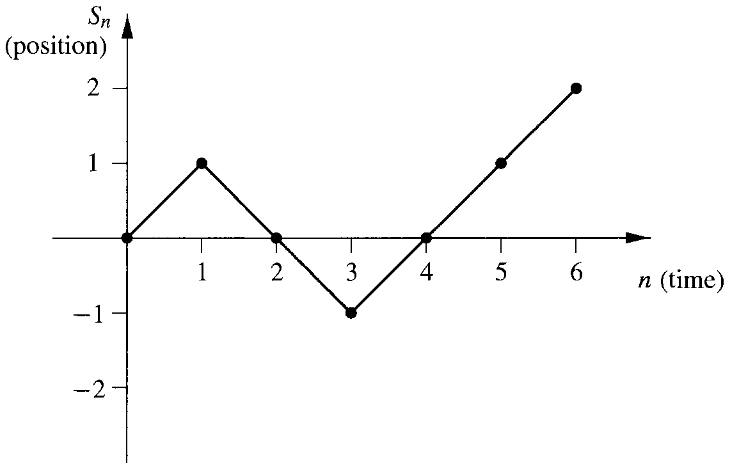

Figure 3.2. A random walk S_n.

(4) Lemma. *The simple random walk is* temporally homogeneous; *that is*

$$\mathbb{P}(S_n = j \mid S_0 = a) = \mathbb{P}(S_{m+n} = j \mid S_m = a).$$

Proof. The left- and right-hand sides satisfy

$$\text{LHS} = \mathbb{P}\left(\sum_1^n X_i = j - a\right) = \mathbb{P}\left(\sum_{m+1}^{m+n} X_i = j - a\right) = \text{RHS}. \qquad \blacksquare$$

(5) Lemma. *The simple random walk has the* Markov property; *that is*

$$\mathbb{P}(S_{m+n} = j \mid S_0, S_1, \dots, S_m) = \mathbb{P}(S_{m+n} = j \mid S_m), \qquad n \ge 0.$$

Statements such as $\mathbb{P}(S = j \mid X, Y) = \mathbb{P}(S = j \mid X)$ are to be interpreted in the obvious way as meaning that

$$\mathbb{P}(S = j \mid X = x,\ Y = y) = \mathbb{P}(S = j \mid X = x) \quad \text{for all } x \text{ and } y;$$

this is a slight abuse of notation.

Proof. If one knows the value of S_m, then the distribution of S_{m+n} depends only on the jumps $X_{m+1}, \dots, X_{m+n}$, and cannot depend on further information concerning the values of $S_0, S_1, \dots, S_{m-1}$. $\blacksquare$

This 'Markov property' is often expressed informally by saying that, conditional upon knowing the value of the process at the mth step, its values after the mth step do not depend on its values before the mth step. More colloquially: conditional upon the present, the future does not depend on the past. We shall meet this property again later.

(6) Example. Absorbing barriers. Let us revisit Example (1.7.4) for general values of p. Equation (1.7.5) gives us the following difference equation for the probabilities $\{p_k\}$ where p_k is the probability of ultimate ruin starting from k:

$$p_k = p \cdot p_{k+1} + q \cdot p_{k-1} \quad \text{if} \quad 1 \le k \le N-1 \tag{7}$$

with boundary conditions $p_0 = 1$, $p_N = 0$. The solution of such a difference equation proceeds as follows. Look for a solution of the form $p_k = \theta^k$. Substitute this into (7) and cancel out the power θ^{k-1} to obtain $p\theta^2 - \theta + q = 0$, which has roots $\theta_1 = 1, \theta_2 = q/p$. If $p \neq \frac{1}{2}$ then these roots are distinct and the general solution of (7) is $p_k = A_1\theta_1^k + A_2\theta_2^k$ for arbitrary constants A_1 and A_2. Use the boundary conditions to obtain

$$p_k = \frac{(q/p)^k - (q/p)^N}{1-(q/p)^N}.$$

If $p = \frac{1}{2}$ then $\theta_1 = \theta_2 = 1$ and the general solution to (7) is $p_k = A_1 + A_2 k$. Use the boundary conditions to obtain $p_k = 1 - (k/N)$.

A more complicated equation is obtained for the mean number D_k of steps before the particle hits one of the absorbing barriers, starting from k. In this case we use conditional expectations and (3.7.4) to find that

$$D_k = p(1 + D_{k+1}) + q(1 + D_{k-1}) \quad \text{if} \quad 1 \le k \le N-1 \tag{8}$$

with the boundary conditions $D_0 = D_N = 0$. Try solving this; you need to find a general solution and a particular solution, as in the solution of second-order linear differential equations. This answer is

$$D_k = \begin{cases} \dfrac{1}{q-p}\left[k - N\left(\dfrac{1-(q/p)^k}{1-(q/p)^N}\right)\right] & \text{if } p \neq \frac{1}{2}, \\ k(N-k) & \text{if } p = \frac{1}{2}. \end{cases} \tag{9}$$

●

(10) Example. Retaining barriers. In Example (1.7.4), suppose that the Jaguar buyer has a rich uncle who will guarantee all his losses. Then the random walk does not end when the particle hits zero, although it cannot visit a negative integer. Instead $\mathbb{P}(S_{n+1} = 0 \mid S_n = 0) = q$ and $\mathbb{P}(S_{n+1} = 1 \mid S_n = 0) = p$. The origin is said to have a 'retaining' barrier (sometimes called 'reflecting').

What now is the expected duration of the game? The mean duration F_k, starting from k, satisfies the same difference equation (8) as before but subject to different boundary conditions. We leave it as an *exercise* to show that the boundary conditions are $F_N = 0$, $pF_0 = 1 + pF_1$, and hence to find F_k. ●

In such examples the techniques of 'conditioning' are supremely useful. The idea is that in order to calculate a probability $\mathbb{P}(A)$ or expectation $\mathbb{E}(Y)$ we condition either on some partition of Ω (and use Lemma (1.4.4)) or on the outcome of some random variable (and use Theorem (3.7.4) or the forthcoming Theorem (4.6.5)). In this section this technique yielded the difference equations (7) and (8). In later sections the same idea will yield differential equations, integral equations, and functional equations, some of which can be solved.

Exercises for Section 3.9

1. Let T be the time which elapses before a simple random walk is absorbed at either of the absorbing barriers at 0 and N, having started at k where $0 \leq k \leq N$. Show that $\mathbb{P}(T < \infty) = 1$ and $\mathbb{E}(T^k) < \infty$ for all $k \geq 1$.

2. For simple random walk S with absorbing barriers at 0 and N, let W be the event that the particle is absorbed at 0 rather than at N, and let $p_k = \mathbb{P}(W \mid S_0 = k)$. Show that, if the particle starts at k where $0 < k < N$, the conditional probability that the first step is rightwards, given W, equals pp_{k+1}/p_k. Deduce that the mean duration J_k of the walk, conditional on W, satisfies the equation

$$pp_{k+1}J_{k+1} - p_k J_k + (p_k - pp_{k+1})J_{k-1} = -p_k, \qquad \text{for } 0 < k < N.$$

Show that we may take as boundary condition $J_0 = 0$. Find J_k in the symmetric case, when $p = \frac{1}{2}$.

3. With the notation of Exercise (2), suppose further that at any step the particle may remain where it is with probability r where $p + q + r = 1$. Show that J_k satisfies

$$pp_{k+1}J_{k+1} - (1-r)p_k J_k + qp_{k-1}J_{k-1} = -p_k$$

and that, when $\rho = q/p \neq 1$,

$$J_k = \frac{1}{p-q} \cdot \frac{1}{\rho^k - \rho^N} \left\{ k(\rho^k + \rho^N) - \frac{2N\rho^N(1-\rho^k)}{1-\rho^N} \right\}.$$

4. Problem of the points. A coin is tossed repeatedly, heads turning up with probability p on each toss. Player A wins the game if m heads appear before n tails have appeared, and player B wins otherwise. Let p_{mn} be the probability that A wins the game. Set up a difference equation for the p_{mn}. What are the boundary conditions?

5. Consider a simple random walk on the set $\{0, 1, 2, \dots, N\}$ in which each step is to the right with probability p or to the left with probability $q = 1 - p$. Absorbing barriers are placed at 0 and N. Show that the number X of positive steps of the walk before absorption satisfies

$$\mathbb{E}(X) = \tfrac{1}{2}\{D_k - k + N(1 - p_k)\}$$

where D_k is the mean number of steps until absorption and p_k is the probability of absorption at 0.

6. (a) "Millionaires should always gamble, poor men never" [J. M. Keynes].
(b) "If I wanted to gamble, I would buy a casino" [P. Getty].
(c) "That the chance of gain is naturally overvalued, we may learn from the universal success of lotteries" [Adam Smith, 1776].
Discuss.

3.10 Random walk: counting sample paths

In the previous section, our principal technique was to condition on the first step of the walk and then solve the ensuing difference equation. Another primitive but useful technique is to count. Let $X_1, X_2, \dots$ be independent variables, each taking the values -1 and 1 with probabilities $q = 1 - p$ and p, as before, and let

(1) $$S_n = a + \sum_{i=1}^{n} X_i$$

be the position of the corresponding random walker after n steps, having started at $S_0 = a$. The set of realizations of the walk is the set of vectors $\mathbf{s} = (s_0, s_1, \dots)$ with $s_0 = a$ and $s_{i+1} - s_i = \pm 1$, and any such vector may be thought of as a 'sample path' of the walk, drawn in the manner of Figure 3.2. The probability that the first n steps of the walk follow a given path $\mathbf{s} = (s_0, s_1, \dots, s_n)$ is $p^r q^l$ where r is the number of steps of s to the right and l is the number to the left†; that is to say, $r = |\{i : s_{i+1} - s_i = 1\}|$ and $l = |\{i : s_{i+1} - s_i = -1\}|$. Any event may be expressed in terms of an appropriate set of paths, and the probability of the event is the sum of the component probabilities. For example, $\mathbb{P}(S_n = b) = \sum_r M_n^r(a, b) p^r q^{n-r}$ where $M_n^r(a, b)$ is the number of paths $(s_0, s_1, \dots, s_n)$ with $s_0 = a$, $s_n = b$, and having exactly r rightward steps. It is easy to see that $r + l = n$, the total number of steps, and $r - l = b - a$, the aggregate rightward displacement, so that $r = \frac{1}{2}(n + b - a)$ and $l = \frac{1}{2}(n - b + a)$. Thus

$$\text{(2)} \qquad \mathbb{P}(S_n = b) = \binom{n}{\frac{1}{2}(n + b - a)} p^{\frac{1}{2}(n+b-a)} q^{\frac{1}{2}(n-b+a)},$$

since there are exactly $\binom{n}{r}$ paths with length n having r rightward steps and $n - r$ leftward steps. Formula (2) is useful only if $\frac{1}{2}(n + b - a)$ is an integer lying in the range $0, 1, \dots, n$; otherwise, the probability in question equals 0.

Natural equations of interest for the walk include:

(a) when does the first visit of the random walk to a given point occur; and

(b) what is the furthest rightward point visited by the random walk by time n?

Such questions may be answered with the aid of certain elegant results and techniques for counting paths. The first of these is the 'reflection principle'. Here is some basic notation. As in Figure 3.2, we keep a record of the random walk S through its path $\{(n, S_n) : n \geq 0\}$.

Suppose we know that $S_0 = a$ and $S_n = b$. The random walk may or may not have visited the origin between times 0 and n. Let $N_n(a, b)$ be the number of possible paths from $(0, a)$ to (n, b), and let $N_n^0(a, b)$ be the number of such paths which contain some point $(k, 0)$ on the x-axis.

(3) Theorem. The reflection principle. *If $a, b > 0$ then $N_n^0(a, b) = N_n(-a, b)$.*

Proof. Each path from $(0, -a)$ to (n, b) intersects the x-axis at some earliest point $(k, 0)$. Reflect the segment of the path with $0 \leq x \leq k$ in the x-axis to obtain a path joining $(0, a)$ to (n, b) which intersects the x-axis (see Figure 3.3). This operation gives a one–one correspondence between the collections of such paths, and the theorem is proved. ■

We have, as before, a formula for $N_n(a, b)$.

(4) Lemma. $N_n(a, b) = \binom{n}{\frac{1}{2}(n + b - a)}$.

Proof. Choose a path from $(0, a)$ to (n, b) and let α and β be the numbers of positive and negative steps, respectively, in this path. Then $\alpha + \beta = n$ and $\alpha - \beta = b - a$, so that $\alpha = \frac{1}{2}(n + b - a)$. The number of such paths is the number of ways of picking α positive steps from the n available. That is

$$\text{(5)} \qquad N_n(a, b) = \binom{n}{\alpha} = \binom{n}{\frac{1}{2}(n + b - a)}. \qquad ■$$

†The words 'right' and 'left' are to be interpreted as meaning in the positive and negative directions respectively, plotted along the y-axis as in Figure 3.2.

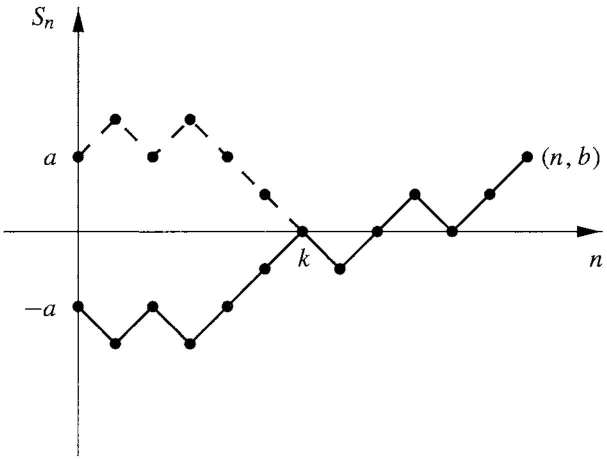

Figure 3.3. A random walk; the dashed line is the reflection of the first segment of the walk.

The famous 'ballot theorem' is a consequence of these elementary results; it was proved first by W. A. Whitworth in 1878.

(6) Corollary†. Ballot theorem. *If $b > 0$ then the number of paths from $(0, 0)$ to (n, b) which do not revisit the x-axis equals $(b/n)N_n(0, b)$.*

Proof. The first step of all such paths is to $(1, 1)$, and so the number of such path is

$$N_{n-1}(1, b) - N^0_{n-1}(1, b) = N_{n-1}(1, b) - N_{n-1}(-1, b)$$

by the reflection principle. We now use (4) and an elementary calculation to obtain the required result. ■

As an application, and an explanation of the title of the theorem, we may easily answer the following amusing question. Suppose that, in a ballot, candidate A scores α votes and candidate B scores β votes where $\alpha > \beta$. What is the probability that, during the ballot, A was always ahead of B? Let X_i equal 1 if the ith vote was cast for A, and -1 otherwise. Assuming that each possible combination of α votes for A and β votes for B is equally likely, we have that the probability is question is the proportion of paths from $(0, 0)$ to $(\alpha + \beta, \alpha - \beta)$ which do not revisit the x-axis. Using the ballot theorem, we obtain the answer $(\alpha - \beta)/(\alpha + \beta)$.

Here are some applications of the reflection principle to random walks. First, what is the probability that the walk does not revisit its starting point in the first n steps? We may as well assume that $S_0 = 0$, so that $S_1 \neq 0, \ldots, S_n \neq 0$ if and only if $S_1 S_2 \cdots S_n \neq 0$.

(7) Theorem. *If $S_0 = 0$ then, for $n \geq 1$,*

$$\mathbb{P}(S_1 S_2 \cdots S_n \neq 0,\ S_n = b) = \frac{|b|}{n}\mathbb{P}(S_n = b), \tag{8}$$

and therefore

$$\mathbb{P}(S_1 S_2 \cdots S_n \neq 0) = \frac{1}{n}\mathbb{E}|S_n|. \tag{9}$$

†Derived from the Latin word 'corollarium' meaning 'money paid for a garland' or 'tip'.

Proof. Suppose that $S_0 = 0$ and $S_n = b\ (> 0)$. The event in question occurs if and only if the path of the random walk does not visit the x-axis in the time interval $[1, n]$. The number of such paths is, by the ballot theorem, $(b/n)N_n(0, b)$, and each such path has $\frac{1}{2}(n+b)$ rightward steps and $\frac{1}{2}(n-b)$ leftward steps. Therefore

$$\mathbb{P}(S_1 S_2 \cdots S_n \neq 0,\ S_n = b) = \frac{b}{n} N_n(0, b) p^{\frac{1}{2}(n+b)} q^{\frac{1}{2}(n-b)} = \frac{b}{n}\mathbb{P}(S_n = b)$$

as required. A similar calculation is valid if $b < 0$. ■

Another feature of interest is the maximum value attained by the random walk. We write $M_n = \max\{S_i : 0 \le i \le n\}$ for the maximum value up to time n, and shall suppose that $S_0 = 0$, so that $M_n \ge 0$. Clearly $M_n \ge S_n$, and the first part of the next theorem is therefore trivial.

(10) Theorem. *Suppose that $S_0 = 0$. Then, for $r \ge 1$,*

$$\text{(11)} \qquad \mathbb{P}(M_n \ge r,\ S_n = b) = \begin{cases} \mathbb{P}(S_n = b) & \text{if } b \ge r, \\ (q/p)^{r-b}\mathbb{P}(S_n = 2r - b) & \text{if } b < r. \end{cases}$$

It follows that, for $r \ge 1$,

$$\text{(12)} \qquad \begin{aligned} \mathbb{P}(M_n \ge r) &= \mathbb{P}(S_n \ge r) + \sum_{b=-\infty}^{r-1} (q/p)^{r-b}\mathbb{P}(S_n = 2r - b) \\ &= \mathbb{P}(S_n = r) + \sum_{c=r+1}^{\infty} \left[1 + (q/p)^{c-r}\right]\mathbb{P}(S_n = c), \end{aligned}$$

yielding in the symmetric case when $p = q = \frac{1}{2}$ that

$$\text{(13)} \qquad \mathbb{P}(M_n \ge r) = 2\mathbb{P}(S_n \ge r + 1) + \mathbb{P}(S_n = r),$$

which is easily expressed in terms of the binomial distribution.

Proof of (10). We may assume that $r \ge 1$ and $b < r$. Let $N_n^r(0, b)$ be the number of paths from $(0, 0)$ to (n, b) which include some point having height r, which is to say some point (i, r) with $0 < i < n$; for such a path π, let (i_π, r) be the earliest such point. We may reflect the segment of the path with $i_\pi \le x \le n$ in the line $y = r$ to obtain a path π' joining $(0, 0)$ to $(n, 2r - b)$. Any such path π' is obtained thus from a unique path π, and therefore $N_n^r(0, b) = N_n(0, 2r - b)$. It follows as required that

$$\begin{aligned} \mathbb{P}(M_n \ge r,\ S_n = b) &= N_n^r(0, b) p^{\frac{1}{2}(n+b)} q^{\frac{1}{2}(n-b)} \\ &= (q/p)^{r-b} N_n(0, 2r - b) p^{\frac{1}{2}(n+2r-b)} q^{\frac{1}{2}(n-2r+b)} \\ &= (q/p)^{r-b}\mathbb{P}(S_n = 2r - b). \end{aligned}$$

■

What is the chance that the walk reaches a new maximum at a particular time? More precisely, what is the probability that the walk, starting from 0, reaches the point b (> 0) for the time time at the nth step? Writing $f_b(n)$ for this probability, we have that

$$\begin{aligned} f_b(n) &= \mathbb{P}(M_{n-1} = S_{n-1} = b-1,\ S_n = b) \\ &= p\Big[\mathbb{P}(M_{n-1} \geq b-1,\ S_{n-1} = b-1) - \mathbb{P}(M_{n-1} \geq b,\ S_{n-1} = b-1)\Big] \\ &= p\big[\mathbb{P}(S_{n-1} = b-1) - (q/p)\mathbb{P}(S_{n-1} = b+1)\big] \quad \text{by (11)} \\ &= \frac{b}{n}\mathbb{P}(S_n = b) \end{aligned}$$

by a simple calculation using (2). A similar conclusion may be reached if $b < 0$, and we arrive at the following.

(14) Hitting time theorem. *The probability $f_b(n)$ that a random walk S hits the point b for the first time at the nth step, having started from 0, satisfies*

(15)
$$f_b(n) = \frac{|b|}{n}\mathbb{P}(S_n = b) \quad \textit{if} \quad n \geq 1.$$

The conclusion here has a close resemblance to that of the ballot theorem, and particularly Theorem (7). This is no coincidence: a closer examination of the two results leads to another technique for random walks, the technique of 'reversal'. If the first n steps of the original random walk are

$$\{0, S_1, S_2, \dots, S_n\} = \left\{0, X_1, X_1 + X_2, \dots, \sum_1^n X_i\right\}$$

then the steps of the *reversed* walk, denoted by $0, T_1, \dots, T_n$, are given by

$$\{0, T_1, T_2, \dots, T_n\} = \left\{0, X_n, X_n + X_{n-1}, \dots, \sum_1^n X_i\right\}.$$

Draw a diagram to see how the two walks correspond to each other. The X_i are independent and identically distributed, and it follows that the two walks have identical distributions even if $p \neq \frac{1}{2}$. Notice that the addition of an extra step to the original walk may change *every* step of the reversed walk.

Now, the original walk satisfies $S_n = b\,(> 0)$ and $S_1 S_2 \cdots S_n \neq 0$ if and only if the reversed walk satisfied $T_n = b$ and $T_n - T_{n-i} = X_1 + \cdots + X_i > 0$ for all $i \geq 1$, which is to say that the first visit of the reversed walk to the point b takes place at time n. Therefore

(16)
$$\mathbb{P}(S_1 S_2 \cdots S_n \neq 0,\ S_n = b) = f_b(n) \quad \text{if} \quad b > 0.$$

This is the 'coincidence' remarked above; a similar argument is valid if $b < 0$. The technique of reversal has other applications. For example, let μ_b be the mean number of visits of the walk to the point b before it returns to its starting point. If $S_0 = 0$ then, by (16),

(17)
$$\mu_b = \sum_{n=1}^{\infty} \mathbb{P}(S_1 S_2 \cdots S_n \neq 0,\ S_n = b) = \sum_{n=1}^{\infty} f_b(n) = \mathbb{P}(S_n = b \text{ for some } n),$$

the probability of ultimately visiting b. This leads to the following result.

(18) Theorem. *If $p = \frac{1}{2}$ and $S_0 = 0$, for any b ($\neq 0$) the mean number μ_b of visits of the walk to the point b before returning to the origin equals* 1.

Proof. Let $f_b = \mathbb{P}(S_n = b \text{ for some } n \geq 0)$. We have, by conditioning on the value of S_1, that $f_b = \frac{1}{2}(f_{b+1} + f_{b-1})$ for $b > 0$, with boundary condition $f_0 = 1$. The solution of this difference equation is $f_b = Ab + B$ for constants A and B. The unique such solution lying in $[0, 1]$ with $f_0 = 1$ is given by $f_b = 1$ for all $b \geq 0$. By symmetry, $f_b = 1$ for $b \leq 0$. However, $f_b = \mu_b$ for $b \neq 0$, and the claim follows. ∎

'The truly amazing implications of this result appear best in the language of fair games. A perfect coin is tossed until the first equalization of the accumulated numbers of heads and tails. The gambler receives one penny for every time that the accumulated number of heads exceeds the accumulated number of tails by m. *The "fair entrance fee" equals* 1 *independently of* m.' (Feller 1968, p. 367).

We conclude with two celebrated properties of the symmetric random walk.

(19) Theorem. Arc sine law for last visit to the origin. *Suppose that $p = \frac{1}{2}$ and $S_0 = 0$. The probability that the last visit to 0 up to time $2n$ occurred at time $2k$ is $\mathbb{P}(S_{2k} = 0)\mathbb{P}(S_{2n-2k} = 0)$.*

In advance of proving this, we note some consequences. Writing $\alpha_{2n}(2k)$ for the probability referred to in the theorem, it follows from the theorem that $\alpha_{2n}(2k) = u_{2k}u_{2n-2k}$ where

$$u_{2k} = \mathbb{P}(S_{2k} = 0) = \binom{2k}{k} 2^{-2k}.$$

In order to understand the behaviour of u_{2k} for large values of k, we use Stirling's formula:

(20) $$n! \sim n^n e^{-n}\sqrt{2\pi n} \quad \text{as} \quad n \to \infty,$$

which is to say that the ratio of the left-hand side to the right-hand side tends to 1 as $n \to \infty$. Applying this formula, we obtain that $u_{2k} \sim 1/\sqrt{\pi k}$ as $k \to \infty$. This gives rise to the approximation

$$\alpha_{2n}(2k) \simeq \frac{1}{\pi\sqrt{k(n-k)}},$$

valid for values of k which are close to neither 0 nor n. With T_{2n} denoting the time of the last visit to 0 up to time $2n$, it follows that

$$\mathbb{P}(T_{2n} \leq 2xn) \simeq \sum_{k \leq xn} \frac{1}{\pi\sqrt{k(n-k)}} \sim \int_{u=0}^{xn} \frac{1}{\pi\sqrt{u(n-u)}}\, du = \frac{2}{\pi}\sin^{-1}\sqrt{x},$$

which is to say that $T_{2n}/(2n)$ has a distribution function which is approximately $(2/\pi)\sin^{-1}\sqrt{x}$ when n is sufficiently large. We have proved a limit theorem.

The arc sine law is rather surprising. One may think that, in a long run of $2n$ tosses of a fair coin, the epochs of time at which there have appeared equal numbers of heads and tails should appear rather frequently. On the contrary, there is for example probability $\frac{1}{2}$ that no such epoch arrived in the final n tosses, and indeed probability approximately $\frac{1}{5}$ that no such epoch occurred after the first $\frac{1}{5}n$ tosses. One may think that, in a long run of $2n$ tosses of a fair coin, the last time at which the numbers of heads and tails were equal tends to be close to the end. On the contrary, the distribution of this time is symmetric around the midpoint.

How much time does a symmetric random walk spend to the right of the origin? More precisely, for how many values of k satisfying $0 \le k \le 2n$ is it the case that $S_k > 0$? Intuitively, one might expect the answer to be around n with large probability, but the truth is quite different. With large probability, the proportion of time spent to the right (or to the left) of the origin is near to 0 or to 1, but not near to $\frac{1}{2}$. That is to say, in a long sequence of tosses of a fair coin, there is large probability that one face (either heads or tails) will lead the other for a disproportionate amount of time.

(21) Theorem. Arc sine law for sojourn times. *Suppose that $p = \frac{1}{2}$ and $S_0 = 0$. The probability that the walk spends exactly $2k$ intervals of time, up to time $2n$, to the right of the origin equals $\mathbb{P}(S_{2k} = 0)\mathbb{P}(S_{2n-2k} = 0)$.*

We say that the interval $(k, k+1)$ is spent to the right of the origin if either $S_k > 0$ or $S_{k+1} > 0$. It is clear that the number of such intervals is even if the total number of steps is even. The conclusion of this theorem is most striking. First, the answer is the same as that of Theorem (19). Secondly, by the calculations following (19) we have that the probability that the walk spends $2xn$ units of time or less to the right of the origin is approximately $(2/\pi)\sin^{-1}\sqrt{x}$.

Proof of (19). The probability in question is

$$\begin{aligned}\alpha_{2n}(2k) &= \mathbb{P}(S_{2k} = 0)\mathbb{P}(S_{2k+1}S_{2k+2}\cdots S_{2n} \neq 0 \mid S_{2k} = 0)\\ &= \mathbb{P}(S_{2k} = 0)\mathbb{P}(S_1S_2\cdots S_{2n-2k} \neq 0).\end{aligned}$$

Now, setting $m = n - k$, we have by (8) that

$$\begin{aligned}\textbf{(22)}\qquad \mathbb{P}(S_1S_2\cdots S_{2m} \neq 0) &= 2\sum_{k=1}^{m}\frac{2k}{2m}\mathbb{P}(S_{2m} = 2k) = 2\sum_{k=1}^{m}\frac{2k}{2m}\binom{2m}{m+k}\left(\frac{1}{2}\right)^{2m}\\ &= 2\left(\frac{1}{2}\right)^{2m}\sum_{k=1}^{m}\left[\binom{2m-1}{m+k-1} - \binom{2m-1}{m+k}\right]\\ &= 2\left(\frac{1}{2}\right)^{2m}\binom{2m-1}{m}\\ &= \binom{2m}{m}\left(\frac{1}{2}\right)^{2m} = \mathbb{P}(S_{2m} = 0).\end{aligned}$$

■

In passing, note the proof in (22) that

$$\textbf{(23)}\qquad \mathbb{P}(S_1S_2\cdots S_{2m} \neq 0) = \mathbb{P}(S_{2m} = 0)$$

for the simple symmetric random walk.

Proof of (21). Let $\beta_{2n}(2k)$ be the probability in question, and write $u_{2m} = \mathbb{P}(S_{2m} = 0)$ as before. We are claiming that, for all $m \ge 1$,

$$\textbf{(24)}\qquad \beta_{2m}(2k) = u_{2k}u_{2m-2k} \quad \text{if} \quad 0 \le k \le m.$$

First,

$$\begin{aligned}\mathbb{P}(S_1 S_2 \cdots S_{2m} > 0) &= \mathbb{P}(S_1 = 1,\ S_2 \geq 1, \dots,\ S_{2m} \geq 1)\\ &= \tfrac{1}{2}\mathbb{P}(S_1 \geq 0,\ S_2 \geq 0, \dots,\ S_{2m-1} \geq 0),\end{aligned}$$

where the second line follows by considering the walk $S_1 - 1, S_2 - 1, \dots, S_{2m} - 1$. Now S_{2m-1} is an odd number, so that $S_{2m-1} \geq 0$ implies that $S_{2m} \geq 0$ also. Thus

$$\mathbb{P}(S_1 S_2 \cdots S_{2m} > 0) = \tfrac{1}{2}\mathbb{P}(S_1 \geq 0, S_2 \geq 0, \dots, S_{2m} \geq 0),$$

yielding by (23) that

$$\tfrac{1}{2}u_{2m} = \mathbb{P}(S_1 S_2 \cdots S_{2m} > 0) = \tfrac{1}{2}\beta_{2m}(2m),$$

and (24) follows for $k = m$, and therefore for $k = 0$ also by symmetry.

Let n be a positive integer, and let T be the time of the first return of the walk to the origin. If $S_{2n} = 0$ then $T \leq 2n$; the probability mass function $f_{2r} = \mathbb{P}(T = 2r)$ satisfies

$$\mathbb{P}(S_{2n} = 0) = \sum_{r=1}^{n} \mathbb{P}(S_{2n} = 0 \mid T = 2r)\mathbb{P}(T = 2r) = \sum_{r=1}^{n} \mathbb{P}(S_{2n-2r} = 0)\mathbb{P}(T = 2r),$$

which is to say that

(25) $$u_{2n} = \sum_{r=1}^{n} u_{2n-2r} f_{2r}.$$

Let $1 \leq k \leq n - 1$, and consider $\beta_{2n}(2k)$. The corresponding event entails that $T = 2r$ for some r satisfying $1 \leq r < n$. The time interval $(0, T)$ is spent entirely either to the right or the left of the origin, and each possibility has probability $\frac{1}{2}$. Therefore,

(26) $$\beta_{2n}(2k) = \sum_{r=1}^{k} \tfrac{1}{2}\mathbb{P}(T = 2r)\beta_{2n-2r}(2k - 2r) + \sum_{r=1}^{n-k} \tfrac{1}{2}\mathbb{P}(T = 2r)\beta_{2n-2r}(2k).$$

We conclude the proof by using induction. Certainly (24) is valid for all k if $m = 1$. Assume (24) is valid for all k and all $m < n$.

From (26),

$$\begin{aligned}\beta_{2n}(2k) &= \tfrac{1}{2}\sum_{r=1}^{k} f_{2r}u_{2k-2r}u_{2n-2k} + \tfrac{1}{2}\sum_{r=1}^{n-k} f_{2r}u_{2k}u_{2n-2k-2r}\\ &= \tfrac{1}{2}u_{2n-2k}u_{2k} + \tfrac{1}{2}u_{2k}u_{2n-2k} = u_{2k}u_{2n-2k}\end{aligned}$$

by (25), as required. ■

Exercises for Section 3.10

1. Consider a symmetric simple random walk S with $S_0 = 0$. Let $T = \min\{n \geq 1 : S_n = 0\}$ be the time of the first return of the walk to its starting point. Show that

$$\mathbb{P}(T = 2n) = \frac{1}{2n-1}\binom{2n}{n}2^{-2n},$$

and deduce that $\mathbb{E}(T^\alpha) < \infty$ if and only if $\alpha < \frac{1}{2}$. You may need Stirling's formula: $n! \sim n^{n+\frac{1}{2}}e^{-n}\sqrt{2\pi}$.

2. For a symmetric simple random walk starting at 0, show that the mass function of the maximum satisfies $\mathbb{P}(M_n = r) = \mathbb{P}(S_n = r) + \mathbb{P}(S_n = r + 1)$ for $r \geq 0$.

3. For a symmetric simple random walk starting at 0, show that the probability that the first visit to S_{2n} takes place at time $2k$ equals the product $\mathbb{P}(S_{2k} = 0)\mathbb{P}(S_{2n-2k} = 0)$, for $0 \leq k \leq n$.

3.11 Problems

1. (a) Let X and Y be independent discrete random variables, and let $g, h : \mathbb{R} \to \mathbb{R}$. Show that $g(X)$ and $h(Y)$ are independent.

(b) Show that two discrete random variables X and Y are independent if and only if $f_{X,Y}(x, y) = f_X(x)f_Y(y)$ for all $x, y \in \mathbb{R}$.

(c) More generally, show that X and Y are independent if and only if $f_{X,Y}(x, y)$ can be factorized as the product $g(x)h(y)$ of a function of x alone and a function of y alone.

2. Show that if $\text{var}(X) = 0$ then X is almost surely constant; that is, there exists $a \in \mathbb{R}$ such that $\mathbb{P}(X = a) = 1$. (First show that if $\mathbb{E}(X^2) = 0$ then $\mathbb{P}(X = 0) = 1$.)

3. (a) Let X be a discrete random variable and let $g : \mathbb{R} \to \mathbb{R}$. Show that, when the sum is absolutely convergent,

$$\mathbb{E}(g(X)) = \sum_x g(x)\mathbb{P}(X = x).$$

(b) If X and Y are independent and $g, h : \mathbb{R} \to \mathbb{R}$, show that $\mathbb{E}(g(X)h(Y)) = \mathbb{E}(g(X))\mathbb{E}(h(Y))$ whenever these expectations exist.

4. Let $\Omega = \{\omega_1, \omega_2, \omega_3\}$, with $\mathbb{P}(\omega_1) = \mathbb{P}(\omega_2) = \mathbb{P}(\omega_3) = \frac{1}{3}$. Define $X, Y, Z : \Omega \to \mathbb{R}$ by

$$\begin{aligned}
X(\omega_1) &= 1, & X(\omega_2) &= 2, & X(\omega_3) &= 3,\\
Y(\omega_1) &= 2, & Y(\omega_2) &= 3, & Y(\omega_3) &= 1,\\
Z(\omega_1) &= 2, & Z(\omega_2) &= 2, & Z(\omega_3) &= 1.
\end{aligned}$$

Show that X and Y have the same mass functions. Find the mass functions of $X + Y$, XY, and X/Y. Find the conditional mass functions $f_{Y|Z}$ and $f_{Z|Y}$.

5. For what values of k and α is f a mass function, where:

(a) $f(n) = k/\{n(n+1)\}$, $n = 1, 2, \ldots$,

(b) $f(n) = kn^\alpha$, $n = 1, 2, \ldots$ (*zeta* or *Zipf distribution*)?

6. Let X and Y be independent Poisson variables with respective parameters λ and μ. Show that:
(a) $X + Y$ is Poisson, parameter $\lambda + \mu$,
(b) the conditional distribution of X, given $X + Y = n$, is binomial, and find its parameters.

7. If X is geometric, show that $\mathbb{P}(X = n+k \mid X > n) = \mathbb{P}(X = k)$ for $k, n \geq 1$. Why do you think that this is called the 'lack of memory' property? Does any other distribution on the positive integers have this property?

8. Show that the sum of two independent binomial variables, $\text{bin}(m, p)$ and $\text{bin}(n, p)$ respectively, is $\text{bin}(m + n, p)$.

9. Let N be the number of heads occurring in n tosses of a biased coin. Write down the mass function of N in terms of the probability p of heads turning up on each toss. Prove and utilize the identity

$$\sum_i \binom{n}{2i} x^{2i} y^{n-2i} = \tfrac{1}{2}\{(x+y)^n + (y-x)^n\}$$

in order to calculate the probability p_n that N is even. Compare with Problem (1.8.20).

10. An urn contains N balls, b of which are blue and r $(= N - b)$ of which are red. A random sample of n balls is withdrawn without replacement from the urn. Show that the number B of blue balls in this sample has the mass function

$$\mathbb{P}(B = k) = \binom{b}{k}\binom{N-b}{n-k} \bigg/ \binom{N}{n}.$$

This is called the *hypergeometric distribution* with parameters N, b, and n. Show further that if N, b, and r approach ∞ in such a way that $b/N \to p$ and $r/N \to 1 - p$, then

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

You have shown that, for small n and large N, the distribution of B barely depends on whether or not the balls are replaced in the urn immediately after their withdrawal.

11. Let X and Y be independent $\text{bin}(n, p)$ variables, and let $Z = X + Y$. Show that the conditional distribution of X given $Z = N$ is the hypergeometric distribution of Problem (3.11.10).

12. Suppose X and Y take values in $\{0, 1\}$, with joint mass function $f(x, y)$. Write $f(0, 0) = a$, $f(0, 1) = b$, $f(1, 0) = c$, $f(1, 1) = d$, and find necessary and sufficient conditions for X and Y to be: (a) uncorrelated, (b) independent.

13. (a) If X takes non-negative integer values show that

$$\mathbb{E}(X) = \sum_{n=0}^{\infty} \mathbb{P}(X > n).$$

(b) An urn contains b blue and r red balls. Balls are removed at random until the first blue ball is drawn. Show that the expected number drawn is $(b + r + 1)/(b + 1)$.
(c) The balls are replaced and then removed at random until all the remaining balls are of the same colour. Find the expected number remaining in the urn.

14. Let $X_1, X_2, \dots, X_n$ be independent random variables, and suppose that X_k is Bernoulli with parameter p_k. Show that $Y = X_1 + X_2 + \cdots + X_n$ has mean and variance given by

$$\mathbb{E}(Y) = \sum_1^n p_k, \quad \operatorname{var}(Y) = \sum_1^n p_k(1 - p_k).$$

Show that, for $\mathbb{E}(Y)$ fixed, $\operatorname{var}(Y)$ is a maximum when $p_1 = p_2 = \cdots = p_n$. That is to say, the variation in the sum is greatest when individuals are most alike. Is this contrary to intuition?

15. Let $\mathbf{X} = (X_1, X_2, \dots, X_n)$ be a vector of random variables. The *covariance matrix* $\mathbf{V}(\mathbf{X})$ of $\mathbf{X}$ is defined to be the symmetric n by n matrix with entries $(v_{ij} : 1 \le i, j \le n)$ given by $v_{ij} = \operatorname{cov}(X_i, X_j)$. Show that $|\mathbf{V}(\mathbf{X})| = 0$ if and only if the X_i are linearly dependent with probability one, in that $\mathbb{P}(a_1X_1 + a_2X_2 + \cdots + a_nX_n = b) = 1$ for some $\mathbf{a}$ and b. ($|\mathbf{V}|$ denotes the determinant of $\mathbf{V}$.)

16. Let X and Y be independent Bernoulli random variables with parameter $\frac{1}{2}$. Show that $X + Y$ and $|X - Y|$ are dependent though uncorrelated.

17. A secretary drops n matching pairs of letters and envelopes down the stairs, and then places the letters into the envelopes in a random order. Use indicators to show that the number X of correctly matched pairs has mean and variance 1 for all $n \ge 2$. Show that the mass function of X converges to a Poisson mass function as $n \to \infty$.

18. Let $\mathbf{X} = (X_1, X_2, \dots, X_n)$ be a vector of independent random variables each having the Bernoulli distribution with parameter p. Let $f : \{0, 1\}^n \to \mathbb{R}$ be *increasing*, which is to say that $f(\mathbf{x}) \le f(\mathbf{y})$ whenever $x_i \le y_i$ for each i.

(a) Let $e(p) = \mathbb{E}(f(\mathbf{X}))$. Show that $e(p_1) \le e(p_2)$ if $p_1 \le p_2$.

(b) **FKG inequality†.** Let f and g be increasing functions from $\{0, 1\}^n$ into $\mathbb{R}$. Show by induction on n that $\operatorname{cov}(f(\mathbf{X}), g(\mathbf{X})) \ge 0$.

19. Let $R(p)$ be the reliability function of a network G with a given source and sink, each edge of which is working with probability p, and let A be the event that there exists a working connection from source to sink. Show that

$$R(p) = \sum_\omega I_A(\omega) p^{N(\omega)}(1 - p)^{m - N(\omega)}$$

where ω is a typical realization (i.e., outcome) of the network, $N(\omega)$ is the number of working edges of ω, and m is the total number of edges of G.

Deduce that $R'(p) = \operatorname{cov}(I_A, N)/\{p(1 - p)\}$, and hence that

$$\frac{R(p)(1 - R(p))}{p(1 - p)} \le R'(p) \le \sqrt{\frac{mR(p)(1 - R(p))}{p(1 - p)}}.$$

20. Let $R(p)$ be the reliability function of a network G, each edge of which is working with probability p.

(a) Show that $R(p_1p_2) \le R(p_1)R(p_2)$ if $0 \le p_1, p_2 \le 1$.

(b) Show that $R(p^\gamma) \le R(p)^\gamma$ for all $0 \le p \le 1$ and $\gamma \ge 1$.

21. DNA fingerprinting. In a certain style of detective fiction, the sleuth is required to declare "the criminal has the unusual characteristics ...; find this person and you have your man". Assume that any given individual has these unusual characteristics with probability 10^{-7} independently of all other individuals, and that the city in question contains 10^7 inhabitants. Calculate the expected number of such people in the city.

†Named after C. Fortuin, P. Kasteleyn, and J. Ginibre (1971), but due in this form to T. E. Harris (1960).

(a) Given that the police inspector finds such a person, what is the probability that there is at least one other?
(b) If the inspector finds two such people, what is the probability that there is at least one more?
(c) How many such people need be found before the inspector can be reasonably confident that he has found them all?
(d) For the given population, how improbable should the characteristics of the criminal be, in order that he (or she) be specified uniquely?

22. In 1710, J. Arbuthnot observed that male births had exceeded female births in London for 82 successive years. Arguing that the two sexes are equally likely, and 2^{-82} is very small, he attributed this run of masculinity to Divine Providence. Let us assume that each birth results in a girl with probability $p = 0.485$, and that the outcomes of different confinements are independent of each other. Ignoring the possibility of twins (and so on), show that the probability that girls outnumber boys in $2n$ live births is no greater than $\binom{2n}{n} p^n q^n \{q/(q-p)\}$, where $q = 1 - p$. Suppose that 20,000 children are born in each of 82 successive years. Show that the probability that boys outnumber girls every year is at least 0.99. You may need Stirling's formula.

23. Consider a symmetric random walk with an absorbing barrier at N and a reflecting barrier at 0 (so that, when the particle is at 0, it moves to 1 at the next step). Let $\alpha_k(j)$ be the probability that the particle, having started at k, visits 0 exactly j times before being absorbed at N. We make the convention that, if $k = 0$, then the starting point counts as one visit. Show that

$$\alpha_k(j) = \frac{N-k}{N^2}\left(1 - \frac{1}{N}\right)^{j-1}, \qquad j \geq 1,\ 0 \leq k \leq N.$$

24. Problem of the points (3.9.4). A coin is tossed repeatedly, heads turning up with probability p on each toss. Player A wins the game if heads appears at least m times before tails has appeared n times; otherwise player B wins the game. Find the probability that A wins the game.

25. A coin is tossed repeatedly, heads appearing on each toss with probability p. A gambler starts with initial fortune k (where $0 < k < N$); he wins one point for each head and loses one point for each tail. If his fortune is ever 0 he is bankrupted, whilst if it ever reaches N he stops gambling to buy a Jaguar. Suppose that $p < \frac{1}{2}$. Show that the gambler can increase his chance of winning by doubling the stakes. You may assume that k and N are even.

What is the corresponding strategy if $p \geq \frac{1}{2}$?

26. A compulsive gambler is never satisfied. At each stage he wins £1 with probability p and loses £1 otherwise. Find the probability that he is ultimately bankrupted, having started with an initial fortune of £k.

27. Range of random walk. Let $\{X_n : n \geq 1\}$ be independent, identically distributed random variables taking integer values. Let $S_0 = 0$, $S_n = \sum_{i=1}^{n} X_i$. The *range* R_n of $S_0, S_1, \dots, S_n$ is the number of distinct values taken by the sequence. Show that $\mathbb{P}(R_n = R_{n-1}+1) = \mathbb{P}(S_1 S_2 \cdots S_n \neq 0)$, and deduce that, as $n \to \infty$,

$$\frac{1}{n}\mathbb{E}(R_n) \to \mathbb{P}(S_k \neq 0 \text{ for all } k \geq 1).$$

Hence show that, for the simple random walk, $n^{-1}\mathbb{E}(R_n) \to |p - q|$ as $n \to \infty$.

28. Arc sine law for maxima. Consider a symmetric random walk S starting from the origin, and let $M_n = \max\{S_i : 0 \leq i \leq n\}$. Show that, for $i = 2k, 2k+1$, the probability that the walk reaches M_{2n} for the first time at time i equals $\frac{1}{2}\mathbb{P}(S_{2k} = 0)\mathbb{P}(S_{2n-2k} = 0)$.

29. Let S be a symmetric random walk with $S_0 = 0$, and let N_n be the number of points that have been visited by S exactly once up to time n. Show that $\mathbb{E}(N_n) = 2$.

30. Family planning. Consider the following fragment of verse entitled 'Note for the scientist'.

People who have three daughters try for more,
And then its fifty–fifty they'll have four,
Those with a son or sons will let things be,
Hence all these surplus women, QED.

(a) What do you think of the argument?

(b) Show that the mean number of children of either sex in a family whose fertile parents have followed this policy equals 1. (You should assume that each delivery yields exactly one child whose sex is equally likely to be male or female.) Discuss.

31. Let $\beta > 1$, let $p_1, p_2, \dots$ denote the prime numbers, and let $N(1), N(2), \dots$ be independent random variables, $N(i)$ having mass function $\mathbb{P}(N(i) = k) = (1-\gamma_i)\gamma_i^k$ for $k \geq 0$, where $\gamma_i = p_i^{-\beta}$ for all i. Show that $M = \prod_{i=1}^{\infty} p_i^{N(i)}$ is a random integer with mass function $\mathbb{P}(M = m) = Cm^{-\beta}$ for $m \geq 1$ (this may be called the *Dirichlet distribution*), where C is a constant satisfying

$$C = \prod_{i=1}^{\infty}\left(1 - \frac{1}{p_i^{\beta}}\right) = \left(\sum_{m=1}^{\infty} \frac{1}{m^{\beta}}\right)^{-1}.$$

32. $N+1$ plates are laid out around a circular dining table, and a hot cake is passed between them in the manner of a symmetric random walk: each time it arrives on a plate, it is tossed to one of the two neighbouring plates, each possibility having probability $\frac{1}{2}$. The game stops at the moment when the cake has visited every plate at least once. Show that, with the exception of the plate where the cake began, each plate has probability $1/N$ of being the last plate visited by the cake.

33. Simplex algorithm. There are $\binom{n}{m}$ points ranked in order of merit with no matches. You seek to reach the best, B. If you are at the jth best, you step to any one of the $j-1$ better points, with equal probability of stepping to each. Let r_j be the expected number of steps to reach B from the jth best vertex. Show that $r_j = \sum_{k=1}^{j-1} k^{-1}$. Give an asymptotic expression for the expected time to reach B from the worst vertex, for large m, n.

34. Dimer problem. There are n unstable molecules in a row, $m_1, m_2, \dots, m_n$. One of the $n-1$ pairs of neighbours, chosen at random, combines to form a stable dimer; this process continues until there remain U_n isolated molecules no two of which are adjacent. Show that the probability that m_1 remains isolated is $\sum_{r=0}^{n-1}(-1)^r/r! \to e^{-1}$ as $n \to \infty$. Deduce that $\lim_{n\to\infty} n^{-1}\mathbb{E}U_n = e^{-2}$.

35. Poisson approximation. Let $\{I_r : 1 \leq r \leq n\}$ be independent Bernoulli random variables with respective parameters $\{p_r : 1 \leq r \leq n\}$ satisfying $p_r \leq c < 1$ for all r and some c. Let $\lambda = \sum_{r=1}^{n} p_r$ and $X = \sum_{r=1}^{n} X_r$. Show that

$$\mathbb{P}(X = k) = \frac{\lambda^k e^{-\lambda}}{k!}\left\{1 + \mathrm{O}\left(\lambda \max_r p_r + \frac{k^2}{\lambda}\max_r p_r\right)\right\}.$$

36. Sampling. The length of the tail of the rth member of a troop of N chimeras is x_r. A random sample of n chimeras is taken (without replacement) and their tails measured. Let I_r be the indicator of the event that the rth chimera is in the sample. Set

$$X_r = x_r I_r, \quad \overline{Y} = \frac{1}{n}\sum_{r=1}^{N} X_r, \quad \overline{x} = \frac{1}{N}\sum_{r=1}^{N} x_r, \quad \sigma^2 = \frac{1}{N}\sum_{r=1}^{N}(x_r - \overline{x})^2.$$

Show that $\mathbb{E}(\overline{Y}) = \mu$, and $\operatorname{var}(\overline{Y}) = (N-n)\sigma^2/\{n(N-1)\}$.

37. Berkson's fallacy. Any individual in a group G contracts a certain disease C with probability γ; such individuals are hospitalized with probability c. Independently of this, anyone in G may be in hospital with probability a, for some other reason. Let X be the number in hospital, and Y the number in hospital who have C (including those with C admitted for any other reason). Show that the correlation between X and Y is

$$\rho(X, Y) = \sqrt{\frac{\gamma p}{1 - \gamma p} \cdot \frac{(1-a)(1-\gamma c)}{a + \gamma c - a\gamma c}},$$

where $p = a + c - ac$.

It has been stated erroneously that, when $\rho(X, Y)$ is near unity, this is evidence for a causal relation between being in G and contracting C.

38. A telephone sales company attempts repeatedly to sell new kitchens to each of the N families in a village. Family i agrees to buy a new kitchen after it has been solicited K_i times, where the K_i are independent identically distributed random variables with mass function $f(n) = \mathbb{P}(K_i = n)$. The value ∞ is allowed, so that $f(\infty) \geq 0$. Let X_n be the number of kitchens sold at the nth round of solicitations, so that $X_n = \sum_{i=1}^{N} I_{\{K_i = n\}}$. Suppose that N is a random variable with the Poisson distribution with parameter ν.

(a) Show that the X_n are independent random variables, X_r having the Poisson distribution with parameter $\nu f(r)$.

(b) The company loses heart after the Tth round of calls, where $T = \inf\{n : X_n = 0\}$. Let $S = X_1 + X_2 + \cdots + X_T$ be the number of solicitations made up to time T. Show further that $\mathbb{E}(S) = \nu\mathbb{E}(F(T))$ where $F(k) = f(1) + f(2) + \cdots + f(k)$.

39. A particle performs a random walk on the non-negative integers as follows. When at the point n (> 0) its next position is uniformly distributed on the set $\{0, 1, 2, \ldots, n+1\}$. When it hits 0 for the first time, it is absorbed. Suppose it starts at the point a.

(a) Find the probability that its position never exceeds a, and prove that, with probability 1, it is absorbed ultimately.

(b) Find the probability that the final step of the walk is from 1 to 0 when $a = 1$.

(c) Find the expected number of steps taken before absorption when $a = 1$.

40. Let G be a finite graph with neither loops nor multiple edges, and write d_v for the degree of the vertex v. An *independent set* is a set of vertices no pair of which is joined by an edge. Let $\alpha(G)$ be the size of the largest independent set of G. Use the probabilistic method to show that $\alpha(G) \geq \sum_v 1/(d_v + 1)$. [This conclusion is sometimes referred to as *Turán's theorem*.]