5

Generating functions and their applications

Summary. A key method for studying distributions is via transforms such as the probability generating function of a discrete random variable, or the moment generating function and characteristic function of a general random variable. Such transforms are particularly suited to the study of sums of independent random variables, and their areas of application include renewal theory, random walks, and branching processes. The inversion theorem tells how to obtain the distribution function from knowledge of its characteristic function. The continuity theorem allows us to use characteristic functions in studying limits of random variables. Two principal applications are to the law of large numbers and the central limit theorem. The theory of large deviations concerns the estimation of probabilities of 'exponentially unlikely' events.

5.1 Generating functions

A sequence $a = \{a_i : i = 0, 1, 2, \dots\}$ of real numbers may contain a lot of information. One concise way of storing this information is to wrap up the numbers together in a 'generating function'. For example, the (ordinary) *generating function* of the sequence a is the function G_a defined by

(1) $$G_a(s) = \sum_{i=0}^{\infty} a_i s^i \quad \text{for } s \in \mathbb{R} \text{ for which the sum converges.}$$

The sequence a may in principle be reconstructed from the function G_a by setting $a_i = G_a^{(i)}(0)/i!$, where $f^{(i)}$ denotes the ith derivative of the function f. In many circumstances it is easier to work with the generating function G_a than with the original sequence a.

(2) Example. De Moivre's theorem. The sequence $a_n = (\cos\theta + i\sin\theta)^n$ has generating function

$$G_a(s) = \sum_{n=0}^{\infty} \big[s(\cos\theta + i\sin\theta)\big]^n = \frac{1}{1 - s(\cos\theta + i\sin\theta)}$$

if $|s| < 1$; here $i = \sqrt{-1}$. It is easily checked by examining the coefficient of s^n that

$$\big[1 - s(\cos\theta + i\sin\theta)\big]\sum_{n=0}^{\infty} s^n\big[\cos(n\theta) + i\sin(n\theta)\big] = 1$$

when $|s| < 1$. Thus

$$\sum_{n=0}^{\infty} s^n\big[\cos(n\theta) + i\sin(n\theta)\big] = \frac{1}{1 - s(\cos\theta + i\sin\theta)}$$

if $|s| < 1$. Equating the coefficients of s^n we obtain the well-known fact that $\cos(n\theta) + i\sin(n\theta) = (\cos\theta + i\sin\theta)^n$. ●

There are several different types of generating function, of which G_a is perhaps the simplest. Another is the *exponential generating function* E_a given by

(3) $$E_a(s) = \sum_{i=0}^{\infty} \frac{a_i s^i}{i!} \quad \text{for } s \in \mathbb{R} \text{ for which the sum converges.}$$

Whilst such generating functions have many uses in mathematics, the ordinary generating function (1) is of greater value when the a_i are probabilities. This is because 'convolutions' are common in probability theory, and (ordinary) generating functions provide an invaluable tool for studying them.

(4) Convolution. The *convolution* of the real sequences $a = \{a_i : i \geq 0\}$ and $b = \{b_i : i \geq 0\}$ is the sequence $c = \{c_i : i \geq 0\}$ defined by

(5) $$c_n = a_0 b_n + a_1 b_{n-1} + \cdots + a_n b_0;$$

we write $c = a * b$. If a and b have generating functions G_a and G_b, then the generating function of c is

(6) $$\begin{aligned} G_c(s) &= \sum_{n=0}^{\infty} c_n s^n = \sum_{n=0}^{\infty}\left(\sum_{i=0}^{n} a_i b_{n-i}\right)s^n \\ &= \sum_{i=0}^{\infty} a_i s^i \sum_{n=i}^{\infty} b_{n-i} s^{n-i} = G_a(s)G_b(s). \end{aligned}$$

Thus we learn that, if $c = a * b$, then $G_c(s) = G_a(s)G_b(s)$; convolutions are numerically complicated operations, and it is often easier to work with generating functions.

(7) Example. The combinatorial identity

$$\sum_i \binom{n}{i}^2 = \binom{2n}{n}$$

may be obtained as follows. The left-hand side is the convolution of the sequence $a_i = \binom{n}{i}$, $i = 0, 1, 2, \ldots$, with itself. However, $G_a(s) = \sum_i \binom{n}{i} s^i = (1+s)^n$, so that

$$G_{a*a}(s) = G_a(s)^2 = (1+s)^{2n} = \sum_i \binom{2n}{i} s^i.$$

Equating the coefficients of s^n yields the required identity. ●

(8) Example. Let X and Y be independent random variables having the Poisson distribution with parameters λ and μ respectively. What is the distribution of $Z = X + Y$?

Solution. We have from equation (3.8.2) that the mass function of Z is the convolution of the mass functions of X and Y, $f_Z = f_X * f_Y$. The generating function of the sequence $\{f_X(i) : i \geq 0\}$ is

(9)
$$G_X(s) = \sum_{i=0}^{\infty} \frac{\lambda^i e^{-\lambda}}{i!} s^i = e^{\lambda(s-1)},$$

and similarly $G_Y(s) = e^{\mu(s-1)}$. Hence the generating function G_Z of $\{f_Z(i) : i \geq 0\}$ satisfies $G_Z(s) = G_X(s)G_Y(s) = \exp[(\lambda+\mu)(s-1)]$, which we recognize from (9) as the generating function of the Poisson mass function with parameter $\lambda + \mu$. ●

The last example is canonical: generating functions provide a basic technique for dealing with sums of independent random variables. With this example in mind, we make an important definition. Suppose that X is a discrete random variable taking values in the non-negative integers $\{0, 1, 2, \dots\}$; its distribution is specified by the sequence of probabilities $f(i) = \mathbb{P}(X = i)$.

(10) Definition. The (probability) generating function of the random variable X is defined to be the generating function $G(s) = \mathbb{E}(s^X)$ of its probability mass function.

Note that G does indeed generate the sequence $\{f(i) : i \geq 0\}$ since

$$\mathbb{E}(s^X) = \sum_i s^i \mathbb{P}(X = i) = \sum_i s^i f(i)$$

by Lemma (3.3.3). We write G_X when we wish to stress the role of X. If X takes values in the non-negative integers, its generating function G_X converges at least when $|s| \leq 1$ and sometimes in a larger interval. Generating functions can be defined for random variables taking negative as well as positive integer values. Such generating functions generally converge for values of s satisfying $\alpha < |s| < \beta$ for some α, β such that $\alpha \leq 1 \leq \beta$. We shall make occasional use of such generating functions, but we do not develop their theory systematically.

In advance of giving examples and applications of the method of generating functions, we recall some basic properties of power series. Let $G(s) = \sum_0^\infty a_i s^i$ where $a = \{a_i : i \geq 0\}$ is a real sequence.

(11) Convergence. There exists a *radius of convergence* R (≥ 0) such that the sum converges absolutely if $|s| < R$ and diverges if $|s| > R$. The sum is uniformly convergent on sets of the form $\{s : |s| \leq R'\}$ for any $R' < R$.

(12) Differentiation. $G_a(s)$ may be differentiated or integrated term by term any number of times at points s satisfying $|s| < R$.

(13) Uniqueness. If $G_a(s) = G_b(s)$ for $|s| < R'$ where $0 < R' \leq R$ then $a_n = b_n$ for all n. Furthermore

(14)
$$a_n = \frac{1}{n!} G_a^{(n)}(0).$$

(15) Abel's theorem. If $a_i \geq 0$ for all i and $G_a(s)$ is finite for $|s| < 1$, then $\lim_{s\uparrow 1} G_a(s) = \sum_{i=0}^{\infty} a_i$, whether the sum is finite or equals $+\infty$. This standard result is useful when the radius of convergence R satisfies $R = 1$, since then one has no *a priori* right to take the limit as $s \uparrow 1$.

Returning to the discrete random variable X taking values in $\{0, 1, 2, \dots\}$ we have that $G(s) = \sum_0^{\infty} s^i \mathbb{P}(X = i)$, so that

(16) $$G(0) = \mathbb{P}(X = 0), \quad G(1) = 1.$$

In particular, the radius of convergence of a probability generating function is at least 1. Here are some examples of probability generating functions.

(17) Examples.

(a) **Constant variables.** If $\mathbb{P}(X = c) = 1$ then $G(s) = \mathbb{E}(s^X) = s^c$.

(b) **Bernoulli variables.** If $\mathbb{P}(X = 1) = p$ and $\mathbb{P}(X = 0) = 1 - p$ then

$$G(s) = \mathbb{E}(s^X) = (1 - p) + ps.$$

(c) **Geometric distribution.** If X is geometrically distributed with parameter p, so that $\mathbb{P}(X = k) = p(1 - p)^{k-1}$ for $k \geq 1$, then

$$G(s) = \mathbb{E}(s^X) = \sum_{k=1}^{\infty} s^k p(1 - p)^{k-1} = \frac{ps}{1 - s(1 - p)}.$$

(d) **Poisson distribution.** If X is Poisson distributed with parameter λ then

$$G(s) = \mathbb{E}(s^X) = \sum_{k=0}^{\infty} s^k \frac{\lambda^k}{k!} e^{-\lambda} = e^{\lambda(s-1)}.$$ ●

Generating functions are useful when working with integer-valued random variables. Problems arise when random variables take negative or non-integer values. Later in this chapter we shall see how to construct another function, called a 'characteristic function', which is very closely related to G_X but which exists for all random variables regardless of their types.

There are two major applications of probability generating functions: in calculating moments, and in calculating the distributions of *sums* of independent random variables. We begin with moments.

(18) Theorem. *If X has generating function $G(s)$ then*

(a) $\mathbb{E}(X) = G'(1)$,

(b) *more generally,* $\mathbb{E}\big[X(X - 1) \cdots (X - k + 1)\big] = G^{(k)}(1)$.

Of course, $G^{(k)}(1)$ is shorthand for $\lim_{s\uparrow 1} G^{(k)}(s)$ whenever the radius of convergence of G is 1. The quantity $\mathbb{E}[X(X - 1) \cdots (X - k + 1)]$ is known as the *kth factorial moment* of X.

Proof of (b). Take $s < 1$ and calculate the kth derivative of G to obtain

$$G^{(k)}(s) = \sum_i s^{i-k} i(i - 1) \cdots (i - k + 1) f(i) = \mathbb{E}\big[s^{X-k} X(X - 1) \cdots (X - k + 1)\big].$$

Let $s \uparrow 1$ and use Abel's theorem (15) to obtain

$$G^{(k)}(s) \to \sum_i i(i-1)\cdots(i-k+1)f(i) = \mathbb{E}\big[X(X-1)\cdots(X-k+1)\big]. \qquad \blacksquare$$

In order to calculate the variance of X in terms of G, we proceed as follows:

(19)
$$\begin{aligned}\operatorname{var}(X) &= \mathbb{E}(X^2) - \mathbb{E}(X)^2 = \mathbb{E}\big(X(X-1)+X\big) - \mathbb{E}(X)^2 \\ &= \mathbb{E}\big(X(X-1)\big) + \mathbb{E}(X) - \mathbb{E}(X)^2 = G''(1) + G'(1) - G'(1)^2.\end{aligned}$$

Exercise. Find the means and variances of the distributions in (17) by this method.

(20) Example. Recall the hypergeometric distribution (3.11.10) with mass function

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

Then $G(s) = \sum_k s^k f(k)$, which can be recognized as the coefficient of x^n in

$$Q(s,x) = (1+sx)^b(1+x)^{N-b}\bigg/\binom{N}{n}.$$

Hence the mean $G'(1)$ is the coefficient of x^n in

$$\frac{\partial Q}{\partial s}(1,x) = xb(1+x)^{N-1}\bigg/\binom{N}{n}$$

and so $G'(1) = bn/N$. Now calculate the variance yourself (*exercise*). ●

If you are more interested in the moments of X than in its mass function, you may prefer to work not with G_X but with the function M_X defined by $M_X(t) = G_X(e^t)$. This change of variable is convenient for the following reason. Expanding $M_X(t)$ as a power series in t, we obtain

(21)
$$\begin{aligned}M_X(t) &= \sum_{k=0}^{\infty} e^{tk}\mathbb{P}(X=k) = \sum_{k=0}^{\infty}\sum_{n=0}^{\infty}\frac{(tk)^n}{n!}\mathbb{P}(X=k) \\ &= \sum_{n=0}^{\infty}\frac{t^n}{n!}\left(\sum_{k=0}^{\infty}k^n\mathbb{P}(X=k)\right) = \sum_{n=0}^{\infty}\frac{t^n}{n!}\mathbb{E}(X^n),\end{aligned}$$

the exponential generating function of the moments $\mathbb{E}(X^0), \mathbb{E}(X^1), \dots$ of X. The function M_X is called the *moment generating function* of the random variable X. We have assumed in (21) that the series in question converge. Some complications can arise in using moment generating functions unless the series $\sum_n t^n\mathbb{E}(X^n)/n!$ has a strictly positive radius of convergence.

(22) Example. We have from (9) that the moment generating function of the Poisson distribution with parameter λ is $M(t) = \exp[\lambda(e^t - 1)]$. ●

We turn next to sums and convolutions. Much of probability theory is concerned with sums of random variables. To study such a sum we need a useful way of describing its distribution in terms of the distributions of its summands, and generating functions prove to be an invaluable asset in this respect. The formula in Theorem (3.8.1) for the mass function of the sum of two independent discrete variables, $\mathbb{P}(X+Y=z)=\sum_x \mathbb{P}(X=x)\mathbb{P}(Y=z-x)$, involves a complicated calculation; the corresponding generating functions provide a more economical way of specifying the distribution of this sum.

(23) Theorem. *If X and Y are independent then $G_{X+Y}(s)=G_X(s)G_Y(s)$.*

Proof. The direct way of doing this is to use equation (3.8.2) to find that $f_Z=f_X*f_Y$, so that the generating function of $\{f_Z(i):i\geq 0\}$ is the product of the generating functions of $\{f_X(i):i\geq 0\}$ and $\{f_Y(i):i\geq 0\}$, by (4). Alternatively, $g(X)=s^X$ and $h(Y)=s^Y$ are independent, by Theorem (3.2.3), and so $\mathbb{E}(g(X)h(Y))=\mathbb{E}(g(X))\mathbb{E}(h(Y))$, as required. ■

(24) Example. Binomial distribution. Let $X_1, X_2, \dots, X_n$ be independent Bernoulli variables, parameter p, with sum $S=X_1+X_2+\dots+X_n$. Each X_i has generating function $G(s)=qs^0+ps^1=q+ps$, where $q=1-p$. Apply (23) repeatedly to find that the $\text{bin}(n,p)$ variable S has generating function

$$G_S(s)=[G(s)]^n=(q+ps)^n.$$

The sum S_1+S_2 of two independent variables, $\text{bin}(n,p)$ and $\text{bin}(m,p)$ respectively, has generating function

$$G_{S_1+S_2}(s)=G_{S_1}(s)G_{S_2}(s)=(q+ps)^{m+n}$$

and is thus $\text{bin}(m+n,p)$. This was Problem (3.11.8). ●

Theorem (23) tells us that the sum $S=X_1+X_2+\dots+X_n$ of independent variables taking values in the non-negative integers has generating function given by

$$G_S=G_{X_1}G_{X_2}\cdots G_{X_n}.$$

If n is itself the outcome of a random experiment then the answer is not quite so simple.

(25) Theorem. *If $X_1, X_2, \dots$ is a sequence of independent identically distributed random variables with common generating function G_X, and N (≥ 0) is a random variable which is independent of the X_i and has generating function G_N, then $S=X_1+X_2+\dots+X_N$ has generating function given by*

(26) $$G_S(s)=G_N(G_X(s)).$$

This has many important applications, one of which we shall meet in Section 5.4. It is an example of a process known as *compounding* with respect to a parameter. Formula (26) is easily remembered; possible confusion about the order in which the functions G_N and G_X are compounded is avoided by remembering that if $\mathbb{P}(N=n)=1$ then $G_N(s)=s^n$ and $G_S(s)=G_X(s)^n$. Incidentally, we adopt the usual convention that, in the case when $N=0$, the sum $X_1+X_2+\dots+X_N$ is the 'empty' sum, and equals 0 also.

Proof. Use conditional expectation and Theorem (3.7.4) to find that

$$\begin{aligned} G_S(s) = \mathbb{E}(s^S) = \mathbb{E}\big(\mathbb{E}(s^S \mid N)\big) &= \sum_n \mathbb{E}(s^S \mid N = n)\mathbb{P}(N = n) \\ &= \sum_n \mathbb{E}(s^{X_1+\cdots+X_n})\mathbb{P}(N = n) \\ &= \sum_n \mathbb{E}(s^{X_1})\cdots\mathbb{E}(s^{X_n})\mathbb{P}(N = n) \quad \text{by independence} \\ &= \sum_n G_X(s)^n \mathbb{P}(N = n) = G_N(G_X(s)). \end{aligned}$$

■

(27) **Example (3.7.5) revisited.** A hen lays N eggs, where N is Poisson distributed with parameter λ. Each egg hatches with probability p, independently of all other eggs. Let K be the number of chicks. Then $K = X_1 + X_2 + \cdots + X_N$ where $X_1, X_2, \ldots$ are independent Bernoulli variables with parameter p. How is K distributed? Clearly

$$G_N(s) = \sum_{n=0}^{\infty} s^n \frac{\lambda^n}{n!} e^{-\lambda} = e^{\lambda(s-1)}, \qquad G_X(s) = q + ps,$$

and so $G_K(s) = G_N(G_X(s)) = e^{\lambda p(s-1)}$, which, by comparison with G_N, we see to be the generating function of a Poisson variable with parameter λp. ●

Just as information about a mass function can be encapsulated in a generating function, so may joint mass functions be similarly described.

(28) **Definition.** The **joint (probability) generating function** of variables X_1 and X_2 taking values in the non-negative integers is defined by

$$G_{X_1,X_2}(s_1, s_2) = \mathbb{E}(s_1^{X_1} s_2^{X_2}).$$

There is a similar definition for the joint generating function of an arbitrary family of random variables. Joint generating functions have important uses also, one of which is the following characterization of independence.

(29) **Theorem.** *Random variables X_1 and X_2 are independent if and only if*

$$G_{X_1,X_2}(s_1, s_2) = G_{X_1}(s_1)G_{X_2}(s_2) \quad \textit{for all } s_1 \textit{ and } s_2.$$

Proof. If X_1 and X_2 are independent then so are $g(X_1) = s_1^{X_1}$ and $h(X_2) = s_2^{X_2}$; then proceed as in the proof of (23). To prove the converse, equate the coefficients of terms such as $s_1^i s_2^j$ to deduce after some manipulation that $\mathbb{P}(X_1 = i,\ X_2 = j) = \mathbb{P}(X_1 = i)\mathbb{P}(X_2 = j)$. ■

So far we have only considered random variables X which take finite values only, and consequently their generating functions G_X satisfy $G_X(1) = 1$. In the near future we shall encounter variables which can take the value $+\infty$ (see the first passage time T_0 of Section 5.3 for example). For such variables X we note that $G_X(s) = \mathbb{E}(s^X)$ converges so long as $|s| < 1$, and furthermore

(30) $$\lim_{s\uparrow 1} G_X(s) = \sum_k \mathbb{P}(X = k) = 1 - \mathbb{P}(X = \infty).$$

We can no longer find the moments of X in terms of G_X; of course, they all equal $+\infty$. If $\mathbb{P}(X = \infty) > 0$ then we say that X is *defective* with defective distribution function F_X.

Exercises for Section 5.1

1. Find the generating functions of the following mass functions, and state where they converge. Hence calculate their means and variances.
(a) $f(m) = \binom{n+m-1}{m} p^n (1-p)^m$, for $m \geq 0$.
(b) $f(m) = \{m(m+1)\}^{-1}$, for $m \geq 1$.
(c) $f(m) = (1-p)p^{|m|}/(1+p)$, for $m = \ldots, -1, 0, 1, \ldots$.
The constant p satisfies $0 < p < 1$.

2. Let X (≥ 0) have probability generating function G and write $t(n) = \mathbb{P}(X > n)$ for the 'tail' probabilities of X. Show that the generating function of the sequence $\{t(n) : n \geq 0\}$ is $T(s) = (1 - G(s))/(1-s)$. Show that $\mathbb{E}(X) = T(1)$ and $\text{var}(X) = 2T'(1) + T(1) - T(1)^2$.

3. Let $G_{X,Y}(s,t)$ be the joint probability generating function of X and Y. Show that $G_X(s) = G_{X,Y}(s,1)$ and $G_Y(t) = G_{X,Y}(1,t)$. Show that

$$\mathbb{E}(XY) = \left.\frac{\partial^2}{\partial s\, \partial t} G_{X,Y}(s,t)\right|_{s=t=1}.$$

4. Find the joint generating functions of the following joint mass functions, and state for what values of the variables the series converge.
(a) $f(j,k) = (1-\alpha)(\beta-\alpha)\alpha^j \beta^{k-j-1}$, for $0 \leq k \leq j$, where $0 < \alpha < 1$, $\alpha < \beta$.
(b) $f(j,k) = (e-1)e^{-(2k+1)}k^j/j!$, for $j, k \geq 0$.
(c) $f(j,k) = \binom{k}{j} p^{j+k}(1-p)^{k-j} / \left[k \log\{1/(1-p)\}\right]$, for $0 \leq j \leq k$, $k \geq 1$, where $0 < p < 1$.
Deduce the marginal probability generating functions and the covariances.

5. A coin is tossed n times, and heads turns up with probability p on each toss. Assuming the usual independence, show that the joint probability generating function of the numbers H and T of heads and tails is $G_{H,T}(x,y) = \{px + (1-p)y\}^n$. Generalize this conclusion to find the joint probability generating function of the multinomial distribution of Exercise (3.5.1).

6. Let X have the binomial distribution $\text{bin}(n, U)$, where U is uniform on $(0,1)$. Show that X is uniformly distributed on $\{0, 1, 2, \ldots, n\}$.

7. Show that

$$G(x,y,z,w) = \tfrac{1}{8}(xyzw + xy + yz + zw + zx + yw + xz + 1)$$

is the joint generating function of four variables that are pairwise and triplewise independent, but are nevertheless *not* independent.

8. Let $p_r > 0$ and $a_r \in \mathbb{R}$ for $1 \leq r \leq n$. Which of the following is a moment generating function, and for what random variable?

$$\text{(a) } M(t) = 1 + \sum_{r=1}^{n} p_r t^r, \qquad \text{(b) } M(t) = \sum_{r=1}^{n} p_r e^{a_r t}.$$

9. Let G_1 and G_2 be probability generating functions, and suppose that $0 \leq \alpha \leq 1$. Show that $G_1 G_2$, and $\alpha G_1 + (1-\alpha) G_2$ are probability generating functions. Is $G(\alpha s)/G(\alpha)$ necessarily a probability generating function?

5.2 Some applications

Generating functions provide a powerful tool, particularly in the presence of difference equations and convolutions. This section contains a variety of examples of this tool in action.

(1) Example. Problem of the points†. A coin is tossed repeatedly and heads turns up with probability p on each toss. Player A wins if m heads appear before n tails, and player B wins otherwise. We have seen, in Exercise (3.9.4) and Problem (3.11.24), two approaches to the problem of determining the probability that A wins. It is elementary, by conditioning on the outcome of the first toss, that the probability p_{mn}, that A wins, satisfies

(2) $$p_{mn} = pp_{m-1,n} + qp_{m,n-1}, \quad \text{for} \quad m, n \geq 1,$$

where $p + q = 1$. The boundary conditions are $p_{m0} = 0$, $p_{0n} = 1$ for $m, n > 0$. We may solve equation (2) by introducing the generating function

$$G(x, y) = \sum_{m=0}^{\infty} \sum_{n=0}^{\infty} p_{mn} x^m y^n$$

subject to the convention that $p_{00} = 0$. Multiplying throughout (2) by $x^m y^n$ and summing over $m, n \geq 1$, we obtain

(3) $$\begin{aligned} G(x, y) - \sum_{m=1}^{\infty} p_{m0} x^m - \sum_{n=1}^{\infty} p_{0n} y^n \\ = px \sum_{m,n=1}^{\infty} p_{m-1,n} x^{m-1} y^n + qy \sum_{m,n=1}^{\infty} p_{m,n-1} x^m y^{n-1}, \end{aligned}$$

and hence, using the boundary conditions,

$$G(x, y) - \frac{y}{1-y} = pxG(x, y) + qy\left(G(x, y) - \frac{y}{1-y}\right), \quad |y| < 1.$$

Therefore,

(4) $$G(x, y) = \frac{y(1 - qy)}{(1 - y)(1 - px - qy)},$$

from which one may derive the required information by expanding in powers of x and y and finding the coefficient of $x^m y^n$. A cautionary note: in passing from (2) to (3), one should be very careful with the limits of the summations. ●

(5) Example. Matching revisited. The famous (mis)matching problem of Example (3.4.3) involves the random placing of n different letters into n differently addressed envelopes. What is the probability p_n that no letter is placed in the correct envelope? Let M be the event that

†First recorded by Pacioli in 1494, and eventually solved by Pascal in 1654. Our method is due to Laplace.

the first letter is put into its correct envelope, and let N be the event that no match occurs. Then

(6) $$p_n = \mathbb{P}(N) = \mathbb{P}(N \mid M^c)\mathbb{P}(M^c),$$

where $\mathbb{P}(M^c) = 1 - n^{-1}$. It is convenient to think of $\alpha_n = \mathbb{P}(N \mid M^c)$ in the following way. It is the probability that, given $n-2$ pairs of matching white letters and envelopes together with a non-matching red letter and blue envelope, there are no colour matches when the letters are inserted randomly into the envelopes. Either the red letter is placed into the blue envelope or it is not, and a consideration of these two cases gives that

(7) $$\alpha_n = \frac{1}{n-1}p_{n-2} + \left(1 - \frac{1}{n-1}\right)\alpha_{n-1}.$$

Combining (6) and (7) we obtain, for $n \geq 3$,

(8) $$\begin{aligned} p_n &= \left(1 - \frac{1}{n}\right)\alpha_n = \left(1 - \frac{1}{n}\right)\left[\frac{1}{n-1}p_{n-2} + \left(1 - \frac{1}{n-1}\right)\alpha_{n-1}\right] \\ &= \left(1 - \frac{1}{n}\right)\left(\frac{1}{n-1}p_{n-2} + p_{n-1}\right) = \frac{1}{n}p_{n-2} + \left(1 - \frac{1}{n}\right)p_{n-1}, \end{aligned}$$

a difference relation subject to the boundary conditions $p_1 = 0$, $p_2 = \frac{1}{2}$. We may solve this difference relation by using the generating function

(9) $$G(s) = \sum_{n=1}^{\infty} p_n s^n.$$

We multiply throughout (8) by ns^{n-1} and sum over all suitable values of n to obtain

$$\sum_{n=3}^{\infty} ns^{n-1}p_n = s\sum_{n=3}^{\infty} s^{n-2}p_{n-2} + s\sum_{n=3}^{\infty}(n-1)s^{n-2}p_{n-1}$$

which we recognize as yielding

$$G'(s) - p_1 - 2p_2 s = sG(s) + s[G'(s) - p_1]$$

or $(1-s)G'(s) = sG(s) + s$, since $p_1 = 0$ and $p_2 = \frac{1}{2}$. This differential equation is easily solved subject to the boundary condition $G(0) = 0$ to obtain $G(s) = (1-s)^{-1}e^{-s} - 1$. Expanding as a power series in s and comparing with (9), we arrive at the conclusion

(10) $$p_n = \frac{(-1)^n}{n!} + \frac{(-1)^{n-1}}{(n-1)!} + \cdots + \frac{(-1)}{1!} + 1, \quad \text{for} \quad n \geq 1,$$

as in the conclusion of Example (3.4.3) with $r = 0$. ●

(11) Example. Matching and occupancy. The matching problem above is one of the simplest of a class of problems involving putting objects randomly into containers. In a general approach to such questions, we suppose that we are given a collection $\mathcal{A} = \{A_i : 1 \le i \le n\}$ of events, and we ask for properties of the random number X of these events which occur (in the previous example, A_i is the event that the ith letter is in the correct envelope, and X is the number of correctly placed letters). The problem is to express the mass function of X in terms of probabilities of the form $\mathbb{P}(A_{i_1} \cap A_{i_2} \cap \cdots \cap A_{i_m})$. We introduce the notation

$$S_m = \sum_{i_1 < i_2 < \cdots < i_m} \mathbb{P}(A_{i_1} \cap A_{i_2} \cap \cdots \cap A_{i_m}),$$

the sum of the probabilities of the intersections of exactly m of the events in question. We make the convention that $S_0 = 1$. It is easily seen as follows that

(12) $$S_m = \mathbb{E}\binom{X}{m},$$

the mean value of the (random) binomial coefficient $\binom{X}{m}$: writing N_m for the number of sub-families of $\mathcal{A}$ having size m, all of whose component events occur, we have that

$$S_m = \sum_{i_1 < \cdots < i_m} \mathbb{E}(I_{A_{i_1}} I_{A_{i_2}} \cdots I_{A_{i_m}}) = \mathbb{E}(N_m),$$

whereas $N_m = \binom{X}{m}$. It follows from (12) that

(13) $$S_m = \sum_{i=0}^{n} \binom{i}{m} \mathbb{P}(X = i).$$

We introduce the generating functions

$$G_S(x) = \sum_{m=0}^{n} x^m S_m, \qquad G_X(x) = \sum_{i=0}^{n} x^i \mathbb{P}(X = i),$$

and we then multiply throughout (13) by x^m and sum over m, obtaining

$$G_S(x) = \sum_i \mathbb{P}(X = i) \sum_m x^m \binom{i}{m} = \sum_i (1+x)^i \mathbb{P}(X = i) = G_X(1+x).$$

Hence $G_X(x) = G_S(x-1)$, and equating coefficients of x^i yields

(14) $$\mathbb{P}(X = i) = \sum_{j=i}^{n} (-1)^{j-i} \binom{j}{i} S_j \quad \text{for} \quad 0 \le i \le n.$$

This formula, sometimes known as 'Waring's theorem', is a complete generalization of certain earlier results, including (10). It may be derived without using generating functions, but at considerable personal cost. ●

(15) **Example. Recurrent events.** Meteorites fall from the sky, your car runs out of fuel, there is a power failure, you fall ill. Each such event recurs at regular or irregular intervals; one cannot generally predict just when such an event will happen next, but one may be prepared to hazard guesses. A simplistic mathematical model is the following. We call the happening in question H, and suppose that, at each time point $1, 2, \dots$, either H occurs or H does not occur. We write X_1 for the first time at which H occurs, $X_1 = \min\{n : H \text{ occurs at time } n\}$, and X_m for the time which elapses between the $(m-1)$th and mth occurrence of H. Thus the mth occurrence of H takes place at time

$$T_m = X_1 + X_2 + \cdots + X_m. \tag{16}$$

Here are our main assumptions. We assume that the 'inter-occurrence' times $X_1, X_2, \dots$ are independent random variables taking values in $\{1, 2, \dots\}$, and furthermore that $X_2, X_3, \dots$ are identically distributed. That is to say, whilst we assume that *inter*-occurrence times are independent and identically distributed, we allow the time to the *first* occurrence to have a special distribution.

Given the distributions of the X_i, how may we calculate the probability that H occurs at some given time? Define $u_n = \mathbb{P}(H \text{ occurs at time } n)$. We have by conditioning on X_1 that

$$u_n = \sum_{i=1}^{n} \mathbb{P}(H_n \mid X_1 = i)\mathbb{P}(X_1 = i), \tag{17}$$

where H_n is the event that H occurs at time n. Now

$$\mathbb{P}(H_n \mid X_1 = i) = \mathbb{P}(H_{n-i+1} \mid X_1 = 1) = \mathbb{P}(H_{n-i+1} \mid H_1),$$

using the 'translation invariance' entailed by the assumption that the $X_i, i \geq 2$, are independent and identically distributed. A similar conditioning on X_2 yields

$$\begin{aligned} \mathbb{P}(H_m \mid H_1) &= \sum_{j=1}^{m-1} \mathbb{P}(H_m \mid H_1, X_2 = j)\mathbb{P}(X_2 = j) \\ &= \sum_{j=1}^{m-1} \mathbb{P}(H_{m-j} \mid H_1)\mathbb{P}(X_2 = j) \end{aligned} \tag{18}$$

for $m \geq 2$, by translation invariance once again. Multiplying through (18) by x^{m-1} and summing over m, we obtain

$$\sum_{m=2}^{\infty} x^{m-1}\mathbb{P}(H_m \mid H_1) = \mathbb{E}(x^{X_2}) \sum_{n=1}^{\infty} x^{n-1}\mathbb{P}(H_n \mid H_1), \tag{19}$$

so that $G_H(x) = \sum_{m=1}^{\infty} x^{m-1}\mathbb{P}(H_m \mid H_1)$ satisfies $G_H(x) - 1 = F(x)G_H(x)$, where $F(x)$ is the common probability generating function of the inter-occurrence times, and hence

$$G_H(x) = \frac{1}{1 - F(x)}. \tag{20}$$

Returning to (17), we obtain similarly that $U(x) = \sum_{n=1}^{\infty} x^n u_n$ satisfies

$$U(x) = D(x)G_H(x) = \frac{D(x)}{1 - F(x)} \tag{21}$$

where $D(x)$ is the probability generating function of X_1. Equation (21) contains much of the information relevant to the process, since it relates the occurrences of H to the generating functions of the elements of the sequence $X_1, X_2, \dots$. We should like to extract information out of (21) about $u_n = \mathbb{P}(H_n)$, the coefficient of x^n in $U(x)$, particularly for large values of n.

In principle, one may expand $D(x)/[1 - F(x)]$ as a polynomial in x in order to find u_n, but this is difficult in practice. There is one special situation in which this may be done with ease, and this is the situation when $D(x)$ is the function $D = D^*$ given by

$$D^*(x) = \frac{1 - F(x)}{\mu(1 - x)} \quad \text{for} \quad |x| < 1, \tag{22}$$

and $\mu = \mathbb{E}(X_2)$ is the mean inter-occurrence time. Let us first check that D^* is indeed a suitable probability generating function. The coefficient of x^n in D^* is easily seen to be $(1 - f_1 - f_2 - \cdots - f_n)/\mu$, where $f_i = \mathbb{P}(X_2 = i)$. This coefficient is non-negative since the f_i form a mass function; furthermore, by L'Hôpital's rule,

$$D^*(1) = \lim_{x \uparrow 1} \frac{1 - F(x)}{\mu(1 - x)} = \lim_{x \uparrow 1} \frac{-F'(x)}{-\mu} = 1$$

since $F'(1) = \mu$, the mean inter-occurrence time. Hence $D^*(x)$ is indeed a probability generating function, and with this choice for D we obtain that $U = U^*$ where

$$U^*(x) = \frac{1}{\mu(1 - x)} \tag{23}$$

from (21). Writing $U^*(x) = \sum_n u_n^* x^n$ we find that $u_n^* = \mu^{-1}$ for all n. That is to say, for the special choice of D^*, the corresponding sequence of the u_n^* is *constant*, so that the density of occurrences of H is constant as time passes. This special process is called a *stationary* recurrent-event process.

How relevant is the choice of D to the behaviour of u_n for large n? Intuitively speaking, the choice of distribution of X_1 should not affect greatly the behaviour of the process over long time periods, and so one might expect that $u_n \to \mu^{-1}$ as $n \to \infty$, irrespective of the choice of D. This is indeed the case, so long as we rule out the possibility that there is 'periodicity' in the process. We call the process *non-arithmetic* if $\gcd\{n : \mathbb{P}(X_2 = n) > 0\} = 1$; certainly the process is non-arithmetic if, for example, $\mathbb{P}(X_2 = 1) > 0$. Note that gcd stands for greatest common divisor.

(24) Renewal theorem. *If the mean inter-occurrence time μ is finite and the process is non-arithmetic, then $u_n = \mathbb{P}(H_n)$ satisfies $u_n \to \mu^{-1}$ as $n \to \infty$.*

Sketch proof. The classical proof of this theorem is a purely analytical approach to the equation (21) (see Feller 1968, pp. 335–8). There is a much neater probabilistic proof using the technique of 'coupling'. We do not give a complete proof at this stage, but merely a sketch. The main idea is to introduce a second recurrent-event process, which is stationary and independent of the first. Let $X = \{X_i : i \geq 1\}$ be the first and inter- occurrence times of the original process, and let $X^* = \{X_i^* : i \geq 1\}$ be another sequence of independent random variables, independent of X, such that $X_2^*, X_3^*, \ldots$ have the common distribution of $X_2, X_3, \ldots$, and X_1^* has probability generating function D^*. Let H_n and H_n^* be the events that H occurs at time n in the first and second process (respectively), and let $T = \min\{n : H_n \cap H_n^*$ occurs$\}$ be the earliest time at which H occurs simultaneously in both processes. It may be shown that $T < \infty$ with probability 1, using the assumptions that $\mu < \infty$ and that the processes are non-arithmetic; it is intuitively natural that a coincidence occurs sooner or later, but this is not quite so easy to prove, and we omit a rigorous proof at this point, returning to complete the job in Example (5.10.21). The point is that, once the time T has passed, the non-stationary and stationary recurrent-event processes are indistinguishable from each other, since they have had simultaneous occurrences of H. That is to say, we have that

$$\begin{aligned} u_n &= \mathbb{P}(H_n \mid T \leq n)\mathbb{P}(T \leq n) + \mathbb{P}(H_n \mid T > n)\mathbb{P}(T > n) \\ &= \mathbb{P}(H_n^* \mid T \leq n)\mathbb{P}(T \leq n) + \mathbb{P}(H_n \mid T > n)\mathbb{P}(T > n) \end{aligned}$$

since, if $T \leq n$, then the two processes have already coincided and the (conditional) probability of H_n equals that of H_n^*. Similarly

$$u_n^* = \mathbb{P}(H_n^* \mid T \leq n)\mathbb{P}(T \leq n) + \mathbb{P}(H_n^* \mid T > n)\mathbb{P}(T > n),$$

so that $|u_n - u_n^*| \leq \mathbb{P}(T > n) \to 0$ as $n \to \infty$. However, $u_n^* = \mu^{-1}$ for all n, so that $u_n \to \mu^{-1}$ as $n \to \infty$. ■

Exercises for Section 5.2

1. Let X be the number of events in the sequence $A_1, A_2, \ldots, A_n$ which occur. Let $S_m = \mathbb{E}\binom{X}{m}$, the mean value of the random binomial coefficient $\binom{X}{m}$, and show that

$$\mathbb{P}(X \geq i) = \sum_{j=i}^{n} (-1)^{j-i} \binom{j-1}{i-1} S_j, \qquad \text{for } 1 \leq i \leq n,$$

$$\text{where} \quad S_m = \sum_{j=m}^{n} \binom{j-1}{m-1} \mathbb{P}(X \geq j), \qquad \text{for } 1 \leq m \leq n.$$

2. Each person in a group of n people chooses another at random. Find the probability:
(a) that exactly k people are chosen by nobody,
(b) that at least k people are chosen by nobody.

3. Compounding.
(a) Let X have the Poisson distribution with parameter Y, where Y has the Poisson distribution with parameter μ. Show that $G_{X+Y}(x) = \exp\{\mu(xe^{x-1} - 1)\}$.
(b) Let $X_1, X_2, \ldots$ be independent identically distributed random variables with the *logarithmic* mass function

$$f(k) = \frac{(1-p)^k}{k \log(1/p)}, \qquad k \geq 1,$$

where $0 < p < 1$. If N is independent of the X_i and has the Poisson distribution with parameter μ, show that $Y = \sum_{i=1}^{N} X_i$ has a negative binomial distribution.

4. Let X have the binomial distribution with parameters n and p, and show that

$$\mathbb{E}\left(\frac{1}{1+X}\right) = \frac{1-(1-p)^{n+1}}{(n+1)p}.$$

Find the limit of this expression as $n \to \infty$ and $p \to 0$, the limit being taken in such a way that $np \to \lambda$ where $0 < \lambda < \infty$. Comment.

5. A coin is tossed repeatedly, and heads turns up with probability p on each toss. Let h_n be the probability of an even number of heads in the first n tosses, with the convention that 0 is an even number. Find a difference equation for the h_n and deduce that they have generating function $\frac{1}{2}\{(1+2ps-s)^{-1} + (1-s)^{-1}\}$.

6. An unfair coin is flipped repeatedly, where $\mathbb{P}(\mathrm{H}) = p = 1-q$. Let X be the number of flips until HTH first appears, and Y the number of flips until either HTH or THT appears. Show that $\mathbb{E}(s^X) = (p^2qs^3)/(1-s+pqs^2-pq^2s^3)$ and find $\mathbb{E}(s^Y)$.

7. Matching again. The pile of (by now dog-eared) letters is dropped again and enveloped at random, yielding X_n matches. Show that $\mathbb{P}(X_n = j) = (j+1)\mathbb{P}(X_{n+1} = j+1)$. Deduce that the derivatives of the $G_n(s) = \mathbb{E}(s^{X_n})$ satisfy $G'_{n+1} = G_n$, and hence derive the conclusion of Example (3.4.3), namely:

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

8. Let X have a Poisson distribution with parameter Λ, where Λ is exponential with parameter μ. Show that X has a geometric distribution.

9. Coupons. Recall from Exercise (3.3.2) that each packet of an overpriced commodity contains a worthless plastic object. There are four types of object, and each packet is equally likely to contain any of the four. Let T be the number of packets you open until you first have the complete set. Find $\mathbb{E}(s^T)$ and $\mathbb{P}(T = k)$.

5.3 Random walk

Generating functions are particularly valuable when studying random walks. As before, we suppose that $X_1, X_2, \ldots$ are independent random variables, each taking the value 1 with probability p, and -1 otherwise, and we write $S_n = \sum_{i=1}^n X_i$; the sequence $S = \{S_i : i \geq 0\}$ is a simple random walk starting at the origin. Natural questions of interest concern the sequence of random times at which the particle subsequently returns to the origin. To describe this sequence we need only find the distribution of the time until the particle returns for the first time, since subsequent times between consecutive visits to the origin are independent copies of this.

Let $p_0(n) = \mathbb{P}(S_n = 0)$ be the probability of being at the origin after n steps, and let $f_0(n) = \mathbb{P}(S_1 \neq 0, \ldots, S_{n-1} \neq 0, S_n = 0)$ be the probability that the first return occurs after n steps. Denote the generating functions of these sequences by

$$P_0(s) = \sum_{n=0}^{\infty} p_0(n)s^n, \quad F_0(s) = \sum_{n=1}^{\infty} f_0(n)s^n.$$

F_0 is the probability generating function of the random time T_0 until the particle makes its first return to the origin. That is $F_0(s) = \mathbb{E}(s^{T_0})$. Take care here: T_0 may be defective, and so it may be the case that $F_0(1) = \mathbb{P}(T_0 < \infty)$ satisfies $F_0(1) < 1$.

(1) Theorem. *We have that:*
(a) $P_0(s) = 1 + P_0(s)F_0(s)$,
(b) $P_0(s) = (1 - 4pqs^2)^{-\frac{1}{2}}$,
(c) $F_0(s) = 1 - (1 - 4pqs^2)^{\frac{1}{2}}$.

Proof. (a) Let A be the event that $S_n = 0$, and let B_k be the event that the first return to the origin happens at the kth step. Clearly the B_k are disjoint and so, by Lemma (1.4.4),

$$\mathbb{P}(A) = \sum_{k=1}^{n} \mathbb{P}(A \mid B_k)\mathbb{P}(B_k).$$

However, $\mathbb{P}(B_k) = f_0(k)$ and $\mathbb{P}(A \mid B_k) = p_0(n-k)$ by temporal homogeneity, giving

(2)
$$p_0(n) = \sum_{k=1}^{n} p_0(n-k)f_0(k) \quad \text{if} \quad n \geq 1.$$

Multiply (2) by s^n, sum over n remembering that $p_0(0) = 1$, and use the convolution property of generating functions to obtain $P_0(s) = 1 + P_0(s)F_0(s)$.

(b) $S_n = 0$ if and only if the particle takes equal numbers of steps to the left and to the right during its first n steps. The number of ways in which it can do this is $\binom{n}{\frac{1}{2}n}$ and each such way occurs with probability $(pq)^{n/2}$, giving

(3)
$$p_0(n) = \binom{n}{\frac{1}{2}n}(pq)^{n/2}.$$

We have that $p_0(n) = 0$ if n is odd. This sequence (3) has the required generating function $P_0(s)$.

(c) This follows immediately from (a) and (b). ∎

(4) Corollary.
(a) *The probability that the particle ever returns to the origin is*

$$\sum_{n=1}^{\infty} f_0(n) = F_0(1) = 1 - |p - q|.$$

(b) *If eventual return is certain, that is $F_0(1) = 1$ and $p = \frac{1}{2}$, then the expected time to the first return is*

$$\sum_{n=1}^{\infty} nf_0(n) = F_0'(1) = \infty.$$

We call the process *persistent* (or *recurrent*) if eventual return to the origin is (almost) certain; otherwise it is called *transient*. It is immediately obvious from (4a) that the process is persistent if and only if $p = \frac{1}{2}$. This is consistent with our intuition, which suggests that if $p > \frac{1}{2}$ or $p < \frac{1}{2}$, then the particle tends to stray a long way to the right or to the left of the origin respectively. Even when $p = \frac{1}{2}$ the time until first return has infinite mean.

Proof. (a) Let $s \uparrow 1$ in (1c), and remember equation (5.1.30).

(b) Eventual return is certain if and only if $p = \frac{1}{2}$. But then the generating function of the time T_0 to the first return is $F_0(s) = 1 - (1 - s^2)^{\frac{1}{2}}$ and $\mathbb{E}(T_0) = \lim_{s \uparrow 1} F_0'(s) = \infty$. ■

Now let us consider the times of visits to the point r. Define

$$f_r(n) = \mathbb{P}(S_1 \neq r, \dots, S_{n-1} \neq r, S_n = r)$$

to be the probability that the first such visit occurs at the nth step, with generating function $F_r(s) = \sum_{n=1}^{\infty} f_r(n) s^n$.

(5) Theorem. *We have that*:

(a) $F_r(s) = [F_1(s)]^r$ *for* $r \geq 1$,

(b) $F_1(s) = \left[1 - (1 - 4pqs^2)^{\frac{1}{2}}\right]/(2qs)$.

Proof. (a) The same argument which yields (2) also shows that

$$f_r(n) = \sum_{k=1}^{n-1} f_{r-1}(n-k) f_1(k) \quad \text{if} \quad r > 1.$$

Multiply by s^n and sum over n to obtain

$$F_r(s) = F_{r-1}(s) F_1(s) = F_1(s)^r.$$

We could have written this out in terms of random variables instead of probabilities, and then used Theorem (5.1.23). To see this, let $T_r = \min\{n : S_n = r\}$ be the number of steps taken before the particle reaches r for the first time (T_r may equal $+\infty$ if $r > 0$ and $p < \frac{1}{2}$ or if $r < 0$ and $p > \frac{1}{2}$). In order to visit r, the particle must first visit the point 1; this requires T_1 steps. After visiting 1 the particle requires a further number, $T_{1,r}$ say, of steps to reach r; $T_{1,r}$ is distributed in the manner of T_{r-1} by 'spatial homogeneity'. Thus

$$T_r = \begin{cases} \infty & \text{if } T_1 = \infty, \\ T_1 + T_{1,r} & \text{if } T_1 < \infty, \end{cases}$$

and the result follows from (5.1.23). Some difficulties arise from the possibility that $T_1 = \infty$, but these are resolved fairly easily (*exercise*).

(b) Condition on X_1 to obtain, for $n > 1$,

$$\begin{aligned} \mathbb{P}(T_1 = n) &= \mathbb{P}(T_1 = n \mid X_1 = 1)p + \mathbb{P}(T_1 = n \mid X_1 = -1)q \\ &= 0 \cdot p + \mathbb{P}(\text{first visit to 1 takes } n-1 \text{ steps} \mid S_0 = -1) \cdot q \\ &\qquad\qquad\qquad\qquad \text{by temporal homogeneity} \\ &= \mathbb{P}(T_2 = n-1)q \qquad\qquad \text{by spatial homogeneity} \\ &= qf_2(n-1). \end{aligned}$$

Therefore $f_1(n) = qf_2(n-1)$ if $n > 1$, and $f_1(1) = p$. Multiply by s^n and sum to obtain

$$F_1(s) = ps + sqF_2(s) = ps + qsF_1(s)^2$$

by (a). Solve this quadratic to find its two roots. Only one can be a probability generating function; why? (Hint: $F_1(0) = 0$.) ■

(6) Corollary. *The probability that the walk ever visits the positive part of the real axis is*

$$F_1(1) = \frac{1 - |p - q|}{2q} = \min\{1, p/q\}.$$

Knowledge of Theorem (5) enables us to calculate $F_0(s)$ directly without recourse to (1). The method of doing this relies upon a symmetry within the collection of paths which may be followed by a random walk. Condition on the value of X_1 as usual to obtain

$$f_0(n) = qf_1(n-1) + pf_{-1}(n-1)$$

and thus

$$F_0(s) = qsF_1(s) + psF_{-1}(s).$$

We need to find $F_{-1}(s)$. Consider any possible path π that the particle may have taken to arrive at the point -1 and replace each step in the path by its mirror image, positive steps becoming negative and negative becoming positive, to obtain a path π^* which ends at $+1$. This operation of reflection provides a one–one correspondence between the collection of paths ending at -1 and the collection of paths ending at $+1$. If $\mathbb{P}(\pi; p, q)$ is the probability that the particle follows π when each step is to the right with probability p, then $\mathbb{P}(\pi; p, q) = \mathbb{P}(\pi^*; q, p)$; thus

$$F_{-1}(s) = \frac{1 - (1 - 4pqs^2)^{\frac{1}{2}}}{2ps},$$

giving that $F_0(s) = 1 - (1 - 4pqs^2)^{\frac{1}{2}}$ as before.

We made use in the last paragraph of a version of the reflection principle discussed in Section 3.10. Generally speaking, results obtained using the reflection principle may also be obtained using generating functions, sometimes in greater generality than before. Consider for example the hitting time theorem (3.10.14): the mass function of the time T_b of the first visit of S to the point b is given by

$$\mathbb{P}(T_b = n) = \frac{|b|}{n}\mathbb{P}(S_n = b) \quad \text{if} \quad n \geq 1.$$

We shall state and prove a version of this for random walks of a more general nature. Consider a sequence $X_1, X_2, \dots$ of independent identically distributed random variables taking values in the integers (positive and negative). We may think of $S_n = X_1 + X_2 + \cdots + X_n$ as being the nth position of a random walk which takes steps X_i; for the simple random walk, each X_i is required to take the values ± 1 only. We call a random walk *right-continuous* (respectively *left-continuous*) if $\mathbb{P}(X_i \leq 1) = 1$ (respectively $\mathbb{P}(X_i \geq -1) = 1$), which is to say that the maximum rightward (respectively leftward) step is no greater than 1. In order to avoid certain situations of no interest, we shall consider only right-continuous walks (respectively left-continuous walks) for which $\mathbb{P}(X_i = 1) > 0$ (respectively $\mathbb{P}(X_i = -1) > 0$).

(7) Hitting time theorem. *Assume that S is a right-continuous random walk, and let T_b be the first hitting time of the point b. Then*

$$\mathbb{P}(T_b = n) = \frac{b}{n}\mathbb{P}(S_n = b) \quad \textit{for} \quad b, n \geq 1.$$

For left-continuous walks, the conclusion becomes

$$\mathbb{P}(T_{-b} = n) = \frac{b}{n}\mathbb{P}(S_n = -b) \quad \text{for} \quad b, n \geq 1. \tag{8}$$

Proof. We introduce the functions

$$G(z) = \mathbb{E}(z^{-X_1}) = \sum_{n=-\infty}^{1} z^{-n}\mathbb{P}(X_1 = n), \quad F_b(z) = \mathbb{E}(z^{T_b}) = \sum_{n=0}^{\infty} z^n \mathbb{P}(T_b = n).$$

These are functions of the complex variable z. The function $G(z)$ has a simple pole at the origin, and the sum defining $F_b(z)$ converges for $|z| < 1$.

Since the walk is assumed to be right-continuous, in order to reach b (where $b > 0$) it must pass through the points $1, 2, \ldots, b-1$. The argument leading to (5a) may therefore be applied, and we find that

$$F_b(z) = F_1(z)^b \quad \text{for} \quad b \geq 1. \tag{9}$$

The argument leading to (5b) may be expressed as

$$F_1(z) = \mathbb{E}(z^{T_1}) = \mathbb{E}\big(\mathbb{E}(z^{T_1} \mid X_1)\big) = \mathbb{E}(z^{1+T_J}) \quad \text{where } J = 1 - X_1$$

since, conditional on X_1, the further time required to reach 1 has the same distribution as T_{1-X_1}. Now $1 - X_1 \geq 0$, and therefore

$$F_1(z) = z\mathbb{E}\big(F_{1-X_1}(z)\big) = z\mathbb{E}\big(F_1(z)^{1-X_1}\big) = zF_1(z)G(F_1(z)),$$

yielding

$$z = \frac{1}{G(w)} \tag{10}$$

where

$$w = w(z) = F_1(z). \tag{11}$$

Inverting (10) to find $F_1(z)$, and hence $F_b(z) = F_1(z)^b$, is a standard exercise in complex analysis using what is called Lagrange's inversion formula.

(12) Theorem. Lagrange's inversion formula. *Let $z = w/f(w)$ where $w/f(w)$ is an analytic function of w on a neighbourhood of the origin. If g is infinitely differentiable, then*

$$g(w(z)) = g(0) + \sum_{n=1}^{\infty} \frac{1}{n!} z^n \left[\frac{d^{n-1}}{du^{n-1}}[g'(u)f(u)^n]\right]_{u=0}. \tag{13}$$

We apply this as follows. Define $w = F_1(z)$ and $f(w) = wG(w)$, so that (10) becomes $z = w/f(w)$. Note that $f(w) = \mathbb{E}(w^{1-X_1})$ which, by the right-continuity of the walk, is a power series in w which converges for $|w| < 1$. Also $f(0) = \mathbb{P}(X_1 = 1) > 0$, and hence

$w/f(w)$ is analytic on a neighbourhood of the origin. We set $g(w) = w^b$ ($= F_1(z)^b = F_b(z)$, by (9)). The inversion formula now yields

(14)
$$F_b(z) = g(w(z)) = g(0) + \sum_{n=1}^{\infty} \frac{1}{n!} z^n D_n$$

where
$$D_n = \frac{d^{n-1}}{du^{n-1}}\left[bu^{b-1}u^n G(u)^n\right]\bigg|_{u=0}.$$

We pick out the coefficient of z^n in (14) to obtain

(15)
$$\mathbb{P}(T_b = n) = \frac{1}{n!} D_n \quad \text{for} \quad n \geq 1.$$

Now $G(u)^n = \sum_{i=-\infty}^{n} u^{-i}\mathbb{P}(S_n = i)$, so that

$$D_n = \frac{d^{n-1}}{du^{n-1}}\left(b \sum_{i=-\infty}^{n} u^{b+n-1-i}\mathbb{P}(S_n = i)\right)\Bigg|_{u=0} = b(n-1)!\mathbb{P}(S_n = b),$$

which may be combined with (15) as required. ■

Once we have the hitting time theorem, we are in a position to derive a magical result called Spitzer's identity, relating the distributions of the maxima of a random walk to those of the walk itself. This identity is valid in considerable generality; the proof given here uses the hitting time theorem, and is therefore valid only for right-continuous walks (and *mutatis mutandis* for left-continuous walks and their minima).

(16) Theorem. Spitzer's identity. *Assume that S is a right-continuous random walk, and let $M_n = \max\{S_i : 0 \leq i \leq n\}$ be the maximum of the walk up to time n. Then, for $|s|, |t| < 1$,*

(17)
$$\log\left(\sum_{n=0}^{\infty} t^n \mathbb{E}(s^{M_n})\right) = \sum_{n=1}^{\infty} \frac{1}{n} t^n \mathbb{E}(s^{S_n^+})$$

where $S_n^+ = \max\{0, S_n\}$ as usual.

This curious and remarkable identity relates the generating function of the probability generating functions of the maxima M_n to the corresponding object for S_n^+. It contains full information about the distributions of the maxima.

Proof. Writing $f_j(n) = \mathbb{P}(T_j = n)$ as in Section 3.10, we have that

(18)
$$\mathbb{P}(M_n = k) = \sum_{j=0}^{n} f_k(j)\mathbb{P}(T_1 > n - j) \quad \text{for} \quad k \geq 0,$$

since $M_n = k$ if the passage to k occurs at some time j ($\leq n$), and in addition the walk does not rise above k during the next $n - j$ steps; remember that $T_1 = \infty$ if no visit to 1 takes place. Multiply throughout (18) by $s^k t^n$ (where $|s|, |t| \leq 1$) and sum over $k, n \geq 0$ to obtain

$$\sum_{n=0}^{\infty} t^n \mathbb{E}(s^{M_n}) = \sum_{k=0}^{\infty} s^k \left(\sum_{n=0}^{\infty} t^n \mathbb{P}(M_n = k)\right) = \sum_{k=0}^{\infty} s^k F_k(t)\left(\frac{1 - F_1(t)}{1 - t}\right),$$

by the convolution formula for generating functions. We have used the result of Exercise (5.1.2) here; as usual, $F_k(t) = \mathbb{E}(t^{T_k})$. Now $F_k(t) = F_1(t)^k$, by (9), and therefore

$$\sum_{n=0}^{\infty} t^n \mathbb{E}(s^{M_n}) = D(s,t) \tag{19}$$

where

$$D(s,t) = \frac{1 - F_1(t)}{(1-t)(1 - sF_1(t))}. \tag{20}$$

We shall find $D(s,t)$ by finding an expression for $\partial D/\partial t$ and integrating with respect to t.

By the hitting time theorem, for $n \geq 0$,

$$n\mathbb{P}(T_1 = n) = \mathbb{P}(S_n = 1) = \sum_{j=0}^{n} \mathbb{P}(T_1 = j)\mathbb{P}(S_{n-j} = 0), \tag{21}$$

as usual; multiply throughout by t^n and sum over n to obtain that $tF_1'(t) = F_1(t)P_0(t)$. Hence

$$\begin{aligned} \frac{\partial}{\partial t} \log[1 - sF_1(t)] &= \frac{-sF_1'(t)}{1 - sF_1(t)} = -\frac{s}{t} F_1(t) P_0(t) \sum_{k=0}^{\infty} s^k F_1(t)^k \\ &= -\sum_{k=1}^{\infty} \frac{s^k}{t} F_k(t) P_0(t) \end{aligned} \tag{22}$$

by (9). Now $F_k(t)P_0(t)$ is the generating function of the sequence

$$\sum_{j=0}^{n} \mathbb{P}(T_k = j)\mathbb{P}(S_{n-j} = 0) = \mathbb{P}(S_n = k)$$

as in (21), which implies that

$$\frac{\partial}{\partial t} \log[1 - sF_1(t)] = -\sum_{n=1}^{\infty} t^{n-1} \sum_{k=1}^{\infty} s^k \mathbb{P}(S_n = k).$$

Hence

$$\begin{aligned} \frac{\partial}{\partial t} \log D(s,t) &= -\frac{\partial}{\partial t} \log(1-t) + \frac{\partial}{\partial t} \log[1 - F_1(t)] - \frac{\partial}{\partial t} \log[1 - sF_1(t)] \\ &= \sum_{n=1}^{\infty} t^{n-1} \left(1 - \sum_{k=1}^{\infty} \mathbb{P}(S_n = k) + \sum_{k=1}^{\infty} s^k \mathbb{P}(S_n = k) \right) \\ &= \sum_{n=1}^{\infty} t^{n-1} \left(\mathbb{P}(S_n \leq 0) + \sum_{k=1}^{\infty} s^k \mathbb{P}(S_n = k) \right) = \sum_{n=1}^{\infty} t^{n-1} \mathbb{E}(s^{S_n^+}). \end{aligned}$$

Integrate over t, noting that both sides of (19) equal 1 when $t = 0$, to obtain (17). ■

For our final example of the use of generating functions, we return to simple random walk, for which each jump equals 1 or -1 with probabilities p and $q = 1 - p$. Suppose that we are told that $S_{2n} = 0$, so that the walk is 'tied down', and we ask for the number of steps of the walk which were not within the negative half-line. In the language of gambling, L_{2n} is the amount of time that the gambler was ahead of the bank. In the arc sine law for sojourn times, Theorem (3.10.21), we explored the distribution of L_{2n} without imposing the condition that $S_{2n} = 0$. Given that $S_{2n} = 0$, we might think that L_{2n} would be about n, but, as can often happen, the contrary turns out to be the case.

(23) Theorem. Leads for tied-down random walk. *For the simple random walk S,*

$$\mathbb{P}(L_{2n} = 2k \mid S_{2n} = 0) = \frac{1}{n+1}, \qquad k = 0, 1, 2, \dots, n.$$

Thus each possible value of L_{2n} is equally likely. Unlike the related results of Section 3.10, we prove this using generating functions. Note that the distribution of L_{2n} does not depend on the value of p. This is not surprising since, conditional on $\{S_{2n} = 0\}$, the joint distribution of $S_0, S_1, \dots, S_{2n}$ does not depend on p (*exercise*).

Proof. Assume $|s|, |t| < 1$, and define $G_{2n}(s) = \mathbb{E}(s^{L_{2n}} \mid S_{2n} = 0)$, $F_0(s) = \mathbb{E}(s^{T_0})$, and the bivariate generating function

$$H(s,t) = \sum_{n=0}^{\infty} t^{2n}\mathbb{P}(S_{2n} = 0)G_{2n}(s).$$

By conditioning on the time of the first return to the origin,

$$G_{2n}(s) = \sum_{r=1}^{n} \mathbb{E}(s^{L_{2n}} \mid S_{2n} = 0,\ T_0 = 2r)\mathbb{P}(T_0 = 2r \mid S_{2n} = 0). \tag{24}$$

We may assume without loss of generality that $p = q = \frac{1}{2}$, so that

$$\mathbb{E}(s^{L_{2n}} \mid S_{2n} = 0,\ T_0 = 2r) = G_{2n-2r}(s)\big(\tfrac{1}{2} + \tfrac{1}{2}s^{2r}\big),$$

since, under these conditions, L_{2r} has (conditional) probability $\frac{1}{2}$ of being equal to either 0 or $2r$. Also

$$\mathbb{P}(T_0 = 2r \mid S_{2n} = 0) = \frac{\mathbb{P}(T_0 = 2r)\mathbb{P}(S_{2n-2r} = 0)}{\mathbb{P}(S_{2n} = 0)},$$

so that (24) becomes

$$G_{2n}(s) = \sum_{r=1}^{n} \frac{\big[G_{2n-2r}(s)\mathbb{P}(S_{2n-2r} = 0)\big]\big[\frac{1}{2}(1 + s^{2r})\mathbb{P}(T_0 = 2r)\big]}{\mathbb{P}(S_{2n} = 0)}.$$

Multiply throughout by $t^{2n}\mathbb{P}(S_{2n} = 0)$ and sum over $n \geq 1$, to find that

$$H(s,t) - 1 = \tfrac{1}{2}H(s,t)[F_0(t) + F_0(st)].$$

Hence

$$H(s,t) = \frac{2}{\sqrt{1-t^2}+\sqrt{1-s^2t^2}} = \frac{2\left[\sqrt{1-s^2t^2}-\sqrt{1-t^2}\right]}{t^2(1-s^2)}$$
$$= \sum_{n=0}^{\infty} t^{2n}\mathbb{P}(S_{2n}=0)\left(\frac{1-s^{2n+2}}{(n+1)(1-s^2)}\right)$$

after a little work using (1b). We deduce that $G_{2n}(s) = \sum_{k=0}^{n}(n+1)^{-1}s^{2k}$, and the proof is finished. ∎

Exercises for Section 5.3

1. For a simple random walk S with $S_0 = 0$ and $p = 1-q < \frac{1}{2}$, show that the maximum $M = \max\{S_n : n \geq 0\}$ satisfies $\mathbb{P}(M \geq r) = (p/q)^r$ for $r \geq 0$.

2. Use generating functions to show that, for a symmetric random walk,
(a) $2kf_0(2k) = \mathbb{P}(S_{2k-2} = 0)$ for $k \geq 1$, and
(b) $\mathbb{P}(S_1S_2\cdots S_{2n} \neq 0) = \mathbb{P}(S_{2n} = 0)$ for $n \geq 1$.

3. A particle performs a random walk on the corners of the square ABCD. At each step, the probability of moving from corner c to corner d equals ρ_{cd}, where

$$\rho_{\mathrm{AB}} = \rho_{\mathrm{BA}} = \rho_{\mathrm{CD}} = \rho_{\mathrm{DC}} = \alpha, \qquad \rho_{\mathrm{AD}} = \rho_{\mathrm{DA}} = \rho_{\mathrm{BC}} = \rho_{\mathrm{CB}} = \beta,$$

and $\alpha, \beta > 0$, $\alpha + \beta = 1$. Let $G_{\mathrm{A}}(s)$ be the generating function of the sequence $(p_{\mathrm{AA}}(n) : n \geq 0)$, where $p_{\mathrm{AA}}(n)$ is the probability that the particle is at A after n steps, having started at A. Show that

$$G_{\mathrm{A}}(s) = \frac{1}{2}\left\{\frac{1}{1-s^2} + \frac{1}{1-|\beta-\alpha|^2s^2}\right\}.$$

Hence find the probability generating function of the time of the first return to A.

4. A particle performs a symmetric random walk in two dimensions starting at the origin: each step is of unit length and has equal probability $\frac{1}{4}$ of being northwards, southwards, eastwards, or westwards. The particle first reaches the line $x + y = m$ at the point (X, Y) and at the time T. Find the probability generating functions of T and $X - Y$, and state where they converge.

5. Derive the arc sine law for sojourn times, Theorem (3.10.21), using generating functions. That is to say, let L_{2n} be the length of time spent (up to time $2n$) by a simple symmetric random walk to the right of its starting point. Show that

$$\mathbb{P}(L_{2n} = 2k) = \mathbb{P}(S_{2k} = 0)\mathbb{P}(S_{2n-2k} = 0) \qquad \text{for } 0 \leq k \leq n.$$

6. Let $\{S_n : n \geq 0\}$ be a simple symmetric random walk with $S_0 = 0$, and let $T = \min\{n > 0 : S_n = 0\}$. Show that

$$\mathbb{E}\big(\min\{T, 2m\}\big) = 2\mathbb{E}|S_{2m}| = 4m\mathbb{P}(S_{2m} = 0) \qquad \text{for } m \geq 0.$$

7. Let $S_n = \sum_{r=0}^{n} X_r$ be a left-continuous random walk on the integers with a retaining barrier at zero. More specifically, we assume that the X_r are identically distributed integer-valued random variables with $X_1 \geq -1$, $\mathbb{P}(X_1 = 0) \neq 0$, and

$$S_{n+1} = \begin{cases} S_n + X_{n+1} & \text{if } S_n > 0, \\ S_n + X_{n+1} + 1 & \text{if } S_n = 0. \end{cases}$$

Show that the distribution of S_0 may be chosen in such a way that $\mathbb{E}(z^{S_n}) = \mathbb{E}(z^{S_0})$ for all n, if and only if $\mathbb{E}(X_1) < 0$, and in this case

$$\mathbb{E}(z^{S_n}) = \frac{(1-z)\mathbb{E}(X_1)\mathbb{E}(z^{X_1})}{1-\mathbb{E}(z^{X_1})}.$$

8. Consider a simple random walk starting at 0 in which each step is to the right with probability $p\ (= 1-q)$. Let T_b be the number of steps until the walk first reaches b where $b > 0$. Show that $\mathbb{E}(T_b \mid T_b < \infty) = b/|p-q|$.

5.4 Branching processes

Besides gambling, many probabilists have been interested in reproduction. Accurate models for the evolution of a population are notoriously difficult to handle, but there are simpler non-trivial models which are both tractable and mathematically interesting. The branching process is such a model. Suppose that a population evolves in generations, and let Z_n be the number of members of the nth generation. Each member of the nth generation gives birth to a family, possibly empty, of members of the $(n+1)$th generation; the size of this family is a random variable. We shall make the following assumptions about these family sizes:

(a) the family sizes of the individuals of the branching process form a collection of independent random variables;

(b) all family sizes have the same probability mass function f and generating function G.

These assumptions, together with information about the distribution of the number Z_0 of founding members, specify the random evolution of the process. We assume here that $Z_0 = 1$. There is nothing notably human about this model, which may be just as suitable a description for the growth of a population of cells, or for the increase of neutrons in a reactor, or for the spread of an epidemic in some population. See Figure 5.1 for a picture of a branching process.

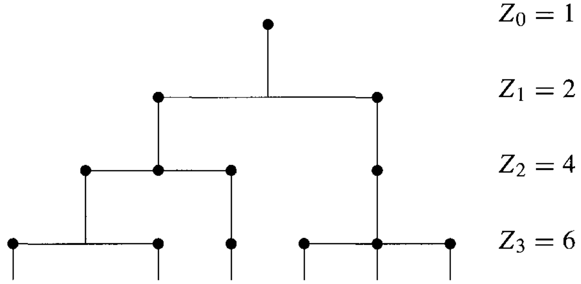


Figure 5.1. The family tree of a branching process.

We are interested in the random sequence $Z_0, Z_1, \dots$ of generation sizes. Let $G_n(s) = \mathbb{E}(s^{Z_n})$ be the generating function of Z_n.

(1) Theorem. *It is the case that $G_{m+n}(s) = G_m(G_n(s)) = G_n(G_m(s))$, and thus $G_n(s) = G(G(\dots(G(s))\dots))$ is the n-fold iterate of G.*

Proof. Each member of the $(m+n)$th generation has a unique ancestor in the mth generation. Thus

$$Z_{m+n} = X_1 + X_2 + \cdots + X_{Z_m}$$

where X_i is the number of members of the $(m+n)$th generation which stem from the ith member of the mth generation. This is the sum of a random number Z_m of variables. These variables are independent by assumption (a); furthermore, by assumption (b) they are identically distributed with the same distribution as the number Z_n of the nth-generation offspring of the first individual in the process. Now use Theorem (5.1.25) to obtain $G_{m+n}(s) = G_m(G_{X_1}(s))$ where $G_{X_1}(s) = G_n(s)$. Iterate this relation to obtain

$$G_n(s) = G_1(G_{n-1}(s)) = G_1(G_1(G_{n-2}(s))) = G_1(G_1(\dots(G_1(s))\dots))$$

and notice that $G_1(s)$ is what we called $G(s)$. ■

In principle, Theorem (1) tells us all about Z_n and its distribution, but in practice $G_n(s)$ may be hard to evaluate. The moments of Z_n, at least, may be routinely computed in terms of the moments of a typical family size Z_1. For example:

(2) Lemma. *Let $\mu = \mathbb{E}(Z_1)$ and $\sigma^2 = \operatorname{var}(Z_1)$. Then*

$$\mathbb{E}(Z_n) = \mu^n, \qquad \operatorname{var}(Z_n) = \begin{cases} n\sigma^2 & \text{if } \mu = 1, \\ \dfrac{\sigma^2(\mu^n - 1)\mu^{n-1}}{\mu - 1} & \text{if } \mu \neq 1. \end{cases}$$

Proof. Differentiate $G_n(s) = G(G_{n-1}(s))$ once at $s = 1$ to obtain $\mathbb{E}(Z_n) = \mu\mathbb{E}(Z_{n-1})$; by iteration, $\mathbb{E}(Z_n) = \mu^n$. Differentiate twice to obtain

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

and use equation (5.1.19) to obtain the second result. ■

(3) Example. Geometric branching. Suppose that each family size has the mass function $f(k) = qp^k$, for $k \geq 0$, where $q = 1 - p$. Then $G(s) = q(1 - ps)^{-1}$, and each family size is one member less than a geometric variable. We can show by induction that

$$G_n(s) = \begin{cases} \dfrac{n - (n-1)s}{n + 1 - ns} & \text{if } p = q = \frac{1}{2}, \\ \dfrac{q\left[p^n - q^n - ps(p^{n-1} - q^{n-1})\right]}{p^{n+1} - q^{n+1} - ps(p^n - q^n)} & \text{if } p \neq q. \end{cases}$$

This result can be useful in providing inequalities for more general distributions. What can we say about the behaviour of this process after many generations? In particular, does it eventually become extinct, or, conversely, do all generations have non-zero size? For this example, we can answer this question from a position of strength since we know $G_n(s)$ in closed form. In fact

$$\mathbb{P}(Z_n = 0) = G_n(0) = \begin{cases} \dfrac{n}{n+1} & \text{if } p = q, \\ \dfrac{q(p^n - q^n)}{p^{n+1} - q^{n+1}} & \text{if } p \neq q. \end{cases}$$

Let $n \to \infty$ to obtain

$$\mathbb{P}(Z_n = 0) \to \mathbb{P}(\text{ultimate extinction}) = \begin{cases} 1 & \text{if } p \leq q, \\ q/p & \text{if } p > q. \end{cases}$$

We have used Lemma (1.3.5) here surreptitiously, since

$$\text{(4)} \qquad \{\text{ultimate extinction}\} = \bigcup_n \{Z_n = 0\}$$

and $A_n = \{Z_n = 0\}$ satisfies $A_n \subseteq A_{n+1}$. ●

We saw in this example that extinction occurs almost surely if and only if $\mu = \mathbb{E}(Z_1) = p/q$ satisfies $\mathbb{E}(Z_1) \leq 1$. This is a very natural condition; it seems reasonable that if $\mathbb{E}(Z_n) = \mathbb{E}(Z_1)^n \leq 1$ then $Z_n = 0$ sooner or later. Actually this result holds in general.

(5) Theorem. *As $n \to \infty$, $\mathbb{P}(Z_n = 0) \to \mathbb{P}(\text{ultimate extinction}) = \eta$, say, where η is the smallest non-negative root of the equation $s = G(s)$. Also, $\eta = 1$ if $\mu < 1$, and $\eta < 1$ if $\mu > 1$. If $\mu = 1$ then $\eta = 1$ so long as the family-size distribution has strictly positive variance.*

Proof†. Let $\eta_n = \mathbb{P}(Z_n = 0)$. Then, by (1),

$$\eta_n = G_n(0) = G(G_{n-1}(0)) = G(\eta_{n-1}).$$

In the light of the remarks about equation (4) we know that $\eta_n \uparrow \eta$, and the continuity of G guarantees that $\eta = G(\eta)$. We show next that if ψ is any non-negative root of the equation $s = G(s)$ then $\eta \leq \psi$. Note that G is non-decreasing on $[0, 1]$ and so

$$\eta_1 = G(0) \leq G(\psi) = \psi.$$

Similarly

$$\eta_2 = G(\eta_1) \leq G(\psi) = \psi$$

and hence, by induction, $\eta_n \leq \psi$ for all n, giving $\eta \leq \psi$. Thus η is the smallest non-negative root of the equation $s = G(s)$.

To verify the second assertion of the theorem, we need the fact that G is convex on $[0, 1]$. This holds because

$$G''(s) = \mathbb{E}\big[Z_1(Z_1 - 1)s^{Z_1-2}\big] \geq 0 \quad \text{if} \quad s \geq 0.$$

So G is convex and non-decreasing on $[0, 1]$ with $G(1) = 1$. We can verify that the two curves $y = G(s)$ and $y = s$ generally have two intersections in $[0, 1]$, and these occur at $s = \eta$ and $s = 1$. A glance at Figure 5.2 (and a more analytical verification) tells us that these intersections are coincident if $\mu = G'(1) < 1$. On the other hand, if $\mu > 1$ then these two intersections are not coincident. In the special case when $\mu = 1$ we need to distinguish between the non-random case in which $\sigma^2 = 0$, $G(s) = s$, and $\eta = 0$, and the random case in which $\sigma^2 > 0$, $G(s) > s$ for $0 \leq s < 1$, and $\eta = 1$. ■

†This method of solution was first attempted by H. W. Watson in 1873 in response to a challenge posed by F. Galton in the April 1 edition of the Educational Times. For this reason, a branching process is sometimes termed a 'Galton–Watson process'. The correct solution in modern format was supplied by J. F. Steffensen in 1930; I. J. Bienaymé and J. B. S. Haldane had earlier realized what the extinction probability should be, but failed to provide the required reasoning.

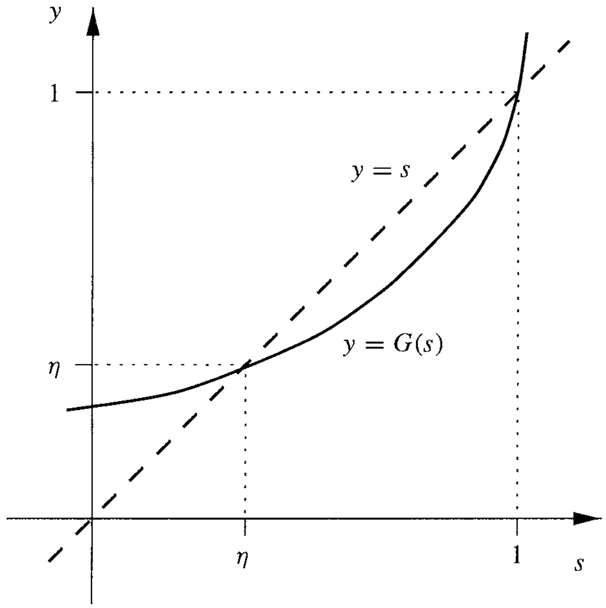


Figure 5.2. A sketch of $G(s)$ showing the roots of the equation $G(s) = s$.

We have seen that, for large n, the nth generation is empty with probability approaching η. However, what if the process does *not* die out? If $\mathbb{E}(Z_1) > 1$ then $\eta < 1$ and extinction is not certain. Indeed $\mathbb{E}(Z_n)$ grows geometrically as $n \to \infty$, and it can be shown that

$$\mathbb{P}(Z_n \to \infty \mid \text{non-extinction}) = 1$$

when this conditional probability is suitably interpreted. To see just how fast Z_n grows, we define $W_n = Z_n/\mathbb{E}(Z_n)$ where $\mathbb{E}(Z_n) = \mu^n$, and we suppose that $\mu > 1$. Easy calculations show that

$$\mathbb{E}(W_n) = 1, \quad \operatorname{var}(W_n) = \frac{\sigma^2(1-\mu^{-n})}{\mu^2 - \mu} \to \frac{\sigma^2}{\mu^2 - \mu} \quad \text{as } n \to \infty,$$

and it seems that W_n may have some non-trivial limit†, called W say. In order to study W, define $g_n(s) = \mathbb{E}(s^{W_n})$. Then

$$g_n(s) = \mathbb{E}(s^{Z_n \mu^{-n}}) = G_n(s^{\mu^{-n}})$$

and (1) shows that g_n satisfies the functional recurrence relation

$$g_n(s) = G\big(g_{n-1}(s^{1/\mu})\big).$$

Now, as $n \to \infty$, we have that $W_n \to W$ and $g_n(s) \to g(s) = \mathbb{E}(s^W)$, and we obtain

(6) $$g(s) = G\big(g(s^{1/\mu})\big)$$

†We are asserting that the sequence $\{W_n\}$ of variables converges to a limit variable W. The convergence of random variables is a complicated topic described in Chapter 7. We overlook the details for the moment.

by abandoning some of our current notions of mathematical rigour. This functional equation can be established rigorously (see Example (7.8.5)) and has various uses. For example, although we cannot solve it for g, we can reach such conclusions as 'if $\mathbb{E}(Z_1^2) < \infty$ then W is continuous, apart from a point mass of size η at zero'.

We have made considerable progress with the theory of branching processes. They are reasonably tractable because they satisfy the Markov condition (see Example (3.9.5)). Can you formulate and prove this property?

Exercises for Section 5.4

1. Let Z_n be the size of the nth generation in an ordinary branching process with $Z_0 = 1, \mathbb{E}(Z_1) = \mu$, and $\text{var}(Z_1) > 0$. Show that $\mathbb{E}(Z_n Z_m) = \mu^{n-m}\mathbb{E}(Z_m^2)$ for $m \le n$. Hence find the correlation coefficient $\rho(Z_m, Z_n)$ in terms of μ.

2. Consider a branching process with generation sizes Z_n satisfying $Z_0 = 1$ and $\mathbb{P}(Z_1 = 0) = 0$. Pick two individuals at random (with replacement) from the nth generation and let L be the index of the generation which contains their most recent common ancestor. Show that $\mathbb{P}(L = r) = \mathbb{E}(Z_r^{-1}) - \mathbb{E}(Z_{r+1}^{-1})$ for $0 \le r < n$. What can be said if $\mathbb{P}(Z_1 = 0) > 0$?

3. Consider a branching process whose family sizes have the geometric mass function $f(k) = qp^k$, $k \ge 0$, where $p + q = 1$, and let Z_n be the size of the nth generation. Let $T = \min\{n : Z_n = 0\}$ be the extinction time, and suppose that $Z_0 = 1$. Find $\mathbb{P}(T = n)$. For what values of p is it the case that $\mathbb{E}(T) < \infty$?

4. Let Z_n be the size of the nth generation of a branching process, and assume $Z_0 = 1$. Find an expression for the generating function G_n of Z_n, in the cases when Z_1 has generating function given by:

(a) $G(s) = 1 - \alpha(1-s)^\beta$, $0 < \alpha, \beta < 1$.

(b) $G(s) = f^{-1}\{P(f(s))\}$, where P is a probability generating function, and f is a suitable function satisfying $f(1) = 1$.

(c) Suppose in the latter case that $f(x) = x^m$ and $P(s) = s\{\gamma - (\gamma - 1)s\}^{-1}$ where $\gamma > 1$. Calculate the answer explicitly.

5. Branching with immigration. Each generation of a branching process (with a single progenitor) is augmented by a random number of immigrants who are indistinguishable from the other members of the population. Suppose that the numbers of immigrants in different generations are independent of each other and of the past history of the branching process, each such number having probability generating function $H(s)$. Show that the probability generating function G_n of the size of the nth generation satisfies $G_{n+1}(s) = G_n(G(s))H(s)$, where G is the probability generating function of a typical family of offspring.

6. Let Z_n be the size of the nth generation in a branching process with $\mathbb{E}(s^{Z_1}) = (2-s)^{-1}$ and $Z_0 = 1$. Let V_r be the total number of generations of size r. Show that $\mathbb{E}(V_1) = \frac{1}{6}\pi^2$, and $\mathbb{E}(2V_2 - V_3) = \frac{1}{6}\pi^2 - \frac{1}{90}\pi^4$.

5.5 Age-dependent branching processes

Here is a more general model for the growth of a population. It incorporates the observation that generations are not contemporaneous in most populations; in fact, individuals in the same generation give birth to families at different times. To model this we attach another random variable, called 'age', to each individual; we shall suppose that the collection of all ages is a set of variables which are independent of each other and of all family sizes, and which

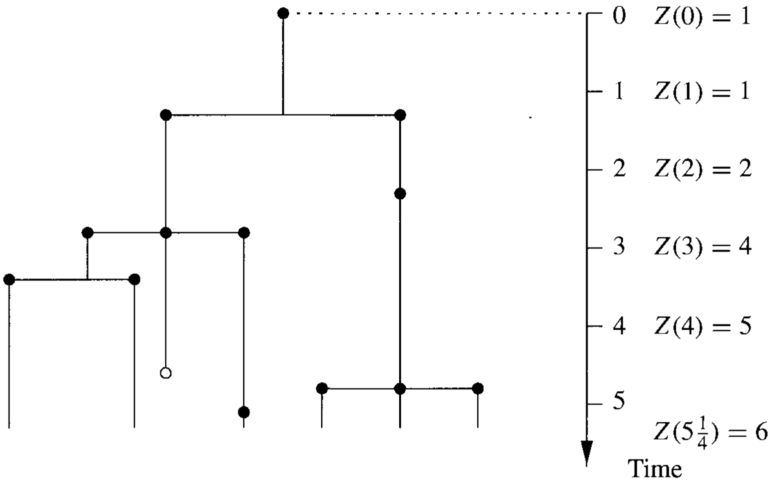


Figure 5.3. The family tree of an age-dependent branching process; • indicates the birth of an individual, and ◦ indicates the death of an individual which has no descendants.

are continuous, positive, and have the common density function f_T. Each individual lives for a period of time, equal to its 'age', before it gives birth to its family of next-generation descendants as before. See Figure 5.3 for a picture of an age-dependent branching process.

Let $Z(t)$ denote the size of the population at time t; we shall assume that $Z(0) = 1$. The population-size generating function $G_t(s) = \mathbb{E}(s^{Z(t)})$ is now a function of t as well. As usual, we hope to find an expression involving G_t by conditioning on some suitable event. In this case we condition on the age of the initial individual in the population.

(1) Theorem. $G_t(s) = \displaystyle\int_0^t G\big(G_{t-u}(s)\big) f_T(u)\,du + \int_t^\infty s f_T(u)\,du.$

Proof. Let T be the age of the initial individual. By the use of conditional expectation,

$$G_t(s) = \mathbb{E}(s^{Z(t)}) = \mathbb{E}\big(\mathbb{E}(s^{Z(t)} \mid T)\big) = \int_0^\infty \mathbb{E}(s^{Z(t)} \mid T = u) f_T(u)\,du. \tag{2}$$

If $T = u$, then at time u the initial individual dies and is replaced by a random number N of offspring, where N has generating function G. Each of these offspring behaves in the future as their ancestor did in the past, and the effect of their ancestor's death is to replace the process by the sum of N independent copies of the process displaced in time by an amount u. Now if $u > t$ then $Z(t) = 1$ and $\mathbb{E}(s^{Z(t)} \mid T = u) = s$, whilst if $u < t$ then $Z(t) = Y_1 + Y_2 + \cdots + Y_N$ is the sum of N independent copies of $Z(t-u)$ and so $\mathbb{E}(s^{Z(t)} \mid T = u) = G(G_{t-u}(s))$ by Theorem (5.1.25). Substitute into (2) to obtain the result. ■

Unfortunately we cannot solve equation (1) except in certain special cases. Possibly the most significant case with which we can make some progress arises when the ages are exponentially distributed. In this case, $f_T(t) = \lambda e^{-\lambda t}$ for $t \geq 0$, and the reader may show (*exercise*) that

$$\frac{\partial}{\partial t} G_t(s) = \lambda\big[G(G_t(s)) - G_t(s)\big]. \tag{3}$$

It is no mere coincidence that this case is more tractable. In this very special instance, and in no other, $Z(t)$ satisfies a Markov condition; it is called a Markov process, and we shall return to the general theory of such processes in Chapter 6.

Some information about the moments of $Z(t)$ is fairly readily available from (1). For example,

$$m(t) = \mathbb{E}(Z(t)) = \lim_{s \uparrow 1} \frac{\partial}{\partial s} G_t(s)$$

satisfies the integral equation

(4) $$m(t) = \mu \int_0^t m(t-u) f_T(u)\, du + \int_t^\infty f_T(u)\, du \quad \text{where} \quad \mu = G'(1).$$

We can find the general solution to this equation only by numerical or series methods. It is reasonably amenable to Laplace transform methods and produces a closed expression for the Laplace transform of m. Later we shall use renewal theory arguments (see Example (10.4.22)) to show that there exist $\delta > 0$ and $\beta > 0$ such that $m(t) \sim \delta e^{\beta t}$ as $t \to \infty$ whenever $\mu > 1$.

Finally observe that, in some sense, the age-dependent process $Z(t)$ contains the old process Z_n. We say that Z_n is *imbedded* in $Z(t)$ in that we can recapture Z_n by aggregating the generation sizes of $Z(t)$. This imbedding enables us to use properties of Z_n to derive corresponding properties of the less tractable $Z(t)$. For instance, $Z(t)$ dies out if and only if Z_n dies out, and so Theorem (5.4.5) provides us immediately with the extinction probability of the age-dependent process. This technique has uses elsewhere as well. With any non-Markov process we can try to find an imbedded Markov process which provides information about the original process. We consider examples of this later.

Exercises for Section 5.5

1. Let Z_n be the size of the nth generation in an age-dependent branching process $Z(t)$, the lifetime distribution of which is exponential with parameter λ. If $Z(0) = 1$, show that the probability generating function $G_t(s)$ of $Z(t)$ satisfies

$$\frac{\partial}{\partial t} G_t(s) = \lambda \big\{ G(G_t(s)) - G_t(s) \big\}.$$

Show in the case of 'exponential binary fission', when $G(s) = s^2$, that

$$G_t(s) = \frac{s e^{-\lambda t}}{1 - s(1 - e^{-\lambda t})}$$

and hence derive the probability mass function of the population size $Z(t)$ at time t.

2. Solve the differential equation of Exercise (1) when $\lambda = 1$ and $G(s) = \frac{1}{2}(1 + s^2)$, to obtain

$$G_t(s) = \frac{2s + t(1-s)}{2 + t(1-s)}.$$

Hence find $\mathbb{P}(Z(t) \geq k)$, and deduce that

$$\mathbb{P}\big(Z(t)/t \geq x \,\big|\, Z(t) > 0\big) \to e^{-2x} \quad \text{as } t \to \infty.$$

5.6 Expectation revisited

This section is divided into parts A and B. All readers must read part A before they proceed to the next section; part B is for people with a keener appreciation of detailed technique. We are about to extend the definition of probability generating functions to more general types of variables than those concentrated on the non-negative integers, and it is a suitable moment to insert some discussion of the expectation of an arbitrary random variable regardless of its type (discrete, continuous, and so on). Up to now we have made only guarded remarks about such variables.

(A) Notation

Remember that the expectations of discrete and continuous variables are given respectively by

(1) $$\mathbb{E}X = \sum xf(x) \quad \text{if } X \text{ has mass function } f,$$

(2) $$\mathbb{E}X = \int xf(x)\,dx \quad \text{if } X \text{ has density function } f.$$

We require a single piece of notation which incorporates both these cases. Suppose X has distribution function F. Subject to a trivial and unimportant condition, (1) and (2) can be rewritten as

(3) $$\mathbb{E}X = \sum x\,dF(x) \quad \text{where } dF(x) = F(x) - \lim_{y\uparrow x} F(y) = f(x),$$

(4) $$\mathbb{E}X = \int x\,dF(x) \quad \text{where } dF(x) = \frac{dF}{dx}dx = f(x)\,dx.$$

This suggests that we denote $\mathbb{E}X$ by

(5) $$\mathbb{E}X = \int x\,dF \quad \text{or} \quad \int x\,dF(x)$$

whatever the type of X, where (5) is interpreted as (3) for discrete variables and as (4) for continuous variables. We adopt this notation forthwith. Those readers who fail to conquer an aversion to this notation should read dF as $f(x)\,dx$. Previous properties of expectation received two statements and proofs which can now be unified. For instance, (3.3.3) and (4.3.3) become

(6) $$\text{if} \quad g : \mathbb{R} \to \mathbb{R} \quad \text{then} \quad \mathbb{E}(g(X)) = \int g(x)\,dF.$$

(B) Abstract integration

The expectation of a random variable X is specified by its distribution function F. But F itself is describable in terms of X and the underlying probability space, and it follows that $\mathbb{E}X$ can be thus described also. This part contains a brief sketch of how to integrate on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. It contains no details, and the reader is left to check up on his or her intuition elsewhere (see Clarke 1975 or Williams 1991 for example). Let $(\Omega, \mathcal{F}, \mathbb{P})$ be some probability space.

(7) The random variable $X : \Omega \to \mathbb{R}$ is called *simple* if it takes only finitely many distinct values. Simple variables can be written in the form

$$X = \sum_{i=1}^{n} x_i I_{A_i}$$

for some partition $A_1, A_2, \dots, A_n$ of Ω and some real numbers $x_1, x_2, \dots, x_n$; we define the *integral* of X, written $\mathbb{E}X$ or $\mathbb{E}(X)$, to be

$$\mathbb{E}(X) = \sum_{i=1}^{n} x_i \mathbb{P}(A_i).$$

(8) Any non-negative random variable $X : \Omega \to [0, \infty)$ is the limit of some increasing sequence $\{X_n\}$ of simple variables. That is, $X_n(\omega) \uparrow X(\omega)$ for all $\omega \in \Omega$. We define the *integral* of X, written $\mathbb{E}(X)$, to be

$$\mathbb{E}(X) = \lim_{n\to\infty} \mathbb{E}(X_n).$$

This is well defined in the sense that two increasing sequences of simple functions, both converging to X, have the same limit for their sequences of integrals. The limit $\mathbb{E}(X)$ can be $+\infty$.

(9) Any random variable $X : \Omega \to \mathbb{R}$ can be written as the difference $X = X^+ - X^-$ of non-negative random variables

$$X^+(\omega) = \max\{X(\omega), 0\}, \quad X^-(\omega) = -\min\{X(\omega), 0\}.$$

If at least one of $\mathbb{E}(X^+)$ and $\mathbb{E}(X^-)$ is finite, then we define the *integral* of X, written $\mathbb{E}(X)$, to be

$$\mathbb{E}(X) = \mathbb{E}(X^+) - \mathbb{E}(X^-).$$

(10) Thus, $\mathbb{E}(X)$ is well defined, at least for any variable X such that

$$\mathbb{E}|X| = \mathbb{E}(X^+ + X^-) < \infty.$$

(11) In the language of measure theory $\mathbb{E}(X)$ is denoted by

$$\mathbb{E}(X) = \int_\Omega X(\omega)\, d\mathbb{P} \quad \text{or} \quad \mathbb{E}(X) = \int_\Omega X(\omega)\mathbb{P}(d\omega).$$

The *expectation operator* $\mathbb{E}$ defined in this way has all the properties which were described in detail for discrete and continuous variables.

(12) Continuity of $\mathbb{E}$. Important further properties are the following. If $\{X_n\}$ is a sequence of variables with $X_n(\omega) \to X(\omega)$ for all $\omega \in \Omega$ then

(a) (*monotone convergence*) if $X_n(\omega) \geq 0$ and $X_n(\omega) \leq X_{n+1}(\omega)$ for all n and ω, then $\mathbb{E}(X_n) \to \mathbb{E}(X)$,

(b) (*dominated convergence*) if $|X_n(\omega)| \le Y(\omega)$ for all n and ω, and $\mathbb{E}|Y| < \infty$, then $\mathbb{E}(X_n) \to \mathbb{E}(X)$,

(c) (*bounded convergence*, a special case of dominated convergence) if $|X_n(\omega)| \le c$ for some constant c and all n and ω then $\mathbb{E}(X_n) \to \mathbb{E}(X)$.

Rather more is true. Events having zero probability (that is, null events) make no contributions to expectations, and may therefore be ignored. Consequently, it suffices to assume above that $X_n(\omega) \to X(\omega)$ for all ω *except possibly on some null event*, with a similar weakening of the hypotheses of (a), (b), and (c). For example, the bounded convergence theorem is normally stated as follows: if $\{X_n\}$ is a sequence of random variables satisfying $X_n \to X$ a.s. and $|X_n| \le c$ a.s. for some constant c, then $\mathbb{E}(X_n) \to \mathbb{E}(X)$. The expression 'a.s.' is an abbreviation for 'almost surely', and means 'except possibly on an event of zero probability'.

Here is a useful consequence of monotone convergence. Let $Z_1, Z_2, \dots$ be non-negative random variables with finite expectations, and let $X = \sum_{i=1}^{\infty} Z_i$. We have by monotone convergence applied to the partial sums of the Z_i that

(13)
$$\mathbb{E}(X) = \sum_{i=1}^{\infty} \mathbb{E}(Z_i),$$

whether or not the summation is finite.

One further property of expectation is called *Fatou's lemma*: if $\{X_n\}$ is a sequence of random variables such that $X_n \ge Y$ a.s. for all n and some Y with $\mathbb{E}|Y| < \infty$, then

(14)
$$\mathbb{E}\Big(\liminf_{n\to\infty} X_n\Big) \le \liminf_{n\to\infty} \mathbb{E}(X_n).$$

This inequality is often applied in practice with $Y = 0$.

(15) Lebesgue–Stieltjes integral. Let X have distribution function F. The function F gives rise to a probability measure μ_F on the Borel sets of $\mathbb{R}$ as follows:

(a) define $\mu_F\big((a, b]\big) = F(b) - F(a)$,

(b) as in the discussion after (4.1.5), the domain of μ_F can be extended to include the Borel σ-field $\mathcal{B}$, being the smallest σ-field containing all half-open intervals $(a, b]$.

So $(\mathbb{R}, \mathcal{B}, \mu_F)$ is a probability space; its completion (see Section 1.6) is denoted by the triple $(\mathbb{R}, \mathcal{L}_F, \mu_F)$, where $\mathcal{L}_F$ is the smallest σ-field containing $\mathcal{B}$ and all subsets of μ_F-null sets. If $g : \mathbb{R} \to \mathbb{R}$ (is $\mathcal{L}_F$-measurable) then the abstract integral $\int g\, d\mu_F$ is called the *Lebesgue–Stieltjes integral* of g with respect to μ_F, and we normally denote it by $\int g(x)\, dF$ or $\int g(x)\, dF(x)$. Think of it as a special case of the abstract integral (11). The purpose of this discussion is the assertion that if $g : \mathbb{R} \to \mathbb{R}$ (and g is suitably measurable) then $g(X)$ is random variable and

$$\mathbb{E}(g(X)) = \int g(x)\, dF,$$

and we adopt this forthwith as the official notation for expectation. Here is a final word of caution. If $g(x) = I_B(x)h(x)$ where I_B is the indicator function of some $B \subseteq \mathbb{R}$ then

$$\int g(x)\, dF = \int_B h(x)\, dF.$$

We do not in general obtain the same result when we integrate over $B_1 = [a, b]$ and $B_2 = (a, b)$ unless F is continuous at a and b, and so we do not use the notation $\int_a^b h(x)\, dF$ unless there is no danger of ambiguity.

Exercises for Section 5.6

1. **Jensen's inequality.** A function $u : \mathbb{R} \to \mathbb{R}$ is called *convex* if for all real a there exists λ, depending on a, such that $u(x) \geq u(a)+\lambda(x-a)$ for all x. (Draw a diagram to illustrate this definition.) Show that, if u is convex and X is a random variable with finite mean, then $\mathbb{E}(u(X)) \geq u(\mathbb{E}(X))$.

2. Let $X_1, X_2, \dots$ be random variables satisfying $\mathbb{E}\left(\sum_{i=1}^{\infty} |X_i|\right) < \infty$. Show that

$$\mathbb{E}\left(\sum_{i=1}^{\infty} X_i\right) = \sum_{i=1}^{\infty} \mathbb{E}(X_i).$$

3. Let $\{X_n\}$ be a sequence of random variables satisfying $X_n \leq Y$ a.s. for some Y with $\mathbb{E}|Y| < \infty$. Show that

$$\mathbb{E}\left(\limsup_{n\to\infty} X_n\right) \geq \limsup_{n\to\infty} \mathbb{E}(X_n).$$

4. Suppose that $\mathbb{E}|X^r| < \infty$ where $r > 0$. Deduce that $x^r \mathbb{P}(|X| \geq x) \to 0$ as $x \to \infty$. Conversely, suppose that $x^r \mathbb{P}(|X| \geq x) \to 0$ as $x \to \infty$ where $r \geq 0$, and show that $\mathbb{E}|X^s| < \infty$ for $0 \leq s < r$.

5. Show that $\mathbb{E}|X| < \infty$ if and only if the following holds: for all $\epsilon > 0$, there exists $\delta > 0$, such that $\mathbb{E}(|X| I_A) < \epsilon$ for all A such that $\mathbb{P}(A) < \delta$.

5.7 Characteristic functions

Probability generating functions proved to be very useful in handling non-negative integral random variables. For more general variables X it is natural to make the substitution $s = e^t$ in the quantity $G_X(s) = \mathbb{E}(s^X)$.

(1) Definition. The moment generating function of a variable X is the function $M : \mathbb{R} \to [0, \infty)$ given by $M(t) = \mathbb{E}(e^{tX})$.

Moment generating functions are related to Laplace transforms† since

$$M(t) = \int e^{tx}\, dF(x) = \int e^{tx} f(x)\, dx$$

if X is continuous with density function f. They have properties similar to those of probability generating functions. For example, if $M(t) < \infty$ on some open interval containing the origin then:

(a) $\mathbb{E}X = M'(0)$, $\mathbb{E}(X^k) = M^{(k)}(0)$;

(b) the function M may be expanded via Taylor's theorem within its circle of convergence,

$$M(t) = \sum_{k=0}^{\infty} \frac{\mathbb{E}(X^k)}{k!} t^k,$$

which is to say that M is the 'exponential generating function' of the sequence of moments of X;

†Note the change of sign from the usual Laplace transform of f, namely $\widehat{f}(t) = \int e^{-tx} f(x)\, dx$.

(c) if X and Y are independent then† $M_{X+Y}(t) = M_X(t)M_Y(t)$.

Moment generating functions provide a very useful technique but suffer the disadvantage that the integrals which define them may not always be finite. Rather than explore their properties in detail we move on immediately to another class of functions that are equally useful and whose finiteness is guaranteed.

(2) Definition. **The characteristic function** of X is the function $\phi : \mathbb{R} \to \mathbb{C}$ defined by

$$\phi(t) = \mathbb{E}(e^{itX}) \quad \text{where} \quad i = \sqrt{-1}.$$

We often write ϕ_X for the characteristic function of the random variable X. Characteristic functions are related to Fourier transforms, since $\phi(t) = \int e^{itx}\, dF(x)$. In the notation of Section 5.6, ϕ is the abstract integral of a complex-valued random variable. It is well defined in the terms of Section 5.6 by $\phi(t) = \mathbb{E}(\cos tX) + i\mathbb{E}(\sin tX)$. Furthermore, ϕ is better behaved than the moment generating function M.

(3) Theorem. *The characteristic function ϕ satisfies*:

(a) $\phi(0) = 1$, $|\phi(t)| \le 1$ *for all* t,

(b) ϕ *is uniformly continuous on* $\mathbb{R}$,

(c) ϕ *is non-negative definite, which is to say that* $\sum_{j,k} \phi(t_j - t_k) z_j \bar{z}_k \ge 0$ *for all real* $t_1, t_2, \dots, t_n$ *and complex* $z_1, z_2, \dots, z_n$.

Proof. (a) Clearly $\phi(0) = \mathbb{E}(1) = 1$. Furthermore

$$|\phi(t)| \le \int |e^{itx}|\, dF = \int dF = 1.$$

(b) We have that

$$|\phi(t+h) - \phi(t)| = \left|\mathbb{E}(e^{i(t+h)X} - e^{itX})\right| \le \mathbb{E}\left|e^{itX}(e^{ihX} - 1)\right| \le \mathbb{E}(Y(h))$$

where $Y(h) = |e^{ihX} - 1|$. However, $|Y(h)| \le 2$ and $Y(h) \to 0$ as $h \to 0$, and so $\mathbb{E}(Y(h)) \to 0$ by bounded convergence (5.6.12).

(c) We have that

$$\begin{aligned} \sum_{j,k} \phi(t_j - t_k) z_j \bar{z}_k &= \sum_{j,k} \int [z_j \exp(it_j x)][\bar{z}_k \exp(-it_k x)]\, dF \\ &= \mathbb{E}\left(\left|\sum_j z_j \exp(it_j X)\right|^2\right) \ge 0. \end{aligned}$$

■

Theorem (3) characterizes characteristic functions in the sense that ϕ is a characteristic function if and only if it satisfies (3a), (3b), and (3c). This result is called Bochner's theorem, for which we offer no proof. Many of the properties of characteristic functions rely for their proofs on a knowledge of complex analysis. This is a textbook on probability theory, and will

†This is essentially the assertion that the Laplace transform of a convolution (see equation (4.8.2)) is the product of the Laplace transforms.

not include such proofs unless they indicate some essential technique. We have asserted that the method of characteristic functions is very useful; however, we warn the reader that we shall not make use of them until Section 5.10. In the meantime we shall establish some of their properties.

First and foremost, from a knowledge of ϕ_X we can recapture the distribution of X. The full power of this statement is deferred until the next section; here we concern ourselves only with the moments of X. Several of the interesting characteristic functions are not very well behaved, and we must move carefully.

(4) Theorem.

(a) *If* $\phi^{(k)}(0)$ *exists then* $\begin{cases} \mathbb{E}|X^k| < \infty & \text{if } k \text{ is even,} \\ \mathbb{E}|X^{k-1}| < \infty & \text{if } k \text{ is odd.} \end{cases}$

(b) *If* $\mathbb{E}|X^k| < \infty$ *then*†

$$\phi(t) = \sum_{j=0}^{k} \frac{\mathbb{E}(X^j)}{j!}(it)^j + \mathrm{o}(t^k),$$

and so $\phi^{(k)}(0) = i^k \mathbb{E}(X^k)$.

Proof. This is essentially Taylor's theorem for a function of a complex variable. For the proof, see Moran (1968) or Kingman and Taylor (1966). ■

One of the useful properties of characteristic functions is that they enable us to handle sums of independent variables with the minimum of fuss.

(5) Theorem. *If X and Y are independent then* $\phi_{X+Y}(t) = \phi_X(t)\phi_Y(t)$.

Proof. We have that

$$\phi_{X+Y}(t) = \mathbb{E}(e^{it(X+Y)}) = \mathbb{E}(e^{itX}e^{itY}).$$

Expand each exponential term into cosines and sines, multiply out, use independence, and put back together to obtain the result. ■

(6) Theorem. *If* $a, b \in \mathbb{R}$ *and* $Y = aX + b$ *then* $\phi_Y(t) = e^{itb}\phi_X(at)$.

Proof. We have that

$$\begin{aligned} \phi_Y(t) &= \mathbb{E}(e^{it(aX+b)}) = \mathbb{E}(e^{itb}e^{i(at)X}) \\ &= e^{itb}\mathbb{E}(e^{i(at)X}) = e^{itb}\phi_X(at). \end{aligned}$$ ■

We shall make repeated use of these last two theorems. We sometimes need to study collections of variables which may be dependent.

(7) Definition. The **joint characteristic function** of X and Y is the function $\phi_{X,Y} : \mathbb{R}^2 \to \mathbb{R}$ given by $\phi_{X,Y}(s,t) = \mathbb{E}(e^{isX}e^{itY})$.

Notice that $\phi_{X,Y}(s,t) = \phi_{sX+tY}(1)$. As usual we shall be interested mostly in independent variables.

†See Subsection (10) of Appendix I for a reminder about Landau's O/o notation.

(8) Theorem. *Random variables X and Y are independent if and only if*

$$\phi_{X,Y}(s,t) = \phi_X(s)\phi_Y(t) \quad \textit{for all } s \textit{ and } t.$$

Proof. If X and Y are independent then the conclusion follows by the argument of (5). The converse is proved by extending the inversion theorem of the next section to deal with joint distributions and showing that the joint distribution function factorizes. ∎

Note particularly that for X and Y to be independent it is not sufficient that

(9) $$\phi_{X,Y}(t,t) = \phi_X(t)\phi_Y(t) \quad \text{for all } t.$$

Exercise. Can you find an example of dependent variables which satisfy (9)?

We have seen in Theorem (4) that it is an easy calculation to find the moments of X by differentiating its characteristic function $\phi_X(t)$ at $t = 0$. A similar calculation gives the 'joint moments' $\mathbb{E}(X^j Y^k)$ of two variables from a knowledge of their joint characteristic function $\phi_{X,Y}(s,t)$ (see Problem (5.12.30) for details).

The properties of moment generating functions are closely related to those of characteristic functions. In the rest of the text we shall use the latter whenever possible, but it will be appropriate to use the former for any topic whose analysis employs Laplace transforms; for example, this is the case for the queueing theory of Chapter 11.

(10) Remark. Moment problem. If I am given a distribution function F, then I can calculate the corresponding moments $m_k(F) = \int_{-\infty}^{\infty} x^k \, dF(x)$, $k = 1, 2, \dots$, whenever these integrals exist. Is the converse true: does the collection of moments $(m_k(F) : k = 1, 2, \dots)$ specify F uniquely? The answer is *no*: there exist distribution functions F and G, all of whose moments exist, such that $F \neq G$ but $m_k(F) = m_k(G)$ for all k. The usual example is obtained by using the log-normal distribution (see Problem (5.12.43)).

Under what conditions on F is it the case that no such G exists? Various sets of conditions are known which guarantee that F is specified by its moments, but no necessary and sufficient condition is known which is easy to apply to a general distribution. Perhaps the simplest sufficient condition is that the moment generating function of F, $M(t) = \int_{-\infty}^{\infty} e^{tx} \, dF(x)$, be finite in some neighbourhood of the point $t = 0$. Those familiar with the theory of Laplace transforms will understand why this is sufficient.

(11) Remark. Moment generating function. The characteristic function of a distribution is closely related to its moment generating function, in a manner made rigorous in the following theorem, the proof of which is omitted. [See Lukacs 1970, pp. 197–198.]

(12) Theorem. *Let $M(t) = \mathbb{E}(e^{tX})$, $t \in \mathbb{R}$, and $\phi(t) = \mathbb{E}(e^{itX})$, $t \in \mathbb{C}$, be the moment generating function and characteristic function, respectively, of a random variable X. For any $a > 0$, the following three statements are equivalent*:

(a) $|M(t)| < \infty$ *for* $|t| < a$,

(b) ϕ *is analytic on the strip* $|\mathrm{Im}(z)| < a$,

(c) *the moments* $m_k = \mathbb{E}(X^k)$ *exist for* $k = 1, 2, \dots$ *and satisfy* $\limsup_{k\to\infty} \{|m_k|/k!\}^{1/k} \le a^{-1}$.

If any of these conditions hold for $a > 0$, the power series expansion for $M(t)$ may be extended analytically to the strip $|\mathrm{Im}(t)| < a$, resulting in a function M with the property that $\phi(t) = M(it)$. [See Moran 1968, p. 260.]

Exercises for Section 5.7

1. Find two dependent random variables X and Y such that $\phi_{X+Y}(t) = \phi_X(t)\phi_Y(t)$ for all t.

2. If ϕ is a characteristic function, show that $\mathrm{Re}\{1 - \phi(t)\} \ge \frac{1}{4}\mathrm{Re}\{1 - \phi(2t)\}$, and deduce that $1 - |\phi(2t)| \le 8\{1 - |\phi(t)|\}$.

3. The **cumulant generating function** $K_X(\theta)$ of the random variable X is defined by $K_X(\theta) = \log \mathbb{E}(e^{\theta X})$, the logarithm of the moment generating function of X. If the latter is finite in a neighbourhood of the origin, then K_X has a convergent Taylor expansion:

$$K_X(\theta) = \sum_{n=1}^{\infty} \frac{1}{n!} k_n(X)\theta^n$$

and $k_n(X)$ is called the nth *cumulant* (or *semi-invariant*) of X.

(a) Express $k_1(X)$, $k_2(X)$, and $k_3(X)$ in terms of the moments of X.

(b) If X and Y are independent random variables, show that $k_n(X+Y) = k_n(X) + k_n(Y)$.

4. Let X be $N(0, 1)$, and show that the cumulants of X are $k_2(X) = 1$, $k_m(X) = 0$ for $m \ne 2$.

5. The random variable X is said to have a *lattice distribution* if there exist a and b such that X takes values in the set $L(a, b) = \{a + bm : m = 0, \pm 1, \dots\}$. The *span* of such a variable X is the maximal value of b for which there exists a such that X takes values in $L(a, b)$.

(a) Suppose that X has a lattice distribution with span b. Show that $|\phi_X(2\pi/b)| = 1$, and that $|\phi_X(t)| < 1$ for $0 < t < 2\pi/b$.

(b) Suppose that $|\phi_X(\theta)| = 1$ for some $\theta \ne 0$. Show that X has a lattice distribution with span $2\pi k/\theta$ for some integer k.

6. Let X be a random variable with density function f. Show that $|\phi_X(t)| \to 0$ as $t \to \pm\infty$.

7. Let $X_1, X_2, \dots, X_n$ be independent variables, X_i being $N(\mu_i, 1)$, and let $Y = X_1^2 + X_2^2 + \cdots + X_n^2$. Show that the characteristic function of Y is

$$\phi_Y(t) = \frac{1}{(1-2it)^{n/2}} \exp\left(\frac{it\theta}{1-2it}\right)$$

where $\theta = \mu_1^2 + \mu_2^2 + \cdots + \mu_n^2$. The random variables Y is said to have the *non-central chi-squared distribution* with n degrees of freedom and non-centrality parameter θ, written $\chi^2(n; \theta)$.

8. Let X be $N(\mu, 1)$ and let Y be $\chi^2(n)$, and suppose that X and Y are independent. The random variable $T = X/\sqrt{Y/n}$ is said to have the *non-central t-distribution* with n degrees of freedom and non-centrality parameter μ. If U and V are independent, U being $\chi^2(m; \theta)$ and V being $\chi^2(n)$, then $F = (U/m)/(V/n)$ is said to have the *non-central F-distribution* with m and n degrees of freedom and non-centrality parameter θ, written $F(m, n; \theta)$.

(a) Show that T^2 is $F(1, n; \mu^2)$.

(b) Show that

$$\mathbb{E}(F) = \frac{n(m+\theta)}{m(n-2)} \quad \text{if } n > 2.$$

9. Let X be a random variable with density function f and characteristic function ϕ. Show, subject to an appropriate condition on f, that

$$\int_{-\infty}^{\infty} f(x)^2\,dx = \frac{1}{2\pi}\int_{-\infty}^{\infty} |\phi(t)|^2\,dt.$$

10. If X and Y are continuous random variables, show that

$$\int_{-\infty}^{\infty} \phi_X(y) f_Y(y) e^{-ity}\, dy = \int_{-\infty}^{\infty} \phi_Y(x-t) f_X(x)\, dx.$$

11. Tilted distributions. (a) Let X have distribution function F and let τ be such that $M(\tau) = \mathbb{E}(e^{\tau X}) < \infty$. Show that $F_\tau(x) = M(\tau)^{-1} \int_{-\infty}^{x} e^{\tau y}\, dF(y)$ is a distribution function, called a 'tilted distribution' of X, and find its moment generating function.

(b) Suppose X and Y are independent and $\mathbb{E}(e^{\tau X})$, $\mathbb{E}(e^{\tau Y}) < \infty$. Find the moment generating function of the tilted distribution of $X + Y$ in terms of those of X and Y.

5.8 Examples of characteristic functions

Those who feel daunted by $i = \sqrt{-1}$ should find it a useful exercise to work through this section using $M(t) = \mathbb{E}(e^{tX})$ in place of $\phi(t) = \mathbb{E}(e^{itX})$. Many calculations here are left as *exercises*.

(1) Example. Bernoulli distribution. If X is Bernoulli with parameter p then

$$\phi(t) = \mathbb{E}(e^{itX}) = e^{it0} \cdot q + e^{it1} \cdot p = q + pe^{it}.$$ ●

(2) Example. Binomial distribution. If X is bin(n, p) then X has the same distribution as the sum of n independent Bernoulli variables $Y_1, Y_2, \dots, Y_n$. Thus

$$\phi_X(t) = \phi_{Y_1}(t)\phi_{Y_2}(t) \cdots \phi_{Y_n}(t) = (q + pe^{it})^n.$$ ●

(3) Example. Exponential distribution. If $f(x) = \lambda e^{-\lambda x}$ for $x \geq 0$ then

$$\phi(t) = \int_0^{\infty} e^{itx} \lambda e^{-\lambda x}\, dx.$$

This is a complex integral and its solution relies on a knowledge of how to integrate around contours in $\mathbb{R}^2$ (the appropriate contour is a sector). Alternatively, the integral may be evaluated by writing $e^{itx} = \cos(tx) + i\sin(tx)$, and integrating the real and imaginary part separately. Do not fall into the trap of treating i as if its were a real number, even though this malpractice yields the correct answer in this case:

$$\phi(t) = \frac{\lambda}{\lambda - it}.$$ ●

(4) Example. Cauchy distribution. If $f(x) = 1/\{\pi(1 + x^2)\}$ then

$$\phi(t) = \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{e^{itx}}{1 + x^2}\, dx.$$

Treating i as a real number will not help you to avoid the contour integral this time. Those who are interested should try integrating around a semicircle with diameter $[-R, R]$ on the real axis, thereby obtaining the required characteristic function $\phi(t) = e^{-|t|}$. Alternatively, you might work backwards from the answer thus: you can calculate the Fourier transform of the function $e^{-|t|}$, and then use the Fourier inversion theorem. ●

(5) Example. Normal distribution. If X is $N(0, 1)$ then

$$\phi(t) = \mathbb{E}(e^{itX}) = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \exp(itx - \tfrac{1}{2}x^2)\, dx.$$

Again, do not treat i as a real number. Consider instead the moment generating function of X

$$M(s) = \mathbb{E}(e^{sX}) = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \exp(sx - \tfrac{1}{2}x^2)\, dx.$$

Complete the square in the integrand and use the hint at the end of Example (4.5.9) to obtain $M(s) = e^{\frac{1}{2}s^2}$. We may not substitute $s = it$ without justification. In this particular instance the theory of analytic continuation of functions of a complex variable provides this justification, see Remark (5.7.11), and we deduce that

$$\phi(t) = e^{-\frac{1}{2}t^2}.$$

By Theorem (5.7.6), the characteristic function of the $N(\mu, \sigma^2)$ variable $Y = \sigma X + \mu$ is

$$\phi_Y(t) = e^{it\mu}\phi_X(\sigma t) = \exp(i\mu t - \tfrac{1}{2}\sigma^2 t^2).$$ ●

(6) Example. Multivariate normal distribution. If $X_1, X_2, \dots, X_n$ has the multivariate normal distribution $N(\mathbf{0}, \mathbf{V})$ then its joint density function is

$$f(\mathbf{x}) = \frac{1}{\sqrt{(2\pi)^n|\mathbf{V}|}} \exp(-\tfrac{1}{2}\mathbf{x}\mathbf{V}^{-1}\mathbf{x}').$$

The joint characteristic function of $X_1, X_2, \dots, X_n$ is the function $\phi(\mathbf{t}) = \mathbb{E}(e^{i\mathbf{t}\mathbf{X}'})$ where $\mathbf{t} = (t_1, t_2, \dots, t_n)$ and $\mathbf{X} = (X_1, X_2, \dots, X_n)$. One way to proceed is to use the fact that $\mathbf{t}\mathbf{X}'$ is univariate normal. Alternatively,

$$\phi(\mathbf{t}) = \int_{\mathbb{R}^n} \frac{1}{\sqrt{(2\pi)^n|\mathbf{V}|}} \exp\big(i\mathbf{t}\mathbf{x}' - \tfrac{1}{2}\mathbf{x}\mathbf{V}^{-1}\mathbf{x}'\big)\, d\mathbf{x}. \tag{7}$$

As in the discussion of Section 4.9, there is a linear transformation $\mathbf{y} = \mathbf{x}\mathbf{B}$ such that

$$\mathbf{x}\mathbf{V}^{-1}\mathbf{x}' = \sum_j \lambda_j y_j^2$$

just as in equation (4.9.3). Make this transformation in (7) to see that the integrand factorizes into the product of functions of the single variables $y_1, y_2, \dots, y_n$. Then use (5) to obtain

$$\phi(t) = \exp(-\tfrac{1}{2}\mathbf{t}\mathbf{V}\mathbf{t}').$$

It is now an easy *exercise* to prove Theorem (4.9.5), that $\mathbf{V}$ is the covariance matrix of $\mathbf{X}$, by using the result of Problem (5.12.30). ●

(8) Example. Gamma distribution. If X is $\Gamma(\lambda, s)$ then

$$\phi(t) = \int_0^\infty \frac{1}{\Gamma(s)} \lambda^s x^{s-1} \exp(itx - \lambda x)\,dx.$$

As in the case of the exponential distribution (3), routine methods of complex analysis give

$$\phi(t) = \left(\frac{\lambda}{\lambda - it}\right)^s.$$

Why is this similar to the result of (3)? This example includes the chi-squared distribution because a $\chi^2(d)$ variable is $\Gamma(\frac{1}{2}, \frac{1}{2}d)$ and thus has characteristic function

$$\phi(t) = (1 - 2it)^{-d/2}.$$

You may try to prove this from the result of Problem (4.14.12). ●

Exercises for Section 5.8

1. If ϕ is a characteristic function, show that $\overline{\phi}$, ϕ^2, $|\phi|^2$, $\mathrm{Re}(\phi)$ are characteristic functions. Show that $|\phi|$ is not necessarily a characteristic function.

2. Show that

$$\mathbb{P}(X \geq x) \leq \inf_{t \geq 0} \{e^{-tx} M_X(t)\},$$

where M_X is the moment generating function of X.

3. Let X have the $\Gamma(\lambda, m)$ distribution and let Y be independent of X with the beta distribution with parameters n and $m - n$, where m and n are non-negative integers satisfying $n \leq m$. Show that $Z = XY$ has the $\Gamma(\lambda, n)$ distribution.

4. Find the characteristic function of X^2 when X has the $N(\mu, \sigma^2)$ distribution.

5. Let $X_1, X_2, \ldots$ be independent $N(0, 1)$ variables. Use characteristic functions to find the distribution of: (a) X_1^2, (b) $\sum_{i=1}^n X_i^2$, (c) X_1/X_2, (d) X_1X_2, (e) $X_1X_2 + X_3X_4$.

6. Let $X_1, X_2, \ldots, X_n$ be such that, for all $a_1, a_2, \ldots, a_n \in \mathbb{R}$, the linear combination $a_1X_1 + a_2X_2 + \cdots + a_nX_n$ has a normal distribution. Show that the joint characteristic function of the X_m is $\exp(i\mathbf{t}\boldsymbol{\mu}' - \frac{1}{2}\mathbf{t}\mathbf{V}\mathbf{t}')$, for an appropriate vector $\boldsymbol{\mu}$ and matrix $\mathbf{V}$. Deduce that the vector $(X_1, X_2, \ldots, X_n)$ has a multivariate normal *density function* so long as $\mathbf{V}$ is invertible.

7. Let X and Y be independent $N(0, 1)$ variables, and let U and V be independent of X and Y. Show that $Z = (UX + VY)/\sqrt{U^2 + V^2}$ has the $N(0, 1)$ distribution. Formulate an extension of this result to cover the case when X and Y have a bivariate normal distribution with zero means, unit variances, and correlation ρ.

8. Let X be exponentially distributed with parameter λ. Show by elementary integration that $\mathbb{E}(e^{itX}) = \lambda/(\lambda - it)$.

9. Find the characteristic functions of the following density functions:

(a) $f(x) = \frac{1}{2}e^{-|x|}$ for $x \in \mathbb{R}$,

(b) $f(x) = \frac{1}{2}|x|e^{-|x|}$ for $x \in \mathbb{R}$.

10. Is it possible for X, Y, and Z to have the same distribution and satisfy $X = U(Y + Z)$, where U is uniform on $[0, 1]$, and Y, Z are independent of U and of one another? (This question arises in modelling energy redistribution among physical particles.)

11. Find the joint characteristic function of two random variables having a bivariate normal distribution with zero means. (No integration is needed.)

5.9 Inversion and continuity theorems

This section contains accounts of two major ways in which characteristic functions are useful. The first of these states that the distribution of a random variable is specified by its characteristic function. That is to say, if X and Y have the same characteristic function then they have the same distribution. Furthermore, there is a formula which tells us how to recapture the distribution function F corresponding to the characteristic function ϕ. Here is a special case first.

(1) Theorem. *If X is continuous with density function f and characteristic function ϕ then*

$$f(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-itx}\phi(t)\,dt$$

at every point x at which f is differentiable.

Proof. This is the Fourier inversion theorem and can be found in any introduction to Fourier transforms. If the integral fails to converge absolutely then we interpret it as its principal value (see Apostol 1974, p. 277). ∎

A sufficient, but not necessary condition that a characteristic function ϕ be the characteristic function of a continuous variable is that

$$\int_{-\infty}^{\infty} |\phi(t)|\,dt < \infty.$$

The general case is more complicated, and is contained in the next theorem.

(2) Inversion theorem. *Let X have distribution function F and characteristic function ϕ. Define $\overline{F} : \mathbb{R} \to [0, 1]$ by*

$$\overline{F}(x) = \frac{1}{2}\left\{F(x) + \lim_{y \uparrow x} F(y)\right\}.$$

Then

$$\overline{F}(b) - \overline{F}(a) = \lim_{N\to\infty}\int_{-N}^{N} \frac{e^{-iat} - e^{-ibt}}{2\pi i t}\phi(t)\,dt.$$

Proof. See Kingman and Taylor (1966). ∎

(3) Corollary. *Random variables X and Y have the same characteristic function if and only if they have the same distribution function.*

Proof. If $\phi_X = \phi_Y$ then, by (2),

$$\overline{F}_X(b) - \overline{F}_X(a) = \overline{F}_Y(b) - \overline{F}_Y(a).$$

Let $a \to -\infty$ to obtain $\overline{F}_X(b) = \overline{F}_Y(b)$; now, for any fixed $x \in \mathbb{R}$, let $b \downarrow x$ and use right-continuity and Lemma (2.1.6c) to obtain $F_X(x) = F_Y(x)$. ■

Exactly similar results hold for jointly distributed random variables. For example, if X and Y have joint density function f and joint characteristic function ϕ then whenever f is differentiable at (x, y)

$$f(x, y) = \frac{1}{4\pi^2} \iint_{\mathbb{R}^2} e^{-isx} e^{-ity} \phi(s, t)\, ds\, dt$$

and Theorem (5.7.8) follows straightaway for this special case.

The second result of this section deals with a sequence $X_1, X_2, \dots$ of random variables. Roughly speaking it asserts that if the distribution functions $F_1, F_2, \dots$ of the sequence approach some limit F then the characteristic functions $\phi_1, \phi_2, \dots$ of the sequence approach the characteristic function of the distribution function F.

(4) **Definition.** We say that the sequence $F_1, F_2, \dots$ of distribution functions **converges** to the distribution function F, written $F_n \to F$, if $F(x) = \lim_{n\to\infty} F_n(x)$ at each point x where F is continuous.

The reason for the condition of continuity of F at x is indicated by the following example. Define the distribution functions F_n and G_n by

$$F_n(x) = \begin{cases} 0 & \text{if } x < n^{-1}, \\ 1 & \text{if } x \ge n^{-1}, \end{cases} \qquad G_n(x) = \begin{cases} 0 & \text{if } x < -n^{-1}, \\ 1 & \text{if } x \ge -n^{-1}. \end{cases}$$

We have as $n \to \infty$ that

$$\begin{aligned} F_n(x) &\to F(x) \qquad \text{if } x \ne 0, \quad F_n(0) \to 0, \\ G_n(x) &\to F(x) \qquad \text{for all } x, \end{aligned}$$

where F is the distribution function of a random variable which is constantly zero. Indeed $\lim_{n\to\infty} F_n(x)$ is not even a distribution function since it is not right-continuous at zero. It is intuitively reasonable to demand that the sequences $\{F_n\}$ and $\{G_n\}$ have the same limit, and so we drop the requirement that $F_n(x) \to F(x)$ at the point of discontinuity of F.

(5) **Continuity theorem.** *Suppose that $F_1, F_2, \dots$ is a sequence of distribution functions with corresponding characteristic functions $\phi_1, \phi_2, \dots$.*

(a) *If $F_n \to F$ for some distribution function F with characteristic function ϕ, then $\phi_n(t) \to \phi(t)$ for all t.*

(b) *Conversely, if $\phi(t) = \lim_{n\to\infty} \phi_n(t)$ exists and is continuous at $t = 0$, then ϕ is the characteristic function of some distribution function F, and $F_n \to F$.*

Proof. As for (2). See also Problem (5.12.35). ■

(6) **Example. Stirling's formula.** This well-known formula† states that $n! \sim n^n e^{-n}\sqrt{2\pi n}$ as $n \to \infty$, which is to say that

$$\frac{n!}{n^n e^{-n}\sqrt{2\pi n}} \to 1 \quad \text{as} \quad n \to \infty.$$

†Due to de Moivre.

A more general form of this relation states that

$$\frac{\Gamma(t)}{t^{t-1}e^{-t}\sqrt{2\pi t}} \to 1 \quad \text{as} \quad t \to \infty \tag{7}$$

where Γ is the gamma function, $\Gamma(t) = \int_0^\infty x^{t-1}e^{-x}\,dx$. Remember that $\Gamma(t) = (t-1)!$ if t is a positive integer; see Example (4.4.6) and Exercise (4.4.1). To prove (7) is an 'elementary' exercise in analysis, (see Exercise (5.9.6)), but it is perhaps amusing to see how simply (7) follows from the Fourier inversion theorem (1).

Let Y be a random variable with the $\Gamma(1,t)$ distribution. Then $X = (Y-t)/\sqrt{t}$ has density function

$$f_t(x) = \frac{1}{\Gamma(t)}\sqrt{t}\big(x\sqrt{t}+t\big)^{t-1}\exp\big[-\big(x\sqrt{t}+t\big)\big], \quad -\sqrt{t} \le x < \infty, \tag{8}$$

and characteristic function

$$\phi_t(u) = \mathbb{E}(e^{iuX}) = \exp\big(-iu\sqrt{t}\big)\left(1 - \frac{iu}{\sqrt{t}}\right)^{-t}.$$

Now $f_t(x)$ is differentiable with respect to x on $(-\sqrt{t}, \infty)$. We apply Theorem (1) at $x = 0$ to obtain

$$f_t(0) = \frac{1}{2\pi}\int_{-\infty}^{\infty}\phi_t(u)\,du. \tag{9}$$

However, $f_t(0) = t^{t-\frac{1}{2}}e^{-t}/\Gamma(t)$ from (8); also

$$\begin{aligned}
\phi_t(u) &= \exp\left[-iu\sqrt{t} - t\log\left(1 - \frac{iu}{\sqrt{t}}\right)\right] \\
&= \exp\left[-iu\sqrt{t} - t\left(-\frac{iu}{\sqrt{t}} + \frac{u^2}{2t} + \mathrm{O}(u^3t^{-\frac{3}{2}})\right)\right] \\
&= \exp\big[-\tfrac{1}{2}u^2 + \mathrm{O}(u^3t^{-\frac{1}{2}})\big] \to e^{-\frac{1}{2}u^2} \quad \text{as} \quad t \to \infty.
\end{aligned}$$

Taking the limit in (9) as $t \to \infty$, we find that

$$\begin{aligned}
\lim_{t\to\infty}\left(\frac{1}{\Gamma(t)}t^{t-\frac{1}{2}}e^{-t}\right) &= \lim_{t\to\infty}\frac{1}{2\pi}\int_{-\infty}^{\infty}\phi_t(u)\,du \\
&= \frac{1}{2\pi}\int_{-\infty}^{\infty}\left(\lim_{t\to\infty}\phi_t(u)\right)du \\
&= \frac{1}{2\pi}\int_{-\infty}^{\infty}e^{-\frac{1}{2}u^2}\,du = \frac{1}{\sqrt{2\pi}}
\end{aligned}$$

as required for (7). A spot of rigour is needed to justify the interchange of the limit and the integral sign above, and this may be provided by the dominated convergence theorem. ●

Exercises for Section 5.9

1. Let X_n be a discrete random variable taking values in $\{1, 2, \dots, n\}$, each possible value having probability n^{-1}. Show that, as $n \to \infty$, $\mathbb{P}(n^{-1}X_n \le y) \to y$, for $0 \le y \le 1$.

2. Let X_n have distribution function

$$F_n(x) = x - \frac{\sin(2n\pi x)}{2n\pi}, \qquad 0 \le x \le 1.$$

(a) Show that F_n is indeed a distribution function, and that X_n has a density function.

(b) Show that, as $n \to \infty$, F_n converges to the uniform distribution function, but that the density function of F_n does not converge to the uniform density function.

3. A coin is tossed repeatedly, with heads turning up with probability p on each toss. Let N be the minimum number of tosses required to obtain k heads. Show that, as $p \downarrow 0$, the distribution function of $2Np$ converges to that of a gamma distribution.

4. If X is an integer-valued random variable with characteristic function ϕ, show that

$$\mathbb{P}(X = k) = \frac{1}{2\pi}\int_{-\pi}^{\pi} e^{-itk}\phi(t)\,dt.$$

What is the corresponding result for a random variable whose distribution is arithmetic with span λ (that is, there is probability one that X is a multiple of λ, and λ is the largest positive number with this property)?

5. Use the inversion theorem to show that

$$\int_{-\infty}^{\infty} \frac{\sin(at)\sin(bt)}{t^2}\,dt = \pi\min\{a, b\}.$$

6. Stirling's formula. Let $f_n(x)$ be a differentiable function on $\mathbb{R}$ with a a global maximum at $a > 0$, and such that $\int_0^\infty \exp\{f_n(x)\}\,dx < \infty$. Laplace's method of steepest descent (related to Watson's lemma and saddlepoint methods) asserts under mild conditions that

$$\int_0^\infty \exp\{f_n(x)\}\,dx \sim \int_0^\infty \exp\{f_n(a) + \tfrac{1}{2}(x-a)^2 f_n''(a)\}\,dx \quad \text{as } n \to \infty.$$

By setting $f_n(x) = n\log x - x$, prove Stirling's formula: $n! \sim n^n e^{-n}\sqrt{2\pi n}$.

7. Let $\mathbf{X} = (X_1, X_2, \dots, X_n)$ have the multivariate normal distribution with zero means, and covariance matrix $\mathbf{V} = (v_{ij})$ satisfying $|\mathbf{V}| > 0$ and $v_{ij} > 0$ for all i, j. Show that

$$\frac{\partial f}{\partial v_{ij}} = \begin{cases} \dfrac{\partial^2 f}{\partial x_i \partial x_j} & \text{if } i \ne j, \\[2ex] \dfrac{1}{2}\dfrac{\partial^2 f}{\partial x_i^2} & \text{if } i = j, \end{cases}$$

and deduce that $\mathbb{P}(\max_{k\le n} X_k \le u) \ge \prod_{k=1}^n \mathbb{P}(X_k \le u)$.

8. Let X_1, X_2 have a bivariate normal distribution with zero means, unit variances, and correlation ρ. Use the inversion theorem to show that

$$\frac{\partial}{\partial\rho}\mathbb{P}(X_1 > 0,\ X_2 > 0) = \frac{1}{2\pi\sqrt{1-\rho^2}}.$$

Hence find $\mathbb{P}(X_1 > 0,\ X_2 > 0)$.

5.10 Two limit theorems

We are now in a position to prove two very celebrated theorems in probability theory, the 'law of large numbers' and the 'central limit theorem'. The first of these explains the remarks of Sections 1.1 and 1.3, where we discussed a heuristic foundation of probability theory. Part of our intuition about chance is that if we perform many repetitions of an experiment which has numerical outcomes then the average of all the outcomes settles down to some fixed number. This observation deals in the convergence of sequences of random variables, the general theory of which is dealt with later. Here it suffices to introduce only one new definition.

(1) Definition. If $X, X_1, X_2, \dots$ is a sequence of random variables with respective distribution functions $F, F_1, F_2, \dots$, we say that X_n **converges in distribution**† to X, written $X_n \xrightarrow{\text{D}} X$, if $F_n \to F$ as $n \to \infty$.

This is just Definition (5.9.4) rewritten in terms of random variables.

(2) Theorem. Law of large numbers. *Let $X_1, X_2, \dots$ be a sequence of independent identically distributed random variables with finite means μ. Their partial sums $S_n = X_1 + X_2 + \cdots + X_n$ satisfy*

$$\frac{1}{n} S_n \xrightarrow{\text{D}} \mu \quad \textit{as} \quad n \to \infty.$$

Proof. The theorem asserts that, as $n \to \infty$,

$$\mathbb{P}(n^{-1} S_n \le x) \to \begin{cases} 0 & \text{if } x < \mu, \\ 1 & \text{if } x > \mu. \end{cases}$$

The method of proof is clear. By the continuity theorem (5.9.5) we need to show that the characteristic function of $n^{-1} S_n$ approaches the characteristic function of the constant random variable μ. Let ϕ be the common characteristic function of the X_i, and let ϕ_n be the characteristic function of $n^{-1} S_n$. By Theorems (5.7.5) and (5.7.6),

$$\phi_n(t) = \{\phi_X(t/n)\}^n. \tag{3}$$

The behaviour of $\phi_X(t/n)$ for large n is given by Theorem (5.7.4) as $\phi_X(t) = 1 + it\mu + \mathrm{o}(t)$. Substitute into (3) to obtain

$$\phi_n(t) = \left\{ 1 + \frac{i\mu t}{n} + \mathrm{o}\left(\frac{t}{n}\right) \right\}^n \to e^{it\mu} \quad \text{as} \quad n \to \infty.$$

However, this limit is the characteristic function of the constant μ, and the result follows. ■

So, for large n, the sum S_n is approximately as big as $n\mu$. What can we say about the difference $S_n - n\mu$? There is an extraordinary answer to this question, valid whenever the X_i have finite variance:

(a) $S_n - n\mu$ is about as big as $\sqrt{n}$,

---

†Also termed *weak convergence* or *convergence in law*. See Section 7.2.

(b) the distribution of $(S_n - n\mu)/\sqrt{n}$ approaches the normal distribution as $n \to \infty$ *irrespective* of the distribution of the X_i.

(4) Central limit theorem. *Let $X_1, X_2, \dots$ be a sequence of independent identically distributed random variables with finite mean μ and finite non-zero variance σ^2, and let $S_n = X_1 + X_2 + \cdots + X_n$. Then*

$$\frac{S_n - n\mu}{\sqrt{n\sigma^2}} \xrightarrow{\mathrm{D}} N(0,1) \quad \textit{as} \quad n \to \infty.$$

Note that the assertion of the theorem is an abuse of notation, since $N(0, 1)$ is a distribution and not a random variable; the above is admissible because convergence in distribution involves only the corresponding distribution functions. The method of proof is the same as for the law of large numbers.

Proof. First, write $Y_i = (X_i - \mu)/\sigma$, and let ϕ_Y be the characteristic function of the Y_i. We have by Theorem (5.7.4) that $\phi_Y(t) = 1 - \frac{1}{2}t^2 + \mathrm{o}(t^2)$. Also, the characteristic function ψ_n of

$$U_n = \frac{S_n - n\mu}{\sqrt{n\sigma^2}} = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} Y_i$$

satisfies, by Theorems (5.7.5) and (5.7.6),

$$\psi_n(t) = \left\{\phi_Y(t/\sqrt{n})\right\}^n = \left\{1 - \frac{t^2}{2n} + \mathrm{o}\left(\frac{t^2}{n}\right)\right\}^n \to e^{-\frac{1}{2}t^2} \quad \text{as} \quad n \to \infty.$$

The last function is the characteristic function of the $N(0, 1)$ distribution, and an application of the continuity theorem (5.9.5) completes the proof. ■

Numerous generalizations of the law of large numbers and the central limit theorem are available. For example, in Chapter 7 we shall meet two stronger versions of (2), involving weaker assumptions on the X_i and more powerful conclusions. The central limit theorem can be generalized in several directions, two of which deal with dependent variables and differently distributed variables respectively. Some of these are within the reader's grasp. Here is an example.

(5) Theorem. *Let $X_1, X_2, \dots$ be independent variables satisfying*

$$\mathbb{E}X_j = 0, \quad \operatorname{var}(X_j) = \sigma_j^2, \quad \mathbb{E}|X_j^3| < \infty,$$

and such that

$$\frac{1}{\sigma(n)^3} \sum_{j=1}^{n} \mathbb{E}|X_j^3| \to 0 \quad \textit{as} \quad n \to \infty,$$

where $\sigma(n)^2 = \operatorname{var}\left(\sum_{j=1}^{n} X_j\right) = \sum_{j=1}^{n} \sigma_j^2$. Then

$$\frac{1}{\sigma(n)} \sum_{j=1}^{n} X_j \xrightarrow{\mathrm{D}} N(0,1).$$

Proof. See Loève (1977, p. 287), and also Problem (5.12.40). ■

The roots of central limit theory are at least 250 years old. The first proof of (4) was found by de Moivre around 1733 for the special case of Bernoulli variables with $p = \frac{1}{2}$. General values of p were treated later by Laplace. Their methods involved the direct estimation of sums of the form

$$\sum_{\substack{k:\\ k \le np + x\sqrt{npq}}} \binom{n}{k} p^k q^{n-k} \quad \text{where} \quad p + q = 1.$$

The first rigorous proof of (4) was discovered by Lyapunov around 1901, thereby confirming a less rigorous proof of Laplace. A glance at these old proofs confirms that the method of characteristic functions is outstanding in its elegance and brevity.

The central limit theorem (4) asserts that the *distribution function* of S_n, suitably normalized to have mean 0 and variance 1, converges to the distribution function of the $N(0, 1)$ distribution. Is the corresponding result valid at the level of density functions and mass functions? Broadly speaking the answer is yes, but some condition of smoothness is necessary; after all, if $F_n(x) \to F(x)$ as $n \to \infty$ for all x, it is not necessarily the case that the derivatives satisfy $F_n'(x) \to F'(x)$. [See Exercise (5.9.2).] The result which follows is called a 'local central limit theorem' since it deals in the local rather than in the cumulative behaviour of the random variables in question. In order to simplify the statement of the theorem, we shall assume that the X_i have zero mean and unit variance.

(6) Local central limit theorem. *Let $X_1, X_2, \dots$ be independent identically distributed random variables with zero mean and unit variance, and suppose further that their common characteristic function ϕ satisfies*

$$\text{(7)} \qquad \int_{-\infty}^{\infty} |\phi(t)|^r \, dt < \infty$$

for some integer $r \ge 1$. The density function g_n of $U_n = (X_1 + X_2 + \cdots + X_n)/\sqrt{n}$ exists for $n \ge r$, and furthermore

$$\text{(8)} \qquad g_n(x) \to \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} \quad \textit{as } n \to \infty, \textit{ uniformly in } x \in \mathbb{R}.$$

A similar result is valid for sums of lattice-valued random variables, suitably adjusted to have zero mean and unit variance. We state this here, leaving its proof as an *exercise*. In place of (7) we assume that the X_i are restricted to take the values $a, a \pm h, a \pm 2h, \dots$, where h is the largest positive number for which such a restriction holds. Then U_n is restricted to values of the form $x = (na + kh)/\sqrt{n}$ for $k = 0, \pm 1, \dots$. For such a number x, we write $g_n(x) = \mathbb{P}(U_n = x)$ and leave $g_n(y)$ undefined for other values of y. It is the case that

$$\text{(9)} \qquad \frac{\sqrt{n}}{h} g_n(x) \to \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} \quad \text{as } n \to \infty, \text{ uniformly in appropriate } x.$$

Proof of (6). A certain amount of analysis is inevitable here. First, the assumption that $|\phi|^r$ is integrable for some $r \ge 1$ implies that $|\phi|^n$ is integrable for $n \ge r$, since $|\phi(t)| \le 1$; hence g_n exists and is given by the Fourier inversion formula

$$\text{(10)} \qquad g_n(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-itx} \psi_n(t) \, dt,$$

where $\psi_n(t) = \phi(t/\sqrt{n})^n$ is the characteristic function of U_n. The Fourier inversion theorem is valid for the normal distribution, and therefore

$$\text{(11)} \qquad \left| g_n(x) - \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} \right| \le \frac{1}{2\pi} \left| \int_{-\infty}^{\infty} e^{-itx} \left[\phi(t/\sqrt{n})^n - e^{-\frac{1}{2}t^2} \right] dt \right| \le I_n$$

where

$$I_n = \frac{1}{2\pi} \int_{-\infty}^{\infty} \left| \phi(t\sqrt{n})^n - e^{-\frac{1}{2}t^2} \right| dt.$$

It suffices to show that $I_n \to 0$ as $n \to \infty$. We have from Theorem (5.7.4) that $\phi(t) = 1 - \frac{1}{2}t^2 + o(t^2)$ as $t \to 0$, and therefore there exists δ (> 0) such that

$$\text{(12)} \qquad |\phi(t)| \le e^{-\frac{1}{4}t^2} \quad \text{if} \quad |t| \le \delta.$$

Now, for any $a > 0$, $\phi(t/\sqrt{n})^n \to e^{-\frac{1}{2}t^2}$ as $n \to \infty$ *uniformly* in $t \in [-a, a]$ (to see this, investigate the proof of (4) slightly more carefully), so that

$$\text{(13)} \qquad \int_{-a}^{a} \left| \phi(t/\sqrt{n})^n - e^{-\frac{1}{2}t^2} \right| dt \to 0 \quad \text{as} \quad n \to \infty,$$

for any a. Also, by (12),

$$\text{(14)} \qquad \int_{a<|t|\le\delta\sqrt{n}} \left| \phi(t/\sqrt{n})^n - e^{-\frac{1}{2}t^2} \right| dt \le 2 \int_a^{\infty} 2e^{-\frac{1}{4}t^2}\, dt$$

which tends to zero as $a \to \infty$.

It remains to deal with the contribution to I_n arising from $|t| > \delta\sqrt{n}$. From the fact that g_n exists for $n \ge r$, we have from Exercises (5.7.5) and (5.7.6) that $|\phi(t)^r| < 1$ for $t \ne 0$, and $|\phi(t)^r| \to 0$ as $t \to \pm\infty$. Hence $|\phi(t)| < 1$ for $t \ne 0$, and $|\phi(t)| \to 0$ as $t \to \pm\infty$, and therefore $\eta = \sup\{|\phi(t)| : |t| \ge \delta\}$ satisfies $\eta < 1$. Now, for $n \ge r$,

$$\begin{aligned} \text{(15)} \quad \int_{|t|>\delta\sqrt{n}} \left| \phi(t/\sqrt{n})^n - e^{-\frac{1}{2}t^2} \right| dt &\le \eta^{n-r} \int_{-\infty}^{\infty} |\phi(t/\sqrt{n})|^r\, dt + 2\int_{\delta\sqrt{n}}^{\infty} e^{-\frac{1}{2}t^2}\, dt \\ &= \eta^{n-r}\sqrt{n} \int_{-\infty}^{\infty} |\phi(u)|^r\, du + 2\int_{\delta\sqrt{n}}^{\infty} e^{-\frac{1}{2}t^2}\, dt \\ &\to 0 \quad \text{as} \quad n \to \infty. \end{aligned}$$

Combining (13)–(15), we deduce that

$$\lim_{n\to\infty} I_n \le 4 \int_a^{\infty} e^{-\frac{1}{4}t^2}\, dt \to 0 \quad \text{as} \quad a \to \infty,$$

so that $I_n \to 0$ as $n \to \infty$ as required. ■

(16) Example. Random walks. Here is an application of the law of large numbers to the persistence of random walks. A simple random walk performs steps of size 1, to the right or left with probability p and $1 - p$. We saw in Section 5.3 that a simple random walk is persistent (that is, returns to its starting point with probability 1) if and only if it is symmetric (which is to say that $p = 1 - p = \frac{1}{2}$). Think of this as saying that the walk is persistent if and only if the mean value of a typical step X satisfies $\mathbb{E}(X) = 0$, that is, each step is 'unbiased'. This conclusion is valid in much greater generality.

Let $X_1, X_2, \dots$ be independent identically distributed integer-valued random variables, and let $S_n = X_1 + X_2 + \dots + X_n$. We think of X_i as being the ith jump of a random walk, so that S_n is the position of the random walker after n jumps, having started at $S_0 = 0$. We call the walk *persistent* (or *recurrent*) if $\mathbb{P}(S_n = 0 \text{ for some } n \geq 1) = 1$ and *transient* otherwise.

(17) Theorem. *The random walk is persistent if the mean size of jumps is* 0.

The converse is valid also: the walk is transient if the mean size of jumps is non-zero (Problem (5.12.44)).

Proof. Suppose that $\mathbb{E}(X_i) = 0$ and let V_i denote the mean number of visits of the walk to the point i,

$$V_i = \mathbb{E}\big|\{n \geq 0 : S_n = i\}\big| = \mathbb{E}\bigg(\sum_{n=0}^{\infty} I_{\{S_n=i\}}\bigg) = \sum_{n=0}^{\infty} \mathbb{P}(S_n = i),$$

where I_A is the indicator function of the event A. We shall prove first that $V_0 = \infty$, and from this we shall deduce the persistence of the walk. Let T be the time of the first visit of the walk to i, with the convention that $T = \infty$ if i is never visited. Then

$$\begin{aligned} V_i = \sum_{n=0}^{\infty} \mathbb{P}(S_n = i) &= \sum_{n=0}^{\infty}\sum_{t=0}^{\infty} \mathbb{P}(S_n = i \mid T = t)\mathbb{P}(T = t) \\ &= \sum_{t=0}^{\infty}\sum_{n=t}^{\infty} \mathbb{P}(S_n = i \mid T = t)\mathbb{P}(T = t) \end{aligned}$$

since $S_n \neq i$ for $n < T$. Now we use the spatial homogeneity of the walk to deduce that

$$V_i = \sum_{t=0}^{\infty} V_0 \mathbb{P}(T = t) = V_0 \mathbb{P}(T < \infty) \leq V_0. \tag{18}$$

The mean number of time points n for which $|S_n| \leq K$ satisfies

$$\sum_{n=0}^{\infty} \mathbb{P}(|S_n| \leq K) = \sum_{i=-K}^{K} V_i \leq (2K + 1)V_0$$

by (18), and hence

$$V_0 \geq \frac{1}{2K + 1} \sum_{n=0}^{\infty} \mathbb{P}(|S_n| \leq K). \tag{19}$$

Now we use the law of large numbers. For $\epsilon > 0$, it is the case that $\mathbb{P}(|S_n| \leq n\epsilon) \to 1$ as $n \to \infty$, so that there exists m such that $\mathbb{P}(|S_n| \leq n\epsilon) > \frac{1}{2}$ for $n \geq m$. If $n\epsilon \leq K$ then $\mathbb{P}(|S_n| \leq n\epsilon) \leq \mathbb{P}(|S_n| \leq K)$, so that

$$\mathbb{P}(|S_n| \leq K) > \tfrac{1}{2} \quad \text{for} \quad m \leq n \leq K/\epsilon. \tag{20}$$

Substituting (20) into (19), we obtain

$$V_0 \geq \frac{1}{2K+1} \sum_{m \leq n \leq K/\epsilon} \mathbb{P}(|S_n| \leq K) > \frac{1}{2(2K+1)} \left(\frac{K}{\epsilon} - m - 1 \right).$$

This is valid for all large K, and we may therefore let $K \to \infty$ and $\epsilon \downarrow 0$ in that order, finding that $V_0 = \infty$ as claimed.

It is now fairly straightforward to deduce that the walk is persistent. Let $T(1)$ be the time of the first return to 0, with the convention that $T(1) = \infty$ if this never occurs. If $T(1) < \infty$, we write $T(2)$ for the subsequent time which elapses until the next visit to 0. It is clear from the homogeneity of the process that, conditional on $\{T(1) < \infty\}$, the random variable $T(2)$ has the same distribution as $T(1)$. Continuing likewise, we see that the times of returns to 0 are distributed in the same way as the sequence $T_1, T_1 + T_2, \dots$, where $T_1, T_2, \dots$ are independent identically distributed random variables having the same distribution as $T(1)$. We wish to exclude the possibility that $\mathbb{P}(T(1) = \infty) > 0$. There are several ways of doing this, one of which is to make use of the recurrent-event analysis of Example (5.2.15). We shall take a slightly more direct route here. Suppose that $\beta = \mathbb{P}(T(1) = \infty)$ satisfies $\beta > 0$, and let $I = \min\{i : T_i = \infty\}$ be the earliest i for which T_i is infinite. The event $\{I = i\}$ corresponds to exactly $i-1$ returns to the origin. Thus, the mean number of returns is $\sum_{i=1}^{\infty}(i-1)\mathbb{P}(I = i)$. However, $I = i$ if and only if $T_j < \infty$ for $1 \leq j < i$ and $T_i = \infty$, an event with probability $(1-\beta)^{i-1}\beta$. Hence the mean number of returns to 0 is $\sum_{i=1}^{\infty}(i-1)(1-\beta)^{i-1}\beta = (1-\beta)/\beta$, which is finite. This contradicts the infiniteness of V_0, and hence $\beta = 0$. ■

We have proved that a walk whose jumps have zero mean must (with probability 1) return to its starting point. It follows that it must return *infinitely often*, since otherwise there exists some T_i which equals infinity, an event having zero probability. ●

(21) Example. Recurrent events. The renewal theorem of Example (5.2.15) is one of the basic results of applied probability, and it will recur in various forms through this book. Our 'elementary' proof in Example (5.2.15) was incomplete, but we may now complete it with the aid of the last theorem (17) concerning the persistence of random walks.

Suppose that we are provided with two sequences $X_1, X_2, \dots$ and $X_1^*, X_2^*, \dots$ of independent identically distributed random variables taking values in the positive integers $\{1, 2, \dots\}$. Let $Y_n = X_n - X_n^*$ and $S_n = \sum_{i=1}^{n} Y_i = \sum_{i=1}^{n} X_i - \sum_{i=1}^{n} X_i^*$. Then $S = \{S_n : n \geq 0\}$ may be thought of as a random walk on the integers with steps $Y_1, Y_2, \dots$; the mean step size satisfies $\mathbb{E}(Y_1) = \mathbb{E}(X_1) - \mathbb{E}(X_1^*) = 0$, and therefore this walk is persistent, by Theorem (17). Furthermore, the walk must revisit its starting point *infinitely often* (with probability 1), which is to say that $\sum_{i=1}^{n} X_i = \sum_{i=1}^{n} X_i^*$ for infinitely many values of n.

What have we proved about recurrent-event processes? Consider two independent recurrent-event processes for which the first occurrence times, X_1 and X_1^*, have the same distribution as the inter-occurrence times. Not only does there exist some finite time T at which the event H occurs simultaneously in both processes, but also: (i) there exist infinitely many such times T, and (ii) there exist infinitely many such times T even if one insists that, by time T, the event H has occurred the *same number of times* in the two processes.

We need to relax the assumption that X_1 and X_1^* have the same distribution as the inter-occurrence times, and it is here that we require that the process be non-arithmetic. Suppose that $X_1 = u$ and $X_1^* = v$. Now $S_n = S_1 + \sum_{i=2}^{n} Y_i$ is a random walk with mean jump size

0 and starting point $S_1 = u - v$. By the foregoing argument, there exist (with probability 1) infinitely many values of n such that $S_n = u - v$, which is to say that

$$\sum_{i=2}^{n} X_i = \sum_{i=2}^{n} X_i^*; \tag{22}$$

we denote these (random) times by the increasing sequence $N_1, N_2, \dots$.

The process is non-arithmetic, and it follows that, for any integer x, there exist integers r and s such that

$$\gamma(r, s; x) = \mathbb{P}\big((X_2 + X_3 + \cdots + X_r) - (X_2^* + X_3^* + \cdots + X_s^*) = x\big) > 0. \tag{23}$$

To check this is an elementary *exercise* (5.10.4) in number theory. The reader may be satisfied with the following proof for the special case when $\beta = \mathbb{P}(X_2 = 1)$ satisfies $\beta > 0$. Then

$$\mathbb{P}\big(X_2 + X_3 + \cdots + X_{x+1} = x\big) \ge \mathbb{P}\big(X_i = 1 \text{ for } 2 \le i \le x + 1\big) = \beta^x > 0$$

if $x \ge 0$, and

$$\mathbb{P}\big(-X_2^* - X_3^* - \cdots - X_{|x|+1}^* = x\big) \ge \mathbb{P}\big(X_i^* = 1 \text{ for } 2 \le i \le |x| + 1\big) = \beta^{|x|} > 0$$

if $x < 0$, so that (23) is valid with $r = x + 1$, $s = 1$ and $r = 1$, $s = |x| + 1$ in these two respective cases. Without more ado we shall accept that such r, s exist under the assumption that the process is non-arithmetic. We set $x = -(u - v)$, choose r and s accordingly, and write $\gamma = \gamma(r, s; x)$.

Suppose now that (22) occurs for some value of n. Then

$$\sum_{i=1}^{n+r-1} X_i - \sum_{i=1}^{n+s-1} X_i^* = (X_1 - X_1^*) + \left(\sum_{i=n+1}^{n+r-1} X_i - \sum_{i=n+1}^{n+s-1} X_i^*\right)$$

which equals $(u - v) - (u - v) = 0$ with strictly positive probability (since the contents of the final parentheses have, by (23), strictly positive probability of equalling $-(u - v)$). Therefore, for each n satisfying (22), there is a strictly positive probability γ that the $(n + r - 1)$th recurrence of the first process coincides with the $(n + s - 1)$th recurrence of the second. There are infinitely many such values N_i for n, and one of infinitely many shots at a target must succeed! More rigorously, define $M_1 = N_1$, and $M_{i+1} = \min\{N_j : N_j > M_i + \max\{r, s\}\}$; the sequence of the M_i is an infinite subsequence of the N_j satisfying $M_{i+1} - M_i > \max\{r, s\}$. Call M_i a *failure* if the $(M_i + r - 1)$th recurrence of the first process does not coincide with the $(M_i + s - 1)$th of the second. Then the events $F_I = \{M_i \text{ is a failure for } 1 \le i \le I\}$ satisfy

$$\mathbb{P}(F_{I+1}) = \mathbb{P}(M_{I+1} \text{ is a failure} \mid F_I)\mathbb{P}(F_I) = (1 - \gamma)\mathbb{P}(F_I),$$

so that $\mathbb{P}(F_I) = (1 - \gamma)^I \to 0$ as $I \to \infty$. However, $\{F_I : I \ge 1\}$ is a decreasing sequence of events with limit $\{M_i \text{ is a failure for all } i\}$, which event therefore has zero probability. Thus one of the M_i is *not* a failure, with probability 1, implying that some recurrence of the first process coincides with some recurrence of the second, as required.

The above argument is valid for all 'initial values' u and v for X_1 and X_1^*, and therefore for all choices of the distribution of X_1 and X_1^*:

$$\begin{aligned}\mathbb{P}(\text{coincident recurrences}) &= \sum_{u,v} \mathbb{P}\big(\text{coincident recurrences} \mid X_1 = u,\ X_1^* = v\big) \\ &\qquad\qquad \times \mathbb{P}(X_1 = u)\mathbb{P}(X_1^* = v) \\ &= \sum_{u,v} 1 \cdot \mathbb{P}(X_1 = u)\mathbb{P}(X_1^* = v) = 1.\end{aligned}$$

In particular, the conclusion is valid when X_1^* has probability generating function D^* given by equation (5.2.22); the proof of the renewal theorem is thereby completed. ●

Exercises for Section 5.10

1. Prove that, for $x \geq 0$, as $n \to \infty$,

(a)
$$\sum_{\substack{k:\\ |k-\frac{1}{2}n| \leq \frac{1}{2}x\sqrt{n}}} \binom{n}{k} \sim 2^n \int_{-x}^{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}u^2}\, du,$$

(b)
$$\sum_{\substack{k:\\ |k-n| \leq x\sqrt{n}}} \frac{n^k}{k!} \sim e^n \int_{-x}^{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}u^2}\, du.$$

2. It is well known that infants born to mothers who smoke tend to be small and prone to a range of ailments. It is conjectured that also they look abnormal. Nurses were shown selections of photographs of babies, one half of whom had smokers as mothers; the nurses were asked to judge from a baby's appearance whether or not the mother smoked. In 1500 trials the correct answer was given 910 times. Is the conjecture plausible? If so, why?

3. Let X have the $\Gamma(1, s)$ distribution; given that $X = x$, let Y have the Poisson distribution with parameter x. Find the characteristic function of Y, and show that

$$\frac{Y - \mathbb{E}(Y)}{\sqrt{\operatorname{var}(Y)}} \xrightarrow{\text{D}} N(0, 1) \qquad \text{as } s \to \infty.$$

Explain the connection with the central limit theorem.

4. Let $X_1, X_2, \dots$ be independent random variables taking values in the positive integers, whose common distribution is non-arithmetic, in that $\gcd\{n : \mathbb{P}(X_1 = n) > 0\} = 1$. Prove that, for all integers x, there exist non-negative integers $r = r(x)$, $s = s(x)$, such that

$$\mathbb{P}\big(X_1 + \cdots + X_r - X_{r+1} - \cdots - X_{r+s} = x\big) > 0.$$

5. Prove the local central limit theorem for sums of random variables taking integer values. You may assume for simplicity that the summands have span 1, in that $\gcd\big\{|x| : \mathbb{P}(X = x) > 0\big\} = 1$.

6. Let $X_1, X_2, \dots$ be independent random variables having common density function $f(x) = 1/\{2|x|(\log|x|)^2\}$ for $|x| < e^{-1}$. Show that the X_i have zero mean and finite variance, and that the density function f_n of $X_1 + X_2 + \cdots + X_n$ satisfies $f_n(x) \to \infty$ as $x \to 0$. Deduce that the X_i do not satisfy the local limit theorem.

7. **First-passage density.** Let X have the density function $f(x) = \sqrt{2\pi x^{-3}} \exp(-\{2x\}^{-1})$, $x > 0$. Show that $\phi(is) = \mathbb{E}(e^{-sX}) = e^{-\sqrt{2s}}$, $s > 0$, and deduce that X has characteristic function

$$\phi(t) = \begin{cases} \exp\{-(1-i)\sqrt{t}\} & \text{if } t \geq 0, \\ \exp\{-(1+i)\sqrt{|t|}\} & \text{if } t \leq 0. \end{cases}$$

[Hint: Use the result of Problem (5.12.18).]

8. Let $\{X_r : r \geq 1\}$ be independent with the distribution of the preceding Exercise (7). Let $U_n = n^{-1} \sum_{r=1}^{n} X_r$, and $T_n = n^{-1} U_n$. Show that:
(a) $\mathbb{P}(U_n < c) \to 0$ for any $c < \infty$,
(b) T_n has the same distribution as X_1.

9. A sequence of biased coins is flipped; the chance that the rth coin shows a head is Θ_r, where Θ_r is a random variable taking values in $(0, 1)$. Let X_n be the number of heads after n flips. Does X_n obey the central limit theorem when:
(a) the Θ_r are independent and identically distributed?
(b) $\Theta_r = \Theta$ for all r, where Θ is a random variable taking values in $(0, 1)$?

5.11 Large deviations

The law of large numbers asserts that, in a certain sense, the sum S_n of n independent identically distributed variables is approximately $n\mu$, where μ is a typical mean. The central limit theorem implies that the deviations of S_n from $n\mu$ are typically of the order $\sqrt{n}$, that is, small compared with the mean. Now, S_n may deviate from $n\mu$ by quantities of greater order than $\sqrt{n}$, say n^α where $\alpha > \frac{1}{2}$, but such 'large deviations' have probabilities which tend to zero as $n \to \infty$. It is often necessary in practice to estimate such probabilities. The theory of large deviations studies the asymptotic behaviour of $\mathbb{P}(|S_n - n\mu| > n^\alpha)$ as $n \to \infty$, for values of α satisfying $\alpha > \frac{1}{2}$; of particular interest is the case when $\alpha = 1$, corresponding to deviations of S_n from its mean $n\mu$ having the same order as the mean. The behaviour of such quantities is somewhat delicate, depending on rather more than the mean and variance of a typical summand.

Let $X_1, X_2, \dots$ be a sequence of independent identically distributed random variables with mean μ and partial sums $S_n = X_1 + X_2 + \cdots + X_n$. It is our target to estimate $\mathbb{P}(S_n > na)$ where $a > \mu$. The quantity central to the required estimate is the moment generating function $M(t) = \mathbb{E}(e^{tX})$ of a typical X_i, or more exactly its logarithm $\Lambda(t) = \log M(t)$. The function Λ is also known as the *cumulant generating function* of the X_i (recall Exercise (5.7.3)).

Before proceeding, we note some properties of Λ. First,

(1) $$\Lambda(0) = \log M(0) = 0, \quad \Lambda'(0) = \frac{M'(0)}{M(0)} = \mu \quad \text{if } M'(0) \text{ exists.}$$

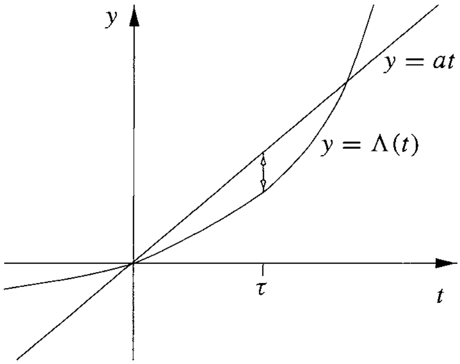
Secondly, $\Lambda(t)$ is convex wherever it is finite, since

(2) $$\Lambda''(t) = \frac{M(t)M''(t) - M'(t)^2}{M(t)^2} = \frac{\mathbb{E}(e^{tX})\mathbb{E}(X^2e^{tX}) - \mathbb{E}(Xe^{tX})^2}{M(t)^2}$$

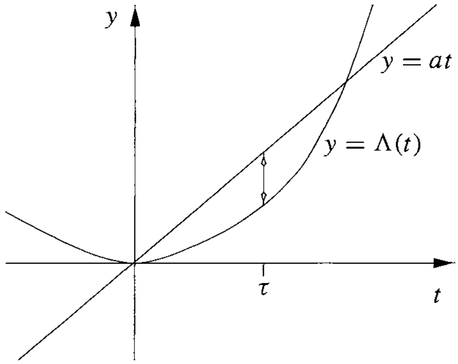
which is non-negative, by the Cauchy–Schwarz inequality (4.5.12) applied to the random variables $Xe^{\frac{1}{2}tX}$ and $e^{\frac{1}{2}tX}$. We define the *Fenchel–Legendre transform* of $\Lambda(t)$ to be the function $\Lambda^*(a)$ given by

(3) $$\Lambda^*(a) = \sup_{t\in\mathbb{R}}\{at - \Lambda(t)\}, \qquad a \in \mathbb{R}.$$

The relationship between Λ and Λ^* is illustrated in Figure 5.4.



(a) The case $\mu > 0$.



(b) The case $\mu = 0$.

Figure 5.4. A sketch of the function $\Lambda(t) = \log M(t)$ in the two cases when $\Lambda'(0) = \mu > 0$ and when $\Lambda'(0) = \mu = 0$. The value of $\Lambda^*(a)$ is found by maximizing the function $g_a(t) = at - \Lambda(t)$, as indicated by the arrows. In the regular case, the supremum is achieved within the domain of convergence of M.

(4) Theorem. Large deviations†. *Let $X_1, X_2, \dots$ be independent identically distributed random variables with mean μ, and suppose that their moment generating function $M(t) = \mathbb{E}(e^{tX})$ is finite in some neighbourhood of the origin $t = 0$. Let a be such that $a > \mu$ and $\mathbb{P}(X > a) > 0$. Then $\Lambda^*(a) > 0$ and*

$$\text{(5)} \qquad \frac{1}{n}\log\mathbb{P}(S_n > na) \to -\Lambda^*(a) \quad \text{as} \quad n \to \infty.$$

Thus, under the conditions of the theorem, $\mathbb{P}(S_n > na)$ decays exponentially in the manner of $e^{-n\Lambda^*(a)}$. We note that $\mathbb{P}(S_n > na) = 0$ if $\mathbb{P}(X > a) = 0$. The theorem may appear to deal only with deviations of S_n in *excess* of its mean; the corresponding result for deviations of S_n below the mean is obtained by replacing X_i by $-X_i$.

Proof. We may assume without loss of generality that $\mu = 0$; if $\mu \neq 0$, we replace X_i by $X_i - \mu$, noting in the obvious notation that $\Lambda_X(t) = \Lambda_{X-\mu}(t) + \mu t$ and $\Lambda_X^*(a) = \Lambda_{X-\mu}^*(a - \mu)$. Assume henceforth that $\mu = 0$.

We prove first that $\Lambda^*(a) > 0$ under the assumptions of the theorem. By the remarks after Definition (5.7.1),

$$at - \Lambda(t) = \log\left(\frac{e^{at}}{M(t)}\right) = \log\left(\frac{1 + at + o(t)}{1 + \frac{1}{2}\sigma^2 t^2 + o(t^2)}\right)$$

for small positive t, where $\sigma^2 = \operatorname{var}(X)$; we have used here the assumption that $M(t) < \infty$ near the origin. For sufficiently small positive t, $1 + at + o(t) > 1 + \frac{1}{2}\sigma^2 t^2 + o(t^2)$, whence $\Lambda^*(a) > 0$ by (3).

†A version of this theorem was first published by Cramér in 1938 using different methods. Such theorems and their ramifications have had a very substantial impact on modern probability theory and its applications.

We make two notes for future use. First, since Λ is convex with $\Lambda'(0) = \mathbb{E}(X) = 0$, and since $a > 0$, the supremum of $at - \Lambda(t)$ over $t \in \mathbb{R}$ is unchanged by the restriction $t > 0$, which is to say that

(6)
$$\Lambda^*(a) = \sup_{t>0}\{at - \Lambda(t)\}, \qquad a > 0.$$

(See Figure 5.4.) Secondly,

(7) Λ is strictly convex wherever the second derivative Λ'' exists.

To see this, note that $\mathrm{var}(X) > 0$ under the hypotheses of the theorem, implying by (2) and Theorem (4.5.12) that $\Lambda''(t) > 0$.

The upper bound for $\mathbb{P}(S_n > na)$ is derived in much the same way as was Bernstein's inequality (2.2.4). For $t > 0$, we have that $e^{tS_n} > e^{nat} I_{\{S_n > na\}}$, so that

$$\mathbb{P}(S_n > na) \le e^{-nat}\mathbb{E}(e^{tS_n}) = \{e^{-at}M(t)\}^n = e^{-n(at-\Lambda(t))}.$$

This is valid for all $t > 0$, whence, by (6),

(8)
$$\frac{1}{n}\log \mathbb{P}(S_n > na) \le -\sup_{t>0}\{at - \Lambda(t)\} = -\Lambda^*(a).$$

More work is needed for the lower bound, and there are two cases which we term the *regular* and *non-regular* cases. The regular case covers most cases of practical interest, and concerns the situation when the supremum defining $\Lambda^*(a)$ in (6) is achieved *strictly* within the domain of convergence of the moment generating function M. Under this condition, the required argument is interesting but fairly straightforward. Let $T = \sup\{t : M(t) < \infty\}$, noting that $0 < T \le \infty$. Assume that we are in the regular case, which is to say that there exists $\tau \in (0, T)$ such that the supremum in (6) is achieved at τ; that is,

(9)
$$\Lambda^*(a) = a\tau - \Lambda(\tau),$$

as sketched in Figure 5.4. Since $at - \Lambda(t)$ has a maximum at τ, and since Λ is infinitely differentiable on $(0, T)$, the derivative of $at - \Lambda(t)$ equals 0 at $t = \tau$, and therefore

(10)
$$\Lambda'(\tau) = a.$$

Let F be the common distribution function of the X_i. We introduce an ancillary distribution function $\widetilde{F}$, sometimes called an 'exponential change of distribution' or a 'tilted distribution' (recall Exercise (5.8.11)), by

(11)
$$d\widetilde{F}(u) = \frac{e^{\tau u}}{M(\tau)}\, dF(u)$$

which some may prefer to interpret as

$$\widetilde{F}(y) = \frac{1}{M(\tau)}\int_{-\infty}^{y} e^{\tau u}\, dF(u).$$

Let $\widetilde{X}_1, \widetilde{X}_2, \dots$ be independent random variables having distribution function $\widetilde{F}$, and write $\widetilde{S}_n = \widetilde{X}_1 + \widetilde{X}_2 + \cdots + \widetilde{X}_n$. We note the following properties of the $\widetilde{X}_i$. The moment generating function of the $\widetilde{X}_i$ is

(12) $$\widetilde{M}(t) = \int_{-\infty}^{\infty} e^{tu}\, d\widetilde{F}(u) = \int_{-\infty}^{\infty} \frac{e^{(t+\tau)u}}{M(\tau)}\, dF(u) = \frac{M(t+\tau)}{M(\tau)}.$$

The first two moments of the $\widetilde{X}_i$ satisfy

(13) $$\begin{aligned} \mathbb{E}(\widetilde{X}_i) &= \widetilde{M}'(0) = \frac{M'(\tau)}{M(\tau)} = \Lambda'(\tau) = a && \text{by (10)},\\ \operatorname{var}(\widetilde{X}_i) &= \mathbb{E}(\widetilde{X}_i^2) - \mathbb{E}(\widetilde{X}_i)^2 = \widetilde{M}''(0) - \widetilde{M}'(0)^2 \\ &= \Lambda''(\tau) \in (0,\infty) && \text{by (2) and (7)}. \end{aligned}$$

Since $\widetilde{S}_n$ is the sum of n independent variables, it has moment generating function

$$\left(\frac{M(t+\tau)}{M(\tau)}\right)^n = \frac{\mathbb{E}(e^{(t+\tau)S_n})}{M(\tau)^n} = \frac{1}{M(\tau)^n}\int_{-\infty}^{\infty} e^{(t+\tau)u}\, dF_n(u)$$

where F_n is the distribution function of S_n. Therefore, the distribution function $\widetilde{F}_n$ of $\widetilde{S}_n$ satisfies

(14) $$d\widetilde{F}_n(u) = \frac{e^{\tau u}}{M(\tau)^n}\, dF_n(u).$$

Let $b > a$. We have that

$$\begin{aligned} \mathbb{P}(S_n > na) &= \int_{na}^{\infty} dF_n(u) \\ &= \int_{na}^{\infty} M(\tau)^n e^{-\tau u}\, d\widetilde{F}_n(u) && \text{by (14)} \\ &\geq M(\tau)^n e^{-\tau nb} \int_{na}^{nb} d\widetilde{F}_n(u) \\ &\geq e^{-n(\tau b - \Lambda(\tau))}\mathbb{P}(na < \widetilde{S}_n < nb). \end{aligned}$$

Since the $\widetilde{X}_i$ have mean a and non-zero variance, we have by the central limit theorem applied to the $\widetilde{X}_i$ that $\mathbb{P}(\widetilde{S}_n > na) \to \frac{1}{2}$ as $n \to \infty$, and by the law of large numbers that $\mathbb{P}(\widetilde{S}_n < nb) \to 1$. Therefore,

$$\begin{aligned} \frac{1}{n}\log \mathbb{P}(S_n > na) &\geq -(\tau b - \Lambda(\tau)) + \frac{1}{n}\log \mathbb{P}(na < \widetilde{S}_n < nb) \\ &\to -(\tau b - \Lambda(\tau)) && \text{as } n \to \infty \\ &\to -(\tau a - \Lambda(\tau)) = -\Lambda^*(a) && \text{as } b \downarrow a, \text{ by (9)}. \end{aligned}$$

This completes the proof in the regular case.

Finally, we consider the non-regular case. Let c be a real number satisfing $c > a$, and write $Z^c = \min\{Z, c\}$, the truncation of the random variable Z at level c. Since $\mathbb{P}(X^c \le c) = 1$, we have that $M^c(t) = \mathbb{E}(e^{tX^c}) \le e^{tc}$ for $t > 0$, and therefore $M(t) < \infty$ for all $t > 0$. Note that $\mathbb{E}(X^c) \le \mathbb{E}(X) = 0$, and $\mathbb{E}(X^c) \to 0$ as $c \to \infty$, by the monotone convergence theorem.

Since $\mathbb{P}(X > a) > 0$, there exists $b \in (a, c)$ such that $\mathbb{P}(X > b) > 0$. It follows that $\Lambda^c(t) = \log M^c(t)$ satisfies

$$at - \Lambda^c(t) \le at - \log\{e^{tb}\mathbb{P}(X > b)\} \to -\infty \quad \text{as } t \to \infty.$$

We deduce that the supremum of $at - \Lambda^c(t)$ over values $t > 0$ is attained at some point $\tau = \tau^c \in (0, \infty)$. The random sequence $X_1^c, X_2^c, \dots$ is therefore a regular case of the large deviation problem, and $a > \mathbb{E}(X^c)$, whence

$$\text{(15)} \qquad \frac{1}{n}\log\mathbb{P}\left(\sum_{i=1}^{n} X_i^c > na\right) \to -\Lambda^{c*}(a) \quad \text{as } n \to \infty,$$

by the previous part of this proof, where

$$\text{(16)} \qquad \Lambda^{c*}(a) = \sup_{t>0}\{at - \Lambda^c(t)\} = a\tau - \Lambda^c(\tau).$$

Now $\Lambda^c(t) = \mathbb{E}(e^{tX^c})$ is non-decreasing in c when $t > 0$, implying that Λ^{c*} is non-increasing. Therefore there exists a real number $\Lambda^{\infty*}$ such that

$$\text{(17)} \qquad \Lambda^{c*}(a) \downarrow \Lambda^{\infty*} \quad \text{as } c \uparrow \infty.$$

Since $\Lambda^{c*}(a) < \infty$ and $\Lambda^{c*}(a) \ge -\Lambda^c(0) = 0$, we have that $0 \le \Lambda^{\infty*} < \infty$.

Evidently $S_n \ge \sum_{i=1}^n X_i^c$, whence

$$\frac{1}{n}\log\mathbb{P}(S_n > na) \ge \frac{1}{n}\log\mathbb{P}\left(\sum_{i=1}^{n} X_i^c > na\right),$$

and it therefore suffices by (15)–(17) to prove that

$$\text{(18)} \qquad \Lambda^{\infty*} \le \Lambda^*(a).$$

Since $\Lambda^{\infty*} \le \Lambda^{c*}(a)$, the set $I_c = \{t \ge 0 : at - \Lambda^c(t) \ge \Lambda^{\infty*}\}$ is non-empty. Using the smoothness of Λ^c, and aided by a glance at Figure 5.4, we see that I_c is a non-empty closed interval. Since $\Lambda^c(t)$ is non-decreasing in c, the sets I_c are non-increasing. Since the intersection of nested compact sets is non-empty, the intersection $\bigcap_{c>a} I_c$ contains at least one real number ζ. By the monotone convergence theorem, $\Lambda^c(\zeta) \to \Lambda(\zeta)$ as $c \to \infty$, whence

$$a\zeta - \Lambda(\zeta) = \lim_{c\to\infty}\{a\zeta - \Lambda^c(\zeta)\} \ge \Lambda^{\infty*}$$

so that

$$\Lambda^*(a) = \sup_{t>0}\{at - \Lambda(t)\} \ge \Lambda^{\infty*}$$

as required in (18). ∎

Exercises for Section 5.11

1. A fair coin is tossed n times, showing heads H_n times and tails T_n times. Let $S_n = H_n - T_n$. Show that

$$\mathbb{P}(S_n > an)^{1/n} \to \frac{1}{\sqrt{(1+a)^{1+a}(1-a)^{1-a}}} \qquad \text{if } 0 < a < 1.$$

What happens if $a \geq 1$?

2. Show that

$$T_n^{1/n} \to \frac{4}{\sqrt{(1+a)^{1+a}(1-a)^{1-a}}}$$

as $n \to \infty$, where $0 < a < 1$ and

$$T_n = \sum_{\substack{k: \\ |k-\frac{1}{2}n| > \frac{1}{2}an}} \binom{n}{k}.$$

Find the asymptotic behaviour of $T_n^{1/n}$ where

$$T_n = \sum_{\substack{k: \\ k>n(1+a)}} \frac{n^k}{k!}, \qquad \text{where } a > 0.$$

3. Show that the moment generating function of X is finite in a neighbourhood of the origin if and only if X has exponentially decaying tails, in the sense that there exist positive constants λ and μ such that $\mathbb{P}(|X| \geq a) \leq \mu e^{-\lambda a}$ for $a > 0$. [Seen in the light of this observation, the condition of the large deviation theorem (5.11.4) is very natural].

4. Let $X_1, X_2, \ldots$ be independent random variables having the Cauchy distribution, and let $S_n = X_1 + X_2 + \cdots + X_n$. Find $\mathbb{P}(S_n > an)$.

5.12 Problems

1. A die is thrown ten times. What is the probability that the sum of the scores is 27?

2. A coin is tossed repeatedly, heads appearing with probability p on each toss.

(a) Let X be the number of tosses until the first occasion by which three heads have appeared successively. Write down a difference equation for $f(k) = \mathbb{P}(X = k)$ and solve it. Now write down an equation for $\mathbb{E}(X)$ using conditional expectation. (Try the same thing for the first occurrence of HTH).

(b) Let N be the number of heads in n tosses of the coin. Write down $G_N(s)$. Hence find the probability that: (i) N is divisible by 2, (ii) N is divisible by 3.

3. A coin is tossed repeatedly, heads occurring on each toss with probability p. Find the probability generating function of the number T of tosses before a run of n heads has appeared for the first time.

4. Find the generating function of the negative binomial mass function

$$f(k) = \binom{k-1}{r-1} p^r (1-p)^{k-r}, \qquad k = r, r+1, \ldots,$$

where $0 < p < 1$ and r is a positive integer. Deduce the mean and variance.

5. For the simple random walk, show that the probability $p_0(2n)$ that the particle returns to the origin at the $(2n)$th step satisfies $p_0(2n) \sim (4pq)^n / \sqrt{\pi n}$, and use this to prove that the walk is persistent if and only if $p = \frac{1}{2}$. You will need Stirling's formula: $n! \sim n^{n+\frac{1}{2}} e^{-n} \sqrt{2\pi}$.

6. A symmetric random walk in two dimensions is defined to be a sequence of points $\{(X_n, Y_n) : n \geq 0\}$ which evolves in the following way: if $(X_n, Y_n) = (x, y)$ then (X_{n+1}, Y_{n+1}) is one of the four points $(x \pm 1, y)$, $(x, y \pm 1)$, each being picked with equal probability $\frac{1}{4}$. If $(X_0, Y_0) = (0, 0)$:

(a) show that $\mathbb{E}(X_n^2 + Y_n^2) = n$,

(b) find the probability $p_0(2n)$ that the particle is at the origin after the $(2n)$th step, and deduce that the probability of ever returning to the origin is 1.

7. Consider the one-dimensional random walk $\{S_n\}$ given by

$$S_{n+1} = \begin{cases} S_n + 2 & \text{with probability } p, \\ S_n - 1 & \text{with probability } q = 1 - p, \end{cases}$$

where $0 < p < 1$. What is the probability of ever reaching the origin starting from $S_0 = a$ where $a > 0$?

8. Let X and Y be independent variables taking values in the positive integers such that

$$\mathbb{P}(X = k \mid X + Y = n) = \binom{n}{k} p^k (1 - p)^{n-k}$$

for some p and all $0 \leq k \leq n$. Show that X and Y have Poisson distributions.

9. In a branching process whose family sizes have mean μ and variance σ^2, find the variance of Z_n, the size of the nth generation, given that $Z_0 = 1$.

10. Waldegrave's problem. A group $\{A_1, A_2, \ldots, A_r\}$ of r (> 2) people play the following game. A_1 and A_2 wager on the toss of a fair coin. The loser puts £1 in the pool, the winner goes on to play A_3. In the next wager, the loser puts £1 in the pool, the winner goes on to play A_4, and so on. The winner of the $(r - 1)$th wager goes on to play A_1, and the cycle recommences. The first person to beat all the others in sequence takes the pool.

(a) Find the probability generating function of the duration of the game.

(b) Find an expression for the probability that A_k wins.

(c) Find an expression for the expected size of the pool at the end of the game, given that A_k wins.

(d) Find an expression for the probability that the pool is intact after the nth spin of the coin.

This problem was discussed by Montmort, Bernoulli, de Moivre, Laplace, and others.

11. Show that the generating function H_n of the *total* number of individuals in the first n generations of a branching process satisfies $H_n(s) = sG(H_{n-1}(s))$.

12. Show that the number Z_n of individuals in the nth generation of a branching process satisfies $\mathbb{P}(Z_n > N \mid Z_m = 0) \leq G_m(0)^N$ for $n < m$.

13. (a) A hen lays N eggs where N is Poisson with parameter λ. The weight of the nth egg is W_n, where $W_1, W_2, \ldots$ are independent identically distributed variables with common probability generating function $G(s)$. Show that the generating function G_W of the total weight $W = \sum_{i=1}^{N} W_i$ is given by $G_W(s) = \exp\{-\lambda + \lambda G(s)\}$. W is said to have a *compound Poisson distribution*. Show further that, for any positive integral value of n, $G_W(s)^{1/n}$ is the probability generating function of some random variable; W (or its distribution) is said to be *infinitely divisible* in this regard.

(b) Show that if $H(s)$ is the probability generating function of some infinitely divisible distribution on the non-negative integers then $H(s) = \exp\{-\lambda + \lambda G(s)\}$ for some λ (> 0) and some probability generating function $G(s)$.

14. The distribution of a random variable X is called *infinitely divisible* if, for all positive integers n, there exists a sequence $Y_1^{(n)}, Y_2^{(n)}, \dots, Y_n^{(n)}$ of independent identically distributed random variables such that X and $Y_1^{(n)} + Y_2^{(n)} + \dots + Y_n^{(n)}$ have the same distribution.
(a) Show that the normal, Poisson, and gamma distributions are infinitely divisible.
(b) Show that the characteristic function ϕ of an infinitely divisible distribution has no real zeros, in that $\phi(t) \neq 0$ for all real t.

15. Let $X_1, X_2, \dots$ be independent variables each taking the values 0 or 1 with probabilities $1-p$ and p, where $0 < p < 1$. Let N be a random variable taking values in the positive integers, independent of the X_i, and write $S = X_1 + X_2 + \dots + X_N$. Write down the conditional generating function of N given that $S = N$, in terms of the probability generating function G of N. Show that N has a Poisson distribution if and only if $\mathbb{E}(x^N)^p = \mathbb{E}(x^N \mid S = N)$ for all p and x.

16. If X and Y have joint probability generating function

$$G_{X,Y}(s,t) = \mathbb{E}(s^X t^Y) = \frac{\{1-(p_1+p_2)\}^n}{\{1-(p_1 s + p_2 t)\}^n} \qquad \text{where } p_1 + p_2 \le 1,$$

find the marginal mass functions of X and Y, and the mass function of $X+Y$. Find also the conditional probability generating function $G_{X|Y}(s \mid y) = \mathbb{E}(s^X \mid Y = y)$ of X given that $Y = y$. The pair X, Y is said to have the *bivariate negative binomial distribution*.

17. If X and Y have joint probability generating function

$$G_{X,Y}(s,t) = \exp\{\alpha(s-1) + \beta(t-1) + \gamma(st-1)\}$$

find the marginal distributions of X, Y, and the distribution of $X+Y$, showing that X and Y have the Poisson distribution, but that $X+Y$ does not unless $\gamma = 0$.

18. Define

$$I(a,b) = \int_0^\infty \exp(-a^2u^2 - b^2u^{-2})\,du$$

for $a, b > 0$. Show that
(a) $I(a,b) = a^{-1}I(1,ab)$, (b) $\partial I/\partial b = -2I(1,ab)$,
(c) $I(a,b) = \sqrt{\pi}e^{-2ab}/(2a)$.
(d) If X has density function $(d/\sqrt{x})e^{-c/x-gx}$ for $x > 0$, then

$$\mathbb{E}(e^{-tX}) = d\sqrt{\frac{\pi}{g+t}}\exp\big(-2\sqrt{c(g+t)}\big), \qquad t > -g.$$

(e) If X has density function $(2\pi x^3)^{-\frac{1}{2}}e^{-1/(2x)}$ for $x > 0$, then X has moment generating function given by $\mathbb{E}(e^{-tX}) = \exp\{-\sqrt{2t}\}$, $t \ge 0$. [Note that $\mathbb{E}(X^n) = \infty$ for $n \ge 1$.]

19. Let X, Y, Z be independent $N(0,1)$ variables. Use characteristic functions and moment generating functions (Laplace transforms) to find the distributions of
(a) $U = X/Y$,
(b) $V = X^{-2}$,
(c) $W = XYZ/\sqrt{X^2Y^2 + Y^2Z^2 + Z^2X^2}$.

20. Let X have density function f and characteristic function ϕ, and suppose that $\int_{-\infty}^{\infty} |\phi(t)|\,dt < \infty$. Deduce that

$$f(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-itx}\phi(t)\,dt.$$

21. Conditioned branching process. Consider a branching process whose family sizes have the geometric mass function $f(k) = qp^k$, $k \ge 0$, where $\mu = p/q > 1$. Let Z_n be the size of the nth generation, and assume $Z_0 = 1$. Show that the conditional distribution of Z_n/μ^n, given that $Z_n > 0$, converges as $n \to \infty$ to the exponential distribution with parameter $1 - \mu^{-1}$.

22. A random variable X is called *symmetric* if X and $-X$ are identically distributed. Show that X is symmetric if and only if the imaginary part of its characteristic function is identically zero.

23. Let X and Y be independent identically distributed variables with means 0 and variances 1. Let $\phi(t)$ be their common characteristic function, and suppose that $X + Y$ and $X - Y$ are independent. Show that $\phi(2t) = \phi(t)^3\phi(-t)$, and deduce that X and Y are $N(0, 1)$ variables.

More generally, suppose that X and Y are independent and identically distributed with means 0 and variances 1, and furthermore that $\mathbb{E}(X - Y \mid X + Y) = 0$ and $\mathrm{var}(X - Y \mid X + Y) = 2$. Deduce that $\phi(s)^2 = \phi'(s)^2 - \phi(s)\phi''(s)$, and hence that X and Y are independent $N(0, 1)$ variables.

24. Show that the average $Z = n^{-1}\sum_{i=1}^{n} X_i$ of n independent Cauchy variables has the Cauchy distribution too. Why does this not violate the law of large numbers?

25. Let X and Y be independent random variables each having the Cauchy density function $f(x) = \{\pi(1 + x^2)\}^{-1}$, and let $Z = \frac{1}{2}(X + Y)$.

(a) Show by using characteristic functions that Z has the Cauchy distribution also.

(b) Show by the convolution formula that Z has the Cauchy density function. You may find it helpful to check first that

$$f(x)f(y - x) = \frac{f(x) + f(y - x)}{\pi(4 + y^2)} + g(y)\{xf(x) + (y - x)f(y - x)\}$$

where $g(y) = 2/\{\pi y(4 + y^2)\}$.

26. Let $X_1, X_2, \dots, X_n$ be independent variables with characteristic functions $\phi_1, \phi_2, \dots, \phi_n$. Describe random variables which have the following characteristic functions:

(a) $\phi_1(t)\phi_2(t)\cdots\phi_n(t)$, (b) $|\phi_1(t)|^2$,
(c) $\sum_1^n p_j\phi_j(t)$ where $p_j \ge 0$ and $\sum_1^n p_j = 1$, (d) $(2 - \phi_1(t))^{-1}$,
(e) $\int_0^\infty \phi_1(ut)e^{-u}\, du$.

27. Find the characteristic functions corresponding to the following density functions on $(-\infty, \infty)$:

(a) $1/\cosh(\pi x)$, (b) $(1 - \cos x)/(\pi x^2)$,
(c) $\exp(-x - e^{-x})$, (d) $\frac{1}{2}e^{-|x|}$.

Show that the mean of the 'extreme-value distribution' in part (c) is Euler's constant γ.

28. Which of the following are characteristic functions:

(a) $\phi(t) = 1 - |t|$ if $|t| \le 1$, $\phi(t) = 0$ otherwise,
(b) $\phi(t) = (1 + t^4)^{-1}$, (c) $\phi(t) = \exp(-t^4)$,
(d) $\phi(t) = \cos t$, (e) $\phi(t) = 2(1 - \cos t)/t^2$.

29. Show that the characteristic function ϕ of a random variable X satisfies $|1 - \phi(t)| \le \mathbb{E}|tX|$.

30. Suppose X and Y have joint characteristic function $\phi(s, t)$. Show that, subject to the appropriate conditions of differentiability,

$$i^{m+n}\mathbb{E}(X^m Y^n) = \left.\frac{\partial^{m+n}\phi}{\partial s^m \partial t^n}\right|_{s=t=0}$$

for any positive integers m and n.

31. If X has distribution function F and characteristic function ϕ, show that for $t > 0$

(a)
$$\int_{[-t^{-1}, t^{-1}]} x^2\, dF \le \frac{3}{t^2}[1 - \mathrm{Re}\,\phi(t)],$$

(b)
$$\mathbb{P}\left(|X| \ge \frac{1}{t}\right) \le \frac{7}{t}\int_0^t [1 - \mathrm{Re}\,\phi(v)]\, dv.$$

32. Let $X_1, X_2, \dots$ be independent variables which are uniformly distributed on $[0, 1]$. Let $M_n = \max\{X_1, X_2, \dots, X_n\}$ and show that $n(1 - M_n) \xrightarrow{\text{D}} X$ where X is exponentially distributed with parameter 1. You need not use characteristic functions.

33. If X is either (a) Poisson with parameter λ, or (b) $\Gamma(1, \lambda)$, show that the distribution of $Y_\lambda = (X - \mathbb{E}X)/\sqrt{\operatorname{var} X}$ approaches the $N(0, 1)$ distribution as $\lambda \to \infty$.
(c) Show that

$$e^{-n}\left(1 + n + \frac{n^2}{2!} + \cdots + \frac{n^n}{n!}\right) \to \frac{1}{2} \qquad \text{as } n \to \infty.$$

34. Coupon collecting. Recall that you regularly buy quantities of some ineffably dull commodity. To attract your attention, the manufacturers add to each packet a small object which is also dull, and in addition useless, but there are n different types. Assume that each packet is equally likely to contain any one of the different types, as usual. Let T_n be the number of packets bought before you acquire a complete set of n objects. Show that $n^{-1}(T_n - n \log n) \xrightarrow{\text{D}} T$, where T is a random variable with distribution function $\mathbb{P}(T \le x) = \exp(-e^{-x})$, $-\infty < x < \infty$.

35. Find a sequence (ϕ_n) of characteristic functions with the property that the limit given by $\phi(t) = \lim_{n\to\infty} \phi_n(t)$ exists for all t, but such that ϕ is not itself a characteristic function.

36. Use generating functions to show that it is not possible to load two dice in such a way that the sum of the values which they show is equally likely to take any value between 2 and 12. Compare with your method for Problem (2.7.12).

37. A biased coin is tossed N times, where N is a random variable which is Poisson distributed with parameter λ. Prove that the total number of heads shown is independent of the total number of tails. Show conversely that if the numbers of heads and tails are independent, then N has the Poisson distribution.

38. A *binary tree* is a tree (as in the section on branching processes) in which each node has exactly two descendants. Suppose that each node of the tree is coloured black with probability p, and white otherwise, independently of all other nodes. For any path π containing n nodes beginning at the root of the tree, let $B(\pi)$ be the number of black nodes in π, and let $X_n(k)$ be the number of such paths π for which $B(\pi) \ge k$. Show that there exists β_c such that

$$\mathbb{E}\{X_n(\beta n)\} \to \begin{cases} 0 & \text{if } \beta > \beta_c, \\ \infty & \text{if } \beta < \beta_c, \end{cases}$$

and show how to determine the value β_c.

Prove that

$$\mathbb{P}\big(X_n(\beta n) \ge 1\big) \to \begin{cases} 0 & \text{if } \beta > \beta_c, \\ 1 & \text{if } \beta < \beta_c. \end{cases}$$

39. Use the continuity theorem (5.9.5) to show that, as $n \to \infty$,
(a) if X_n is bin$(n, \lambda/n)$ then the distribution of X_n converges to a Poisson distribution,
(b) if Y_n is geometric with parameter $p = \lambda/n$ then the distribution of Y_n/n converges to an exponential distribution.

40. Let $X_1, X_2, \dots$ be independent random variables with zero means and such that $\mathbb{E}|X_j^3| < \infty$ for all j. Show that $S_n = X_1 + X_2 + \cdots + X_n$ satisfies $S_n/\sqrt{\operatorname{var}(S_n)} \xrightarrow{\text{D}} N(0, 1)$ as $n \to \infty$ if

$$\sum_{j=1}^{n} \mathbb{E}|X_j^3| = \mathrm{o}\Big(\{\operatorname{var}(S_n)\}^{-\frac{3}{2}}\Big).$$

The following steps may be useful. Let $\sigma_j^2 = \operatorname{var}(X_j)$, $\sigma(n)^2 = \operatorname{var}(S_n)$, $\rho_j = \mathbb{E}|X_j^3|$, and ϕ_j and ψ_n be the characteristic functions of X_j and $S_n/\sigma(n)$ respectively.

(i) Use Taylor's theorem to show that $|\phi_j(t) - 1| \le 2t^2\sigma_j^2$ and $|\phi_j(t) - 1 + \frac{1}{2}\sigma_j^2 t^2| \le |t|^3\rho_j$ for $j \ge 1$.

(ii) Show that $|\log(1+z) - z| \le |z|^2$ if $|z| \le \frac{1}{2}$, where the logarithm has its principal value.

(iii) Show that $\sigma_j^3 \le \rho_j$, and deduce from the hypothesis that $\max_{1\le j\le n}\sigma_j/\sigma(n) \to 0$ as $n \to \infty$, implying that $\max_{1\le j\le n}|\phi_j(t/\sigma(n)) - 1| \to 0$.

(iv) Deduce an upper bound for $\left|\log\phi_j(t/\sigma(n)) - \frac{1}{2}t^2\sigma_j^2/\sigma(n)^2\right|$, and sum to obtain that $\log\psi_n(t) \to -\frac{1}{2}t^2$.

41. Let $X_1, X_2, \ldots$ be independent variables each taking values $+1$ or -1 with probabilities $\frac{1}{2}$ and $\frac{1}{2}$. Show that

$$\sqrt{\frac{3}{n^3}}\sum_{k=1}^{n} kX_k \xrightarrow{\mathrm{D}} N(0,1) \qquad \text{as } n \to \infty.$$

42. Normal sample. Let $X_1, X_2, \ldots, X_n$ be independent $N(\mu, \sigma^2)$ random variables. Define $\overline{X} = n^{-1}\sum_1^n X_i$ and $Z_i = X_i - \overline{X}$. Find the joint characteristic function of $\overline{X}, Z_1, Z_2, \ldots, Z_n$, and hence prove that $\overline{X}$ and $S^2 = (n-1)^{-1}\sum_1^n(X_i - \overline{X})^2$ are independent.

43. Log-normal distribution. Let X be $N(0,1)$, and let $Y = e^X$; Y is said to have the *log-normal* distribution. Show that the density function of Y is

$$f(x) = \frac{1}{x\sqrt{2\pi}}\exp\left\{-\tfrac{1}{2}(\log x)^2\right\}, \qquad x > 0.$$

For $|a| \le 1$, define $f_a(x) = \{1 + a\sin(2\pi\log x)\}f(x)$. Show that f_a is a density function with finite moments of all (positive) orders, none of which depends on the value of a. The family $\{f_a : |a| \le 1\}$ contains density functions which are not specified by their moments.

44. Consider a random walk whose steps are independent and identically distributed integer-valued random variables with non-zero mean. Prove that the walk is transient.

45. Recurrent events. Let $\{X_r : r \ge 1\}$ be the integer-valued identically distributed intervals between the times of a recurrent event process. Let L be the earliest time by which there has been an interval of length a containing no occurrence time. Show that, for integral a,

$$\mathbb{E}(s^L) = \frac{s^a\mathbb{P}(X_1 > a)}{1 - \sum_{r=1}^{a} s^r\mathbb{P}(X_1 = r)}.$$

46. A biased coin shows heads with probability $p\ (= 1-q)$. It is flipped repeatedly until the first time W_n by which it has shown n consecutive heads. Let $\mathbb{E}(s^{W_n}) = G_n(s)$. Show that $G_n = psG_{n-1}/(1 - qsG_{n-1})$, and deduce that

$$G_n(s) = \frac{(1-ps)p^n s^n}{1 - s + qp^n s^{n+1}}.$$

47. In n flips of a biased coin which shows heads with probability $p\ (= 1-q)$, let L_n be the length of the longest run of heads. Show that, for $r \ge 1$,

$$1 + \sum_{n=1}^{\infty} s^n\mathbb{P}(L_n < r) = \frac{1 - p^r s^r}{1 - s + qp^r s^{r+1}}.$$

48. The random process $\{X_n : n \ge 1\}$ decays geometrically fast in that, in the absence of external input, $X_{n+1} = \frac{1}{2}X_n$. However, at any time n the process is also increased by Y_n with probability

$\frac{1}{2}$, where $\{Y_n : n \geq 1\}$ is a sequence of independent exponential random variables with parameter λ. Find the limiting distribution of X_n as $n \to \infty$.

49. Let $G(s) = \mathbb{E}(s^X)$ where $X \geq 0$. Show that $\mathbb{E}\{(X+1)^{-1}\} = \int_0^1 G(s)\,ds$, and evaluate this when X is (a) Poisson with parameter λ, (b) geometric with parameter p, (c) binomial bin(n, p), (d) logarithmic with parameter p (see Exercise (5.2.3)). Is there a non-trivial choice for the distribution of X such that $\mathbb{E}\{(X+1)^{-1}\} = \{\mathbb{E}(X+1)\}^{-1}$?

50. Find the density function of $\sum_{r=1}^{N} X_r$, where $\{X_r : r \geq 1\}$ are independent and exponentially distributed with parameter λ, and N is geometric with parameter p and independent of the X_r.

51. Let X have finite non-zero variance and characteristic function $\phi(t)$. Show that

$$\psi(t) = -\frac{1}{\mathbb{E}(X^2)} \frac{d^2\phi}{dt^2}$$

is a characteristic function, and find the corresponding distribution.

52. Let X and Y have joint density function

$$f(x, y) = \tfrac{1}{4}\{1 + xy(x^2 - y^2)\}, \qquad |x| < 1,\ |y| < 1.$$

Show that $\phi_X(t)\phi_Y(t) = \phi_{X+Y}(t)$, and that X and Y are dependent.