3

Transforms

1 Introduction

In Chapter 1 we learned how to handle transformations in order to find the distribution of new (constructed) random variables. Since the arithmetic mean or average of a set of (independent) random variables is a very important object in probability theory as well as in statistics, we focus in this chapter on sums of independent random variables (from which one easily finds corresponding results for the average). We know from earlier work that the convolution formula may be used but also that the sums or integrals involved may be difficult or even impossible to compute. In particular, this is the case if the number of summands is “large.” In that case, however, the central limit theorem is (frequently) applicable. This result will be proved in the chapter on convergence; see Theorem 6.5.2.

Exercise 1.1. Let $X_1, X_2, \ldots$ be independent $U(0,1)$-distributed random variables.

(a) Find the distribution of $X_1 + X_2$.
(b) Find the distribution of $X_1 + X_2 + X_3$.
(c) Show that the distribution of $S_n = X_1 + X_2 + \cdots + X_n$ is given by

$$F_{S_n}(x) = \frac{1}{n!}\sum_{k=0}^{n-1}(-1)^k\binom{n}{k}(x-k)_+^n, \quad 0 \le x \le n,$$

where $x_+ = \max\{x, 0\}$. □

Even if, in theory, we have solved this problem, we face new problems if we actually wish to compute $P(S_n \le x)$ for some given x already for moderately sized values of n; for example, what is $P(S_5 \le \pi)$?

In this chapter we shall learn how such problems can be transformed into new problems, how the new (simpler) problems are solved, and finally that these solutions can be retransformed or inverted to provide a solution to the original problems.

A. Gut, *An Intermediate course in Probabilty,* Springer Texts in Statistics,
DOI: 10.1007/978-1-4419-0162-0_3,

Remark 1.1. In order to determine the distribution of sums of independent random variables we mentioned the convolution formula. From analysis we recall that the problem of convolving functions can be transformed to the problem of multiplying their Laplace transforms or Fourier transforms (which is a simpler task). □

We begin, however, with an example from a different area.

Example 1.1. Let $a_1, a_2, \ldots, a_n$ be positive reals. We want to know their product.

This is a "difficult" problem. We therefore find the logarithms of the numbers, add them to yield $\sum_{k=1}^{n} \log a_k$, and then invert. □

Figure 1.1 illustrates the procedure.

$$\begin{tikzcd}
\{a_k\} \arrow[r] \arrow[d, dashed] & \{\log a_k\} \arrow[d] \\
\Pi\, a_k & \Sigma \log a_k \arrow[l]
\end{tikzcd}$$

Figure 1.1

We obtained the correct answer since $\exp\{\sum_{k=1}^{n} \log a_k\} = \prod_{k=1}^{n} a_k$.

The central ideas of the solution thus are

(a) addition is easier to perform than multiplication;
(b) the logarithm has a unique inverse (i.e., if $\log x = \log y$, then $x = y$), namely, the exponential function.

As for sums of independent random variables, the topic of this chapter, we shall introduce three transforms: the (probability) generating function, the moment generating function, and the characteristic function. Two common features of these transforms are that

(a) summation of independent random variables (convolution) corresponds to multiplication of the transforms;
(b) the transformation is 1-to-1, namely, there is a uniqueness theorem to the effect that if two random variables have the same transform then they also have the same distribution.

Notation: The notation

$$X \stackrel{d}{=} Y$$

means that the random variables X and Y are equidistributed. □

Remark 1.2. It is worth pointing out that two random variables, X and Y, may well have the property $X \overset{d}{=} Y$ and yet $X(\omega) \neq Y(\omega)$ for *all* ω. A very simple example is the following: Toss a fair coin once and set

$$X = \begin{cases} 1, & \text{if the outcome is heads,} \\ 0, & \text{if the outcome is tails,} \end{cases}$$

and

$$Y = \begin{cases} 1, & \text{if the outcome is tails,} \\ 0, & \text{if the outcome is heads.} \end{cases}$$

Clearly, $X \in \mathrm{Be}(1/2)$ and $Y \in \mathrm{Be}(1/2)$, in particular, $X \overset{d}{=} Y$. But $X(\omega)$ and $Y(\omega)$ differ for every ω. □

2 The Probability Generating Function

Definition 2.1. *Let X be a nonnegative, integer-valued random variable. The* (probability) generating function *of X is*

$$g_X(t) = E\,t^X = \sum_{n=0}^{\infty} t^n \cdot P(X = n).$$ □

Remark 2.1. The generating function is defined at least for $|t| \leq 1$, since it is a power series with coefficients in $[0, 1]$. Note also that $g_X(1) = \sum_{n=0}^{\infty} P(X = n) = 1$. □

Theorem 2.1. *Let X and Y be nonnegative, integer-valued random variables. If $g_X = g_Y$, then $p_X = p_Y$.* □

The theorem states that if two nonnegative, integer-valued random variables have the same generating function then they follow the same probability law. It is thus the uniqueness theorem mentioned in the previous section. The result is a special case of the uniqueness theorem for power series. We refer to the literature cited in Appendix A for a complete proof.

Theorem 2.2. *Let $X_1, X_2, \ldots, X_n$ be independent, nonnegative, integer-valued random variables, and set $S_n = X_1 + X_2 + \cdots + X_n$. Then*

$$g_{S_n}(t) = \prod_{k=1}^{n} g_{X_k}(t).$$

Proof. Since $X_1, X_2, \ldots, X_n$ are independent, the same is true for t^{X_1}, t^{X_2}, $\ldots$, t^{X_n}, which yields

$$g_{S_n}(t) = E\,t^{X_1+X_2+\cdots+X_n} = E\prod_{k=1}^{n} t^{X_k} = \prod_{k=1}^{n} E\,t^{X_k} = \prod_{k=1}^{n} g_{X_k}(t).$$ □

This result asserts that adding independent, nonnegative, integer-valued random variables corresponds to multiplying their generating functions (recall Example 1.1(a)).

A case of particular importance is given next.

Corollary 2.2.1. *If, in addition,* $X_1, X_2, \ldots, X_n$ *are equidistributed, then*

$$g_{S_n}(t) = \big(g_X(t)\big)^n. \qquad \square$$

Termwise differentiation of the generating function (this is permitted (at least) for $|t| < 1$) yields

$$g'_X(t) = \sum_{n=1}^{\infty} n t^{n-1} P(X = n), \tag{2.1}$$

$$g''_X(t) = \sum_{n=2}^{\infty} n(n-1) t^{n-2} P(X = n), \tag{2.2}$$

and, in general, for $k = 1, 2, \ldots,$

$$g_X^{(k)}(t) = \sum_{n=k}^{\infty} n(n-1)\cdots(n-k+1) t^{n-k} P(X = n)\,. \tag{2.3}$$

By putting $t = 0$ in (2.1)–(2.3), we obtain $g_X^{(n)}(0) = n! \cdot P(X = n)$, that is,

$$P(X = n) = \frac{g_X^{(n)}(0)}{n!}. \tag{2.4}$$

The probability generating function thus generates the probabilities; hence the name of the transform.

By letting $t \nearrow 1$ in (2.1)–(2.3) (this requires a little more care), the following result is obtained.

Theorem 2.3. *Let* X *be a nonnegative, integer-valued random variable, and suppose that* $E\,|X|^k < \infty$ *for some* $k = 1, 2, \ldots.$ *Then*

$$E\,X(X-1)\cdots(X-k+1) = g_X^{(k)}(1). \qquad \square$$

Remark 2.2. Derivatives at $t = 1$ are throughout to be interpreted as limits as $t \nearrow 1$. For simplicity, however, we use the simpler notation $g'(1)$, $g''(1)$, and so on. $\square$

The following example illustrates the relevance of this remark.

Example 2.1. Suppose that X has the probability function

$$p(k) = \frac{C}{k^2}, \quad k = 1, 2, 3, \ldots,$$

(where, to be precise, $C^{-1} = \sum_{k=1}^{\infty} 1/k^2 = \pi^2/6$). The divergence of the harmonic series tells us that the distribution does not have a finite mean.

Now, the generating function is

$$g(t) = \frac{6}{\pi^2} \sum_{k=1}^{\infty} \frac{t^k}{k^2}, \quad \text{for} \quad |t| \le 1,$$

so that

$$g'(t) = \frac{6}{\pi^2} \sum_{k=1}^{\infty} \frac{t^{k-1}}{k} = -\frac{6}{\pi^2} \cdot \frac{\log(1-t)}{t} \nearrow +\infty \quad \text{as} \quad t \nearrow 1.$$

This shows that although the generating function itself exists for $t = 1$, the derivative only exists for all t *strictly smaller* than 1, but *not* for the boundary value $t = 1$. □

For $k = 1$ and $k = 2$ we have, in particular, the following result:

Corollary 2.3.1 *Let X be as before. Then*

(a) $E\,|X| < \infty \quad \Longrightarrow \quad E\,X = g'_X(1)$, *and*
(b) $E\,X^2 < \infty \quad \Longrightarrow \quad \operatorname{Var} X = g''_X(1) + g'_X(1) - \big(g'_X(1)\big)^2.$ □

Exercise 2.1. Prove Corollary 2.3.1. □

Next we consider some special distributions:

The Bernoulli distribution. Let $X \in \mathrm{Be}(p)$. Then

$$g_X(t) = q \cdot t^0 + p \cdot t^1 = q + pt, \quad \text{for all} \quad t,$$
$$g'_X(t) = p, \quad \text{and} \quad g''_X(t) = 0,$$

which yields

$$E\,X = g'_X(1) = p$$

and

$$\operatorname{Var} X = g''_X(1) + g'_X(1) - (g'_X(1))^2 = 0 + p - p^2 = p(1-p) = pq.$$

The binomial distribution. Let $X \in \mathrm{Bin}(n, p)$. Then

$$g_X(t) = \sum_{k=0}^{n} t^k \binom{n}{k} p^k q^{n-k} = \sum_{k=0}^{n} \binom{n}{k} (pt)^k q^{n-k} = (q + pt)^n,$$

for all t. Furthermore,

$$g_X'(t) = n(q+pt)^{n-1} \cdot p \quad \text{and} \quad g_X''(t) = n(n-1)(q+pt)^{n-2} \cdot p^2,$$

which yields

$$E\,X = np \quad \text{and} \quad \operatorname{Var} X = n(n-1)p^2 + np - (np)^2 = npq.$$

We further observe that

$$g_{\text{Bin}(n,p)}(t) = \big(g_{\text{Be}(p)}(t)\big)^n,$$

which, according to Corollary 2.2.1, shows that if $Y_1, Y_2, \ldots, Y_n$ are independent, Be(p)-distributed random variables, and $X_n = Y_1 + Y_2 + \cdots + Y_n$, then

$$g_{X_n}(t) = g_{\text{Bin}(n,p)}(t).$$

By Theorem 2.1 (uniqueness) it follows that $X_n \in \text{Bin}(n,p)$, a conclusion that, alternatively, could be proved by the convolution formula and induction.

Similarly, if $X_1 \in \text{Bin}(n_1,p)$ and $X_2 \in \text{Bin}(n_2,p)$ are independent, then, by Theorem 2.2,

$$g_{X_1+X_2}(t) = (q+pt)^{n_1+n_2} = g_{\text{Bin}(n_1+n_2,p)}(t)\,,$$

which proves that $X_1+X_2 \in \text{Bin}(n_1+n_2,p)$ and hence establishes, in a simple manner, the addition theorem for the binomial distribution.

Remark 2.3. It is instructive to reprove the last results by actually using the convolution formula. We stress, however, that the simplicity of the method of generating functions is illusory, since it in fact exploits various results on generating functions and their derivatives. □

The geometric distribution. Let $X \in \text{Ge}(p)$. Then

$$g_X(t) = \sum_{k=0}^{\infty} t^k p q^k = p \sum_{k=0}^{\infty} (tq)^k = \frac{p}{1-qt}, \quad |t| < \frac{1}{q}.$$

Moreover,

$$g_X'(t) = -\frac{p}{(1-qt)^2} \cdot (-q) = \frac{pq}{(1-qt)^2}$$

and

$$g_X''(t) = -\frac{2pq}{(1-qt)^3} \cdot (-q) = \frac{2pq^2}{(1-qt)^3}\,,$$

from which it follows that $E\,X = q/p$ and $\operatorname{Var} X = q/p^2$.

Exercise 2.2. Let $X_1, X_2, \ldots, X_n$ be independent Ge(p)-distributed random variables. Determine the distribution of $X_1 + X_2 + \cdots + X_n$. □

The Poisson distribution. Let $X \in \text{Po}(m)$. Then

$$g_X(t) = \sum_{k=0}^{\infty} t^k e^{-m} \frac{m^k}{k!} = e^{-m} \sum_{k=0}^{\infty} \frac{(mt)^k}{k!} = e^{m(t-1)}.$$

Exercise 2.3. (a) Let $X \in \text{Po}(m)$. Show that $E\,X = \text{Var}\,X = m$.
(b) Let $X_1 \in \text{Po}(m_1)$ and $X_2 \in \text{Po}(m_2)$ be independent random variables. Show that $X_1 + X_2 \in \text{Po}(m_1 + m_2)$. □

3 The Moment Generating Function

In spite of their usefulness, probability generating functions are of limited use in that they are only defined for nonnegative, integer-valued random variables. Important distributions, such as the normal distribution and the exponential distribution, cannot be handled with this transform. This inconvenience is overcome as follows:

Definition 3.1. *The* moment generating function *of a random variable X is*

$$\psi_X(t) = E\,e^{tX},$$

provided there exists $h > 0$, such that the expectation exists and is finite for $|t| < h$. □

Remark 3.1. As a first observation we mention the close connection between moment generating functions and Laplace transforms of real-valued functions. Indeed, for a nonnegative random variable X, one may define the Laplace transform

$$E\,e^{-sX} \quad \text{for} \quad s \geq 0,$$

which thus always exist (why?). Analogously, one may view the moment generating function as a two-sided Laplace transform.

Remark 3.2. Note that for nonnegative, integer-valued random variables we have $\psi(t) = g(e^t)$, for $|t| < h$, provided the moment generating function exists (for $|t| < h$). □

The uniqueness and multiplication theorems are presented next. The proofs are analogous to those for the generating function.

Theorem 3.1. *Let X and Y be random variables. If there exists $h > 0$, such that $\psi_X(t) = \psi_Y(t)$ for $|t| < h$, then $X \overset{d}{=} Y$.* □

Theorem 3.2. *Let $X_1, X_2, \ldots, X_n$ be independent random variables whose moment generating functions exist for $|t| < h$ for some $h > 0$, and set $S_n = X_1 + X_2 + \cdots + X_n$. Then*

$$\psi_{S_n}(t) = \prod_{k=1}^{n} \psi_{X_k}(t)\,, \quad |t| < h. \qquad \square$$

Corollary 3.2.1. *If, in addition, $X_1, X_2, \ldots, X_n$ are equidistributed, then*

$$\psi_{S_n}(t) = \big(\psi_X(t)\big)^n, \quad |t| < h. \qquad \square$$

For the probability generating function we found that the derivatives at zero produced the probabilities (which motivated the name of the transform). The derivatives at 0 of the moment generating function produce the moments (hence the name of the transform).

Theorem 3.3. *Let X be a random variable whose moment generating function $\psi_X(t)$, exists for $|t| < h$ for some $h > 0$. Then*

(a) *all moments exist, that is, $E\,|X|^r < \infty$ for all $r > 0$;*
(b) $E\,X^n = \psi_X^{(n)}(0)$ *for $n = 1, 2, \ldots$.*

Proof. We prove the theorem in the continuous case, leaving the completely analogous proof in the discrete case as an exercise.

By assumption,

$$\int_{-\infty}^{\infty} e^{tx} f_X(x)\,dx < \infty \quad \text{for} \quad |t| < h.$$

Let t, $0 < t < h$, be given. The assumption implies that, for every $x_1 > 0$,

$$\int_{x_1}^{\infty} e^{tx} f_X(x)\,dx < \infty \quad \text{and} \quad \int_{-\infty}^{-x_1} e^{-tx} f_X(x)\,dx < \infty. \tag{3.1}$$

Since $|x|^r / e^{|tx|} \to 0$ as $x \to \infty$ for all $r > 0$, we further have

$$|x|^r \le e^{|tx|} \quad \text{for} \quad |x| > x_2. \tag{3.2}$$

Now, let $x_0 > x_2$. It follows from (3.1) and (3.2) that

$$\begin{aligned}
&\int_{-\infty}^{\infty} |x|^r f_X(x)\,dx \\
&= \int_{-\infty}^{-x_0} |x|^r f_X(x)\,dx + \int_{-x_0}^{x_0} |x|^r f_X(x)\,dx + \int_{x_0}^{\infty} |x|^r f_X(x)\,dx \\
&\le \int_{-\infty}^{-x_0} e^{-tx} f_X(x)\,dx + |x_0|^r \cdot P(|X| \le x_0) + \int_{x_0}^{\infty} e^{tx} f_X(x)\,dx < \infty.
\end{aligned}$$

This proves (a), from which (b) follows by differentiation:

$$\psi_X^{(n)}(t) = \int_{-\infty}^{\infty} x^n e^{tx} f_X(x)\,dx$$

and, hence,

$$\psi_X^{(n)}(0) = \int_{-\infty}^{\infty} x^n f_X(x)\,dx = E\,X^n. \qquad \square$$

Remark 3.3. The idea in part (a) is that the exponential function grows more rapidly than every polynomial. As a consequence, $|x|^r \leq e^{|tx|}$ as soon as $|x| > x_2$ (say). On the other hand, for $|x| < x_2$ we trivially have $|x|^r \leq Ce^{|tx|}$ for some constant C. It follows that for all x

$$|x|^r \leq (C+1)e^{|tx|}\,,$$

and hence that

$$E\,|X|^r \leq (C+1)E\,e^{|tX|} < \infty \quad \text{for} \quad |t| < h.$$

Note that this, in fact, proves Theorem 3.2(a) in the continuous case as well as in the discrete case.

Remark 3.4. Taylor expansion of the exponential function yields

$$e^{tX} = 1 + \sum_{n=1}^{\infty} \frac{t^n X^n}{n!} \quad \text{for} \quad |t| < h.$$

By taking expectation termwise (this is permitted), we obtain

$$\psi_X(t) = E\,e^{tX} = 1 + \sum_{n=1}^{\infty} \frac{t^n}{n!} E\,X^n \quad \text{for} \quad |t| < h.$$

Termwise differentiation (which is also permitted) yields the result of part (b). A special feature with the series expansion is that if the moment generating function is given in that form we may simply read off the moments; $E\,X^n$ is the coefficient of $t^n/n!$, $n = 1, 2, \ldots$, in the series expansion. □

Let us now, as in the previous section, study some known distributions. First, some discrete ones:

The Bernoulli distribution. Let $X \in \text{Be}(p)$. Then $\psi_X(t) = q + pe^t$. Differentiation yields $E\,X = p$ and $\text{Var}\,X = pq$. Taylor expansion of e^t leads to

$$\psi_X(t) = q + p\sum_{n=0}^{\infty} \frac{t^n}{n!} = 1 + \sum_{n=1}^{\infty} \frac{t^n}{n!} \cdot p\,,$$

from which it follows that $E\,X^n = p$, $n = 1, 2, \ldots$. In particular, $E\,X = p$ and $\text{Var}\,X = p - p^2 = pq$.

The binomial distribution. Let $X \in \text{Bin}(n, p)$. Then

$$\psi_X(t) = \sum_{k=0}^{n} e^{tk} \binom{n}{k} p^k q^{n-k} = \sum_{k=0}^{n} \binom{n}{k} (pe^t)^k q^{n-k} = (q + pe^t)^n.$$

Differentiation yields $E\,X = np$ and $\text{Var}\,X = npq$.

Taylor expansion can also be performed in this case, but it is more cumbersome. If, however, we only wish to find $E\,X$ and $\operatorname{Var} X$ it is not too hard:

$$\begin{aligned}\psi_X(t) &= \left(q+pe^t\right)^n = \Bigl(q+p\sum_{k=0}^{\infty}\frac{t^k}{k!}\Bigr)^n = \Bigl(1+pt+p\frac{t^2}{2!}+\cdots\Bigr)^n\\ &= 1+npt+\binom{n}{2}p^2t^2+np\frac{t^2}{2}+\cdots\\ &= 1+npt+\bigl(n(n-1)p^2+np\bigr)\frac{t^2}{2}+\cdots\,.\end{aligned}$$

Here the ellipses mean that the following terms contain t raised to at least the third degree. By identifying the coefficients we find that $E\,X = np$ and that $E\,X^2 = n(n-1)p^2 + np$, which yields $\operatorname{Var} X = npq$.

Remark 3.5. Let us immediately point out that in this particular case this is not a very convenient procedure for determining $E\,X$ and $\operatorname{Var} X$; the purpose was merely to illustrate the method. □

Exercise 3.1. Prove, with the aid of moment generating functions, that if $Y_1, Y_2, \ldots, Y_n$ are independent $\mathrm{Be}(p)$-distributed random variables, then $Y_1 + Y_2 + \cdots + Y_n \in \mathrm{Bin}(n,p)$.

Exercise 3.2. Prove, similarly, that if $X_1 \in \mathrm{Bin}(n_1,p)$ and $X_2 \in \mathrm{Bin}(n_2,p)$ are independent, then $X_1 + X_2 \in \mathrm{Bin}(n_1+n_2,p)$. □

The geometric distribution. For $X \in \mathrm{Ge}(p)$ computations like those made for the generating function yield $\psi_X(t) = p/(1-qe^t)$ (for $qe^t < 1$). Differentiation yields $E\,X$ and $\operatorname{Var} X$.

The Poisson distribution. For $X \in \mathrm{Po}(m)$ we obtain $\psi_X(t) = e^{m(e^t-1)}$ for all t, and so forth.

Next we compute the moment generating function for some continuous distributions.

The uniform (rectangular) distribution. Let $X \in U(a,b)$. Then

$$\psi_X(t) = \int_a^b e^{tx}\frac{1}{b-a}\,dx = \frac{1}{b-a}\left[\frac{1}{t}e^{tx}\right]_a^b = \frac{e^{tb}-e^{ta}}{t(b-a)}$$

for all t. In particular,

$$\psi_{U(0,1)}(t) = \frac{e^t-1}{t} \quad\text{and}\quad \psi_{U(-1,1)}(t) = \frac{e^t-e^{-t}}{2t} = \frac{\sinh t}{t}.$$

The moments can be obtained by differentiation. If, instead, we use Taylor expansion, then

$$
\begin{aligned}
\psi_X(t) &= \frac{1}{t(b-a)}\Big[1+\sum_{n=1}^{\infty}\frac{(tb)^n}{n!}-\Big(1+\sum_{n=1}^{\infty}\frac{(ta)^n}{n!}\Big)\Big]\\
&= \frac{1}{t(b-a)}\sum_{n=1}^{\infty}\Big(\frac{(tb)^n}{n!}-\frac{(ta)^n}{n!}\Big)=\frac{1}{b-a}\sum_{n=1}^{\infty}\frac{b^n-a^n}{n!}t^{n-1}\\
&= 1+\sum_{n=1}^{\infty}\frac{b^{n+1}-a^{n+1}}{(b-a)(n+1)!}t^n = 1+\sum_{n=1}^{\infty}\frac{b^{n+1}-a^{n+1}}{(b-a)(n+1)}\cdot\frac{t^n}{n!}\,,
\end{aligned}
$$

from which we conclude that

$$
E\,X^n = \frac{b^{n+1}-a^{n+1}}{(b-a)(n+1)} \quad \text{for} \quad n=1,\,2,\,\ldots,
$$

and thus, in particular, the known expressions for mean and variance, via

$$
\begin{aligned}
E\,X &= \frac{b^2-a^2}{2(b-a)} = \frac{a+b}{2},\\
E\,X^2 &= \frac{b^3-a^3}{3(b-a)} = \frac{b^2+ab+a^2}{3},\\
\operatorname{Var} X &= \frac{b^2+ab+a^2}{3}-\Big(\frac{a+b}{2}\Big)^2 = \frac{(b-a)^2}{12}\,.
\end{aligned}
$$

The exponential distribution. Let $X \in \operatorname{Exp}(a)$. Then

$$
\begin{aligned}
\psi_X(t) &= \int_0^\infty e^{tx}\frac{1}{a}e^{-x/a}\,dx = \frac{1}{a}\int_0^\infty e^{-x(\frac{1}{a}-t)}\,dx\\
&= \frac{1}{a}\cdot\frac{1}{\frac{1}{a}-t} = \frac{1}{1-at} \quad \text{for} \quad t<\frac{1}{a}.
\end{aligned}
$$

Furthermore, $\psi_X'(t) = a/(1-at)^2$, $\psi_X''(t) = 2a^2/(1-at)^3$, and, in general, $\psi_X^{(n)}(t) = n!a^n/(1-at)^{n+1}$. It follows that $E\,X^n = n!a^n$, $n = 1, 2, \ldots$, and, in particular, that $E\,X = a$ and $\operatorname{Var} X = a^2$.

Exercise 3.3. Perform a Taylor expansion of the moment generating function, and verify the expressions for the moments. □

The gamma distribution. For $X \in \Gamma(p,a)$, we have

$$
\begin{aligned}
\psi_X(t) &= \int_0^\infty e^{tx}\frac{1}{\Gamma(p)}x^{p-1}\frac{1}{a^p}e^{-x/a}\,dx\\
&= \frac{1}{a^p}\cdot\frac{1}{(\frac{1}{a}-t)^p}\int_0^\infty \frac{1}{\Gamma(p)}x^{p-1}\Big(\frac{1}{a}-t\Big)^p e^{-x(\frac{1}{a}-t)}\,dx\\
&= \frac{1}{a^p}\frac{1}{(\frac{1}{a}-t)^p}\cdot 1 = \frac{1}{(1-at)^p} \quad \text{for} \quad t<\frac{1}{a}.
\end{aligned}
$$

As is standard by now, the moments may be obtained via differentiation. Note also that $\psi(t) = (\psi_{\mathrm{Exp}(a)}(t))^p$. Thus, for $p = 1, 2, \ldots,$ we conclude from Corollary 3.2.1 and Theorem 3.1 that if $Y_1, Y_2, \ldots, Y_p$ are independent, $\mathrm{Exp}(a)$-distributed random variables then $Y_1 + Y_2 + \cdots + Y_p \in \Gamma(p, a)$.

Exercise 3.4. (a) Check the details of the last statement.
(b) Show that if $X_1 \in \Gamma(p_1, a)$ and $X_2 \in \Gamma(p_2, a)$ are independent random variables then $X_1 + X_2 \in \Gamma(p_1 + p_2, a)$. □

The standard normal distribution. Suppose that $X \in N(0, 1)$. Then

$$\begin{aligned}\psi_X(t) &= \int_{-\infty}^{\infty} e^{tx} \frac{1}{\sqrt{2\pi}} \exp\{-x^2/2\}\,dx \\ &= e^{t^2/2} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\{-(x-t)^2/2\}\,dx = e^{t^2/2}, \quad -\infty < t < \infty.\end{aligned}$$

The general normal (Gaussian) distribution. Suppose that $X \in N(\mu, \sigma^2)$. Then

$$\begin{aligned}\psi_X(t) &= \int_{-\infty}^{\infty} e^{tx} \frac{1}{\sigma\sqrt{2\pi}} \exp\Big\{-\frac{(x-\mu)^2}{2\sigma^2}\Big\}\,dx \\ &= e^{t\mu+\sigma^2t^2/2} \int_{-\infty}^{\infty} \frac{1}{\sigma\sqrt{2\pi}} \exp\Big\{-\frac{(x-\mu-\sigma^2 t)^2}{2\sigma^2}\Big\}\,dx \\ &= e^{t\mu+\sigma^2t^2/2}, \quad -\infty < t < \infty.\end{aligned}$$

The computations in the special case and the general case are essentially the same; it is a matter of completing squares. However, this is a bit more technical in the general case.

This leads to the following useful result, which shows how to derive the moment generating function of a linear transformation of a random variable.

Theorem 3.4. *Let X be a random variable and a and b be real numbers. Then*

$$\psi_{aX+b}(t) = e^{tb}\psi_X(at).$$

Proof. $\psi_{aX+b}(t) = E\,e^{t(aX+b)} = e^{tb} \cdot E\,e^{(at)X} = e^{tb} \cdot \psi_X(at)$. □

As an illustration we show how the moment generating function for a general normal distribution can be derived from the moment generating function of the standard normal one.

Thus, suppose that $X \in N(\mu, \sigma^2)$. We then know that $X \stackrel{d}{=} \sigma Y + \mu$, where $Y \in N(0, 1)$. An application of Theorem 3.4 thus tells us that

$$\psi_X(t) = e^{t\mu}\psi_Y(\sigma t) = e^{t\mu+\sigma^2t^2/2},$$

as expected.

Exercise 3.5. (a) Show that if $X \in N(\mu, \sigma^2)$ then $E\,X = \mu$ and $\operatorname{Var} X = \sigma^2$.

(b) Let $X_1 \in N(\mu_1, \sigma_1^2)$ and $X_2 \in N(\mu_2, \sigma_2^2)$ be independent random variables. Show that $X_1 + X_2$ is normally distributed, and find the parameters.

(c) Let $X \in N(0, \sigma^2)$. Show that $E\,X^{2n+1} = 0$ for $n = 0, 1, 2, \ldots$, and that $E\,X^{2n} = [(2n)!/2^n n!] \cdot \sigma^{2n} = (2n-1)!!\sigma^{2n} = 1 \cdot 3 \cdots (2n-1)\sigma^{2n}$ for $n = 1, 2, \ldots$.

Exercise 3.6. (a) Show that if $X \in N(0,1)$ then $X^2 \in \chi^2(1)$ by computing the moment generating function of X^2, that is, by showing that

$$\psi_{X^2}(t) = E\exp\{tX^2\} = \frac{1}{\sqrt{1-2t}} \quad \text{for} \quad t < \frac{1}{2}.$$

(b) Show that if $X_1 \in N(0,1)$ and $X_2 \in N(0,1)$ are independent then $X_1^2 + X_2^2 \in \chi^2(2) \quad (= \operatorname{Exp}(2))$. □

For two-dimensional analogs to Exercise 3.6, see Problems 5.10.36 and 37.

The Cauchy distribution. The moment generating function does not exist for the Cauchy distribution, since $\int [e^{tx}/(1+x^2)]\,dx$ is divergent for all $t \neq 0$. Note also that the nonexistence of the moment generating function follows from Theorem 3.3(a), since no moments of order 1 and above exist.

According to Theorem 3.3(a), it is conceivable that there might exist distributions with moments of all orders and, yet, the moment generating function does not exist in any neighborhood around zero. In fact, the *log-normal distribution* is one such example. To see this we first note that if $X \in \mathrm{LN}(\mu, \sigma^2)$, then $X \overset{d}{=} e^Y$, where $Y \in N(\mu, \sigma^2)$, which implies that

$$f_X(x) = \begin{cases} \dfrac{1}{\sigma x \sqrt{2\pi}} \exp\{-\frac{(\log x - \mu)^2}{2\sigma^2}\}, & \text{for} \quad x > 0, \\ 0, & \text{otherwise.} \end{cases}$$

It follows that

$$E\,X^r = E\,e^{rY} = \psi_Y(r) = \exp\{r\mu + \tfrac{1}{2}\sigma^2 r^2\},$$

for any $r > 0$, that is, all moments exist.

However, since $e^x \geq x^n/n!$ for any n, it follows that, for any $t > 0$,

$$\begin{aligned} E\exp\{tX\} &= E\exp\{te^Y\} \geq E\frac{(te^Y)^n}{n!} = \frac{t^n}{n!}E\,e^{nY} \\ &= \frac{t^n}{n!}\psi_Y(n) = \frac{t^n}{n!}\exp\{n\mu + \tfrac{1}{2}\sigma^2 n^2\} \\ &= \frac{1}{n!}\exp\{n(\log t + \mu + \tfrac{1}{2}\sigma^2 n)\}, \end{aligned}$$

which can be made arbitrarily large by choosing n sufficiently large, since $\log t + \mu + \frac{1}{2}\sigma^2 n \geq \frac{1}{4}\sigma^2 n$ for any fixed $t > 0$ as $n \to \infty$ and $\exp\{cn^2\}/n! \to \infty$

as $n \to \infty$ for any positive constant c. The moment generating function thus does not exist for any positive t.

Another class of distributions that possesses moments of all orders but not a moment generating function is the class of *generalized gamma distributions* whose densities are

$$f(x) = Cx^{\beta-1}e^{-x^{\alpha}}, \quad x > 0,$$

where $\beta > -1$, $0 < \alpha < 1$, and C is a normalizing constant (that is chosen such that the total mass equals 1).

It is clear that all moments exist, but, since $\alpha < 1$, we have

$$\int_{-\infty}^{\infty} e^{tx}x^{\beta-1}e^{-x^{\alpha}}\,dx = +\infty$$

for *all* $t > 0$, so that the moment generating function does not exist.

Remark 3.6. The fact that the integral is finite for all $t < 0$ is no contradiction, since for a moment generating function to exist we require finiteness of the integral in a neighborhood of zero, that is, for $|t| < h$ for some $h > 0$. □

We close this section by defining the moment generating function for random vectors.

Definition 3.2. *Let* $\mathbf{X} = (X_1, X_2, \ldots, X_n)'$ *be a random vector. The* moment generating function *of* $\mathbf{X}$ *is*

$$\psi_{X_1,\ldots,X_n}(t_1,\ldots,t_n) = E\,e^{t_1X_1+\cdots+t_nX_n},$$

provided there exist $h_1, h_2, \ldots, h_n > 0$ *such that the expectation exists for* $|t_k| < h_k$, $k = 1, 2, \ldots, n$. □

Remark 3.7. In vector notation (where, thus, $\mathbf{X}, \mathbf{t}$, and $\mathbf{h}$ are column vectors) the definition may be rewritten in the more compact form

$$\psi_{\mathbf{X}}(\mathbf{t}) = E\,e^{\mathbf{t}'\mathbf{X}},$$

provided there exists $\mathbf{h} > \mathbf{0}$, such that the expectation exists for $|\mathbf{t}| < \mathbf{h}$ (the inequalities being interpreted componentwise). □

4 The Characteristic Function

So far we have introduced two transforms: the generating function and the moment generating function. The advantage of moment generating functions over generating functions is that they can be defined for all kinds of random variables. However, the moment generating function does not exist for all distributions; the Cauchy and the log-normal distributions are two such examples. In this section we introduce a third transform, the characteristic function, which exists for *all* distributions. A minor technical complication, however, is that this transform is complex-valued and therefore requires somewhat more sophisticated mathematics in order to be dealt with stringently.

Definition 4.1. *The* characteristic function *of a random variable X is*

$$\varphi_X(t) = E\, e^{itX} = E(\cos tX + i \sin tX).$$ □

As mentioned above, the characteristic function is complex-valued. Since

$$|E\, e^{itX}| \leq E\, |e^{itX}| = E\, 1 = 1, \tag{4.1}$$

it follows that the characteristic function exists for all t and for all random variables.

Remark 4.1. Apart from a minus sign in the exponent (and, possibly, a factor $\sqrt{1/2\pi}$), characteristic functions coincide with Fourier transforms in the continuous case and with Fourier series in the discrete case. □

We begin with some basic facts and properties.

Theorem 4.1. *Let X be a random variable. Then*

(a) $|\varphi_X(t)| \leq \varphi_X(0) = 1$;
(b) $\overline{\varphi_X(t)} = \varphi_X(-t)$;
(c) $\varphi_X(t)$ *is (uniformly) continuous.*

Proof. (a) $\varphi_X(0) = E\, e^{i \cdot 0 \cdot X} = 1$. This, together with (4.1), proves (a).
(b) We have

$$\begin{aligned}\overline{\varphi_X(t)} &= E(\cos tX - i \sin tX) = E(\cos(-t)X + i \sin(-t)X) \\ &= E\, e^{i(-t)X} = \varphi_X(-t).\end{aligned}$$

(c) Let t be arbitrary and $h > 0$ (a similar argument works for $h < 0$). Then

$$\begin{aligned}|\varphi_X(t+h) - \varphi_X(t)| &= |E\, e^{i(t+h)X} - E\, e^{itX}| \\ &= |E\, e^{itX}(e^{ihX} - 1)| \leq E|e^{itX}(e^{ihX} - 1)| \\ &= E\, |e^{ihX} - 1|. \end{aligned} \tag{4.2}$$

Now, suppose that X has a continuous distribution; the discrete case is treated analogously.

For the function e^{ix} we have the trivial estimate $|e^{ix} - 1| \leq 2$, but also the more delicate one $|e^{ix} - 1| \leq |x|$. With the aid of these estimates we obtain, for $A > 0$,

$$\begin{aligned}E\, |e^{ihX} - 1| &= \int_{-\infty}^{-A} |e^{ihx} - 1| f_X(x)\, dx + \int_{-A}^{A} |e^{ihx} - 1| f_X(x)\, dx \\ &\qquad + \int_{A}^{\infty} |e^{ihx} - 1| f_X(x)\, dx \\ &\leq \int_{-\infty}^{-A} 2 f_X(x)\, dx + \int_{-A}^{A} |hx| f_X(x)\, dx + \int_{A}^{\infty} 2 f_X(x)\, dx \\ &\leq 2P(|X| \geq A) + hAP(|X| \leq A) \\ &\leq 2P(|X| \geq A) + hA. \end{aligned} \tag{4.3}$$

Let $\varepsilon > 0$ be arbitrarily small. It follows from (4.2) and (4.3) that

$$|\varphi_X(t+h) - \varphi_X(t)| \le 2P(|X| \ge A) + hA < \varepsilon, \tag{4.4}$$

provided we first choose A so large that $2P(|X| \ge A) < \varepsilon/2$, and then h so small that $hA < \varepsilon/2$. This proves the continuity of φ_X. Since the estimate in (4.4) does not depend on t, we have, in fact, shown that φ_X is uniformly continuous. □

Theorem 4.2. *Let X and Y be random variables. If $\varphi_X = \varphi_Y$, then $X \stackrel{d}{=} Y$.* □

This is the uniqueness theorem for characteristic functions. Next we present, without proof, some inversion theorems.

Theorem 4.3. *Let X be a random variable with distribution function F and characteristic function φ. If F is continuous at a and b, then*

$$F(b) - F(a) = \lim_{T\to\infty} \frac{1}{2\pi} \int_{-T}^{T} \frac{e^{-itb} - e^{-ita}}{-it} \cdot \varphi(t)\,dt.$$ □

Remark 4.2. Observe that Theorem 4.2 is an immediate corollary of Theorem 4.3. This is due to the fact that the former theorem is an existence result (only), whereas the latter provides a formula for explicitly computing the distribution function in terms of the characteristic function. □

Theorem 4.4. *If, in addition, $\int_{-\infty}^{\infty} |\varphi(t)|\,dt < \infty$, then X has a continuous distribution with density*

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

Theorem 4.5. *If the distribution of X is discrete, then*

$$P(X = x) = \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} e^{-itx} \cdot \varphi(t)\,dt.$$ □

As for the name of the transform, we have just seen that every random variable possesses a unique characteristic function; the characteristic function characterizes the distribution uniquely.

The proof of the following result, the multiplication theorem for characteristic functions, is similar to those for the other transforms and is therefore omitted.

Theorem 4.6. *Let $X_1, X_2, \ldots, X_n$ be independent random variables, and set $S_n = X_1 + X_2 + \cdots + X_n$. Then*

$$\varphi_{S_n}(t) = \prod_{k=1}^{n} \varphi_{X_k}(t).$$ □

Corollary 4.6.1. *If, in addition, $X_1, X_2, \ldots, X_n$ are equidistributed, then*

$$\varphi_{S_n}(t) = \big(\varphi_X(t)\big)^n. \qquad \square$$

Since we have derived the transform of several known distributions in the two previous sections, we leave some of them as exercises in this section.

Exercise 4.1. Show that $\varphi_{\mathrm{Be}(p)}(t) = q + pe^{it}$, $\varphi_{\mathrm{Bin}(n,p)}(t) = (q + pe^{it})^n$, $\varphi_{\mathrm{Ge}(p)}(t) = p/(1 - qe^{it})$, and $\varphi_{\mathrm{Po}(m)}(t) = \exp\{m(e^{it} - 1)\}$. $\square$

Note that for the computation of these characteristic functions one seems to perform the same work as for the computation of the corresponding moment generating function, the only difference being that t is replaced by it. In fact, in the discrete cases we considered in the previous sections, the computations are really completely analogous. The binomial theorem, convergence of geometric series, and Taylor expansion of the exponential function hold unchanged in the complex case.

The situation is somewhat more complicated for continuous distributions.

The uniform (rectangular) distribution. Let $X \in U(a, b)$. Then

$$\begin{aligned}
\varphi_X(t) &= \int_a^b e^{itx} \frac{1}{b-a}\, dx = \frac{1}{b-a} \int_a^b (\cos tx + i \sin tx)\, dx \\
&= \frac{1}{b-a} \cdot \Big[\frac{1}{t} \sin tx - i\frac{1}{t} \cos tx\Big]_a^b \\
&= \frac{1}{b-a} \cdot \frac{1}{t}(\sin bt - \sin at - i \cos bt + i \cos at) \\
&= \frac{1}{it(b-a)}(i \sin bt - i \sin at + \cos bt - \cos at) \\
&= \frac{e^{itb} - e^{ita}}{it(b-a)} \quad \big(= \psi_X(it)\big).
\end{aligned}$$

In particular,

$$\varphi_{U(0,1)}(t) = \frac{e^{it} - 1}{it} \quad \text{and} \quad \varphi_{U(-1,1)}(t) = \frac{e^{it} - e^{-it}}{2it} = \frac{\sin t}{t}. \tag{4.5}$$

The (mathematical) complication is that we cannot integrate as easily as we could before. However, in this case we observe that the derivative of e^{ix} equals ie^{ix}, which justifies the integration and hence implies that the computations here are “the same” as for the moment generating function.

For the exponential and gamma distributions, the complication arises in the following manner:

The exponential distribution. Let $X \in \mathrm{Exp}(a)$. Then

$$\varphi_X(t) = \int_0^\infty e^{itx}\frac{1}{a}e^{-x/a}\,dx = \frac{1}{a}\int_0^\infty e^{-x(\frac{1}{a}-it)}\,dx$$
$$= \frac{1}{a}\cdot\frac{1}{\frac{1}{a}-it} = \frac{1}{1-ait}\,.$$

The gamma distribution. Let $X \in \Gamma(p,a)$. We are faced with the same problems as for the exponential distribution. The conclusion is that $\varphi_{\Gamma(p,a)}(t) = (1-ait)^{-p}$.

The standard normal (Gaussian) distribution. Let $X \in N(0,1)$. Then

$$\varphi_X(t) = \int_{-\infty}^\infty e^{itx}\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}x^2}\,dx$$
$$= e^{-t^2/2}\int_{-\infty}^\infty \frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}(x-it)^2}\,dx = e^{-t^2/2}.$$

In this case one cannot argue as before, since there is no primitive function. Instead we observe that the moment generating function can be extended into a function that is analytic in the complex plane. The characteristic function equals the thus extended function along the imaginary axis, from which we conclude that $\varphi_X(t) = \psi_X(it)$ $(= e^{(it)^2/2} = e^{-t^2/2})$.

It is now possible to prove the addition theorems for the various distributions just as for generating functions and moment generating functions.

Exercise 4.2. Prove the addition theorems for the binomial, Poisson, and gamma distributions. □

In Remark 3.4 we gave a series expansion of the moment generating function. Following is the counterpart for characteristic functions:

Theorem 4.7. *Let X be a random variable. If $E\,|X|^n < \infty$ for some $n = 1, 2, \ldots,$ then*

(a) $\varphi_X^{(k)}(0) = i^k \cdot E\,X^k \quad \textit{for} \quad k = 1, 2, \ldots, n;$

(b) $\varphi_X(t) = 1 + \sum_{k=1}^n E\,X^k \cdot (it)^k/k! + o(|t|^n) \quad \textit{as} \quad t \to 0.$ □

Remark 4.3. For $n = 2$ we obtain, in particular,

$$\varphi_X(t) = 1 + itE\,X - \frac{t^2}{2}E\,X^2 + o(t^2) \quad \text{as} \quad t \to 0.$$

If, moreover, $E\,X = 0$ and $\operatorname{Var} X = \sigma^2$, then

$$\varphi_X(t) = 1 - \tfrac{1}{2}t^2\sigma^2 + o(t^2) \quad \text{as} \quad t \to 0.$$ □

Exercise 4.3. Find the mean and variance of the binomial, Poisson, uniform, exponential, and standard normal distributions. □

The conclusion of Theorem 4.7 is rather natural in view of Theorem 3.3 and Remark 3.4. Note, however, that a random variable whose moment generating function exists has moments of all orders (Theorem 3.3(a)), which implies that the series expansion can be carried out as an infinite sum. Since, however, all random variables (in particular, those without (higher order) moments) possess a characteristic function, it is reasonable to expect that the expansion here can only be carried out as long as moments exist. The order of magnitude of the remainder follows from estimating the difference of e^{ix} and the first part of its (complex) Taylor expansion.

Furthermore, a comparison between Theorems 3.3(b) and 4.7(a) tempts one to guess that these results could be derived from one another; once again the relation $\varphi_X(t) = \psi_X(it)$ seems plausible. This relation is, however, not true in general—recall that there are random variables, such as the Cauchy distribution, for which the moment generating function does not exist. In short, the validity of the relation depends on to what extent (if at all) the function $E\, e^{izX}$, where z is *complex-valued*, is an analytic function of z, a problem that will not be considered here (recall, however, the earlier arguments for the standard normal distribution).

Theorem 4.7 states that if the moment of a given order exists, then the characteristic function is differentiable, and the moments up to that order can be computed via the derivatives of the characteristic function as stated in the theorem. A natural question is whether a converse holds. The answer is yes, but only for moments of *even* order.

Theorem 4.8. *Let X be a random variable. If, for some $n = 0, 1, 2, \ldots$, the characteristic function φ has a finite derivative of order $2n$ at $t = 0$, then $E|X|^{2n} < \infty$ (and the conclusions of Theorem 4.7 hold).*

The "problem" with the converse is that if we want to apply Theorem 4.8 to show that the mean is finite we must first show that the second derivative of the characteristic function exists. Since there exist distributions with finite mean whose characteristic functions are not twice differentiable (such as the so-called stable distributions with index between 1 and 2), the theorem is not always applicable.

Next we present the analog of Theorem 3.4 on how to find the transform of a linearly transformed random variable.

Theorem 4.9. *Let X be a random variable and a and b be real numbers. Then*

$$\varphi_{aX+b}(t) = e^{ibt} \cdot \varphi_X(at).$$

Proof. $\varphi_{aX+b}(t) = E\, e^{it(aX+b)} = e^{itb} \cdot E\, e^{i(at)X} = e^{itb} \cdot \varphi_X(at)$. □

Exercise 4.4. Let $X \in N(\mu, \sigma^2)$. Use the expression above for the characteristic function of the standard normal distribution and Theorem 4.9 to show that $\varphi_X(t) = e^{it\mu - \sigma^2 t^2/2}$.

Exercise 4.5. Prove the addition theorem for the normal distribution. □

The Cauchy distribution. For $X \in C(0,1)$, one can show that

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

A device for doing this is the following: If we "already happen to know" that the difference between two independent, Exp(1)-distributed random variables is $L(1)$-distributed, then we know that

$$\varphi_{L(1)}(t) = \frac{1}{1-it} \cdot \frac{1}{1+it} = \frac{1}{1+t^2}$$

(use Theorem 4.6 and Theorem 4.9 (with $a=-1$ and $b=0$)). We thus have

$$\frac{1}{1+t^2} = \int_{-\infty}^{\infty} e^{itx} \tfrac{1}{2} e^{-|x|}\, dx.$$

A change of variables, such that $x \to t$ and $t \to x$, yields

$$\frac{1}{1+x^2} = \int_{-\infty}^{\infty} e^{itx} \tfrac{1}{2} e^{-|t|}\, dt,$$

and, by symmetry,

$$\frac{1}{1+x^2} = \int_{-\infty}^{\infty} e^{-itx} \tfrac{1}{2} e^{-|t|}\, dt,$$

which can be rewritten as

$$\frac{1}{\pi} \cdot \frac{1}{1+x^2} = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-itx} e^{-|t|}\, dt. \tag{4.6}$$

A comparison with the inversion formula given in Theorem 4.4 shows that since the left-hand side of (4.6) is the density of the $C(0,1)$-distribution, it necessarily follows that $e^{-|t|}$ is the characteristic function of this distribution.

Exercise 4.6. Use Theorem 4.9 to show that $\varphi_{C(m,a)}(t) = e^{itm}\varphi_X(at) = e^{itm-a|t|}$. □

Our final result in this section is a consequence of Theorems 4.9 and 4.1(b).

Theorem 4.10. *Let X be a random variable. Then*

$$\varphi_X \textit{ is real} \quad \Longleftrightarrow \quad X \stackrel{d}{=} -X$$

(i.e., iff the distribution of X is symmetric).

Proof. Theorem 4.9 (with $a = -1$ and $b = 0$) and Theorem 4.1(b) together yield

$$\varphi_{-X}(t) = \varphi_X(-t) = \overline{\varphi_X(t)}. \tag{4.7}$$

First suppose that φ_X is real-valued, that is, that $\overline{\varphi_X(t)} = \varphi_X(t)$. It follows that $\varphi_{-X}(t) = \varphi_X(t)$, or that X and $-X$ have the same characteristic function. By the uniqueness theorem they are equidistributed.

Now suppose that $X \stackrel{d}{=} -X$. Then $\varphi_X(t) = \varphi_{-X}(t)$, which, together with (4.7), yields $\varphi_X(t) = \overline{\varphi_X(t)}$, that is, φ_X is real-valued. $\square$

Exercise 4.7. Show that if X and Y are i.i.d. random variables then $X - Y$ has a symmetric distribution.

Exercise 4.8. Show that one cannot find i.i.d. random variables X and Y such that $X - Y \in U(-1, 1)$. $\square$

We conclude by defining the characteristic function for random vectors.

Definition 4.2. *Let* $\mathbf{X} = (X_1, X_2 \ldots, X_n)'$ *be a random vector. The characteristic function of* $\mathbf{X}$ *is*

$$\varphi_{X_1,\ldots,X_n}(t_1, \ldots, t_n) = E\, e^{i(t_1 X_1 + \cdots + t_n X_n)}.$$

In the more compact vector notation (cf. Remark 3.7) this may be rewritten as

$$\varphi_{\mathbf{X}}(\mathbf{t}) = E\, e^{i\mathbf{t}'\mathbf{X}}.$$

$\square$

In particular, the following special formulas, which are useful at times, can be obtained:

$$\varphi_{X_1,\ldots,X_n}(t, t, \ldots, t) = E\, e^{it(X_1 + \cdots + X_n)} = \varphi_{X_1 + \cdots + X_n}(t)$$

and

$$\varphi_{X_1,\ldots,X_n}(t, 0, \ldots, 0) = \varphi_{X_1}(t).$$

Characteristic functions of random vectors are an important tool in the treatment of the multivariate normal distribution in Chapter 5.

5 Distributions with Random Parameters

This topic was treated in Section 2.3 by conditioning methods. Here we show how Examples 2.3.1 and 2.3.2 (in the reverse order) can be tackled with the aid of transforms. Let us begin by saying that transforms are often easier to work with computationally than the conditioning methods. However, one reason for this is that behind the transform approach there are theorems that sometimes are rather sophisticated.

Example 2.3.2 (continued). Recall that the point of departure was

$$X \mid N = n \in \text{Bin}(n, p) \quad \text{with} \quad N \in \text{Po}(\lambda). \tag{5.1}$$

An application of Theorem 2.2.1 yields

$$g_X(t) = E\big(E(t^X \mid N)\big) = E\,h(N)\,,$$

where

$$h(n) = E(t^X \mid N = n) = (q + pt)^n,$$

from which it follows that

$$g_X(t) = E(q + pt)^N = g_N(q + pt) = e^{\lambda((q+pt)-1)} = e^{\lambda p(t-1)}\,,$$

that is, $X \in \text{Po}(\lambda p)$ (why?). Note also that $g_N(q + pt) = g_N(g_{\text{Be}(p)}(t))$.

Example 2.3.1 (continued). We had

$$X \mid M = m \in \text{Po}(m) \quad \text{with} \quad M \in \text{Exp}(1).$$

By using the moment generating function (for a change) and Theorem 2.2.1, we obtain

$$\psi_X(t) = E\,e^{tX} = E\big(E(e^{tX} \mid M)\big) = E\,h(M),$$

where

$$h(m) = E(e^{tX} \mid M = m) = \psi_{X|M=m}(t) = e^{m(e^t-1)}.$$

Thus,

$$\begin{aligned}\psi_X(t) &= E\,e^{M(e^t-1)} = \psi_M(e^t - 1) = \frac{1}{1 - (e^t - 1)}\\ &= \frac{1}{2 - e^t} = \frac{\frac{1}{2}}{1 - \frac{1}{2}e^t} = \psi_{\text{Ge}(1/2)}(t)\,,\end{aligned}$$

and we conclude that $X \in \text{Ge}(1/2)$. □

Remark 5.1. It may be somewhat faster to use generating functions, but it is useful to practise another transform. □

Exercise 5.1. Solve Exercise 2.3.1 using transforms. □

In Section 2.3 we also considered the situation

$$X \mid \Sigma^2 = y \in N(0, y) \quad \text{with} \quad \Sigma^2 \in \text{Exp}(1),$$

which is the normal distribution with mean zero and an exponentially distributed variance. After hard work we found that $X \in L(1/\sqrt{2})$. The alternative, using characteristic functions and Theorem 2.2.1, yields

$$\varphi_X(t) = E\, e^{itX} = E\big(E(e^{itX} \mid \Sigma^2)\big) = E\, h(\Sigma^2)\,,$$

where

$$h(y) = \varphi_{X \mid \Sigma^2 = y}(t) = e^{-t^2 y/2},$$

and so

$$\begin{aligned}\varphi_X(t) &= E\, e^{-t^2\Sigma^2/2} = \psi_{\Sigma^2}(-\tfrac{t^2}{2}) \\ &= \frac{1}{1-(-\frac{t^2}{2})} = \frac{1}{1+(\frac{1}{\sqrt{2}})^2 t^2} = \varphi_{L(1/\sqrt{2})}(t),\end{aligned}$$

and the desired conclusion follows. At this point, however, let us stress once again that the price of the simpler computations here are some general theorems (Theorem 2.2.1 and the uniqueness theorem for characteristic functions), the proofs of which are all the more intricate.

Exercise 5.2. Solve Exercise 2.3.3 using transforms. □

6 Sums of a Random Number of Random Variables

An important generalization of the theory of sums of independent random variables is the theory of sums of *a random number of* (independent) random variables. Apart from being a theory in its own right, it has several interesting and important applications. In this section we study this problem under the additional assumption that *the number of terms in the sum is independent of the summands*; in the following section we present an important application to branching processes (the interested reader might pause here for a moment and read the first few paragraphs of that section).

Before proceeding, however, here are some examples that will be solved after some theory has been presented.

Example 6.1. Consider a roulette wheel with the numbers $0, 1, \ldots, 36$. Charlie bets one dollar on number 13 until it appears. He then bets one dollar the same number of times on number 36. We wish to determine his expected loss in the second round (in which he bets on number 36).

Example 6.2. Let $X_1, X_2, \ldots$ be independent, Exp(1)-distributed random variables, and let $N \in \mathrm{Fs}(p)$ be independent of $X_1, X_2, \ldots$. We wish to find the distribution of $X_1 + X_2 + \cdots + X_N$.

In Section 5 we presented a solution of Example 2.3.2 based on transforms. Next we present another solution based on transforms where, instead, we consider the random variable in focus as a sum of a random number of Be(p)-distributed random variables.

Example 2.3.2 (continued). As before, let N be the number of emitted particles during a given hour. We introduce the following indicator random variables:

$$Y_k = \begin{cases} 1, & \text{if the } k\text{th particle is registered,} \\ 0, & \text{otherwise.} \end{cases}$$

Then

$$X = Y_1 + Y_2 + \cdots + Y_N$$

equals the number of registered particles during this particular hour. □

Thus, the general idea is that we are given a set $X_1, X_2, \ldots$ of i.i.d. random variables with partial sums $S_n = X_1 + X_2 + \cdots + X_n$, for $n \geq 1$. Furthermore, N is a nonnegative, integer-valued random variable that is independent of $X_1, X_2, \ldots$. Our aim is to investigate the random variable

$$S_N = X_1 + X_2 + \cdots + X_N\,, \tag{6.1}$$

where $S_N = S_0 = 0$ when $N = 0$.

For $A \subset (-\infty, \infty)$, we have

$$P(S_N \in A \mid N = n) = P(S_n \in A \mid N = n) = P(S_n \in A), \tag{6.2}$$

where the last equality is due to the independence of N and $X_1, X_2, \ldots$. The interpretation of (6.2) is that the distribution of S_N, given $N = n$, is the same as that of S_n.

Remark 6.1. Let $N = \min\{n : S_n > 0\}$. Clearly, $P(S_N > 0) = 1$. This implies that if the summands are allowed to assume negative values (with positive probability) then so will S_n, whereas S_N is always positive. However, in this case N is not independent of the summands; on the contrary, N is defined in terms of the summands. □

In case the summands are nonnegative and integer-valued, the generating function of S_N can be derived as follows:

Theorem 6.1. *Let $X_1, X_2, \ldots$ be i.i.d. nonnegative, integer-valued random variables, and let N be a nonnegative, integer-valued random variable, independent of $X_1, X_2, \ldots$. Set $S_0 = 0$ and $S_n = X_1 + X_2 + \cdots + X_n$, for $n \geq 1$. Then*

$$g_{S_N}(t) = g_N\big(g_X(t)\big). \tag{6.3}$$

Proof. We have

$$\begin{aligned} g_{S_N}(t) &= E\,t^{S_N} = \sum_{n=0}^{\infty} E\left(t^{S_N} \mid N = n\right) \cdot P(N = n) \\ &= \sum_{n=0}^{\infty} E\left(t^{S_n} \mid N = n\right) \cdot P(N = n) = \sum_{n=0}^{\infty} E\left(t^{S_n}\right) \cdot P(N = n) \\ &= \sum_{n=0}^{\infty} \big(g_X(t)\big)^n \cdot P(N = n) = g_N\big(g_X(t)\big). \end{aligned}$$

□

Remark 6.2. In the notation of Chapter 2 and with the aid of Theorem 2.2.1, we may alternatively write

$$g_{S_N}(t) = E\, t^{S_N} = E\big(E\,(t^{S_N} \mid N)\big) = E\, h(N)\,,$$

where

$$h(n) = E\,(t^{S_N} \mid N = n) = \cdots = \big(g_X(t)\big)^n,$$

which yields

$$g_{S_N}(t) = E\,\big(g_X(t)\big)^N = g_N\big(g_X(t)\big). \qquad \square$$

Theorem 6.2. *Suppose that the conditions of Theorem 6.1 are satisfied.*

(a) *If, moreover,*

$$E\,N < \infty \quad \textit{and} \quad E\,|X| < \infty,$$

then

$$E\,S_N = E\,N \cdot E\,X.$$

(b) *If, in addition,*

$$\operatorname{Var} N < \infty \quad \textit{and} \quad \operatorname{Var} X < \infty,$$

then

$$\operatorname{Var} S_N = E\,N \cdot \operatorname{Var} X + (E\,X)^2 \cdot \operatorname{Var} N.$$

Proof. It follows from Corollary 2.3.1 that

$$E\,S_N = g'_{S_N}(1) \tag{6.4}$$

and that

$$\operatorname{Var} S_N = g''_{S_N}(1) + g'_{S_N}(1) - \big(g'_{S_N}(1)\big)^2. \tag{6.5}$$

Furthermore, by differentiating the right-hand side of (6.3), using the chain rule, we obtain

$$g'_{S_N}(t) = g'_N\big(g_X(t)\big) \cdot g'_X(t),$$

which, after letting $t \nearrow 1$, yields

$$E\,S_N = g'_{S_N}(1) = g'_N(1) \cdot g'_X(1) = E\,N \cdot E\,X.$$

This proves (a).

A further differentiation shows that

$$g''_{S_N}(t) = g''_N\big(g_X(t)\big) \cdot \big(g'_X(t)\big)^2 + g'_N\big(g_X(t)\big) \cdot g''_X(t),$$

which yields

$$\begin{aligned} g''_{S_N}(1) &= g''_N(1) \cdot \big(g'_X(1)\big)^2 + g'_N(1) \cdot g''_X(1) \\ &= E\,N(N-1) \cdot (E\,X)^2 + E\,N \cdot E\,X(X-1). \end{aligned}$$

It finally follows that

$$\begin{aligned}\operatorname{Var} S_N &= g''_{S_N}(1) + g'_{S_N}(1) - \big(g'_{S_N}(1)\big)^2 \\ &= E\,N(N-1)\cdot(E\,X)^2 + E\,N\cdot E\,X(X-1) \\ &\quad + E\,N\cdot E\,X - (E\,N\cdot E\,X)^2 \\ &= E\,N\cdot \operatorname{Var} X + (E\,X)^2\cdot \operatorname{Var} N. \end{aligned}$$ □

Theorem 6.2 can also be proved directly by modifying the proof of Theorem 6.1 in the obvious manner. As for (a) we then have

$$\begin{aligned} E\,S_N &= \sum_{n=0}^{\infty} E\,(S_N \mid N=n)\cdot P(N=n) \\ &= \sum_{n=0}^{\infty} E\,(S_n \mid N=n)\cdot P(N=n) \\ &= \sum_{n=0}^{\infty} E\,(S_n)\cdot P(N=n) = \sum_{n=0}^{\infty} nE\,X\cdot P(N=n) \\ &= E\,X\cdot \sum_{n=0}^{\infty} nP(N=n) = E\,X\cdot E\,N. \end{aligned}$$

Note in particular that this proof is valid for arbitrary X_1, X_2, ... (some argument concerning the absolute convergence is needed).

Exercise 6.1. Compute $E\,S_N^2$ similarly and prove Theorem 6.2(b). □

In the notation of Chapter 2 we have, for Theorem 6.2(a) (cf. Remark 6.2),

$$E\,S_N = E\big(E(S_N \mid N)\big) = E\,h(N)\,,$$

where

$$h(n) = E(S_N \mid N=n) = E(S_n \mid N=n) = E\,S_n = nE\,X,$$

that is,

$$E\,S_N = E(N\,E\,X) = E\,X\cdot E\,N.$$

For an alternative proof of Theorem 6.2(b), we use Corollary 2.2.3.1, according to which

$$\operatorname{Var} S_N = E \operatorname{Var}(S_N \mid N) + \operatorname{Var}\big(E(S_N \mid N)\big).$$

Since (check!)

$$\operatorname{Var}(S_N \mid N=n) = \operatorname{Var}(S_n \mid N=n) = \operatorname{Var} S_n = n\operatorname{Var} X,$$

it follows that

$$E \operatorname{Var}(S_N \mid N) = E(N \operatorname{Var} X) = E\,N\cdot \operatorname{Var} X.$$

Furthermore, $E(S_N \mid N = n) = nE\,X$, which yields

$$\operatorname{Var}\big(E(S_N \mid N)\big) = \operatorname{Var}(N \cdot E\,X) = (E\,X)^2 \cdot \operatorname{Var} N,$$

and the desired conclusion follows.

Let us now use these results in order to obtain another solution of Example 2.3.2 and to solve the problem posed in Example 6.1.

Example 2.3.2 (continued). Recall that N was the number of emitted particles during a given hour, that we kept track of whether particles were registered or not by the indicator variables $Y_1, Y_2, \ldots,$ and that the number of registered particles during this particular hour was given by $X = Y_1 + Y_2 + \cdots + Y_N$.

An application of Theorem 6.1 now yields

$$g_X(t) = g_N\big(g_Y(t)\big) = \exp\{\lambda(g_Y(t) - 1)\} = e^{\lambda(q+pt-1)} = e^{\lambda p(t-1)},$$

which is the generating function of a $\operatorname{Po}(\lambda p)$-distribution. It follows from the uniqueness theorem for generating functions that $X \in \operatorname{Po}(\lambda p)$.

Moreover, by Theorem 6.2,

$$\begin{aligned} E\,X &= E\,N \cdot E\,Y = \lambda \cdot p, \\ \operatorname{Var} X &= E\,N \cdot \operatorname{Var} Y + (E\,Y)^2 \operatorname{Var} N = \lambda \cdot pq + p^2 \cdot \lambda = \lambda p. \end{aligned}$$ □

Remark 6.3. The answers here and in Section 5 are obviously the same, but they are obtained somewhat differently. Analogous arguments can be made in other examples. This provides a link between the two sections. □

As for Example 6.1, let $N \in \operatorname{Fs}(1/37)$ equal the number of bets on number 13, and let $Y_1, Y_2, \ldots$ be the losses in the bets on number 36. Thus

$$Y_k = \begin{cases} 1, & \text{if number 36 does not appear,} \\ -35, & \text{(i.e., } -36+1) \quad \text{otherwise,} \end{cases}$$

and $Y_1, Y_2, \ldots$ are independent with $P(Y_k = 1) = 36/37$ and $P(Y_k = -35) = 1/37$ (note that a negative loss is a gain). With this notation Charlie's total loss in the second round equals $X = Y_1 + Y_2 + \cdots + Y_N$, and an application of Theorem 6.2(a) yields

$$E\,X = E\,N \cdot E\,Y = 37 \cdot \left(1 \cdot \frac{36}{37} - 35 \cdot \frac{1}{37}\right) = 1.$$

If we wish to determine his overall loss, we have to add $(N-1) \cdot 1 - 35$ (or $N \cdot 1 - 36$) to X, in which case we find that the expected overall loss equals 2.

Although this does not seem so terrible, we must remember that this game requires access to an infinite amount of money to start with.

Exercise 6.2. Find the generating function of his loss in the second round. Try also to find it for his overall loss. □

If, as in Example 6.2, the summands have a continuous distribution, then Theorem 6.1 no longer applies, since the generating function is not defined for such random variables. However, the following result holds.

Theorem 6.3. *Let $X_1, X_2, \ldots$ be i.i.d. random variables, whose moment generating function exists for $|t| < h$ for some $h > 0$. Furthermore, let N be a nonnegative, integer-valued random variable independent of $X_1, X_2, \ldots$. Set $S_0 = 0$ and $S_n = X_1 + X_2 + \cdots + X_n$, for $n \geq 1$. Then*

$$\psi_{S_N}(t) = g_N\big(\psi_X(t)\big). \qquad \square$$

The proof is completely analogous to the proof of Theorem 6.1 and is therefore left as an exercise.

Exercise 6.3. Prove Theorem 6.2 by starting from Theorem 6.3. Note, however, that this requires the existence of the moment generating function of the summands, a restriction that we know from above is not necessary for Theorem 6.2 to hold. $\square$

Next we solve the problem posed in Example 6.2. Recall from there that we were given $X_1, X_2, \ldots$ independent, Exp(1)-distributed random variables and $N \in \text{Fs}(p)$ independent of $X_1, X_2, \ldots$ and that we wish to find the distribution of $X_1 + X_2 + \cdots + X_N$.

With the (by now) usual notation we have, by Theorem 6.3, for $t < p$,

$$\begin{aligned}
\psi_{S_N}(t) = g_N\big(\psi_X(t)\big) &= \frac{p \cdot \frac{1}{1-t}}{1 - q\frac{1}{1-t}} = \frac{p}{1-t-q} = \\
&= \frac{p}{p-t} = \frac{1}{1-\frac{t}{p}} = \psi_{\text{Exp}(1/p)}(t)\,,
\end{aligned}$$

which, by the uniqueness theorem for moment generating functions, shows that $S_N \in \text{Exp}(1/p)$. $\square$

Remark 6.4. If in Example 6.2 we had assumed that $N \in \text{Ge}(p)$, we would have obtained

$$\psi_{S_N}(t) = \frac{p}{1 - q\frac{1}{1-t}} = \frac{p(1-t)}{p-t} = p + q\frac{1}{1-\frac{t}{p}}\,.$$

This means that S_N is a mixture of a $\delta(0)$-distribution and an Exp$(1/p)$-distribution, the weights being p and q, respectively. An intuitive argument supporting this is that $P(S_N = 0) = P(N = 0) = p$. If $N \geq 1$, then S_N behaves as in Example 6.2. The distribution of S_N thus is neither discrete nor continuous; it is a mixture. Note also that a geometric random variable that is known to be positive is, in fact, Fs-distributed; if $Z \in \text{Ge}(p)$, then $Z \mid Z > 0 \in \text{Fs}(p)$. $\square$

Finally, if the summands do not possess a moment generating function, then characteristic functions can be used in the obvious way.

Theorem 6.4. *Let X_1, X_2, ... be i.i.d. random variables, and let N be a nonnegative, integer-valued random variable independent of X_1, X_2, ... Set $S_0 = 0$ and $S_n = X_1 + X_2 + \cdots + X_n$, for $n \geq 1$. Then*

$$\varphi_{S_N}(t) = g_N\big(\varphi_X(t)\big). \qquad \square$$

Exercise 6.4. Prove Theorem 6.4.

Exercise 6.5. Use Theorem 6.4 to prove Theorem 6.2. $\square$

7 Branching Processes

An important application for the results of the previous section is provided by the theory of branching processes, which is described by the following model:

At time $t = 0$ there exists an initial population (a group of ancestors or founding members) $X(0)$. During its lifespan, every individual gives birth to a random number of children. During their lifespans, these children give birth to a random number of children, and so on. The reproduction rules for the simplest case, which is the only one we shall consider, are

(a) all individuals give birth according to the same probability law, independently of each other;
(b) the number of children produced by an individual is independent of the number of individuals in their generation.

Such branching processes are called *Galton–Watson processes* after Sir Francis Galton (1822–1911)—a cousin of Charles Darwin—who studied the decay of English peerage and other family names of distinction (he contested the hypothesis that distinguished family names are more likely to become extinct than names of ordinary families) and Rev. Henry William Watson (1827–1903). They met via problem 4001 posed by Galton in the *Educational Times*, 1 April 1873, for which Watson proposed a solution in the same journal, 1 August 1873. Another of Galton's achievements was that he established the use of fingerprints in the police force.

In the sequel we also assume that $X(0) = 1$; this is a common assumption, made in order to simplify some nonsignificant matters. Furthermore, since individuals give birth, we attribute the female sex to them. Finally, to avoid certain trivialities, we exclude, throughout, the degenerate case—when each individual always gives birth to exactly one child.

Example 7.1. Family names. Assume that men and women who live together actually marry and that the woman changes her last name to that of her husband (as in the old days). A family name thus survives only through sons. If sons are born according to the rules above, the evolution of a family name may be described by a branching process. In particular, one might be interested in whether or not a family name will live on forever or become extinct.

Instead of family names, one might consider some mutant gene and its survival or otherwise.

Example 7.2. Nuclear reactions. The fission caused by colliding neutrons results in a (random) number of new neutrons, which, when they collide produce new neutrons, and so on.

Example 7.3. Waiting lines. A customer who arrives at an empty server (or a telephone call that arrives at a switchboard) may be viewed as an ancestor. The customers (or calls) arriving while he is being served are his children, and so on. The process continues as long as there are people waiting to be served.

Example 7.4. The laptometer. When the sprows burst in a laptometer we are faced with failures of the first kind. Now, every sprow that bursts causes failures of the second kind (independently of the number of failures of the first kind and of the other sprows). Suppose the number of failures of the first kind during one hour follows the $\mathrm{Po}(\lambda)$-distribution and that the number of failures of the second kind caused by one sprow follows the $\mathrm{Bin}(n, p)$-distribution. Find the mean and variance of the total number of failures during one hour. □

We shall solve the problem posed in Example 7.4 later.

Now, let, for $n \geq 1$,

$$X(n) = \# \text{ individuals in generation } n,$$

let Y and $\{Y_k,\ k \geq 1\}$ be generic random variables denoting the number of children obtained by individuals, and set $p_k = P(Y = k)$, $k = 0, 1, 2, \ldots$. Recall that we exclude the case $P(Y = 1) = 1$.

Consider the initial population or the ancestor $X(0)$ $(= 1 = \text{Eve})$. Then $X(1)$ equals the number of children of the ancestor and $X(1) \stackrel{d}{=} Y$. Next, let Y_1, Y_2, … be the number of children obtained by the first, second, … child. It follows from the assumptions that Y_1, Y_2, … are i.i.d. and, furthermore, independent of $X(1)$. Since

$$X(2) = Y_1 + \cdots + Y_{X(1)}, \tag{7.1}$$

we may apply the results from Section 6. An application of Theorem 6.1 yields

$$g_{X(2)}(t) = g_{X(1)}\big(g_{Y_1}(t)\big)\,. \tag{7.2}$$

If we introduce the notations

$$g_n(t) = g_{X(n)}(t) \quad \text{for} \quad n = 1, 2, \ldots$$

and $g(t) = g_1(t)$ $(= g_{X(1)}(t) = g_Y(t))$, (7.2) may be rewritten as

$$g_2(t) = g\big(g(t)\big)\,. \tag{7.3}$$

Next, let $Y_1, Y_2, \ldots$ be the number of children obtained by the first, second, $\ldots$ individuals in generation $n-1$. By arguing as before, we obtain

$$g_{X(n)}(t) = g_{X(n-1)}\big(g_{Y_1}(t)\big)\,,$$

that is,

$$g_n(t) = g_{n-1}\big(g(t)\big)\,. \tag{7.4}$$

This corresponds to the case $k = 1$ in the following result.

Theorem 7.1. *For a branching process as above we have*

$$g_n(t) = g_{n-k}\big(g_k(t)\big) \quad \textit{for} \quad k = 1, 2, \ldots, n-1. \qquad \square$$

If, in addition, $E\,Y_1 < \infty$, it follows from Theorem 6.2(a) that

$$E\,X(2) = E\,X(1) \cdot E\,Y_1 = (E\,Y_1)^2,$$

which, after iteration, yields

$$E\,X(n) = (E\,Y_1)^n. \tag{7.5}$$

Since every individual is expected to produce $E\,Y_1$ children, this is, intuitively, a very reasonable relation.

An analogous, although slightly more complicated, formula for the variance can also be obtained.

Theorem 7.2. (a) *Suppose that $m = E\,Y_1 < \infty$. Then*

$$E\,X(n) = m^n.$$

(b) *Suppose, further, that $\sigma^2 = \operatorname{Var} Y_1 < \infty$. Then*

$$\operatorname{Var} X(n) = \sigma^2(m^{n-1} + m^n + \cdots + m^{2n-2}). \qquad \square$$

Exercise 7.1. Prove Theorems 7.1 and 7.2(b). $\square$

Remark 7.1. Theorem 7.2 may, of course, also be derived from Theorem 7.1 by differentiation (cf. Corollary 2.3.1). $\square$

Asymptotics

Suppose that $\sigma^2 = \operatorname{Var} Y_1 < \infty$. It follows from Theorem 7.2 that

$$E\,X(n) \to \begin{cases} 0, & \text{when} \quad m < 1, \\ (=)1, & \text{when} \quad m = 1, \\ +\infty, & \text{when} \quad m > 1, \end{cases} \tag{7.6}$$

and that

$$\operatorname{Var} X(n) \to \begin{cases} 0, & \text{when} \quad m < 1, \\ +\infty, & \text{when} \quad m \geq 1 \end{cases} \tag{7.7}$$

as $n \to \infty$. It is easy to show that $P(X(n) > 0) \to 0$ as $n \to \infty$ when $m < 1$. Although we have not defined any concept of convergence yet (this will be done in Chapter 6), our intuition tells us that $X(n)$ should converge to zero as $n \to \infty$ in some sense in this case. Furthermore, (7.6) tells us that $X(n)$ increases indefinitely (on average) when $m > 1$. In this case, however, one might imagine that since the variance also grows the population may, by chance, die out at some finite time (in particular, at some early point in time). For the boundary case $m = 1$, it may be a little harder to guess what will happen in the long run. The following result puts our speculations into a stringent formulation.

Denote by η the probability of *(ultimate) extinction* of a branching process. For future reference we note that

$$\begin{aligned} \eta &= P(\text{ultimate extinction}) = P(X(n) = 0 \text{ for some } n) \\ &= P\Big(\bigcup_{n=1}^{\infty} \{X(n) = 0\}\Big). \end{aligned} \tag{7.8}$$

For obvious reasons we assume in the following that $P(X(1) = 0) > 0$.

Theorem 7.3. (a) *η satisfies the equation $t = g(t)$.*
(b) *η is the smallest nonnegative root of the equation $t = g(t)$.*
(c) *$\eta = 1$ for $m \leq 1$ and $\eta < 1$ for $m > 1$.*

Proof. (a) Let $A_k = \{\text{the founding member produces } k \text{ children}\}$, $k \geq 0$. By the law of total probability we have

$$\eta = \sum_{k=0}^{\infty} P(\text{extinction} \mid A_k) \cdot P(A_k). \tag{7.9}$$

Now, $P(A_k) = p_k$, and by the independence assumptions we have

$$P(\text{extinction} \mid A_k) = \eta^k. \tag{7.10}$$

These facts and (7.9) yield

$$\eta = \sum_{k=0}^{\infty} \eta^k p_k = g(\eta), \tag{7.11}$$

which proves (a).

(b) Set $\eta_n = P(X(n) = 0)$ and suppose that η^* is *some* nonnegative root of the equation $t = g(t)$ (since $g(1) = 1$, such a root exists always). Since g is nondecreasing for $t \geq 0$, we have, by Theorem 7.1,

$$\eta_1 = g(0) \leq g(\eta^*) = \eta^*,$$
$$\eta_2 = g(\eta_1) \leq g(\eta^*) = \eta^*,$$

and, by induction,

$$\eta_{n+1} = g(\eta_n) \leq g(\eta^*) = \eta^*,$$

that is, $\eta_n \leq \eta^*$ for all n. Finally, in view of (7.8) and the fact that $\{X(n) = 0\} \subset \{X(n+1) = 0\}$ for all n, it follows that $\eta_n \nearrow \eta$ and hence that $\eta \leq \eta^*$, which was to be proved.

(c) Since g is an infinite series with nonnegative coefficients, it follows that $g'(t) \geq 0$ and $g''(t) \geq 0$ for $0 \leq t \leq 1$. This implies that g is convex and nondecreasing on [0,1]. Furthermore, $g(1) = 1$. By comparing the graphs of the functions $y = g(t)$ and $y = t$ in the three cases $m < 1$, $m = 1$, and $m > 1$, respectively, it follows that they intersect at $t = 1$ only when $m \leq 1$ (tangentially when $m = 1$) and at $t = \eta$ and $t = 1$ when $m > 1$ (see Figure 7.1).

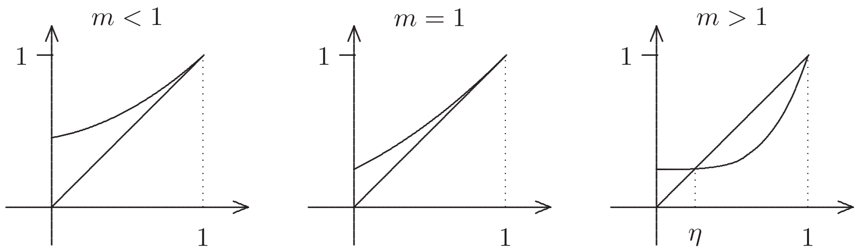

Figure 7.1

The proof of the theorem is complete. □

We close this section with some computations to illustrate the theory. Given first is an example related to Example 7.2 as well as to a biological phenomenon called binary splitting.

Example 7.5. In this branching process, the neutrons or cells either split into two new "individuals" during their lifetime or die. Suppose that the probabilities for these alternatives are p and $q = 1 - p$, respectively.

Since $m = 0 \cdot q + 2 \cdot p = 2p$, it follows that the population becomes extinct with probability 1 when $p \leq 1/2$. For $p > 1/2$ we use Theorem 7.3. The equation $t = g(t)$ then becomes

$$t = q + p \cdot t^2,$$

the solutions of which are $t_1 = 1$ and $t_2 = q/p < 1$. Thus $\eta = q/p$ in this case.

Example 7.6. A branching process starts with one individual, who reproduces according to the following principle:

# children	0	1	2
probability	$\frac{1}{6}$	$\frac{1}{2}$	$\frac{1}{3}$

The children reproduce according to the same rule, independently of each other, and so on.

(a) What is the probability of extinction?
(b) Determine the distribution of the number of grandchildren.

Solution. (a) We wish to apply Theorem 7.3. Since

$$m = \frac{1}{6} \cdot 0 + \frac{1}{2} \cdot 1 + \frac{1}{3} \cdot 2 = \frac{7}{6} > 1\,,$$

we solve the equation $t = g(t)$, that is,

$$t = \frac{1}{6} + \frac{1}{2}t + \frac{1}{3}t^2.$$

The roots are $t_1 = 1$ and $t_2 = 1/2$ (recall that $t = 1$ is always a solution). It follows that $\eta = 1/2$.
(b) According to Theorem 7.1, we have

$$g_2(t) = g\big(g(t)\big) = \frac{1}{6} + \frac{1}{2} \cdot \Big(\frac{1}{6} + \frac{1}{2}t + \frac{1}{3}t^2\Big) + \tfrac{1}{3} \cdot \Big(\frac{1}{6} + \frac{1}{2}t + \frac{1}{3}t^2\Big)^2 .$$

The distribution of $X(2)$ is obtained by simplifying the expression on the right-hand side, noting that $P(X(2) = k)$ is the coefficient of t^k. We omit the details. □

Remark 7.2. The distribution may, of course, also be found by combinatorial methods (try it and check that the results are the same!). □

Finally, let us solve the problems posed in Example 7.4.

Regard failures of the first kind as children and failures of the second kind as grandchildren. Thus, $X(1) \in \text{Po}(\lambda)$ and $X(2) = Y_1 + Y_2 + \cdots + Y_{X(1)}$, where $Y_1, Y_2, \ldots \in \text{Bin}(n, p)$ are independent and independent of $X(1)$. We wish to find the expected value and the variance of $X(1) + X(2)$. Note, however, a discrepancy from the usual model in that the failures of the second kind do not have the same distribution as $X(1)$.

Since $E\,X(1) = \lambda$ and $E\,X(2) = E\,X(1) \cdot E\,Y_1 = \lambda np$, we obtain

$$E\big(X(1) + X(2)\big) = \lambda + \lambda np.$$

The computation of the variance is a little more tricky, since $X(1)$ and $X(2)$ are not independent. But

$$\begin{aligned}X(1)+X(2) &= X(1)+Y_1+\cdots+Y_{X(1)}\\ &= (1+Y_1)+(1+Y_2)+\cdots+(1+Y_{X(1)})\\ &= \sum_{k=1}^{X(1)}(1+Y_k),\end{aligned}$$

and so

$$E\big(X(1)+X(2)\big) = E\,X(1)E(1+Y_1) = \lambda(1+np)$$

(as above) and

$$\begin{aligned}\operatorname{Var}\big(X(1)+X(2)\big) &= E\,X(1)\operatorname{Var}(1+Y_1)+\big(E(1+Y_1)\big)^2\operatorname{Var}X(1)\\ &= \lambda npq+(1+np)^2\lambda = \lambda\big(npq+(1+np)^2\big).\end{aligned}$$

The same device can be used to find the generating function. Namely,

$$g_{X(1)+X(2)}(t) = g_{X(1)}\big(g_{1+Y_1}(t)\big)\,,$$

which, together with the fact that

$$g_{1+Y_1}(t) = E\,t^{1+Y_1} = tE\,t^{Y_1} = tg_{Y_1}(t) = t(q+pt)^n\,,$$

yields

$$g_{X(1)+X(2)}(t) = e^{\lambda(t(q+pt)^n-1)}. \qquad \square$$

8 Problems

1. The nonnegative, integer-valued, random variable X has generating function $g_X(t) = \log\big(1/(1-qt)\big)$. Determine $P(X=k)$ for $k = 0,\ 1,\ 2,\ \ldots,$ $E\,X$, and $\operatorname{Var}X$.
2. The random variable X has the property that all moments are equal, that is, $E\,X^n = c$ for all $n \geq 1$, for some constant c. Find the distribution of X (no proof of uniqueness is required).
3. The random variable X has the property that

$$E\,X^n = \frac{2^n}{n+1}, \quad n = 1,\ 2, \ldots\,.$$

   Find some (in fact, the unique) distribution of X having these moments.
4. Suppose that Y is a random variable such that

$$E\,Y^k = \frac{1}{4}+2^{k-1}, \quad k = 1, 2, \ldots.$$

   Determine the distribution of Y.

5. Let $Y \in \beta(n, m)$ (n, m integers).
   (a) Compute the moment generating function of $-\log Y$.
   (b) Show that $-\log Y$ has the same distribution as $\sum_{k=1}^{m} X_k$, where X_1, $X_2 \cdots$ are independent, exponentially distributed random variables.

   Remark. The formula $\Gamma(r+s)/\Gamma(r) = (r+s-1)\cdots(r+1)r$, which holds when s is an integer, might be useful.
6. Show, by using moment generating functions, that if $X \in L(1)$, then $X \stackrel{d}{=} Y_1 - Y_2$, where Y_1 and Y_2 are independent, exponentially distributed random variables.
7. In the previous problem we found that a standard Laplace-distributed random variable has the same distribution as the difference between two standard exponential random variables. It is therefore reasonable to believe that if Y_1 and Y_2 are independent $L(1)$-distributed, then
$$Y_1 + Y_2 \stackrel{d}{=} X_1 - X_2,$$
   where X_1 and X_2 are independent $\Gamma(2, 1)$-distributed random variables. Prove, by checking moment generating functions, that this is in fact true.
8. Let $X \in \Gamma(p, a)$. Compute the (two-dimensional) moment generating function of $(X, \log X)$.
9. Let $X \in \text{Bin}(n, p)$. Compute $E\,X^4$ with the aid of the moment generating function.
10. Let X_1, X_2, ..., X_n be independent random variables with expectation 0 and finite third moments. Show, with the aid of characteristic functions, that
$$E(X_1 + X_2 + \cdots + X_n)^3 = E\,X_1^3 + E\,X_2^3 + \cdots + E\,X_n^3.$$
11. Let X and Y be independent random variables and suppose that Y is symmetric (around zero). Show that XY is symmetric.
12. The aim of the problem is to prove the double-angle formula
$$\sin 2t = 2 \sin t \cos t.$$
   Let X and Y be independent random variables, where $X \in U(-1, 1)$ and Y assumes the values $+1$ and -1 with probabilities $1/2$.
   (a) Show that $Z = X + Y \in U(-2, 2)$ by finding the distribution function of Z.
   (b) Translate this fact into a statement about the corresponding characteristic functions.
   (c) Rearrange.
13. Let X_1, X_2 ... be independent $C(0, 1)$-distributed random variables, and set $S_n = \sum_{k=1}^{n} X_k$, $n \geq 1$. Show that
   (a) $S_n/n \in C(0, 1)$,
   (b) $(1/n)\sum_{k=1}^{n} S_k/k \in C(0, 1)$.

   Remark. If $\{S_k/k,\ k \geq 1\}$ were independent, then (b) would follow immediately from (a).

14. For a positive, (absolutely) continuous random variable X we define the Laplace transform as

$$L_X(s) = E\, e^{-sX} = \int_0^\infty e^{-sx} f_X(x)\, dx, \quad s > 0.$$

Suppose that X is positive and stable with index $\alpha \in (0,1)$, which means that

$$L_X(s) = e^{-s^\alpha}, \quad s > 0.$$

Further, let $Y \in \text{Exp}(1)$ be independent of X. Show that

$$\left(\frac{Y}{X}\right)^\alpha \in \text{Exp}(1) \qquad \text{(which means that } \left(\frac{Y}{X}\right)^\alpha \stackrel{d}{=} Y\,).$$

15. Another transform: For a random variable X we define the *cumulant generating function*, $K_X(t) = \log \psi_X(t)$ as

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

where $k_n = k_n(X)$ is the so called nth *cumulant* or *semi-invariant* of X.
(a) Show that, if X and Y are independent random variables, then

$$k_n(X+Y) = k_n(X) + k_n(Y).$$

(b) Express k_1, k_2, and k_3 in terms of the moments $E\,X^k$, $k = 1,2,3$, of X.

16. Suppose that X_1, X_2, ... are independent, identically Linnik(α)-distributed random variables, that $N \in \text{Fs}(p)$, and that N and X_1, X_2, ... are independent. Show that $p^{1/\alpha}(X_1 + X_2 + \cdots + X_N)$ is, again, Linnik(α)-distributed.
Remark. The characteristic function of the Linnik(α)-distribution ($\alpha > 0$) is $\varphi(t) = (1 + |t|^\alpha)^{-1}$.

17. Suppose that the joint generating function of X and Y equals

$$g_{X,Y}(s,t) = E\, s^X t^Y = \exp\{\alpha(s-1) + \beta(t-1) + \gamma(st-1)\},$$

with $\alpha > 0$, $\beta > 0$, $\gamma \neq 0$.
(a) Show that X and Y both have a Poisson distribution, but that $X + Y$ does not.
(b) Are X and Y independent?

18. Let the random variables Y, X_1, X_2, ... be independent, suppose that $Y \in \text{Fs}(p)$, where $0 < p < 1$, and suppose that X_1, X_2, X_3, ... are all $\text{Exp}(1/a)$-distributed. Find the distribution of

$$Z = \sum_{j=1}^Y X_j.$$

19. Let $X_1, X_2, \ldots$ be Ge(α)-distributed random variables, let $N \in \text{Fs}(p)$, suppose that all random variables are independent, and set
$$Y = X_1 + X_2 + \cdots + X_N.$$
(a) Show that $Y \in \text{Ge}(\beta)$, and determine β.
(b) Compute $E\,Y$ and $\operatorname{Var} Y$ with "the usual formulas", and check that the results agree with mean and variance of the distribution in (a).
20. Let $0 < p = 1 - q < 1$. Suppose that $X_1, X_2, \ldots$ are independent Ge(q)-distributed random variables and that $N \in \text{Ge}(p)$ is independent of $X_1, X_2, \ldots$.
(a) Find the distribution of $Z = X_1 + X_2 + \cdots + X_N$.
(b) Show that $Z \mid Z > 0 \in \text{Fs}(\alpha)$, and determine α.
21. Suppose that $X_1, X_2, \ldots$ are independent $L(a)$-distributed random variables, let $N_p \in \text{Fs}(p)$ be independent of $X_1, X_2, \ldots$, and set $Y_p = \sum_{k=1}^{N_p} X_k$. Show that
$$\sqrt{p}Y_p \in L(a)\,.$$
22. Let $N, X_1, X_2, \ldots$ be independent random variables such that $N \in \text{Po}(1)$ and $X_k \in \text{Po}(2)$ for all k. Set $Z = \sum_{k=1}^{N} X_k$ (and $Z = 0$ when $N = 0$). Compute $E\,Z$, $\operatorname{Var} Z$, and $P(Z = 0)$.
23. Let $Y_1, Y_2, \ldots$ be i.i.d. random variables, and let N be a nonnegative, integer-valued random variable that is independent of $Y_1, Y_2, \ldots$. Compute $\operatorname{Cov}\left(\sum_{k=1}^{N} Y_k, N\right)$.
24. Let, for $m \neq 1$, $X_1, X_2, \ldots$ be independent random variables with $E\,X_n = m^n$, $n \geq 1$, let $N \in \text{Po}(\lambda)$ be independent of $X_1, X_2, \ldots$, and set
$$Z = X_1 + X_2 + \cdots + X_N.$$
Determine $E\,Z$.
Remark. Note that $X_1, X_2, \ldots$ are NOT identically distributed, that is, the usual "$E\,S_N = E\,N \cdot E\,X$" does NOT work; you have to modify the proof of that formula.
25. Let $N \in \text{Bin}(n, 1 - e^{-m})$, and let $X_1, X_2, \ldots$ have the same 0-truncated Poisson distribution, namely,
$$P(X_1 = x) = \frac{m^x/x!}{e^m - 1}, \quad x = 1, 2, 3, \ldots.$$
Further, assume that $N, X_1, X_2, \ldots$ are independent,
(a) Find the distribution of $Y = \sum_{k=1}^{N} X_k$ ($Y = 0$ when $N = 0$).
(b) Compute $E\,Y$ and $\operatorname{Var} Y$ without using (a).
26. The number of cars passing a road crossing during an hour is Po(b)-distributed. The number of passengers in each car is Po(p)-distributed. Find the generating function of the total number of passengers, Y, passing the road crossing during one hour, and compute $E\,Y$ and $\operatorname{Var} Y$.

27. A miner has been trapped in a mine with three doors. One takes him to freedom after one hour, one brings him back to the mine after 3 hours and the third one brings him back after 5 hours. Suppose that he wishes to get out of the mine and that he does so by choosing one of the three doors uniformly at random and continues to do so until he is free. Find the generating function, the mean and the variance for the time it takes him to reach freedom.
28. Lisa shoots at a target. The probability of a hit in each shot is 1/2. Given a hit, the probability of a bull's-eye is p. She shoots until she misses the target. Let X be the total number of bull's-eyes Lisa has obtained when she has finished shooting; find its distribution.
29. Karin has an unfair coin; the probability of heads is p $(0 < p < 1)$. She tosses the coin until she obtains heads. She then tosses a fair coin as many times as she tossed the unfair one. For every head she has obtained with the fair coin she finally throws a symmetric die. Determine the expected number and variance of the total number of dots Karin obtains by this procedure.
30. Philip throws a fair die until he obtains a four. Diane then tosses a coin as many times as Philip threw his die. Determine the expected value and variance of the number of
    (a) heads,
    (b) tails, and
    (c) heads and tails obtained by Diane.
31. Let p be the probability that the tip points downward after a person throws a drawing pin once. Miriam throws a drawing pin until it points downward for the first time. Let X be the number of throws for this to happen. She then throws the drawing pin another X times. Let Y be the number of times the drawing pin points downward in the latter series of throws. Find the distribution of Y (cf. Problem 2.6.38).
32. Let $X_1, X_2, \ldots$ be independent observations of a random variable X, whose density function is

$$f_X(x) = \tfrac{1}{2}e^{-|x|}, \quad -\infty < x < \infty.$$

Suppose we continue sampling until a negative observation appears. Let Y be the sum of the observations thus obtained (including the negative one). Show that the density function of Y is

$$f_Y(x) = \begin{cases} \frac{2}{3}e^{x}, & \text{for} \quad x < 0, \\ \frac{1}{6}e^{-x/2}, & \text{for} \quad x > 0. \end{cases}$$

33. At a certain black spot, the number of traffic accidents per year follows a Po(10, 000)-distribution. The number of deaths per accident follows a Po(0.1)-distribution, and the number of casualties per accidents follows a Po(2)-distribution. The correlation coefficient between the number of

casualties and the number of deaths per accidents is 0.5. Compute the expectation and variance of the total number of deaths and casualties during a year.

34. Suppose that X is a nonnegative, integer-valued random variable, and let n and m be nonnegative integers. Show that

$$g_{nX+m}(t) = t^m \cdot g_X(t^n).$$

35. Suppose that the offspring distribution in a branching process is the Ge(p)-distribution, and let $X(n)$ be the number of individuals in generation n, $n = 0, 1, 2, \ldots$.
(a) What is the probability of extinction?
(b) Find the probability that the population is extinct in the second generation.
36. Consider a branching process whose offspring distribution is Bin(n, p)-distributed. Compute the expected value, the variance and the probability that there are 0 or 1 grandchild, that is, find, in the usual notation, $E\,X(2)$, $\operatorname{Var} X(2)$, $P(X(2) = 0)$, and $P(X(2) = 1)$.
37. Consider a branching process where the individuals reproduce according to the following pattern:

# of children	0	1	2
probability	$\frac{1}{6}$	$\frac{1}{3}$	$\frac{1}{2}$

Individuals reproduce independently of each other and independently of the number of their sisters and brothers. Determine
(a) the probability that the population becomes extinct;
(b) the probability that the population has become extinct in the second generation;
(c) the expected number of children given that there are no grandchildren.
38. One bacterium each of the two dangerous Alphomylia and Klaipeda tribes have escaped from a laboratory. They reproduce according to a standard branching process as follows:

# of children	0	1	2
probability Alphomylia	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{2}$
probability Klaipeda	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{2}{3}$

The two cultures reproduce independently of each other. Determine the probability that 0, 1, and 2 of the cultures, respectively, become extinct.
39. Suppose that the offspring distribution in a branching process is the Ge(p)-distribution, and let $X(n)$ be the number of individuals in generation n, $n = 0, 1, 2, \ldots$.
(a) What is the probability of extinction?
Now suppose that $p = 1/2$, and set $g_n(t) = g_{X(n)}(t)$.

(b) Show that

$$g_n(t) = \frac{n-(n-1)t}{n+1-nt}, \quad n = 1, 2, \ldots .$$

(c) Show that

$$P(X(n) = k) = \begin{cases} \dfrac{n}{n+1}, & \text{for} \quad k = 0, \\ \dfrac{n^{k-1}}{(n+1)^{k+1}}, & \text{for} \quad k = 1, 2, \ldots . \end{cases}$$

(d) Show that

$$P(X(n) = k \mid X(n) > 0) = \frac{1}{n+1}\Big(\frac{n}{n+1}\Big)^{k-1}, \quad \text{for} \quad k = 1, 2, \ldots,$$

that is, show that the number of individuals in generation n, given that the population is not yet extinct, follows an $\mathrm{Fs}(1/(n+1))$-distribution. Finally, suppose that the population becomes extinct at generation number N.

(e) Show that

$$P(N = n) = g_{n-1}(\tfrac{1}{2}) - g_{n-1}(0), \quad n = 1, 2, \ldots .$$

(f) Show that $P(N = n) = 1/(n(n+1))$, $\quad n = 1, 2, \ldots$ (and hence that $P(N < \infty) = 1$, i.e., $\eta = 1$).

(g) Compute $E\,N$. Why is this a reasonable answer?

40. The growth dynamics of pollen cells can be modeled by binary splitting as follows: After one unit of time, a cell either splits into two or dies. The new cells develop according to the same law independently of each other. The probabilities of dying and splitting are 0.46 and 0.54, respectively.
 (a) Determine the maximal initial size of the population in order for the probability of extinction to be at least 0.3.
 (b) What is the probability that the population is extinct after two generations if the initial population is the maximal number obtained in (a)?

41. Consider binary splitting, that is, the branching process where the distribution of Y = the number of children is given by

$$P(Y = 2) = 1 - P(Y = 0) = p, \quad 0 < p < 1.$$

However, suppose that p is not known, that p is random, viz., consider the following setup: Assume that

$$P(Y = 2 \mid P = p) = p, \quad P(Y = 0 \mid P = p) = 1 - p, \quad \text{with}$$

$$f_P(x) = \begin{cases} 2x, & \text{for} \quad 0 < x < 1, \\ 0, & \text{otherwise.} \end{cases}$$

 (a) Find the distribution of Y.
 (b) Determine the probability of extinction.

42. Consider the following modification of a branching process: A mature individual produces children according to the generating function $g(t)$. However, an individual becomes mature with probability α and dies before maturity with probability $1-\alpha$. Throughout $X(0)=1$, that is, we start with one immature individual.
    (a) Find the generating function of the number of individuals in the first two generations.
    (b) Suppose that the offspring distribution is geometric with parameter p. Determine the extinction probability.
43. Let $\{X(n),\, n \geq 0\}$ be the usual Galton–Watson process, starting with $X(0)=1$. Suppose, in addition, that immigration is allowed in the sense that in addition to the children born in generation n there are Z_n individuals immigrating, where $\{Z_n,\, n \geq 1\}$ are i.i.d. random variables with the same distribution as $X(1)$.
    (a) What is the expected number of individuals in generation 1?
    (b) Find the generating function of the number of individuals in generations 1 and 2, respectively.
    (c) Determine/express the probability that the population is extinct after two generations.

    Remark. It may be helpful to let p_0 denote the probability that an individual does not have any children (which, in particular, means that $P(X(1)=0)=p_0$).
44. Consider a branching process with reproduction mean $m<1$. Suppose also, as before, that $X(0)=1$.
    (a) What is the probability of extinction?
    (b) Determine the expected value of the total progeny.
    (c) Now suppose that $X(0)=k$, where k is an integer ≥ 2. What are the answers to the questions in (a) and (b) now?
45. The following model can be used to describe the number of women (mothers and daughters) in a given area. The number of mothers is a random variable $X \in \text{Po}(\lambda)$. Independently of the others, every mother gives birth to a $\text{Po}(\mu)$-distributed number of daughters. Let Y be the total number of daughters and hence $Z = X+Y$ be the total number of women in the area.
    (a) Find the generating function of Z.
    (b) Compute $E\,Z$ and $\operatorname{Var} Z$.
46. Let $X(n)$ be the number of individuals in the nth generation of a branching process ($X(0)=1$), and set $T_n = 1+X(1)+\cdots+X(n)$, that is, T_n equals the total progeny up to and including generation number n. Let $g(t)$ and $G_n(t)$ be the generating functions of $X(1)$ and T_n, respectively. Prove the following formula:
$$G_n(t) = t \cdot g\bigl(G_{n-1}(t)\bigr).$$

47. Consider a branching process with a Po(m)-distributed offspring. Let $X(1)$ and $X(2)$ be the number of individuals in generations 1 and 2, respectively. Determine the generating function of
    (a) $X(1)$,
    (b) $X(2)$,
    (c) $X(1) + X(2)$,
    (d) Determine $\mathrm{Cov}(X(1), X(2))$.
48. Let X be the number of coin tosses until heads is obtained. Suppose that the probability of heads is unknown in the sense that we consider it to be a random variable $Y \in U(0, 1)$. Find the distribution of X (cf. Problem 2.6.37).