1

Generating functions

The shortest path between two truths in the real domain passes through the complex domain. (Jacques Hadamard)

There are various integrals of the form

$$\int_{-\infty}^{\infty} g(t, x)\, \mathrm{d}F_X(x) = \mathbb{E}[g(t, X)] \tag{1.1}$$

which are often of great value for studying r.v.s. For example, taking $g(n, x) = x^n$ and $g(n, x) = |x|^n$, for $n \in \mathbb{N}$, give the algebraic and absolute moments, respectively, while $g(n, x) = x_{[n]} = x(x-1)\cdots(x-n+1)$ yields the factorial moments of X, which are of use for lattice r.v.s. Also important (if not essential) for working with lattice distributions with nonnegative support is the *probability generating function*, obtained by taking $g(t, x) = t^x$ in (1.1), i.e., $\mathbb{P}_X(t) := \sum_{x=0}^{\infty} t^x p_x$, where $p_x = \Pr(X = x)$.[1]

For our purposes, we will concentrate on the use of the two forms $g(t, x) = \exp(tx)$ and $g(t, x) = \exp(itx)$, which are not only applicable to both discrete and continuous r.v.s, but also, as we shall see, of enormous theoretical and practical use.

1.1 The moment generating function

The *moment generating function* (m.g.f.), of random variable X is the function $\mathbb{M}_X$: $\mathbb{R} \mapsto \mathbb{X}_{\geq 0}$ (where $\mathbb{X}$ is the extended real line) given by $t \mapsto \mathbb{E}\left[\mathrm{e}^{tX}\right]$. The m.g.f. $\mathbb{M}_X$ is said to exist if it is finite on a neighbourhood of zero, i.e., if there is an $h > 0$ such that, $\forall t \in (-h, h)$, $\mathbb{M}_X(t) < \infty$. If $\mathbb{M}_X$ exists, then the largest (open) interval $\mathcal{I}$

[1] Probability generating functions arise ubiquitously in the study of stochastic processes (often the 'next course' after an introduction to probability such as this). There are numerous books, at various levels, on stochastic processes; three highly recommended 'entry-level' accounts which make generous use of probability generating functions are Kao (1996), Jones and Smith (2001), and Stirzaker (2003). See also Wilf (1994) for a general account of generating functions.

Intermediate Probability: A Computational Approach M. Paolella

around zero such that $\mathbb{M}_X(t) < \infty$ for $t \in \mathcal{I}$ is referred to as the *convergence strip (of the m.g.f.) of* X.

1.1.1 Moments and the m.g.f.

A fundamental result is that, if $\mathbb{M}_X$ exists, then all positive moments of X exist. This is worth emphasizing:

$$\boxed{\text{If } \mathbb{M}_X \text{ exists, then } \forall r \in \mathbb{R}_{>0},\ \mathbb{E}\left[|X|^r\right] < \infty.} \tag{1.2}$$

To prove (1.2), let r be an arbitrary positive (real) number, and recall that $\lim_{x\to\infty} x^r/\mathrm{e}^x = 0$, as shown in (I.7.3) and (I.A.36). This implies that, $\forall\, t \in \mathbb{R} \setminus 0$, $\lim_{x\to\infty} x^r/\mathrm{e}^{|tx|} = 0$. Choose an $h > 0$ such that $(-h, h)$ is in the convergence strip of X, and a value t such that $0 < t < h$ (so that $\mathbb{E}\left[\mathrm{e}^{tX}\right]$ and $\mathbb{E}\left[\mathrm{e}^{-tX}\right]$ are finite). Then there must exist an x_0 such that $|x|^r < \mathrm{e}^{|tx|}$ for $|x| > x_0$. For $|x| \le x_0$, there exists a finite constant K_0 such that $|x|^r < K_0\mathrm{e}^{|tx|}$. Thus, there exists a K such that $|x|^r < K\mathrm{e}^{|tx|}$ for all x, so that, from the inequality-preserving nature of expectation (see Section I.4.4.2), $\mathbb{E}\left[|X|^r\right] \le K\mathbb{E}\left[\mathrm{e}^{|tX|}\right]$. Finally, from the trivial identity $\mathrm{e}^{|tx|} \le \mathrm{e}^{tx} + \mathrm{e}^{-tx}$ and the linearity of the expectation operator, $\mathbb{E}\left[\mathrm{e}^{|tX|}\right] \le \mathbb{E}\left[\mathrm{e}^{tX}\right] + \mathbb{E}\left[\mathrm{e}^{-tX}\right] < \infty$, showing that $\mathbb{E}\left[|X|^r\right]$ is finite.

***Remark*:** This previous argument also shows that, if the m.g.f. of X is finite on the interval $(-h, h)$ for some $h > 0$, then so is the m.g.f. of r.v. $|X|$ on the same neighbourhood. Let $|t| < h$, so that $\mathbb{E}\left[\mathrm{e}^{t|X|}\right]$ is finite, and let $k \in \mathbb{N} \cup 0$. From the Taylor series of e^x, it follows that $0 \le |tX|^k/k! \le \mathrm{e}^{|tX|}$, implying $\mathbb{E}\left[|tX|^k\right] \le k!\,\mathbb{E}\left[\mathrm{e}^{|tX|}\right] < \infty$. Moreover, for all $N \in \mathbb{N}$,

$$S(N) := \sum_{k=0}^{N} \left| \frac{\mathbb{E}\left[|tX|^k\right]}{k!} \right| = \sum_{k=0}^{N} \frac{\mathbb{E}\left[|tX|^k\right]}{k!} = \mathbb{E}\left(\sum_{k=0}^{N} \frac{|tX|^k}{k!} \right) \le \mathbb{E}\left[\mathrm{e}^{|tX|}\right],$$

so that

$$\lim_{N\to\infty} S(N) = \sum_{k=0}^{\infty} \frac{\mathbb{E}\left[|tX|^k\right]}{k!} \le \mathbb{E}\left[\mathrm{e}^{|tX|}\right]$$

and the infinite series converges absolutely. Now, as $|\mathbb{E}\left[(tX)^k\right]| \le \mathbb{E}\left[|tX|^k\right] < \infty$, it follows that the series $\sum_{k=0}^{\infty} \mathbb{E}\left[(tX)^k\right]/k!$ also converges. As $\sum_{k=0}^{\infty} (tX)^k/k!$ converges pointwise to e^{tX}, and $|\mathrm{e}^{tX}| \le \mathrm{e}^{|tX|}$, the dominated convergence theorem applied to the integral of the expectation operator implies

$$\lim_{N\to\infty} \mathbb{E}\left[\sum_{k=0}^{N} \frac{(tX)^k}{k!} \right] = \mathbb{E}\left[\mathrm{e}^{tX}\right].$$

That is,

$$\mathbb{M}_X(t) = \mathbb{E}\left[e^{tX}\right] = \mathbb{E}\left[\sum_{k=0}^{\infty} \frac{(tX)^k}{k!}\right] = \sum_{k=0}^{\infty} \frac{t^k}{k!} \mathbb{E}\left[X^k\right], \tag{1.3}$$

which is important for the next result. ■

It can be shown that termwise differentiation of (1.3) is valid, so that the jth derivative with respect to t is

$$\begin{aligned}
\mathbb{M}_X^{(j)}(t) &= \sum_{i=j}^{\infty} \frac{t^{i-j}}{(i-j)!} \mathbb{E}\left[X^i\right] = \sum_{n=0}^{\infty} \frac{t^n}{n!} \mathbb{E}\left[X^{n+j}\right] \\
&= \mathbb{E}\left[\sum_{n=0}^{\infty} \frac{(tX)^n X^j}{n!}\right] = \mathbb{E}\left[X^j \sum_{n=0}^{\infty} \frac{(tX)^n}{n!}\right] = \mathbb{E}\left[X^j e^{tX}\right],
\end{aligned} \tag{1.4}$$

or

$$\boxed{\mathbb{M}_X^{(j)}(t)\Big|_{t=0} = \mathbb{E}\left[X^j\right].}$$

Similarly, it can be shown that we are justified in arriving at (1.4) by simply writing

$$\mathbb{M}_X^{(j)}(t) = \frac{d^j}{dt^j} \mathbb{E}\left[e^{tX}\right] = \mathbb{E}\left[\frac{d^j}{dt^j} e^{tX}\right] = \mathbb{E}\left[X^j e^{tX}\right].$$

In general, if $\mathbb{M}_Z(t)$ is the m.g.f. of r.v. Z and $X = \mu + \sigma Z$, then it is easy to show that

$$\mathbb{M}_X(t) = \mathbb{E}\left[e^{tX}\right] = \mathbb{E}\left[e^{t(\mu+\sigma Z)}\right] = e^{t\mu} \mathbb{M}_Z(t\sigma). \tag{1.5}$$

The next two examples illustrates the computation of the m.g.f. in a discrete and continuous case, respectively.

⊖ ***Example 1.1*** Let $X \sim \text{DUnif}(\theta)$ with p.m.f. $f_X(x;\theta) = \theta^{-1}\mathbb{I}_{\{1,2,\ldots,\theta\}}(x)$. The m.g.f. of X is

$$\mathbb{M}_X(t) = \mathbb{E}\left[e^{tX}\right] = \frac{1}{\theta}\sum_{j=1}^{\theta} e^{tj},$$

so that

$$\mathbb{M}_X'(t) = \frac{1}{\theta}\sum_{j=1}^{\theta} j e^{tj}, \qquad \mathbb{E}[X] = \mathbb{M}_X'(0) = \frac{1}{\theta}\sum_{j=1}^{\theta} j = \frac{\theta+1}{2},$$

and

$$\mathbb{M}_X''(t) = \frac{1}{\theta}\sum_{j=1}^{\theta} j^2 e^{tj}, \qquad \mathbb{E}\left[X^2\right] = \mathbb{M}_X''(0) = \frac{1}{\theta}\sum_{j=1}^{\theta} j^2 = \frac{(\theta+1)(2\theta+1)}{6},$$

from which it follows that

$$\mathbb{V}(X) = \mu_2' - \mu^2 = \frac{(\theta+1)(2\theta+1)}{6} - \left(\frac{\theta+1}{2}\right)^2 = \frac{(\theta-1)(\theta+1)}{12},$$

recalling (I.4.40). More generally, letting $X \sim \text{DUnif}(\theta_1, \theta_2)$ with p.d.f. $f_X(x; \theta_1, \theta_2) = (\theta_2 - \theta_1 + 1)^{-1} \mathbb{I}_{\{\theta_1, \theta_1+1, \ldots, \theta_2\}}(x)$,

$$\mathbb{E}[X] = \frac{1}{2}(\theta_1 + \theta_2) \quad \text{and} \quad \mathbb{V}(X) = \frac{1}{12}(\theta_2 - \theta_1)(\theta_2 - \theta_1 + 2),$$

which can be shown directly using the m.g.f., or by simple symmetry arguments. ■

⊖ ***Example 1.2*** Let $U \sim \text{Unif}(0, 1)$. Then,

$$\mathbb{M}_U(t) = \mathbb{E}\left[e^{tU}\right] = \int_0^1 e^{tu}\,du = \frac{e^t - 1}{t}, \quad t \neq 0,$$

which is finite in any neighbourhood of zero, and continuous at zero, as, via l'Hôpital's rule,

$$\lim_{t\to 0} \frac{e^t - 1}{t} = \lim_{t\to 0} \frac{e^t}{1} = 1 = \int_0^1 e^{0u}\,du = \mathbb{M}_U(0).$$

The Taylor series expansion of $\mathbb{M}_U(t)$ around zero is

$$\frac{e^t - 1}{t} = \frac{1}{t}\left(t + \frac{t^2}{2} + \frac{t^3}{6} + \cdots\right) = 1 + \frac{t}{2} + \frac{t^2}{6} + \cdots = \sum_{j=0}^{\infty} \frac{1}{j+1}\frac{t^j}{j!}$$

so that, from (1.3),

$$\mathbb{E}\left[U^r\right] = (r+1)^{-1}, \quad r = 1, 2, \ldots. \tag{1.6}$$

In particular,

$$\mathbb{E}[U] = \frac{1}{2}, \quad \mathbb{E}\left[U^2\right] = \frac{1}{3}, \quad \mathbb{V}(U) = \frac{1}{3} - \frac{1}{4} = \frac{1}{12}.$$

Of course, (1.6) could have been derived with much less work and in more generality, as

$$\mathbb{E}\left[U^r\right] = \int_0^1 u^r\,du = (r+1)^{-1}, \qquad r \in \mathbb{R}_{>0}.$$

For $X \sim \mathrm{Unif}(a,b)$, write $X = U(b-a)+a$ so that, from the binomial theorem and (1.6),

$$\mathbb{E}\left[X^r\right] = \sum_{j=0}^{r} \binom{r}{j} a^{r-j} (b-a)^j \frac{1}{j+1} = \frac{b^{r+1}-a^{r+1}}{(b-a)(r+1)}, \tag{1.7}$$

where the last equality is given in (I.1.57). Alternatively, we can use the location–scale relationship (1.5) with $\mu = a$ and $\sigma = b-a$ to get

$$\mathbb{M}_X(t) = \frac{\mathrm{e}^{tb}-\mathrm{e}^{ta}}{t(b-a)}, \quad t \neq 0, \qquad \mathbb{M}_X(0) = 1.$$

Then, with $j = i-1$ and $t \neq 0$,

$$\begin{aligned} \mathbb{M}_X(t) &= \frac{1}{t(b-a)}\left(\sum_{i=0}^{\infty}\frac{(tb)^i}{i!} - \sum_{k=0}^{\infty}\frac{(ta)^k}{k!}\right) = \sum_{i=1}^{\infty}\frac{b^i-a^i}{i!\,(b-a)}t^{i-1} \\ &= \sum_{j=0}^{\infty}\frac{b^{j+1}-a^{j+1}}{(j+1)!\,(b-a)}t^j = \sum_{j=0}^{\infty}\frac{b^{j+1}-a^{j+1}}{(j+1)(b-a)}\frac{t^j}{j!}, \end{aligned}$$

which, from (1.3), yields the result in (1.7). ■

1.1.2 The cumulant generating function

The *cumulant generating function* (c.g.f.), is defined as

$$\boxed{\mathbb{K}_X(t) = \log \mathbb{M}_X(t)\,.} \tag{1.8}$$

The terms κ_i in the series expansion $\mathbb{K}_X(t) = \sum_{r=0}^{\infty} \kappa_r t^r / r!$ are referred to as the *cumulants* of X, so that the ith derivative of $\mathbb{K}_X(t)$ evaluated at $t=0$ is κ_i, i.e.,

$$\boxed{\kappa_i = \mathbb{K}_X^{(i)}(t)\Big|_{t=0}\,.}$$

It is straightforward to show that

$$\boxed{\kappa_1 = \mu, \qquad \kappa_2 = \mu_2, \qquad \kappa_3 = \mu_3, \qquad \kappa_4 = \mu_4 - 3\mu_2^2} \tag{1.9}$$

(see Problem 1.1), with higher-order terms given in Stuart and Ord (1994, Section 3.14).

⊛ ***Example 1.3*** From Problem I.7.17, the m.g.f. of $X \sim \mathrm{N}\left(\mu, \sigma^2\right)$ is given by

$$\mathbb{M}_X(t) = \exp\left\{\mu t + \frac{1}{2}\sigma^2 t^2\right\}, \quad \mathbb{K}_X(t) = \mu t + \frac{1}{2}\sigma^2 t^2. \tag{1.10}$$

Thus,

$$\mathbb{K}'_X(t) = \mu + \sigma^2 t, \quad \mathbb{E}[X] = \mathbb{K}'_X(0) = \mu, \quad \mathbb{K}''_X(t) = \sigma^2, \quad \mathbb{V}(X) = \mathbb{K}''_X(0) = \sigma^2,$$

and $\mathbb{K}_X^{(i)}(t) = 0,\ i \geq 3$, so that $\mu_3 = 0$ and $\mu_4 = \kappa_4 + 3\mu_2^2 = 3\sigma^4$, as also determined directly in Example I.7.3. This also shows that X has skewness $\mu_3/\mu_2^{3/2} = 0$ and kurtosis $\mu_4/\mu_2^2 = 3$. ■

◎ ***Example 1.4*** For $X \sim \text{Poi}(\lambda)$,

$$\mathbb{M}_X(t) = \mathbb{E}\left[e^{tX}\right] = \sum_{x=0}^{\infty} \frac{e^{tx} e^{-\lambda} \lambda^x}{x!} = e^{-\lambda} \sum_{x=0}^{\infty} \frac{\left(\lambda e^t\right)^x}{x!}$$
$$= \exp\left(-\lambda + \lambda e^t\right). \tag{1.11}$$

As $\mathbb{K}_X^{(r)}(t) = \lambda e^t$ for $r \geq 1$, it follows that $\mathbb{E}[X] = \mathbb{K}'_X(t)\big|_{t=0} = \lambda$ and $\mathbb{V}(X) = \mathbb{K}''_X(t)\big|_{t=0} = \lambda$. This calculation should be compared with that in (I.4.34). Once the m.g.f. is available, higher moments are easily obtained, in particular,

$$\text{skew}(X) = \mu_3/\mu_2^{3/2} = \lambda/\lambda^{3/2} = \lambda^{-1/2} \to 0$$

and

$$\text{kurt}(X) = \mu_4/\mu_2^2 = \left(\kappa_4 + 3\mu_2^2\right)/\mu_2^2 = \left(\lambda + 3\lambda^2\right)/\lambda^2 \to 3,$$

as $\lambda \to \infty$. That is, as λ increases, the skewness and kurtosis of a Poisson random variable tend towards the skewness and kurtosis of a normal random variable. ■

◎ ***Example 1.5*** For $X \sim \text{Gam}(a, b)$, the m.g.f. is, with $y = x(b-t)$,

$$\begin{aligned}
\mathbb{M}_X(t) &= \mathbb{E}\left[e^{tX}\right] \\
&= \frac{b^a}{\Gamma(a)} \int_0^\infty x^{a-1} e^{-x(b-t)}\, dx = (b-t)^{-a} b^a \int_0^\infty \frac{1}{\Gamma(a)} y^{a-1} e^{-y}\, dy \\
&= \left(\frac{b}{b-t}\right)^a, \qquad t < b.
\end{aligned}$$

From this,

$$\mathbb{E}[X] = \left.\frac{d\mathbb{M}_X(t)}{dt}\right|_{t=0} = a\left(\frac{b}{b-t}\right)^{a-1} b\,(b-t)^{-2}\Big|_{t=0} = \frac{a}{b}$$

or, more easily, with $\mathbb{K}_X(t) = a(\ln b - \ln(b-t))$, (1.9) implies

$$\kappa_1 = \mathbb{E}[X] = \left.\frac{d\mathbb{K}_X(t)}{dt}\right|_{t=0} = \left.\frac{a}{b-t}\right|_{t=0} = \frac{a}{b} \tag{1.12}$$

and

$$\kappa_2 = \mu_2 = \mathbb{V}(X) = \left.\frac{\mathrm{d}\mathbb{K}_X^2(t)}{\mathrm{d}t^2}\right|_{t=0} = \left.\frac{a}{(b-t)^2}\right|_{t=0} = \frac{a}{b^2}.$$

Similarly,

$$\mu_3 = \frac{2a}{b^3} \quad \text{and} \quad \kappa_4 = \frac{6a}{b^4},$$

i.e., $\mu_4 = \kappa_4 + 3\mu_2^2 = 3a\,(2+a)\,/b^4$, so that the skewness and kurtosis are

$$\frac{\mu_3}{\mu_2^{3/2}} = \frac{2a/b^3}{\left(a/b^2\right)^{3/2}} = \frac{2}{\sqrt{a}} \quad \text{and} \quad \frac{\mu_4}{\mu_2^2} = \frac{3a\,(2+a)\,/b^4}{\left(a/b^2\right)^2} = \frac{3\,(2+a)}{a}. \tag{1.13}$$

These converge to 0 and 3, respectively, as a increases. ■

⊖ ***Example 1.6*** From density (I.7.51), the m.g.f. of a location-zero, scale-one logistic random variable is (with $y = \left(1+\mathrm{e}^{-x}\right)^{-1}$), for $|t| < 1$,

$$\begin{aligned}
\mathbb{M}_X(t) = \mathbb{E}\left[\mathrm{e}^{tX}\right] &= \int_{-\infty}^{\infty} \left(\mathrm{e}^{-x}\right)^{1-t} \left(1+\mathrm{e}^{-x}\right)^{-2} \mathrm{d}x \\
&= \int_0^1 \left(\frac{1-y}{y}\right)^{1-t} y^2 y^{-1} (1-y)^{-1}\, \mathrm{d}y = \int_0^1 (1-y)^{-t}\, y^t\, \mathrm{d}y \\
&= B\,(1-t, 1+t) = \Gamma\,(1-t)\,\Gamma\,(1+t)\,.
\end{aligned}$$

If, in addition, $t \neq 0$, the m.g.f. can also be expressed as

$$\mathbb{M}_X(t) = t\Gamma\,(t)\,\Gamma\,(1-t) = t\frac{\pi}{\sin \pi t}, \tag{1.14}$$

where the second identity is *Euler's reflection formula*.[2] ■

For certain problems, the m.g.f. can be expressed recursively, as the next example shows.

⊖ ***Example 1.7*** Let $N_m \sim \mathrm{Consec}(m, p)$, i.e., N_m is the random number of Bernoulli trials, each with success probability p, which need to be conducted until m successes in a row occur. The mean of N_m was computed in Example I.8.13 and the variance

[2] Andrews, Askey and Roy (1999, pp. 9–10) provide four different methods for proving Euler's reflection formula; see also Jones (2001, pp. 217–18), Havil (2003, p. 59), or Schiff (1999, p. 174). As an aside, from (1.14) with $t = 1/2$, it follows that $\Gamma\,(1/2) = \sqrt{\pi}$.

and m.g.f. in Problem I.8.13. In particular, from (I.8.52), with $\mathbb{M}_m(t) := \mathbb{M}_{N_m}(t)$ and $q = 1 - p$,

$$\mathbb{M}_m(t) = \frac{p\mathrm{e}^t \mathbb{M}_{m-1}(t)}{1 - q\mathbb{M}_{m-1}(t)\,\mathrm{e}^t}. \tag{1.15}$$

This can be recursively evaluated with $\mathbb{M}_1(t) = p\mathrm{e}^t / \left(1 - q\mathrm{e}^t\right)$ for $t \neq -\ln(1-p)$, from the geometric distribution. Example 1.20 below illustrates how to use (1.15) to obtain the p.m.f. Problem 1.10 uses (1.15) to compute $\mathbb{E}\left[N_m\right]$. ■

Calculation of the m.g.f. can also be useful for determining the expected value of particular functions of random variables, as illustrated next.

⊖ ***Example 1.8*** To determine $\mathbb{E}[\ln X]$ when $X \sim \chi^2_v$, we could try to directly integrate, i.e.,

$$\mathbb{E}[\ln X] = \frac{1}{2^{v/2}\Gamma(v/2)} \int_0^\infty (\ln x)\, x^{v/2-1} \mathrm{e}^{-x/2}\, \mathrm{d}x, \tag{1.16}$$

but this seems to lead nowhere. Note instead that the m.g.f. of $Z = \ln X$ is

$$\mathbb{M}_Z(t) = \mathbb{E}\left[\mathrm{e}^{tZ}\right] = \mathbb{E}\left[X^t\right] = \frac{1}{2^{v/2}\Gamma(v/2)} \int_0^\infty x^{t+v/2-1} \mathrm{e}^{-x/2}\, \mathrm{d}x$$

or, with $y = x/2$,

$$\mathbb{M}_Z(t) = \frac{2^{t+v/2-1+1}}{2^{v/2}\Gamma(v/2)} \int_0^\infty y^{t+v/2-1} \mathrm{e}^{-y}\, \mathrm{d}y = 2^t \frac{\Gamma(t+v/2)}{\Gamma(v/2)}.$$

Then, with $\mathrm{d}2^t/\mathrm{d}t = 2^t \ln 2$,

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathbb{M}_Z(t) = \frac{1}{\Gamma(v/2)} \left(2^t\, \Gamma'(t+v/2) + 2^t\, \ln 2\, \Gamma(t+v/2)\right)$$

and

$$\mathbb{E}[\ln X] = \left.\frac{\mathrm{d}}{\mathrm{d}t}\mathbb{M}_Z(t)\right|_{t=0} = \frac{\Gamma'(v/2)}{\Gamma(v/2)} + \ln 2 = \psi(v/2) + \ln 2.$$

Having seen the answer, the integral (1.16) is easy; differentiating $\Gamma(\nu/2)$ with respect to $\nu/2$, using (I.A.43), and setting $y = 2x$,

$$\begin{aligned}
\Gamma'\left(\frac{v}{2}\right) &= \int_0^\infty \frac{\mathrm{d}}{\mathrm{d}(v/2)} x^{v/2-1} \mathrm{e}^{-x}\, \mathrm{d}x = \int_0^\infty x^{v/2-1} (\ln x)\, \mathrm{e}^{-x}\, \mathrm{d}x \\
&= \int_0^\infty \left(\frac{y}{2}\right)^{v/2-1} \left(\ln \frac{y}{2}\right) \mathrm{e}^{-y/2} \frac{\mathrm{d}y}{2} \\
&= \frac{1}{2^{v/2}} \int_0^\infty y^{v/2-1} (\ln y)\, \mathrm{e}^{-y/2}\, \mathrm{d}y - \frac{\ln 2}{2^{v/2}} \int_0^\infty y^{v/2-1} \mathrm{e}^{-y/2}\, \mathrm{d}y \\
&= \Gamma(v/2)\, \mathbb{E}[\ln X] - (\ln 2)\, \Gamma(v/2),
\end{aligned}$$

giving $\mathbb{E}[\ln X] = \Gamma'(v/2) / \Gamma(v/2) + \ln 2$. ■

1.1.3 Uniqueness of the m.g.f.

Under certain conditions, the m.g.f. uniquely determines or *characterizes* the distribution. To be more specific, we need the concept of equality in distribution: Let r.v.s X and Y be defined on the (induced) probability space $\{\mathbb{R}, \mathcal{B}, \Pr(\cdot)\}$, where $\mathcal{B}$ is the Borel σ-field generated by the collection of intervals $(a, b]$, $a, b \in \mathbb{R}$. Then X and Y are said to be *equal in distribution*, written $X \stackrel{d}{=} Y$, if

$$\Pr(X \in A) = \Pr(Y \in A) \quad \forall A \in \mathcal{B}. \tag{1.17}$$

The uniqueness result states that for r.v.s X and Y and some $h > 0$,

$$\mathbb{M}_X(t) = \mathbb{M}_Y(t) \; \forall \, |t| < h \quad \Rightarrow \quad X \stackrel{d}{=} Y. \tag{1.18}$$

See Section 1.2.4 below for some insight into why this result is true. As a concrete example, if the m.g.f. of an r.v. X is the same as, say, that of an exponential r.v., then one can conclude that X is exponentially distributed.

A similar notion applies to sequences of r.v.s, for which we need the concept of convergence in distribution, For a sequence of r.v.s X_n, $n = 1, 2, \ldots$, we say that X_n *converges in distribution* to X, written $X_n \stackrel{d}{\to} X$, if $F_{X_n}(x) \to F_X(x)$ as $n \to \infty$, for all points x such that $F_X(x)$ is continuous. Section 4.3.4 provides much more detail. It is important to note that if F_X is continuous, then it need not be the case that the F_{X_n} are continuous.

If X_n converges in distribution to a random variable which is, say, normally distributed, we will write $X_n \stackrel{d}{\to} \mathrm{N}(\cdot, \cdot)$, where the mean and variance of the specified normal distribution are constants, and do not depend on n. Observe that $X_n \stackrel{d}{\to} \mathrm{N}\left(\mu, \sigma^2\right)$ implies that, for n sufficiently large, the distribution of X_n can be adequately approximated by that of a $\mathrm{N}\left(\mu, \sigma^2\right)$ random variable. We will denote this by writing $X_n \stackrel{\text{app}}{\sim} \mathrm{N}\left(\mu, \sigma^2\right)$. This notation also allows the right-hand-side (r.h.s.) variable to depend on n; for example, we will write $S_n \stackrel{\text{app}}{\sim} \mathrm{N}(n, n)$ to indicate that, as n increases, the distribution of S_n can be adequately approximated by a $\mathrm{N}(n, n)$ random variable. In this case, we cannot write $S_n \stackrel{d}{\to} \mathrm{N}(n, n)$, but it is true that $n^{-1/2}(S_n - n) \stackrel{d}{\to} \mathrm{N}(0, 1)$.

We are now ready to state the convergence result for m.g.f.s. Let X_n be a sequence of r.v.s such that the corresponding m.g.f.s $\mathbb{M}_{X_n}(t)$ exist for $|t| < h$, for some $h > 0$, and all $n \in \mathbb{N}$. If X is a random variable whose m.g.f. $\mathbb{M}_X(t)$ exists for $|t| \leq h_1 < h$ for some $h_1 > 0$ and $\mathbb{M}_{X_n}(t) \to \mathbb{M}_X(t)$ as $n \to \infty$ for $|t| < h_1$, then $X_n \stackrel{d}{\to} X$. This convergence result also applies to the c.g.f. (1.8).

⊖ ***Example 1.9***
(**a**) Let X_n, $n = 1, 2, \ldots$, be a sequence of r.v.s such that $X_n \sim \mathrm{Bin}(n, p_n)$, with $p_n = \lambda/n$, for some constant value $\lambda \in \mathbb{R}_{>0}$, so that $\mathbb{M}_{X_n}(t) = \left(p_n \mathrm{e}^t + 1 - p_n\right)^n$ (see Problem 1.4), or

$$\mathbb{M}_{X_n}(t) = \left(\frac{\lambda}{n}\mathrm{e}^t + 1 - \frac{\lambda}{n}\right)^n = \left(1 + \frac{\lambda}{n}\left(\mathrm{e}^t - 1\right)\right)^n .$$

For all $h > 0$ and $|t| < h$, $\lim_{n\to\infty} \mathbb{M}_{X_n}(t) = \exp\left\{\lambda\left(e^t - 1\right)\right\} = \mathbb{M}_P(t)$, where $P \sim \text{Poi}(\lambda)$. That is, $X_n \xrightarrow{d} \text{Poi}(\lambda)$. Informally speaking, the binomial distribution with increasing n and decreasing p, such that np is a constant, approaches a Poisson distribution. This was also shown in Chapter I.4 by using the p.m.f. of a binomial random variable.

(**b**) Let $P_\lambda \sim \text{Poi}(\lambda)$, $\lambda \in \mathbb{R}_{>0}$, and $Y_\lambda = (P_\lambda - \lambda)/\sqrt{\lambda}$. From (1.5),

$$\mathbb{M}_{Y_\lambda}(t) = \exp\left\{\lambda\left(e^{t/\sqrt{\lambda}} - 1\right) - t\sqrt{\lambda}\right\}.$$

Writing

$$e^{t/\sqrt{\lambda}} = 1 + \frac{t}{\lambda^{1/2}} + \frac{t^2}{2\lambda} + \frac{t^3}{3!\lambda^{3/2}} + \cdots,$$

we see that

$$\lim_{\lambda\to\infty}\left[\lambda\left(e^{t/\sqrt{\lambda}} - 1\right) - t\sqrt{\lambda}\right] = \frac{t^2}{2},$$

or $\lim_{\lambda\to\infty} \mathbb{M}_{Y_\lambda}(t) = \exp\left(t^2/2\right)$, which is the m.g.f. of a standard normal random variable. That is, $Y_\lambda \xrightarrow{d} \text{N}(0, 1)$ as $\lambda \to \infty$. This should not be too surprising in light of the skewness and kurtosis results of Example 1.4.

(**c**) Let $P_\lambda \sim \text{Poi}(\lambda)$ with $\lambda \in \mathbb{N}$, and $Y_\lambda = (P_\lambda - \lambda)/\sqrt{\lambda}$. Then

$$p_{1,\lambda} := \frac{e^{-\lambda}\lambda^\lambda}{\lambda!} = \Pr(P_\lambda = \lambda) = \Pr(\lambda - 1 < P_\lambda \le \lambda) = \Pr\left(\frac{-1}{\sqrt{\lambda}} < Y_\lambda \le 0\right).$$

From the result in part (**b**) above, the limiting distribution of Y_λ is standard normal, motivating the conjecture that

$$\Pr\left(\frac{-1}{\sqrt{\lambda}} < Y_\lambda \le 0\right) \approx \Phi(0) - \Phi\left(-\lambda^{-1/2}\right) =: p_{2,\lambda}, \tag{1.19}$$

where $\approx$ means that, as $\lambda \to \infty$, the ratio of the two sides approaches unity. To informally verify (1.19), Figure 1.1 plots the relative percentage error (RPE), $100(p_{2,\lambda} - p_{1,\lambda})/p_{1,\lambda}$, on a log scale, as a function of λ.

The mean value theorem (Section I.A.2.2.2) implies the existence of an $x_\lambda \in \left(-\lambda^{-1/2}, 0\right)$ such that

$$\frac{\Phi(0) - \Phi\left(-\lambda^{-1/2}\right)}{0 - \left(-\lambda^{-1/2}\right)} = \Phi'(x_\lambda) = \phi(x_\lambda).$$

Clearly, $x_\lambda \in \left(-\lambda^{-1/2}, 0\right) \to 0$ as $\lambda \to \infty$, so that

$$\Phi(0) - \Phi\left(-\lambda^{-1/2}\right) = \frac{\lambda^{-1/2}}{\sqrt{2\pi}} \exp\left\{-\frac{1}{2}x_\lambda^2\right\} \approx \frac{\lambda^{-1/2}}{\sqrt{2\pi}}.$$

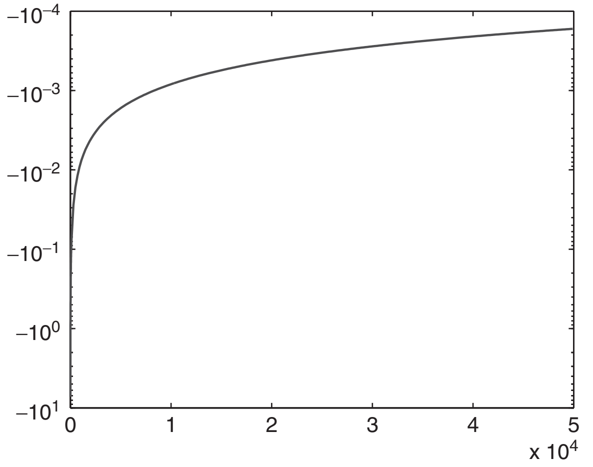

Figure 1.1 The relative percentage error of (1.19) as a function of λ

Combining these results yields

$$\frac{e^{-\lambda}\lambda^{\lambda}}{\lambda!} \approx \frac{\lambda^{-1/2}}{\sqrt{2\pi}},$$

or, rearranging, $\lambda! \approx \sqrt{2\pi}\lambda^{\lambda+1/2}e^{-\lambda}$. We understand this to mean that, for large λ, $\lambda!$ can be accurately approximated by the r.h.s. quantity, which is Stirling's approximation. ■

⊖ ***Example 1.10***
(**a**) Let $b > 0$ be a fixed value and, for any $a > 0$, let $X_a \sim \text{Gam}\,(a, b)$ and $Y_a = (X_a - a/b)\,/\sqrt{a/b^2}$. Then, for $t < a^{1/2}$,

$$\mathbb{M}_{Y_a}(t) = e^{-t\sqrt{a}}\mathbb{M}_{X_a}\left(\frac{b}{\sqrt{a}}t\right) = e^{-t\sqrt{a}}\left(\frac{1}{1-a^{-1/2}t}\right)^{a},$$

or $\mathbb{K}_{Y_a}(t) = -t\sqrt{a} - a\log\left(1 - a^{-1/2}t\right)$. From (I.A.114),

$$\log\,(1 + x) = \sum_{i=1}^{\infty}(-1)^{i+1}\frac{x^i}{i},$$

so that

$$\log\left(1 - a^{-1/2}t\right) = -\frac{t}{a^{1/2}} - \frac{t^2}{2a} - \frac{t^3}{3a^{3/2}} - \cdots$$

and $\lim_{a\to\infty}\mathbb{K}_{Y_a}(t) = t^2/2$. Thus, as $a \to \infty$, $Y_a \xrightarrow{d} \text{N}\,(0, 1)$, or, for large a, $X_a \overset{\text{app}}{\sim} \text{N}\left(a/b, a/b^2\right)$. Again, recall the skewness and kurtosis results of Example 1.5.

(**b**) Now let $S_n \sim \mathrm{Gam}(n, 1)$ for $n \in \mathbb{N}$, so that, for large n, $S_n \overset{\text{app}}{\sim} \mathrm{N}(n, n)$. The definition of convergence in distribution, and the continuity of the c.d.f. of S_n and that of its limiting distribution, informally suggest the limiting behaviour of the p.d.f. of S_n, i.e.,

$$f_{S_n}(s) = \frac{1}{\Gamma(n)} s^{n-1} \exp(-s) \approx \frac{1}{\sqrt{2\pi n}} \exp\left(-\frac{(s-n)^2}{2n^2}\right).$$

Choosing $s = n$ leads to $\Gamma(n+1) = n! \approx \sqrt{2\pi}\,(n+1)^{n+1/2} \exp(-n-1)$. From (I.A.46), $\lim_{n\to\infty}(1+\lambda/n)^n = \mathrm{e}^{\lambda}$, so

$$(n+1)^{n+1/2} = n^{n+1/2}\left(1+\frac{1}{n}\right)^{n+1/2} \approx n^{n+1/2}\mathrm{e},$$

and substituting this into the previous expression for $n!$ yields Stirling's approximation $n! \approx \sqrt{2\pi}\, n^{n+1/2}\, \mathrm{e}^{-n}$. ■

1.1.4 Vector m.g.f.

Analogous to the univariate case, the (joint) m.g.f. of the vector $\mathbf{X} = (X_1, \ldots, X_n)$ is defined as

$$\boxed{\mathbb{M}_{\mathbf{X}}(\mathbf{t}) = \mathbb{E}\,[\mathrm{e}^{\mathbf{t}'\mathbf{X}}], \quad \mathbf{t} = (t_1, \ldots, t_n)\,,}$$

and exists if the expectation is finite on an open rectangle of $\mathbf{0}$ in $\mathbb{R}^n$, i.e., if there is a $\varepsilon > 0$ such that $\mathbb{E}[\mathrm{e}^{\mathbf{t}'\mathbf{X}}]$ is finite for all $\mathbf{t}$ such that $|t_i| < \varepsilon$ for $i = 1, \ldots, n$.

As in the univariate case, if the joint m.g.f. exists, then it characterizes the distribution of $\mathbf{X}$ and, thus, all the marginals as well. In particular,

$$\mathbb{M}_{\mathbf{X}}((0, \ldots, 0, t_i, 0, \ldots, 0)) = \mathbb{E}\left[\mathrm{e}^{t_i X_i}\right] = \mathbb{M}_{X_i}(t_i)\,, \quad i = 1, \ldots, n.$$

Generalizing (1.4) and assuming the validity of exchanging derivative and integral,

$$\frac{\partial^k \mathbb{M}_{\mathbf{X}}(\mathbf{t})}{\partial t_1^{k_1} \partial t_2^{k_2} \cdots \partial t_n^{k_n}} = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} x_1^{k_1} x_2^{k_2} \cdots x_n^{k_n} \exp\{t_1 x_1 + t_2 x_2 + \cdots + t_n x_n\}\, f_{\mathbf{X}}(\mathbf{x})\, \mathrm{d}\mathbf{x},$$

so that the integer product moments of $\mathbf{X}$, $\mathbb{E}\left[\prod_{i=1}^n X_i^{k_i}\right]$ for $k_i \in \mathbb{N}$, are given by

$$\left.\frac{\partial^k \mathbb{M}_{\mathbf{X}}(\mathbf{t})}{\partial t_1^{k_1} \partial t_2^{k_2} \cdots \partial t_n^{k_n}}\right|_{\mathbf{t}=\mathbf{0}} = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} x_1^{k_1} x_2^{k_2} \cdots x_n^{k_n} f_{\mathbf{X}}(\mathbf{x})\, \mathrm{d}x_1\, \mathrm{d}x_2 \cdots \mathrm{d}x_n \qquad (1.20)$$

for $k = \sum_{i=1}^{n} k_i$ and such that $k_i = 0$ means that the derivative with respect to t_i is not taken. For example, if X and Y are r.v.s with m.g.f. $\mathbb{M}_{X,Y}(t_1, t_2)$, then

$$\mathbb{E}[XY] = \left.\frac{\partial^2 \mathbb{M}_{X,Y}(t_1, t_2)}{\partial t_1 \partial t_2}\right|_{t_1=t_2=0}$$

and

$$\mathbb{E}\left[X^2\right] = \left.\frac{\partial^2 \mathbb{M}_{X,Y}(t_1, t_2)}{\partial t_1^2}\right|_{t_1=t_2=0} = \left.\frac{\partial^2 \mathbb{M}_{X,Y}(t_1, 0)}{\partial t_1^2}\right|_{t_1=0}.$$

Example 1.11 (Example I.8.12 cont.) Let $f_{X,Y}(x, y) = \mathrm{e}^{-y}\mathbb{I}_{(0,\infty)}(x)\,\mathbb{I}_{(x,\infty)}(y)$ be the joint density of r.v.s X and Y. The m.g.f. is

$$\mathbb{M}_{X,Y}(t_1, t_2) = \int_0^\infty \int_x^\infty \exp\{t_1 x + t_2 y - y\}\,\mathrm{d}y\,\mathrm{d}x \tag{1.21}$$

$$= \int_0^\infty \frac{1}{1-t_2}\exp\{x(t_1 + t_2 - 1)\}\,\mathrm{d}x$$

$$= \frac{1}{(1-t_1-t_2)(1-t_2)}, \quad t_1 + t_2 < 1, \quad t_2 < 1, \tag{1.22}$$

so that $\mathbb{M}_{X,Y}(t_1, 0) = (1-t_1)^{-1}$, $t_1 < 1$, and $\mathbb{M}_{X,Y}(0, t_2) = (1-t_2)^{-2}$, $t_2 < 1$. From Example 1.5, this implies that $X \sim \mathrm{Exp}(1)$ and $Y \sim \mathrm{Gam}(2, 1)$. Also,

$$\left.\frac{\partial \mathbb{M}_{X,Y}(t_1, 0)}{\partial t_1}\right|_{t_1=0} = \left.(1-t_1)^{-2}\right|_{t_1=0} = 1,$$

$$\left.\frac{\partial \mathbb{M}_{X,Y}(0, t_2)}{\partial t_2}\right|_{t_2=0} = \left.2(1-t_2)^{-3}\right|_{t_2=0} = 2,$$

and

$$\frac{\partial^2 \mathbb{M}_{X,Y}(t_1, t_2)}{\partial t_1 \partial t_2} = \frac{3t_2 + t_1 - 3}{(t_1 + t_2 - 1)^3 (t_2 - 1)^2}, \qquad \left.\frac{\partial^2 \mathbb{M}_{X,Y}(t_1, t_2)}{\partial t_1 \partial t_2}\right|_{t_1=t_2=0} = 3,$$

so that $\mathbb{E}[X] = 1$, $\mathbb{E}[Y] = 2$ and $\mathrm{Cov}(X, Y) = \mathbb{E}[XY] - \mathbb{E}[X]\,\mathbb{E}[Y] = 1$. ■

The following result is due to Sawa (1972, p. 658), , and he used it for evaluating the moments of an estimator arising in an important class of econometric models; see also Sawa (1978). Let X_1 and X_2 be r.v.s such that $\Pr(X_1 > 0) = 1$, with joint m.g.f. $\mathbb{M}_{X_1,X_2}(t_1, t_2)$ which exists for $t_1 < \epsilon$ and $|t_2| < \epsilon$, for $\epsilon > 0$. Then, if it exists, the kth-order moment, $k \in \mathbb{N}$, of X_2/X_1 is given by

$$\mathbb{E}\left[\left(\frac{X_2}{X_1}\right)^k\right] = \frac{1}{\Gamma(k)} \int_{-\infty}^0 (-t_1)^{k-1} \left[\frac{\partial^k}{\partial t_2^k} \mathbb{M}_{X_1,X_2}(t_1, t_2)\right]_{t_2=0} \mathrm{d}t_1. \tag{1.23}$$

To informally verify this, assume we may reverse the order of the expectation with either the derivative or integral with respect to t_1 and t_2, so that the r.h.s. of (1.23) is

$$\frac{1}{\Gamma(k)}\int_{-\infty}^{0}(-t_1)^{k-1}\left[\frac{\partial^k}{\partial t_2^k}\mathbb{E}\left[\mathrm{e}^{t_1X_1}\mathrm{e}^{t_2X_2}\right]\right]_{t_2=0}\mathrm{d}t_1$$

$$=\frac{1}{\Gamma(k)}\mathbb{E}\left[\left[\frac{\partial^k}{\partial t_2^k}\mathrm{e}^{t_2X_2}\right]_{t_2=0}\int_{-\infty}^{0}(-t_1)^{k-1}\,\mathrm{e}^{t_1X_1}\,\mathrm{d}t_1\right]$$

$$=\frac{1}{\Gamma(k)}\mathbb{E}\left[X_2^k\int_0^\infty u^{k-1}\mathrm{e}^{-uX_1}\,\mathrm{d}u\right]=\mathbb{E}\left[\left(\frac{X_2}{X_1}\right)^k\right].$$

By working with $\mathbb{M}_{X_2,X_1}(t_2,t_1)$ instead of $\mathbb{M}_{X_1,X_2}(t_1,t_2)$, an expression for $\mathbb{E}[(X_1/X_2)^k]$ immediately results, though in terms of the more natural $\mathbb{M}_{X_1,X_2}(t_1,t_2)$, we get the following. Similar to (1.23), let X_1 and X_2 be r.v.s such that $\Pr(X_2>0)=1$, with joint m.g.f. $\mathbb{M}_{X_1,X_2}(t_1,t_2)$ which exists for $|t_1|<\epsilon$ and $t_2>-\epsilon$, for $\epsilon>0$. Then the kth-order moment, $k\in\mathbb{N}$, of X_1/X_2 is given by

$$\mathbb{E}\left[\left(\frac{X_1}{X_2}\right)^k\right]=\frac{1}{\Gamma(k)}\int_0^\infty t_2^{k-1}\left[\frac{\partial^k}{\partial t_1^k}\mathbb{M}_{X_1,X_2}(t_1,-t_2)\right]_{t_1=0}\mathrm{d}t_2, \tag{1.24}$$

if it exists. To confirm this, the r.h.s. of (1.24) is (indulging in complete lack of rigour),

$$\frac{1}{\Gamma(k)}\int_0^\infty t_2^{k-1}\left[\frac{\partial^k}{\partial t_1^k}\mathbb{E}\left[\mathrm{e}^{t_1X_1}\mathrm{e}^{-t_2X_2}\right]\right]_{t_1=0}\mathrm{d}t_2$$

$$=\frac{1}{\Gamma(k)}\mathbb{E}\left[\left[\frac{\partial^k}{\partial t_1^k}\mathrm{e}^{t_1X_1}\right]_{t_1=0}\int_0^\infty t_2^{k-1}\mathrm{e}^{-t_2X_2}\,\mathrm{d}t_2\right]=\mathbb{E}\left[\left(\frac{X_1}{X_2}\right)^k\right].$$

***Remark*:** A rigorous derivation of (1.23) and (1.24) is more subtle than it might appear. A flaw in Sawa's derivation is noted by Mehta and Swamy (1978), who provide a more rigorous derivation of this result. However, even the latter authors did not correctly characterize Sawa's error, as pointed out by Meng (2005), who provides the (so far) definitive conditions and derivation of the result for the more general case of $\mathbb{E}[X_1^k/X_2^b]$, $k\in\mathbb{N}$, $b\in\mathbb{R}_{>0}$, and also references to related results and applications.[3] Meng also provides several interesting examples of the utility of working with the joint m.g.f., including relationships to earlier work by R. A. Fisher. An important use of (1.24) arises in the study of ratios of quadratic forms.

The inequality

$$\mathbb{E}\left[\left(\frac{X_2}{X_1}\right)^k\right]\geq\frac{\mathbb{E}\left[X_2^k\right]}{\mathbb{E}\left[X_1^k\right]} \tag{1.25}$$

is shown in Mullen (1967). ■

[3] Lange (2003, p. 39) also provides an expression for $\mathbb{E}[X_1^k/X_2^b]$.

⊖ ***Example 1.12*** (Example 1.11 cont.) From (1.22) and (1.24),

$$\mathbb{E}\left[\frac{X}{Y}\right] = \int_0^\infty \left[\frac{\partial}{\partial t_1} \frac{1}{(1-t_1+t_2)(1+t_2)}\right]_{t_1=0} \mathrm{d}t_2 = \int_0^\infty (1+t_2)^{-3} \, \mathrm{d}t_2 = \frac{1}{2}$$

and

$$\mathbb{E}\left[\left(\frac{X}{Y}\right)^2\right] = 2\int_0^\infty t_2 (1+t_2)^{-4} \, \mathrm{d}t_2 = \frac{1}{3},$$

so that $\mathbb{V}(X/Y) = 1/12$. This is confirmed another way in Problem 2.6. ■

1.2 Characteristic functions

Similar to the m.g.f., the *characteristic function* (c.f.) of r.v. X is defined as $\mathbb{E}\left[\mathrm{e}^{itX}\right]$, where $i^2 = -1$, and is usually denoted as $\varphi_X(t)$. The c.f. is fundamental to probability theory and of much greater importance than the m.g.f. Its widespread use in introductory expositions of probability theory, however, is hampered because its involves notions from complex analysis, with which not all students are familiar. This is remedied to some extent via Section 1.2.1, which provides enough material for the reader to understand the rest of the chapter. More detailed treatments of c.f.s can be found in textbooks on advanced probability theory such as Wilks (1963), Billingsley (1995), Shiryaev (1996), Fristedt and Gray (1997), Gut (2005), or the book by Lukacs (1970), which is dedicated to the topic.

While it may not be too shocking that complex analysis arises in the theoretical underpinnings of probability theory, it might come as a surprise that it greatly assists *numerical* aspects by giving rise to expressions for real quantities which would otherwise not have been at all obvious. This, in fact, is true in general in mathematics (see the quote by Jacques Hadamard at the beginning of this chapter).

1.2.1 Complex numbers

> Should I refuse a good dinner simply because I do not understand the process of digestion?
> (Oliver Heaviside)

The *imaginary unit* i is defined to be a number having the property that

$$i^2 = -1. \tag{1.26}$$

One can use i in calculations as one does any ordinary real number such as $1, -1$ or $\sqrt{2}$, so expressions such as $1+i$, i^5 or $3-5i$ can be interpreted naively. We define the set of all complex numbers to be $\mathbb{C} := \{a+bi \mid a, b \in \mathbb{R}\}$. If $z = a+bi$, then $\mathrm{Re}(z) := a$ and $\mathrm{Im}(z) := b$ are the real and imaginary parts of z.

The set of complex numbers is closed under addition and multiplication, i.e., sums and products of complex numbers are also complex numbers. In particular,

$$(a+bi)+(c+di)=(a+c)+(b+d)i$$
$$(a+bi)\cdot(c+di)=(ac-bd)+(bc+ad)i.$$

As a special case, note that $i^3=-i$ and $i^4=1$. Therefore we have $i=i^5=i^9=\ldots$.

For each complex number $z=a+bi$, its *complex conjugate* is defined as $\overline{z}=a-bi$. The product $z\cdot\overline{z}=(a+bi)(a-bi)=a^2-b^2i^2=a^2+b^2$ is always a non-negative real number. The sum is

$$z+\overline{z}=(a+bi)+(a-bi)=2a=2\,\mathrm{Re}(z). \tag{1.27}$$

The absolute value of z, or its *(complex) modulus*, is defined to be

$$|z|=|a+bi|=\sqrt{z\overline{z}}=\sqrt{a^2+b^2}. \tag{1.28}$$

Simple calculations show that

$$|z_1z_2|=|z_1||z_2|,\quad |z_1+z_2|\le|z_1|+|z_2|,\quad \overline{z_1z_2}=\overline{z_1}\,\overline{z_2},\qquad z_1,z_2\in\mathbb{C}. \tag{1.29}$$

A sequence z_n of complex numbers is said to converge to some complex number $z\in\mathbb{C}$ iff the sequences $\mathrm{Re}\,z_n$ and $\mathrm{Im}\,z_n$ converge to $\mathrm{Re}\,z$ and $\mathrm{Im}\,z$, respectively. Hence, the series $\sum_{n=1}^{\infty}z_n$ converges if the series $\sum_{n=1}^{\infty}\mathrm{Re}\,z_n$ and $\sum_{n=1}^{\infty}\mathrm{Im}\,z_n$ converge separately.

As in $\mathbb{R}$, define the exponential function by

$$\exp(z)=\sum_{k=0}^{\infty}\frac{z^k}{k!},\qquad z\in\mathbb{C}.$$

It can be shown that, as in $\mathbb{R}$, $\exp(z_1+z_2)=\exp(z_1)\exp(z_2)$ for every $z_1,z_2\in\mathbb{C}$.

The definitions of the fundamental trigonometric functions in (I.A.28), i.e.,

$$\cos(z)=\sum_{k=0}^{\infty}(-1)^k\frac{z^{2k}}{(2k)!}\qquad\text{and}\qquad \sin(z)=\sum_{k=0}^{\infty}(-1)^k\frac{z^{2k+1}}{(2k+1)!},$$

also hold for complex numbers. In particular, if z takes the form $z=it$, where $t\in\mathbb{R}$, then $\exp(z)$ can be expressed as

$$\exp(it)=\sum_{k=0}^{\infty}\frac{(it)^k}{k!}=\sum_{k=0}^{\infty}\frac{i^kt^k}{k!}=\sum_{k=0}^{\infty}(-1)^k\frac{t^{2k}}{(2k)!}+i\sum_{k=0}^{\infty}(-1)^k\frac{t^{2k+1}}{(2k+1)!}, \tag{1.30}$$

i.e., from (I.A.28),

$$\boxed{\exp(it)=\cos(t)+i\sin(t).} \tag{1.31}$$

This relation is of fundamental importance, and is known as the *Euler formula*.[4]

[4] Named after the prolific Leonhard Euler (1707–1783), though (as often with naming conventions) it was actually discovered and published years before, in 1714, by Roger Cotes (1682–1716).

It easily follows from (1.31) that

$$\sin z = \frac{e^{iz} - e^{-iz}}{2i} \quad \text{and} \quad \cos z = \frac{e^{iz} + e^{-iz}}{2}. \tag{1.32}$$

Also, from (1.31) using $t = \pi$, we have $\cos \pi + i \sin \pi = -1$, or $e^{i\pi} + 1 = 0$, which is a simple but famous equation because it contains five of the most important quantities in mathematics. Similarly, $\exp(2\pi i) = 1$, so that, for $z \in \mathbb{C}$,

$$\exp(z + 2\pi i) = \exp(z)\exp(2\pi i) = \exp(z),$$

and one says that exp is a $2\pi i$-*cyclic* function. Lastly, with $z = a + ib \in \mathbb{C}$, (1.31) gives

$$\begin{aligned}\exp(\overline{z}) &= \exp(a - bi) = \exp(a)\exp(-bi) = \exp(a)\big[\cos(-b) + i\sin(-b)\big] \\ &= \exp(a)\big[\cos(b) - i\sin(b)\big] = \exp(a)\,\overline{\exp(ib)} = \overline{\exp(a + ib)} = \overline{\exp(z)}.\end{aligned}$$

As a shorthand for $\cos(t) + i\sin(t)$, one sometimes sees $\operatorname{cis}(t) := \cos(t) + i\sin(t)$, i.e., $\operatorname{cis}(t) = \exp(it)$.

A complex-valued function can also be integrated: the Riemann integral of a complex-valued function is the sum of the Riemann integrals of its real and imaginary parts.

⊛ ***Example 1.13*** For $s \in \mathbb{R} \setminus 0$, we know that $\int e^{st}\,dt = s^{-1}e^{st}$, but what if $s \in \mathbb{C}$? Let $s = x + iy$, and use (1.31) and the integral results in Example I.A.24 to write

$$\begin{aligned}\int e^{(x+iy)t}\,dt &= \int e^{xt}\cos(yt)\,dt + i\int e^{xt}\sin(yt)\,dt \\ &= \frac{e^{xt}}{x^2 + y^2}\left(x\cos(yt) + y\sin(yt)\right) + i\frac{e^{xt}}{x^2 + y^2}\left(x\sin(yt) - y\cos(yt)\right).\end{aligned}$$

This, however, is the same as $s^{-1}e^{st}$, as can be seen by writing

$$\frac{e^{st}}{s} = \frac{e^{xt}\left(\cos(yt) + i\sin(yt)\right)}{x + iy}\,\frac{x - iy}{x - iy},$$

with $(x + iy)(x - iy) = x^2 + y^2$ and multiplying out the numerator. Thus,

$$\int e^{st}\,dt = s^{-1}e^{st}, \quad s \in \mathbb{C} \setminus 0, \tag{1.33}$$

a result which will be used below. ■

A geometric approach to the complex numbers represents them as vectors in the plane, with the real term on the horizontal axis and the imaginary term on the vertical axis. Thus, the sum of two complex numbers can be interpreted as the sum of two vectors, and the modulus of $z \in \mathbb{C}$ is the length from 0 to z in the complex plane, recalling Pythagoras' theorem. The *unit circle* is the circle in the complex plane of radius 1 centred at 0, and includes all complex numbers of absolute value 1, i.e., such that $|z| = 1$; see Figure 1.2(a). If $t \in \mathbb{R}$, then the number $\exp(it)$ is contained in the unit circle, because

$$|\exp(it)| = \sqrt{\cos^2(t) + \sin^2(t)} = 1, \quad t \in \mathbb{R}. \tag{1.34}$$

For example, if $z = a + bi \in \mathbb{C}$, $a, b \in \mathbb{R}$, then (1.31) implies

$$\exp(z) = \exp(a + bi) = \exp(a)\exp(bi) = \exp(a)\big[\cos(b) + i\sin(b)\big],$$

and from (1.34),

$$|\exp(z)| = |\exp(a)||\exp(bi)| = \exp(a) = \exp(\mathrm{Re}(z)). \tag{1.35}$$

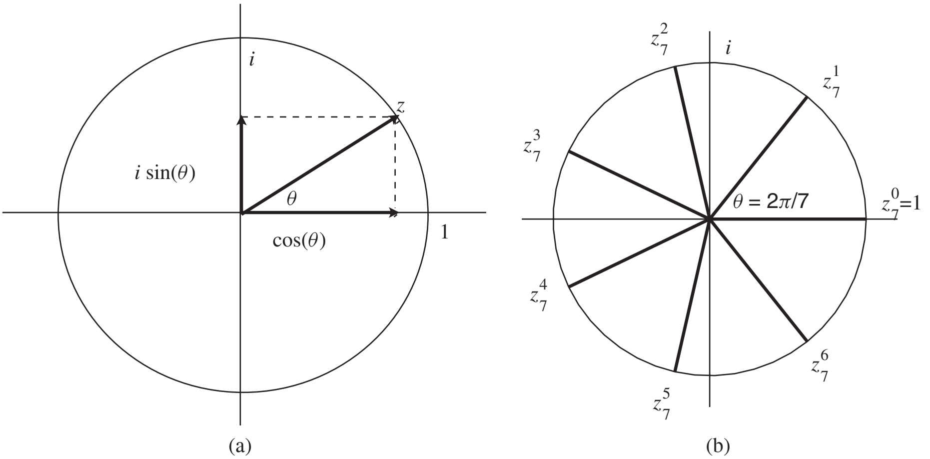

Figure 1.2 (a) Geometric representation of complex number $z = \cos(\theta) + i\sin(\theta)$ in the complex plane. (b) Plot of powers of $z_n = \exp(2\pi i/n)$ for $n = 7$, demonstrating that $\sum_{j=0}^{n-1} z_n^j = 0$

From the depiction of z as a vector in the complex plane, polar coordinates can also be used to represent z when $z \neq 0$. Let $r = |z| = \sqrt{a^2 + b^2}$ and define the *(complex) argument*, or *phase angle*, of z, denoted $\arg(z)$, to be the angle, say θ (in radians, measured counterclockwise from the positive real axis, modulo 2π), such that $a = r\cos(\theta)$ and $b = r\sin(\theta)$, i.e., for $a \neq 0$, $\arg(z) := \arctan(b/a)$. This is shown in Figure 1.2(a) for $r = 1$. From (1.31),

$$z = a + bi = r\cos(\theta) + ir\sin(\theta) = r\,\mathrm{cis}(\theta) = re^{i\theta},$$

and, as $r = |z|$ and $\theta = \arg(z)$, we can write

$$\mathrm{Re}(z) = a = |z|\cos(\arg z) \quad \text{and} \quad \mathrm{Im}(z) = b = |z|\sin(\arg z). \tag{1.36}$$

Now observe that, if $z_j = r_j\exp(i\theta_j) = r_j\,\mathrm{cis}(\theta_j)$, then

$$z_1 z_2 = r_1 r_2 \exp(i(\theta_1 + \theta_2)) = r_1 r_2\,\mathrm{cis}(\theta_1 + \theta_2), \tag{1.37}$$

so that

$$\arg(z_1 z_2 \cdots z_n) = \arg(z_1) + \arg(z_2) + \cdots + \arg(z_n) \quad \text{and} \quad \arg z^n = n \arg(z).$$

The following two examples illustrate very simple results which are used below in Example 1.25.

⊖ ***Example 1.14*** Let $z = 1 - ik$ for some $k \in \mathbb{R}$. Set $z = re^{i\theta}$ so that $r = \sqrt{1+k^2}$ and $\theta = \arctan(-k/1) = -\arctan(k)$. Then, with $z^m = r^m e^{i\theta m}$, $|z^m| = r^m = \left(1+k^2\right)^{m/2}$ and $\arg(z^m) = \theta m = -m \arctan(k)$. ■

⊖ ***Example 1.15*** Let $z = \exp\{ia/(1-ib)\}$ for $a, b \in \mathbb{R}$. As

$$\frac{ia}{1-ib} = -\frac{ab}{1+b^2} + i\frac{a}{1+b^2},$$

we can write

$$re^{i\theta} = z = \exp\left(-\frac{ab}{1+b^2} + i\frac{a}{1+b^2}\right) = \exp\left(-\frac{ab}{1+b^2}\right)\exp\left(i\frac{a}{1+b^2}\right),$$

for

$$r = \exp\left(-\frac{ab}{1+b^2}\right) \quad \text{and} \quad \theta = \frac{a}{1+b^2} \text{ modulo } 2\pi.$$

■

The next example derives a simple but highly useful result which we will require when working with the discrete Fourier transform.

⊛ ***Example 1.16*** Recall that the length of the unit circle is 2π, and let θ be the phase angle of the complex number z measured in radians (the arc length of the piece of the unit circle from $1 + 0i$ to z in Figure 1.2(a)). Then the quantity $z_n := \exp(2\pi i/n)$, $n \in \mathbb{N}$, plotted as a vector, will 'carve out' an nth of the unit circle, and, from (1.37), n equal pieces of the unit circle are obtained by plotting $z_n^0, z_n^1, \ldots, z_n^{n-1}$. This is shown in Figure 1.2(b) for $n = 7$. When seen as vectors emanating from the centre, it is clear that their sum is zero, i.e., for any $n \in \mathbb{N}$, $\sum_{j=0}^{n-1} z_n^j = 0$. More generally, for $k \in \{1, \ldots, n-1\}$, $\sum_{j=0}^{n-1} \left(z_n^k\right)^j = 0$, and clearly, $\sum_{j=0}^{n-1} \left(z_n^0\right)^j = n$. Because $z_n^n = 1 = z_n^{-n}$, this can be written as

$$\sum_{j=0}^{n-1} \left(z_n^k\right)^j = \begin{cases} n, & \text{if } k \in n\mathbb{Z} := \{0, n, -n, 2n, -2n, \ldots\}, \\ 0, & \text{if } k \in \mathbb{Z} \setminus n\mathbb{Z}. \end{cases} \tag{1.38}$$

The first part of (1.38) is trivial. For the second part, let $k \in \mathbb{Z} \setminus n\mathbb{Z}$ and $b := \sum_{j=0}^{n-1} (z_n^k)^j$. Note that $(z_n^k)^n = (z_n^n)^k = 1$. It follows that

$$z_n^k b = \sum_{j=0}^{n-1} (z_n^k)^{j+1} = \sum_{i=1}^{n-1} (z_n^k)^i + (z_n^k)^n = \sum_{i=0}^{n-1} (z_n^k)^i = b.$$

As $z_n^k \neq 1$ it follows that $b = 0$. See also Problem 1.23. ■

1.2.2 Laplace transforms

The *Laplace transform* of a real or complex function g of a real variable is denoted by $\mathcal{L}\{g\}$, and defined by

$$G(s) := \mathcal{L}\{g\}(s) := \int_0^\infty g(t)\, \mathrm{e}^{-st}\, \mathrm{d}t, \tag{1.39}$$

for all real or complex numbers s, if the integral exists (see below). From the form of (1.39), there is clearly a relationship between the Laplace transform and the moment generating function, and indeed, the m.g.f. is sometimes referred to as a two-sided Laplace transform. We study it here instead of in Section 1.1 above because we allow s to be complex. The Laplace transform is also related to the Fourier transform, which is discussed below in Section 1.3 and Problem 1.19.

1.2.2.1 Existence of the Laplace transform

The integral (1.39) exists for $\mathrm{Re}(s) > \alpha$ if g is continuous on $[0, \infty)$ and g has *exponential order* α, i.e., $\exists\, \alpha \in \mathbb{R}$, $\exists\, M > 0, \exists\, t_0 \geq 0$ such that $|g(t)| \leq M\mathrm{e}^{\alpha t}$ for $t \geq t_0$.[5] To see this, let g be of exponential order α and (piecewise) continuous. Then (as g is bounded on all subintervals on $\mathbb{R}_{\geq 0}$), $\exists M > 0$ such that $|g(t)| \leq M\mathrm{e}^{\alpha t}$ for $t \geq 0$, and, with $s = x + iy$,

$$\int_0^u \left| g(t)\, \mathrm{e}^{-st} \right| \mathrm{d}t \leq M \int_0^u \left| \mathrm{e}^{-(s-\alpha)t} \right| \mathrm{d}t = M \int_0^u \left| \mathrm{e}^{-(x-\alpha)t} \right| \mathrm{d}t = M \int_0^u \mathrm{e}^{-(x-\alpha)t}\, \mathrm{d}t,$$

where the second to last equality follows from (1.35), i.e.,

$$\left| \mathrm{e}^{-st} \mathrm{e}^{\alpha t} \right| = \left| \mathrm{e}^{-xt} \right| \left| \mathrm{e}^{-iyt} \right| \left| \mathrm{e}^{\alpha t} \right| = \left| \mathrm{e}^{-xt} \right| \left| \mathrm{e}^{\alpha t} \right|.$$

As $x = \mathrm{Re}(s) > \alpha$,

$$\lim_{u \to \infty} M \int_0^u \mathrm{e}^{-(x-\alpha)t}\, \mathrm{d}t = M \lim_{u \to \infty} \frac{1 - \mathrm{e}^{-(x-\alpha)u}}{x - \alpha} = \frac{M}{x - \alpha} < \infty, \tag{1.40}$$

showing that, under the stated conditions on g and s, the integral defining $\mathcal{L}\{g\}(s)$ converges absolutely, and thus exists.

1.2.2.2 Inverse Laplace transform

If G is a function defined on some part of the real line or the complex plane, and there exists a function g such that $\mathcal{L}\{g\}(s) = G(s)$ then, rather informally, this function g is referred to as the *inverse Laplace transform* of G, denoted by $\mathcal{L}^{-1}\{G\}$. Such an inverse Laplace transform need not exist, and if it exists, it will not be unique. If g is a function of a real variable such that $\mathcal{L}\{g\} = G$ and h is another function which is almost everywhere identical to g but differs on a finite set (or, more generally, on a set of measure zero), then, from properties of the Riemann (or Lebesgue) integral,

[5] Continuity of g on $[0, \infty)$ can be weakened to *piecewise continuity* on $[0, \infty)$. This means that $\lim_{t \downarrow 0} g(t)$ exists, and g is continuous on every finite interval $(0, b)$, except at a finite number of points in $(0, b)$ at which g has a *jump discontinuity*, i.e., g has a jump discontinuity at x if the limits $\lim_{t \uparrow x} g(t)$ and $\lim_{t \downarrow x} g(t)$ are finite, but differ. Notice that a piecewise continuous function is bounded on every bounded subinterval of $[0, \infty)$.

their Laplace transforms are identical, i.e., $\mathcal{L}\{g\} = G = \mathcal{L}\{h\}$. So both g and h could be regarded as versions of $\mathcal{L}^{-1}\{G\}$. If, however, functions g and h are continuous on $[0, \infty)$, such that $\mathcal{L}\{g\} = G = \mathcal{L}\{h\}$, then it can be proven that $g = h$, so in this case, there is a distinct choice of $\mathcal{L}^{-1}\{G\}$. See Beerends *et al.* (2003, p. 304) for a more rigorous discussion.

The linearity property of the Riemann integral implies the linearity property of Laplace transforms, i.e., for constants c_1 and c_2, and two functions $g_1(t)$ and $g_2(t)$ with Laplace transforms $\mathcal{L}\{g_1\}$ and $\mathcal{L}\{g_2\}$, respectively,

$$\mathcal{L}\{c_1 g_1 + c_2 g_2\} = c_1 \mathcal{L}\{g_1\} + c_2 \mathcal{L}\{g_2\}. \tag{1.41}$$

Also, by applying $\mathcal{L}^{-1}$ to both sides of (1.41),

$$c_1 g_1(t) + c_2 g_2(t) = \mathcal{L}^{-1}\{\mathcal{L}\{c_1 g_1 + c_2 g_2\}\} = \mathcal{L}^{-1}\{c_1 \mathcal{L}\{g_1\} + c_2 \mathcal{L}\{g_2\}\},$$

we see that $\mathcal{L}^{-1}$ is also a linear operator. Problem 1.17 proves a variety of further results involving Laplace transforms.

◎ ***Example 1.17*** Let $g\colon [0, \infty) \to \mathbb{C}$, $t \mapsto \mathrm{e}^{it}$. Then its Laplace transform at $s \in \mathbb{C}$ with $\mathrm{Re}(s) > 0$ is, from (1.33),

$$\begin{aligned}
\mathcal{L}\{g\}(s) &= \int_0^\infty \mathrm{e}^{it}\mathrm{e}^{-st}\,\mathrm{d}t = \int_0^\infty \mathrm{e}^{t(i-s)}\mathrm{d}t = \left.\frac{\mathrm{e}^{t(i-s)}}{i-s}\right|_0^\infty = \frac{1}{i-s}\left(\lim_{t\to\infty} \mathrm{e}^{-t(s-i)} - 1\right) \\
&= \frac{1}{s-i} = \frac{s+i}{(s+i)(s-i)} = \frac{s}{s^2+1} + i\frac{1}{s^2+1}.
\end{aligned}$$

Now, (1.31) and (1.41) imply $\mathcal{L}\{g\} = \mathcal{L}\{\cos\} + i\mathcal{L}\{\sin\}$ or

$$\int_0^\infty \cos(t)\,\mathrm{e}^{-st}\,\mathrm{d}t = \frac{s}{s^2+1}, \qquad \int_0^\infty \sin(t)\,\mathrm{e}^{-st}\,\mathrm{d}t = \frac{1}{s^2+1}. \tag{1.42}$$

Relations (1.42) are derived directly in Example I.A.24. See Example 1.22 for their use. ■

1.2.3 Basic properties of characteristic functions

For the c.f. of r.v. X, using (1.31) and the notation defined in (I.4.31),

$$\begin{aligned}
\varphi_X(t) &= \int_{-\infty}^\infty \mathrm{e}^{itx}\,\mathrm{d}F_X(x) \\
&= \int_{-\infty}^\infty \cos(tx)\,\mathrm{d}F_X(x) + i\int_{-\infty}^\infty \sin(tx)\,\mathrm{d}F_X(x) \\
&= \mathbb{E}[\cos(tX)] + i\,\mathbb{E}[\sin(tX)].
\end{aligned}$$

As

$$\cos(-\theta) = \cos(\theta) \quad \text{and} \quad \sin(-\theta) = -\sin(\theta), \tag{1.43}$$

it follows that

$$\varphi_X(-t) = \overline{\varphi}_X(t), \tag{1.44}$$

where $\overline{\varphi}_X$ is the complex conjugate of φ_X. Also, from (1.27) and (1.44),

$$\varphi_X(t) + \varphi_X(-t) = 2\,\mathrm{Re}(\varphi_X(t)). \tag{1.45}$$

Contrary to the m.g.f., the c.f. will always exist: from (1.34),

$$|\varphi_X(t)| = \left| \int_{-\infty}^{\infty} \mathrm{e}^{itx}\, \mathrm{d}F_X(x) \right| \leq \int_{-\infty}^{\infty} \left|\mathrm{e}^{itx}\right| \mathrm{d}F_X(x) = \int_{-\infty}^{\infty} \mathrm{d}F_X(x) = 1. \tag{1.46}$$

***Remark*:** A set of necessary and sufficient conditions for a function to be a c.f. is given by *Bochner's theorem*: A complex-valued function φ of a real variable t is a characteristic function iff (i) $\varphi(0) = 1$, (ii) φ is continuous, and (iii) for any positive integer n, real values $t_1, \ldots, t_n$, and complex values $\xi_1, \ldots, \xi_n$, the sum

$$S = \sum_{j=1}^{n} \sum_{k=1}^{n} \varphi(t_j - t_k)\xi_j\overline{\xi}_k \geq 0, \tag{1.47}$$

i.e., S is real and nonnegative. (The latter two conditions are equivalent to stating that f is a nonnegative definite function.) Note that, if φ_X is the c.f. of r.v. X, then, from (1.28) and (1.29),

$$S = \mathbb{E}\left[\sum_{j=1}^{n} \sum_{k=1}^{n} \left[\exp\left(i(t_j - t_k)X\right)\right] \xi_j\overline{\xi}_k \right] = \mathbb{E}\left[\left| \sum\nolimits_{j=1}^{n} \mathrm{e}^{it_jX}\xi_j \right|^2 \right] \geq 0,$$

which shows that if φ is a c.f., then it satisfies (1.47). See also Fristedt and Gray (1997, p. 227). The proof of the converse is more advanced; for it, and alternative criteria, see Lukacs (1970, Section 4.2) and Berger (1993, pp. 58–59), and the references therein. ■

The *uniqueness theorem*, first proven in Lévy (1925), states that, for r.v.s X and Y,

$$\varphi_X = \varphi_Y \quad \Leftrightarrow \quad X \overset{d}{=} Y. \tag{1.48}$$

Proofs can be found in Lukacs (1970, Section 3.1) or Gut (2005, p. 160 and 250).

⊖ ***Example 1.18*** Recall the probability integral transform at the end of Section I.7.3. If X is a continuous random variable with c.d.f. F_X, the c.f. of $Y = F_X(X)$ is

$$\phi_Y(s) = \mathbb{E}\left[\mathrm{e}^{isF_X(X)}\right] = \int_{-\infty}^{\infty} \mathrm{e}^{isF_X(t)} f_X(t)\,\mathrm{d}t$$

so that, with $u = F_X(t)$ and $\mathrm{d}u = f_X(t)\,\mathrm{d}t$,

$$\phi_Y(s) = \int_0^1 \mathrm{e}^{isu}\,\mathrm{d}u = \frac{\mathrm{e}^{is} - 1}{is},$$

which is the c.f. of a uniform random variable, implying $Y \sim \mathrm{Unif}\,(0, 1)$. ■

If φ_X is real, then $\varphi_X(t) = \int_{-\infty}^{\infty} \cos(tx)\,\mathrm{d}F(x)$ and $\int_{-\infty}^{\infty} \sin(tx)\,\mathrm{d}F_X(x) = 0$. This is the case, for example, if X is a continuous r.v. with p.d.f. $f_X(x)$ which is symmetric about zero because, with $u = -x$,

$$\int_0^{\infty} \sin(tx)\, f(x)\,\mathrm{d}x = -\int_0^{-\infty} \sin(-tu)\, f(-u)\,\mathrm{d}u = -\int_{-\infty}^{0} \sin(tu)\, f(u)\,\mathrm{d}u.$$

In fact, it can be proven that, for r.v. X with c.f. φ_X and p.m.f. or p.d.f. f_X, φ_X is real iff f_X is symmetric about zero (see Resnick, 1999, p. 297).

1.2.4 Relation between the m.g.f. and c.f.

1.2.4.1 If the m.g.f. exists on a neighbourhood of zero

Let X be an r.v. whose m.g.f. exists (i.e., $\mathbb{E}\left[\mathrm{e}^{tX}\right]$ is finite on a neighbourhood of zero). At first glance, comparing the definitions of the m.g.f. and the c.f., it seems that we could just write

$$\varphi_X(t) = \mathbb{M}_X(it)\,. \tag{1.49}$$

As an illustration of how to apply this formula, consider the case $X \sim \mathrm{N}(0, 1)$. Then $\mathbb{M}_X(t) = \mathrm{e}^{t^2/2}$. Plugging *it* into the right-hand side of this formula yields $\varphi_X(t) = \mathrm{e}^{-t^2/2}$. This can be shown rigorously using complex analysis.

Note that, formally, (1.49) does not make sense because we are plugging a complex variable into a function that only admits real arguments. But in the vast majority of real applications, such as the calculation in the Gaussian case, the m.g.f. will usually have a functional form which allows for complex arguments in an obvious manner. In the remainder of the text, we shall use the relation $\varphi_X(t) = \mathbb{M}_X(it)$ if $\mathbb{M}_X$ is finite on a neighbourhood of zero.

This approach has a nontrivial mathematical background, an understanding of which is not required to be able to apply the formula. The following discussion sheds some light on this issue, and can be skipped.

If $\mathbb{M}_X$ is finite on the interval $(-h, h)$ for some $h > 0$, then the complex function $z \mapsto \mathbb{E}\left[\mathrm{e}^{zX}\right]$ is well defined on the strip $\{z \in \mathbb{C} : |\mathrm{Re}(z)| < h\}$ (see Gut, 2005, p. 190). Note that, on the imaginary line, this function is the c.f. of X, while, on the real interval $(-h, h) \subseteq \mathbb{R}$, it is the m.g.f. It can be shown that the function $z \mapsto \mathbb{E}\left[\mathrm{e}^{zX}\right]$ is what is called 'complex analytic',[6] which might be translated as 'nice and smooth'. The so-called *identity theorem* asserts in our case that a complex analytic function on the strip is uniquely determined by its values on the interval $(-h, h)$.

On the one hand, this shows that the m.g.f. completely determines the function $z \mapsto \mathbb{E}\left[\mathrm{e}^{zX}\right]$ and, hence, the c.f. As a consequence, *if the m.g.f. of X exists on a*

[6] Complex analytic functions are also called 'holomorphic' and are the main objects of study in complex analysis; see Palka (1991) for an introduction.

neighbourhood of 0, then it uniquely determines the distribution of X. This was stated in (1.18).

On the other hand, this theorem can be used to actually *find* the c.f. As stated above, the m.g.f. will usually come in a functional form which suggests an extension to the complex plane (or at least to the strip). In our Gaussian example, the extension $z \mapsto e^{z^2/2}$ of $t \mapsto e^{t^2/2}$ immediately comes to mind. As a matter of fact, the extension thus constructed will normally be 'nice and smooth', so an application of the identity theorem tells us that it is the only 'nice and smooth' extension to the strip and thus coincides with the function $z \mapsto \mathbb{E}\left[e^{zX}\right]$. Hence plugging in the argument it really does result in $\varphi_X(t)$.

In the Gaussian example, $z \mapsto e^{z^2/2}$ is indeed a complex analytic function, so this shows that $\mathbb{E}\left[e^{zX}\right] = e^{z^2/2}$ and, in particular, that $\mathbb{E}\left[e^{itX}\right] = e^{-t^2/2}$.

1.2.4.2 The m.g.f. and c.f. for nonnegative X

Up to this point, only the case where the m.g.f. $\mathbb{M}_X$ is defined on a neighbourhood of zero has been considered. There is another important case where we can apply (1.49). If X is a nonnegative r.v., then $\mathbb{M}_X(t)$ is finite for all $t \in (-\infty, 0]$. It might well be defined for a larger set, but to give a meaning to (1.49), one only needs the fact that the m.g.f. is defined for all negative t.

If $X \geq 0$, then, similarly to the above digression on complex analysis, the function $z \mapsto \mathbb{E}\left[e^{zX}\right]$ is well defined on the left half-plane $\{z \in \mathbb{C} : \operatorname{Re}(z) < 0\}$. Again, this function is 'complex analytic' and hence uniquely determined by the m.g.f. (defined on $(-\infty, 0]$). It can be shown that $z \mapsto \mathbb{E}\left[e^{zX}\right]$ is at least continuous on $\{z \in \mathbb{C} : \operatorname{Re}(z) \leq 0\}$, which is the left half-plane plus the imaginary axis. So, a more involved application of the identity theorem tells us that the m.g.f. uniquely determines the c.f. in this case, too. As a consequence, *the m.g.f. of an r.v. $X \geq 0$ uniquely determines the distribution of X.* A rigorous proof of this result can be found in Fristedt and Gray (1997, p. 218, Theorem 6).

Equation (1.49) is still applicable as a rule of thumb, and the more complicated background that has just been outlined entails some inconveniences which are illustrated by the following example.

⊖ ***Example 1.19*** The Lévy distribution was introduced in Examples I.7.6 and I.7.15, and will be revisited in Section 8.2 in the context of the asymmetric stable Paretian distribution and Section 9.4.2 in the context of the generalized inverse Gaussian distribution. Its p.d.f. is, for $\chi > 0$,

$$f_X(x) = \left(\frac{\chi}{2\pi}\right)^{1/2} x^{-3/2} \exp\left[-\frac{\chi}{2x}\right] \mathbb{I}_{(0,\infty)}(x),$$

and its m.g.f. is shown in (9.34) to be, for $t \leq 0$ (and not on a neighbourhood of zero),

$$\mathbb{M}_X(t) = \exp\left[-\sqrt{\chi}\sqrt{-2t}\right], \qquad t \leq 0.$$

Evaluating $\mathbb{M}_X(it)$ requires knowing the square root of $-2it$. Consider the case $t > 0$. Then we have a choice because $\left((1-i)\sqrt{t}\right)^2 = (1-i)^2 t = -2it$ and also $\left(-(1-i)\sqrt{t}\right)^2 = -2it$. For negative t, we have a similar choice.

To see which is the right choice, we have to identify the extension $z \mapsto \mathbb{E}\left[e^{zX}\right]$ of the m.g.f. to the left half-plane: what is the nice and smooth extension of $t \mapsto \exp\left[-\sqrt{\chi}\sqrt{-2t}\right]$ for all $z \in \mathbb{C}$ such that $\mathrm{Re}(z) < 0$? If $\mathrm{Re}(z) < 0$, then $\mathrm{Re}(-2z) > 0$, in other words, $-2z$ lies in the right half-plane. Therefore $-2z$ has two square roots, one in the right half-plane and one in the left half-plane. Call the one in the right half-plane $\sqrt{-2z}$ (implying that the other is $-\sqrt{-2z}$). Then the function $z \mapsto \exp\left[-\sqrt{\chi}\sqrt{-2z}\right]$ is nice and smooth on the left half-plane and extends the m.g.f.

It can be continuously extended to $z \in \mathbb{C}$ with $\mathrm{Re}(z) \le 0$ (and further) by always taking the proper root of $-2z$, which means that, for $t > 0$, the proper choice of $\sqrt{-2it}$ will be $(1-i)\sqrt{t}$ (being in the right half-plane), while, for $t < 0$, the proper choice of $\sqrt{-2it}$ will be $(1+i)\sqrt{-t}$ (also being the root in the right half-plane). This can be combined as

$$\sqrt{-2it} = \sqrt{|t|}\,(1 - i\,\mathrm{sgn}\,(t)), \qquad t \in \mathbb{R},$$

so that

$$\varphi_X(t) = \exp\left\{-\chi^{1/2}|t|^{1/2}\left[1 - i\,\mathrm{sgn}(t)\right]\right\}.$$

This result is also seen in the more general context of the c.f. of the stable Paretian distribution. In particular, taking $c = \chi$, $\mu = 0$, $\beta = 1$ and $\alpha = 1/2$ in (8.8) yields the same expression. ■

1.2.5 Inversion formulae for mass and density functions

If you are going through hell, keep going. (Sir Winston Churchill)

By *inversion formula* we mean an integral expression whose integrand involves the c.f. of a random variable X and a number x, and whose result is the p.m.f., p.d.f. or c.d.f. of X at x.

Some of the proofs below make use of the following results for constants $k \in \mathbb{R} \setminus 0$ and $T \in \mathbb{R}_{\ge 0}$:

$$\int_{-T}^{T} \cos(kt)\,\mathrm{d}t = \frac{2\sin(kT)}{k} \quad \text{and} \quad \int_{-T}^{T} \sin(kt)\,\mathrm{d}t = 0. \tag{1.50}$$

The first follows from (I.A.27), i.e., that $\mathrm{d}\sin x/\mathrm{d}x = \cos x$, the fundamental theorem of calculus, a simple change of variable, and (1.43). The second can be derived similarly and/or confirmed quickly from a plot of sin.

We begin with expressions for the probability mass function. Let X be a discrete r.v. with support $\{x_j\}_{j=1}^{\infty}$, such that $x_j = x_k$ iff $j = k$, and mass function f_X. Then

$$\boxed{f_X(x_j) = \lim_{T\to\infty} \frac{1}{2T}\int_{-T}^{T} e^{-itx_j}\varphi_X(t)\,\mathrm{d}t.} \tag{1.51}$$

Proof (outline):[7] First observe that, from Euler's formula (1.31) and (1.50),

$$\frac{1}{2T}\int_{-T}^{T} \mathrm{e}^{it(x_k-x_j)}\,\mathrm{d}t = \frac{1}{2T}\int_{-T}^{T} \cos t(x_k-x_j)\,\mathrm{d}t + \frac{i}{2T}\int_{-T}^{T} \sin t(x_k-x_j)\,\mathrm{d}t$$

$$= \frac{1}{2T}\int_{-T}^{T} \cos t(x_k-x_j)\,\mathrm{d}t \tag{1.52}$$

which, if $x_k = x_j$, is 1, and otherwise, using (1.50), is $\sin(A)/A$, where $A = T(x_k - x_j)$. Clearly, $\lim_{T\to\infty} \sin(A)/A = 0$, so that

$$\lim_{T\to\infty} \frac{1}{2T}\int_{-T}^{T} \mathrm{e}^{it(x_k-x_j)}\,\mathrm{d}t = \begin{cases} 1, & \text{if} \quad x_k = x_j, \\ 0, & \text{if} \quad x_k \neq x_j. \end{cases}$$

Next,

$$\frac{1}{2T}\int_{-T}^{T} \mathrm{e}^{-itx_j}\varphi_X(t)\,\mathrm{d}t = \frac{1}{2T}\int_{-T}^{T} \mathrm{e}^{-itx_j}\sum_{k=1}^{\infty} \mathrm{e}^{itx_k} f_X(x_k)\,\mathrm{d}t$$

$$= \sum_{k=1}^{\infty} f_X(x_k)\,\frac{1}{2T}\int_{-T}^{T} \mathrm{e}^{it(x_k-x_j)}\,\mathrm{d}t$$

and

$$\lim_{T\to\infty} \frac{1}{2T}\int_{-T}^{T} \mathrm{e}^{-itx_j}\varphi_X(t)\,\mathrm{d}t = \sum_{k=1}^{\infty} f(x_k) \lim_{T\to\infty}\frac{1}{2T}\int_{-T}^{T} \mathrm{e}^{it(x_k-x_j)}\,\mathrm{d}t = f_X(x_j),$$

where the interchange of integral and summation, and limit and summation, can be shown to be valid. ■

To illustrate (1.51), let $X \sim \mathrm{Bin}(10, 0.3)$ and consider calculating the p.m.f. at $x = 4$, the true value of which is 0.200121. The program in Listing 1.1 calculates the p.m.f. via (1.51), using the c.f. $\varphi_X(t) = \mathbb{M}_X(it)$. The binomial m.g.f. is shown in Problem 1.4 to be $\mathbb{M}_X(t) = (p\mathrm{e}^t + 1 - p)^n$. The discrepancies of the computed values from the true p.m.f. value, as a function of T, are plotted in Figure 1.3.

```
function pmf = schlecht(x,n,p,T)
pmf=quadl(@ff,-T,T,1e-6,0,n,p,x) / 2 / T;

function I=ff(tvec,n,p,x);
q=1-p; I=zeros(size(tvec));
for loop=1:length(tvec)
  t=tvec(loop); cf=(p*exp(i*t)+q)^n; I(loop)= ( exp(-i*t*x) * cf );
end
```

Program Listing 1.1 Uses (1.51) to compute the p.m.f. of $X \sim \mathrm{Bin}(n, p)$ at x

[7] The proofs given in this section are not rigorous, but should nevertheless convince most readers of the validity of the results. Complete proofs are of course similar, and are available in advanced textbooks such as those mentioned at the beginning of Section 1.2. See, for example, Shiryaev (1996, p. 283) regarding this proof, and, for even more detail, Lukacs (1970, Section 3.2).

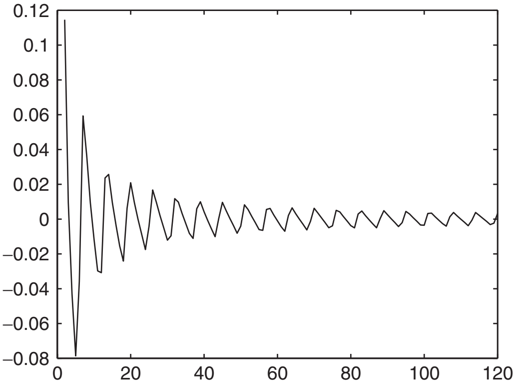

Figure 1.3 Discrepancy of the probability calculated via (1.51) (disregarding the complex part, which is nonzero for $T < \infty$) for $X \sim \text{Bin}(n, p)$ at $x = 4$ from the true p.m.f. value as a function of T, $T = 2, 3, \ldots, 120$

It does indeed appear that the magnitude of the discrepancy becomes smaller as T increases. More importantly, however, the discrepancy appears to be exactly zero on a regularly spaced set of T-values. These values occur for values of T which are integral multiples of π. This is verified in Figure 1.4, which is the same as Figure 1.3, but with T chosen as $\pi, 2\pi, \ldots$. This motivates the following claim, which is indeed a well-known fact and far more useful for numeric purposes:

$$\boxed{f_X(x_j) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \mathrm{e}^{-itx_j} \varphi_X(t)\, \mathrm{d}t.} \tag{1.53}$$

See, for example, Feller (1971, p. 511) or Gut (2005, p. 164).

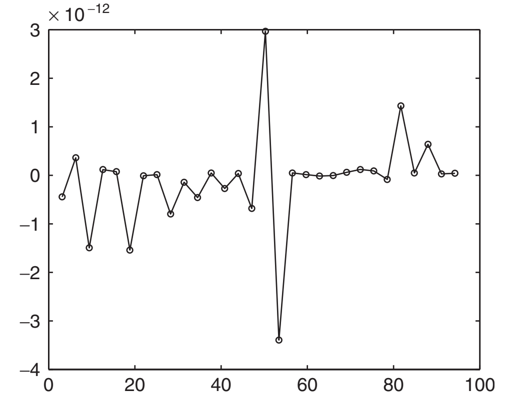

Figure 1.4 Same as Figure 1.3 but using values $T = \pi, 2\pi, \ldots, 30\pi$

⊖ ***Example 1.20*** (Example 1.7 cont.) The m.g.f. of N_m (number of i.i.d. Bernoulli trials with success probability p required until m consecutive successes occur) was found to be, with $q = 1 - p$,

$$\mathbb{M}_m(t) = \frac{pe^t \mathbb{M}_{m-1}(t)}{1 - q\mathbb{M}_{m-1}(t)\,e^t}, \quad \mathbb{M}_1(t) = pe^t / \left(1 - qe^t\right), \quad t \neq -\ln(1-p),$$

and the c.f. is $\mathbb{M}_m(it)$. Using (1.53), the program in Listing 1.2 computes the p.m.f., which is plotted in Figure 1.5 for $m = 3$ and $p = 1/2$. This is indeed the same as in Figure I.2.1(a), which used a recursive expression for the p.m.f. ■

```
function f = consecpmf (xvec,m,p)
xl=length(xvec); f=zeros(xl,1); tol=1e-5;
for j=1:xl
  x=xvec(j); f(j)=quadl(@fun,-pi,pi,tol,0,x,m,p);
end
f=real(f); % roundoff error sometimes gives an imaginary
           % component of the order of 1e-16.

function I=fun(t,x,m,p), I=exp(-i*t*x) .* cf(t,m,p) / (2*pi);

function g = cf(t,m,p) % cf
q=1-p;
if m==1
  g = p*exp(i*t) ./ (1-q*exp(i*t));
else
  kk = exp(i*t) .* cf(t,m-1,p); g = (p*kk) ./ (1-q*kk);
end
```

Program Listing 1.2 Computes the p.m.f. via inversion formula (1.53) of the number of i.i.d. Ber(p) trials required until m consecutive successes occur

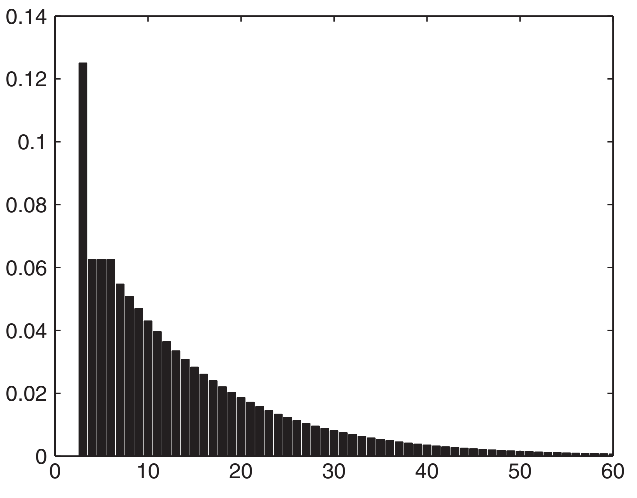

Figure 1.5 The p.m.f. of the number of children (assuming equal boy and girl probabilities) until one obtains three girls in a row, obtained by computing (1.53). Compare with Figure I.2.1(a)

Now consider the continuous case. If X is a continuous random variable with p.d.f. f_X and $\int_{-\infty}^{\infty} |\varphi_X(t)|\, \mathrm{d}t < \infty$, then

$$\boxed{f_X(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \mathrm{e}^{-itx} \varphi_X(t)\, \mathrm{d}t.} \tag{1.54}$$

Proof (outline): Take $T > 0$ so that, assuming we can change the order of integration and using (1.52) and (1.50) (and ignoring the 'measure zero' case of $x = y$),

$$\begin{aligned}
\int_{-T}^{T} \mathrm{e}^{-itx} \varphi_X(t)\, \mathrm{d}t &= \int_{-T}^{T} \mathrm{e}^{-itx} \int_{-\infty}^{\infty} \mathrm{e}^{ity} f_X(y)\, \mathrm{d}y\, \mathrm{d}t \\
&= \int_{-\infty}^{\infty} f_X(y) \int_{-T}^{T} \mathrm{e}^{it(y-x)}\, \mathrm{d}t\, \mathrm{d}y \\
&= \int_{-\infty}^{\infty} f_X(y) \frac{2 \sin T(y-x)}{y-x}\, \mathrm{d}y.
\end{aligned}$$

Then, with $A = T(y - x)$ and $\mathrm{d}y = \mathrm{d}A/T$,

$$\begin{aligned}
\int_{-\infty}^{\infty} \mathrm{e}^{-itx} \varphi_X(t)\, \mathrm{d}t &= 2 \lim_{T\to\infty} \int_{-\infty}^{\infty} f_X(y) \frac{\sin T(y-x)}{y-x}\, \mathrm{d}y \\
&= 2 \lim_{T\to\infty} \int_{-\infty}^{\infty} f_X\left(x + \frac{A}{T}\right) \frac{\sin A}{A}\, \mathrm{d}A \\
&= 2\pi f_X(x),
\end{aligned}$$

where the last equality makes the assumption (which can be justified) that interchanging the limit and integral is valid, and because

$$\int_0^{\infty} \frac{\sin x}{x}\, \mathrm{d}x = \frac{\pi}{2}, \tag{1.55}$$

as shown in Example I.A.30. ■

Remark: If the cumulant generating function $\mathbb{K}_X(s) = \ln \mathbb{M}_X(s)$ of X exists, then (1.54) can be expressed as

$$\boxed{f_X(x) = \frac{1}{2\pi i} \int_{-i\infty}^{i\infty} \exp\{\mathbb{K}_X(s) - sx\}\, \mathrm{d}s,} \tag{1.56}$$

as shown in Problem 1.12. ■

⊖ ***Example 1.21*** From (1.10), if $Z \sim \mathrm{N}(0, 1)$, then $\mathbb{M}_Z(t) = \mathrm{e}^{t^2/2}$, and $\varphi_Z(t) = \mathbb{M}_Z(it) = \mathrm{e}^{-t^2/2}$, which is real, as Z is symmetric about zero. Going the other way, from (1.54),

$$f_Z(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \mathrm{e}^{-itx} \mathrm{e}^{-t^2/2}\, \mathrm{d}t = \frac{1}{2\pi} \int_{-\infty}^{\infty} \mathrm{e}^{-\frac{1}{2}(t^2 + 2itx)}\, \mathrm{d}t$$

$$= \frac{1}{2\pi}\int_{-\infty}^{\infty} \mathrm{e}^{-\frac{1}{2}\left(t^2+2t(ix)+(ix)^2\right)} \mathrm{e}^{\frac{1}{2}(ix)^2}\,\mathrm{d}t$$

$$= \frac{1}{\sqrt{2\pi}}\mathrm{e}^{\frac{1}{2}(ix)^2}\,\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} \mathrm{e}^{-\frac{1}{2}(t+ix)^2}\,\mathrm{d}t$$

$$= \frac{1}{\sqrt{2\pi}}\mathrm{e}^{-x^2/2},$$

using $u = t + ix$ and $\mathrm{d}t = \mathrm{d}u$ to resolve the integral. ■

◎ ***Example 1.22*** Consider deriving the density corresponding to $\varphi_X(t) = \mathrm{e}^{-c|t|}$ for $c > 0$. Using the fact that $\mathrm{e}^{-ixt} = \cos(tx) - i\sin(tx)$ and separating negative and positive t,

$$2\pi\, f_X(x) = \int_{-\infty}^{\infty} \mathrm{e}^{-ixt}\mathrm{e}^{-c|t|}\,\mathrm{d}t$$

$$= \int_{-\infty}^{0} \cos(tx)\,\mathrm{e}^{ct}\,\mathrm{d}t - i\int_{-\infty}^{0} \sin(tx)\,\mathrm{e}^{ct}\,\mathrm{d}t$$

$$+ \int_{0}^{\infty} \cos(tx)\,\mathrm{e}^{-ct}\,\mathrm{d}t - i\int_{0}^{\infty} \sin(tx)\,\mathrm{e}^{-ct}\,\mathrm{d}t.$$

With the substitution $u = -ct$ in the first two integrals, $u = ct$ in the last two, and the basic facts (1.43), this simplifies to

$$f_X(x; c) = \frac{1}{c\pi}\int_0^{\infty} \cos(ux/c)\,\mathrm{e}^{-u}\,\mathrm{d}u. \tag{1.57}$$

To resolve this for $x \neq 0$, first set $t = ux/c$ and $s = 1/x$, and then use (1.42) to get

$$f_X(x; c) = \frac{s}{\pi}\int_0^{\infty} \cos(t)\,\mathrm{e}^{-cst}\,\mathrm{d}t = \frac{s}{\pi}\frac{cs}{(cs)^2+1}$$

$$= \frac{1}{\pi}\frac{cx^{-2}}{c^2x^{-2}+1} = \frac{1}{\pi}\frac{c}{c^2+x^2} = \frac{1}{c\pi}\frac{1}{1+(x/c)^2},$$

which is the Cauchy distribution with scale c. This integral also follows from (I.A.24) with $a = -1$ and $b = x/c$.

Observe that φ_X is real and f_X is symmetric about zero. ■

We showed above that $\exp(\overline{z}) = \overline{\exp(z)}$, which implies $\mathrm{e}^{-itx} = \overline{\mathrm{e}^{itx}}$. We also saw above that $\overline{z_1 z_2} = \overline{z_1}\,\overline{z_2}$, $\varphi_X(-t) = \overline{\varphi}_X(t)$, and $z + \overline{z} = 2\,\mathrm{Re}(z)$. Then, with $v = -t$,

$$\int_{-\infty}^{0} \mathrm{e}^{-itx}\varphi_X(t)\,\mathrm{d}t = \int_0^{\infty} \mathrm{e}^{ivx}\varphi_X(-v)\,\mathrm{d}v = \int_0^{\infty} \overline{\mathrm{e}^{-ivx}}\,\overline{\varphi}_X(v)\,\mathrm{d}v$$

$$= \int_0^{\infty} \overline{\mathrm{e}^{-ivx}\varphi_X(v)}\mathrm{d}v$$

so that, for X continuous, (1.54) is

$$\begin{aligned} f_X(x) = \int_{-\infty}^{\infty} \mathrm{e}^{-itx}\varphi_X(t)\,\mathrm{d}t &= \int_0^{\infty} \mathrm{e}^{-itx}\varphi_X(t)\,\mathrm{d}t + \int_{-\infty}^{0} \mathrm{e}^{-itx}\varphi_X(t)\,\mathrm{d}t \\ &= \int_0^{\infty} \mathrm{e}^{-itx}\varphi_X(t)\,\mathrm{d}t + \int_0^{\infty} \overline{\mathrm{e}^{-ivx}\varphi_X(v)}\mathrm{d}v \\ &= 2\int_0^{\infty} \mathrm{Re}\left[\mathrm{e}^{-itx}\varphi_X(t)\right]\mathrm{d}t. \end{aligned}$$

A similar calculation is done for the discrete case. Thus, the inversion formulae can also be written as

$$\boxed{f_X(x_j) = \frac{1}{\pi}\int_0^{\pi} \mathrm{Re}\left[\mathrm{e}^{-itx_j}\varphi_X(t)\right]\mathrm{d}t} \tag{1.58}$$

in the discrete case and

$$\boxed{f_X(x) = \frac{1}{\pi}\int_0^{\infty} \mathrm{Re}\left[\mathrm{e}^{-itx}\varphi_X(t)\right]\mathrm{d}t} \tag{1.59}$$

in the continuous case. These could be more suitable for numeric work than their counterparts (1.53) and (1.54) if a 'canned' integration routine is available but which does not support complex arithmetic.

***Remark*:** In practice, the integral in (1.59) needs to be truncated at a point such that a specified degree of accuracy will be obtained. Ideally, one would accomplish this by algebraically determining an upper limit on the tail value of the integral. Less satisfactory, but still useful, is to verify that the absolute value of the integrand is, after some point, monotonically decreasing at a fast enough rate, and one then evaluates and accumulates the integral over contiguous pieces of the real line (possibly doubling in size each step) until the contribution to the integral is below a given threshold.

To avoid having to determine the upper bound in the integral, one can transform the integrand such that the range of integration is over a finite interval, say (0, 1). One way of achieving this is via the substitution $u = 1/(1+t)$ (so that $t = (1-u)/u$ and $\mathrm{d}t = -u^{-2}\mathrm{d}u$), which leads to

$$f_X(x) = \frac{1}{\pi}\int_0^1 h_x\left(\frac{1-u}{u}\right)u^{-2}\,\mathrm{d}u, \qquad h_x(t) = \mathrm{Re}\left[\mathrm{e}^{-itx}\varphi_X(t)\right]. \tag{1.60}$$

Alternatively, use of $u = t/(t+1)$ in (1.59) leads to

$$f_X(x) = \frac{1}{\pi}\int_0^1 h_x\left(\frac{u}{1-u}\right)(1-u)^{-2}\,\mathrm{d}u. \tag{1.61}$$

Use of (1.61) is actually equivalent to (1.60) and just results in the same integrand, flipped over the $u = 1/2$ axis. There are two caveats to this approach. First, as the integrand in (1.60) cannot be evaluated at zero, a similar problem remains in the sense of determining how close to zero the lower integration limit should be chosen (and similarly for (1.61) at the upper bound of one). Second, the transformation (1.60) or (1.61) could induce pathological oscillatory behaviour in the integrand, making it difficult, or impossible, to accurately integrate. This is the case in Example 1.23 and Problem 1.11. ■

⊖ ***Example 1.23*** Problem 1.13 shows that the c.f. of $L \sim \text{Lap}(0, 1)$ is given by $\varphi_L(t) = \left(1 + t^2\right)^{-1}$. Thus, $L/\lambda \sim \text{Lap}(0, \lambda)$ has c.f.

$$\varphi_{L/\lambda}(t) = \mathbb{E}\left[\mathrm{e}^{itL/\lambda}\right] = \mathbb{E}\left[\mathrm{e}^{i(t/\lambda)L}\right] = \varphi_L(t/\lambda) = \left(1 + t^2/\lambda^2\right)^{-1}.$$

From Example I.9.5, we know that $X - Y \sim \text{Lap}(0, \lambda)$, where $X, Y \overset{\text{i.i.d.}}{\sim} \text{Exp}(\lambda)$, so that $\varphi_{X-Y}(t) = \left(1 + t^2/\lambda^2\right)^{-1}$. This result is generalized in Example 2.10, which shows that if $X, Y \overset{\text{i.i.d.}}{\sim} \text{Gam}(a, b)$, then the c.f. of $D = X - Y$ is $\varphi_D(t) = \left(1 + t^2/b^2\right)^{-a}$. From (I.6.3) and (I.7.9), $\text{Var}(D) = 2a/b^2$, so that setting $b = \sqrt{2a}$ gives a unit variance. The p.d.f. is symmetric about zero, which is obvious from the definition of D, and also because φ_D is real.

For $a \neq 1$, $\varphi_D(t)$ is not recognized to be the c.f. of any common random variable, so the inversion formulae need to be applied to compute the p.d.f. of D. To use (1.59), a finite value for the upper limit of integration is required, and a plot of the integrand for $a = 1$ and a few x-values confirms that 200 is adequate for reasonable accuracy. The program in Listing 1.3 can be used to compute the p.d.f. over a grid of x-values. As a decreases, the tails of the distribution become fatter. The reader is encouraged to confirm that the use of transformation (1.60) is problematic in this case. See also Example 7.19 for another way of computing the p.d.f. ■

```
function f = pdfinvgamdiff(xvec,a,b)
lo=1e-8; hi=200; tol=1e-8;
xl=length(xvec); f=zeros(xl,1);
for loop=1:xl
  x=xvec(loop); f(loop)=quadl(@gammdiffcf,lo,hi,tol,[],x,a,b) / pi;
end;

function I=gammdiffcf(tvec,x,a,b);
for ii=1:length(tvec)
  t=tvec(ii); cf=(1+(t/b).^2).^(-a); z =exp(-i*t*x).*cf; I(ii)=real(z);
end
```

Program Listing 1.3 Computes the p.d.f., based on (1.59) with an upper integral limit of 200, of the difference of two i.i.d. Gam(a, b) r.v.s at the values in the vector `xvec`

⊚ ***Example 1.24*** Let X_1, X_2 be continuous r.v.s with joint p.d.f. f_{X_1,X_2}, joint c.f. φ_{X_1,X_2}, and such that $\Pr(X_2 > 0) = 1$ and $\mathbb{E}[X_2] = \mu_2 < \infty$. Geary (1944) showed that the density of $R = X_1/X_2$ can be written

$$f_R(r) = \frac{1}{2\pi i}\int_{-\infty}^{\infty}\left[\frac{\partial\varphi_{X_1,X_2}(s,t)}{\partial t}\right]_{t=-rs}\mathrm{d}s. \tag{1.62}$$

To prove this, as in Daniels (1954), consider a new density

$$f_{Y_1,Y_2}(y_1,y_2) = \frac{y_2}{\mu_2} f_{X_1,X_2}(y_1,y_2),$$

and observe that

$$\begin{aligned}\int_0^{\infty}\int_{-\infty}^{\infty} f_{Y_1,Y_2}(y_1,y_2)\,\mathrm{d}y_1\,\mathrm{d}y_2 &= \frac{1}{\mu_2}\int_0^{\infty} y_2 \int_{-\infty}^{\infty} f_{X_1,X_2}(y_1,y_2)\,\mathrm{d}y_1\,\mathrm{d}y_2\\ &= \frac{1}{\mu_2}\int_0^{\infty} y_2 f_{X_2}(y_2)\,\mathrm{d}y_2 = 1.\end{aligned}$$

Next, define the r.v. $W = Y_1 - rY_2$, and use a result in Section I.8.2.6 (or a straightforward multivariate transformation) to get

$$f_W(w) = \int_0^{\infty} f_{Y_1,Y_2}(w + ry_2, y_2)\,\mathrm{d}y_2.$$

Then

$$f_W(0) = \int_0^{\infty} f_{Y_1,Y_2}(ry_2,y_2)\,\mathrm{d}y_2 = \frac{1}{\mu_2}\int_0^{\infty} y_2 f_{X_1,X_2}(ry_2,y_2)\,\mathrm{d}y_2 = \frac{1}{\mu_2} f_R(r),$$

where the last equality follows from (2.10), as derived later in Section 2.2. Hence, from (1.54),

$$f_R(r) = \mu_2 f_W(0) = \frac{\mu_2}{2\pi}\int_{-\infty}^{\infty}\varphi_W(s)\,\mathrm{d}s,$$

where

$$\varphi_W(s) = \mathbb{E}\left[\mathrm{e}^{isW}\right] = \mathbb{E}\left[\mathrm{e}^{isY_1 - isrY_2}\right] = \varphi_{Y_1,Y_2}(s,t)\big|_{t=-rs}.$$

It remains to show that

$$\varphi_{Y_1,Y_2}(s,t) = \frac{1}{i\mu_2}\frac{\partial}{\partial t}\varphi_{X_1,X_2}(s,t),$$

which follows because

$$\begin{aligned}\varphi_{Y_1,Y_2}(s,t) = \mathbb{E}\left[\mathrm{e}^{isY_1+itY_2)}\right] &= \int_0^{\infty}\int_{-\infty}^{\infty}\mathrm{e}^{isy_1+ity_2} f_{Y_1,Y_2}(y_1,y_2)\,\mathrm{d}y_1\,\mathrm{d}y_2\\ &= \int_0^{\infty}\int_{-\infty}^{\infty}\mathrm{e}^{isy_1+ity_2}\frac{y_2}{\mu_2} f_{X_1,X_2}(y_1,y_2)\,\mathrm{d}y_1\,\mathrm{d}y_2\\ &= \frac{1}{i\mu_2}\int_0^{\infty}\int_{-\infty}^{\infty}\frac{\partial}{\partial t}\left[\mathrm{e}^{isy_1+ity_2} f_{X_1,X_2}(y_1,y_2)\right]\mathrm{d}y_1\,\mathrm{d}y_2\end{aligned}$$

$$= \frac{1}{i\mu_2}\frac{\partial}{\partial t}\int_0^{\infty}\int_{-\infty}^{\infty}\left[e^{isy_1+ity_2}\right] f_{X_1,X_2}(y_1, y_2)\, dy_1\, dy_2$$

$$= \frac{1}{i\mu_2}\frac{\partial}{\partial t}\varphi_{X_1,X_2}(s,t),$$

and the exchange of differentiation and integration can be justified by the results given in Section I.A.3.4 and (1.46).

A similar derivation shows that the m.g.f. of W, if it exists, is

$$\mathbb{M}_W(s) = \frac{1}{\mu_2}\frac{\partial}{\partial t}\mathbb{M}_{X_1,X_2}(s,t)\Big|_{t=-rs}, \tag{1.63}$$

so that

$$f_R(r) = \mu_2 f_W(0), \tag{1.64}$$

where W has m.g.f. (1.63). ■

1.2.6 Inversion formulae for the c.d.f.

We begin with a result which has limited computational value, but is useful for establishing the uniqueness theorem and also can be used to establish (1.54).

For the c.f. φ_X of r.v. X with c.d.f. F_X, Lévy (1925) showed that, if F_X is continuous at the two points $a \pm h$, $h > 0$, then

$$\boxed{F_X(a+h) - F_X(a-h) = \lim_{T\to\infty}\frac{1}{\pi}\int_{-T}^{T}\frac{\sin(ht)}{t}e^{-ita}\varphi_X(t)\, dt.} \tag{1.65}$$

Proof (outline): This follows Wilks (1963, p. 116). We will require the identity

$$\sin(b) - \sin(c) = 2\sin\left(\frac{b-c}{2}\right)\cos\left(\frac{b+c}{2}\right) \tag{1.66}$$

from (I.A.30), and the results from Example I.A.21, namely

$$\int_{-T}^{T} t^{-1}\sin(bt)\sin(ct)\, dt = 0 \tag{1.67a}$$

and

$$\int_{-T}^{T} t^{-1}\sin(bt)\cos(ct)\, dt = 2\int_0^T t^{-1}\sin(bt)\cos(ct)\, dt. \tag{1.67b}$$

Let

$$G(a,T,h) = \frac{1}{\pi}\int_{-T}^{T}\frac{\sin(ht)}{t}e^{-ita}\varphi_X(t)\, dt$$

$$= \frac{1}{\pi}\int_{-T}^{T}\frac{\sin(ht)}{t}e^{-ita}\int_{-\infty}^{\infty}e^{itx}\, dF_X(x)\, dt$$

$$= \int_{-\infty}^{\infty} \frac{1}{\pi} \int_{-T}^{T} \frac{\sin(ht)}{t} \mathrm{e}^{-ita} \mathrm{e}^{itx} \, \mathrm{d}t \, \mathrm{d}F_X(x)$$

$$= \int_{-\infty}^{\infty} m(x, a, T, h) \, \mathrm{d}F_X(x),$$

where the exchange of integrals can be justified, and

$$\begin{aligned} m(x, a, T, h) &= \frac{1}{\pi} \int_{-T}^{T} \frac{\sin(ht)}{t} \mathrm{e}^{it(x-a)} \mathrm{d}t \\ &\overset{(1.31)}{=} \frac{1}{\pi} \int_{-T}^{T} \frac{\sin(ht)}{t} \left[\cos(t(x-a)) + i \sin(t(x-a))\right] \mathrm{d}t \\ &\overset{(1.67)}{=} \frac{2}{\pi} \int_{0}^{T} \frac{\sin(ht)}{t} \cos(t(x-a)) \, \mathrm{d}t. \end{aligned}$$

Now, with $b = t(x - a + h)$ and $c = t(x - a - h)$,

$$\frac{b-c}{2} = ht, \qquad \frac{b+c}{2} = t(x-a),$$

and (1.66) implies

$$m(x, a, T, h) = \frac{1}{\pi} \int_{0}^{T} \frac{\sin(x-a+h)t}{t} \mathrm{d}t - \frac{1}{\pi} \int_{0}^{T} \frac{\sin(x-a-h)t}{t} \mathrm{d}t.$$

It is easy to show (see Problem 1.14) that

$$\lim_{T\to\infty} m(x, a, T, h) = \begin{cases} 0, & \text{if } x \notin [a-h, a+h], \\ 1/2, & \text{if } x = a-h \text{ or } x = a+h, \\ 1, & \text{if } x \in (a-h, a+h), \end{cases} \tag{1.68}$$

so that, assuming the exchange of limit and integral and using the continuity of F_X at the two points $a \pm h$,

$$\begin{aligned} \lim_{T\to\infty} G(a, T, h) &= \int_{-\infty}^{\infty} \lim_{T\to\infty} m(x, a, T, h) \, \mathrm{d}F_X(x) \\ &= \int_{a-h}^{a+h} \mathrm{d}F_X(x) = F_X(a+h) - F_X(a-h), \end{aligned}$$

which is (1.65). ■

***Remarks*:**
(a) If two r.v.s X and Y have the same c.f., then, from (1.65),

$$F_X(a+h) - F_X(a-h) = F_Y(a+h) - F_Y(a-h)$$

for all pairs of points $(a \pm h)$ for which F_X and F_Y are continuous. This, together with the fact that F_X and F_Y are right-continuous c.d.f.s, implies that $F_X(x) = F_Y(x)$, i.e., the uniqueness theorem.

(b) It is also straightforward to derive the p.d.f. inversion formula (1.54) from (1.65); see Problem 1.15. ■

We now turn to expressions for the c.d.f. which lend themselves to computation. If the cumulant generating function exists, as is often the case, a useful integral expression for the c.d.f. is derived in Problem 1.12. More generally, Gil-Peleaz (1951) derived the inversion formula for continuous r.v. X with c.f. φ_X,

$$\boxed{F_X(x) = \frac{1}{2} + \frac{1}{2\pi}\int_0^\infty \frac{e^{itx}\varphi_X(-t) - e^{-itx}\varphi_X(t)}{it}\,dt,} \tag{1.69}$$

which can be proved as follows.

Proof (outline): From the definition of φ_X and expression (1.32) for $\sin z$, let

$$\begin{aligned} I(a,b) &= \int_a^b \frac{e^{itx}\varphi_X(-t) - e^{-itx}\varphi_X(t)}{it}\,dt \\ &= \int_a^b \int_{-\infty}^\infty \frac{e^{-it(y-x)} - e^{it(y-x)}}{it} f_X(y)\,dy\,dt \\ &= -2\int_a^b \int_{-\infty}^\infty \frac{\sin(t(y-x))}{t} f_X(y)\,dy\,dt \end{aligned}$$

for $0 < a < b < \infty$. Also let

$$s(z) = \frac{2}{\pi}\int_0^\infty \frac{\sin tz}{t}\,dt = \begin{cases} -1, & \text{if } z < 0, \\ 0, & \text{if } z = 0, \\ 1, & \text{if } z > 0 \end{cases}$$

(which easily follows from (1.55) after substituting $y = tz$), so that

$$\begin{aligned} \int_{-\infty}^\infty s(y-x)\,f_X(y)\,dy &= \int_{-\infty}^x s(y-x)\,f_X(y)\,dy + \int_x^\infty s(y-x)\,f_X(y)\,dy \\ &= -F_X(x) + (1 - F_X(x)) = 1 - 2F_X(x). \end{aligned}$$

Assuming the order of integration can be changed,[8]

$$\frac{1}{\pi} I(a,b) = -\int_{-\infty}^\infty \frac{2}{\pi}\int_a^b \frac{\sin t(y-x)}{t}\,dt\,f_X(y)\,dy$$

and

$$\lim_{\substack{a\to 0\\ b\to\infty}} \frac{1}{\pi} I(a,b) = -\int_{-\infty}^\infty s(y-x)\,f_X(y)\,dy = 2F_X(x) - 1,$$

from which (1.69) follows. ■

[8] To verify that the order can be changed, one needs

$$\left|\int_a^b \frac{\sin t(y-x)}{t}\,dt\right| \le \int_a^b \left|\frac{\sin t(y-x)}{t}\right| dt \le \int_a^b \frac{1}{t}\,dt \le \int_a^b \frac{1}{a}\,dt \le \frac{b-a}{a}.$$

For numeric inversion purposes, an expression related to (1.69) is even more useful. As $\sin(-\theta) = -\sin(\theta)$,

$$\begin{aligned} e^{-itx}\varphi_X(t) &= \int_{-\infty}^{\infty} e^{-it(x-y)} f_X(y)\, dy \\ &= \int_{-\infty}^{\infty} \cos(t(x-y)) f_X(y)\, dy - i \int_{-\infty}^{\infty} \sin(t(x-y)) f_X(y)\, dy \end{aligned}$$

and

$$\begin{aligned} I(0,\infty) &= +2 \int_0^{\infty} \frac{1}{t} \int_{-\infty}^{\infty} \sin(t(x-y)) f_X(y)\, dy\, dt \\ &= -2 \int_0^{\infty} \frac{1}{t} \operatorname{Im}\left(e^{-itx}\varphi_X(t)\right) dt, \end{aligned}$$

recalling $\operatorname{Im}(a+bi) = b$. That is,

$$F_X(x) = \frac{1}{2} - \frac{1}{\pi} \int_0^{\infty} g_x(t)\, dt, \tag{1.70}$$

where, using (1.36),

$$g_x(t) = \frac{\operatorname{Im} z_x(t)}{t} = \frac{|z_x(t)| \sin(\arg z_x(t))}{t} \quad \text{and} \quad z_x(t) = e^{-itx}\varphi_X(t).$$

As in (1.60) and (1.61), (1.70) can sometimes be more effectively computed as either

$$F_X(x) = \frac{1}{2} - \frac{1}{\pi} \int_0^{1} g_x\left(\frac{1-u}{u}\right) u^{-2}\, du \tag{1.71}$$

or

$$F_X(x) = \frac{1}{2} - \frac{1}{\pi} \int_0^{1} g_x\left(\frac{u}{1-u}\right) (1-u)^{-2}\, du. \tag{1.72}$$

⊖ ***Example 1.25*** For $n, \theta \in \mathbb{R}_{>0}$, let X be the r.v. with

$$\varphi_X(t) = (1-2it)^{-n/2} \exp\left\{\frac{it\theta}{1-2it}\right\}, \tag{1.73}$$

which is the c.f. of a so-called *noncentral* χ^2 *distribution* with n degrees of freedom and noncentrality parameter θ. Let $z_x(t) = e^{-itx}\varphi_X(t)$. From (1.37) and the results in Examples 1.14 and 1.15,

$$\begin{aligned} |z_x(t)| &= \left|e^{-itx}\right| |\varphi_X(t)| = |\varphi_X(t)| = \left|(1-2it)^{-n/2}\right| \left|\exp\left\{\frac{it\theta}{1-2it}\right\}\right| \\ &= \left(1+4t^2\right)^{-n/4} \exp\left(-\frac{2\theta t^2}{1+4t^2}\right), \end{aligned}$$

and

$$\arg z_x(t) = \arg e^{-itx} + \arg \varphi_X(t) = -tx + \frac{n}{2}\arctan(2t) + \frac{t\theta}{1+4t^2},$$

so that (1.70) yields the impressive looking

$$F_X(x) = \frac{1}{2} - \frac{1}{\pi}\int_0^\infty \frac{\sin\left(-tx + \frac{n}{2}\arctan(2t) + t\theta/(1+4t^2)\right)}{t\left(1+4t^2\right)^{n/4}\exp\left(2\theta t^2/(1+4t^2)\right)}\mathrm{d}t. \tag{1.74}$$

(Other, more preferable, methods to compute F_X will be discussed in Chapter 10.) The next section discusses a method of computing the p.d.f. ■

1.3 Use of the fast Fourier transform

The fast Fourier transform (FFT) is a numerical procedure which vastly speeds up the calculation of the discrete Fourier transform (DFT), to be discussed below. The FFT is rightfully considered to be one of the most important scientific developments in the twentieth century because of its applicability to a vast variety of fields, from medicine to financial option pricing (see, for example, Carr and Madan, 1999), not to mention signal processing and electrical engineering. For our purposes, the application of interest is the fast numeric inversion of the characteristic function of a random variable for a (potentially large) grid of points. For a small number of points, the numeric integration methods of Sections 1.2.5 and 1.2.6 are certainly adequate, but if, say, a plot of the p.d.f. is required, or, more importantly, a likelihood calculation (which involves the evaluation of the p.d.f. at all the observed data points), then such methods are far too slow for practical use. For example, in financial applications, one often has thousands of observations.

A full appreciation of the FFT requires a book-length treatment, as well as a deeper understanding of the mathematics of Fourier analysis. Our goal here is far more modest; we wish to gain a heuristic, basic understanding of Fourier series, Fourier transforms, and the FFT, and concentrate on their use for inverting the c.f.

1.3.1 Fourier series

We begin with a rudimentary discussion of Fourier series, based on the presentations in Dyke (2001) and Schiff (1999), which are both highly recommended starting points for students who wish to pursue the subject further. More advanced but still accessible treatments are provided in Bachman, Narici and Beckenstein (2000) and Beerends *et al.* (2003).

Consider the *periodic function* shown in Figure 1.6, as a function of time t in, say, seconds. The *period* is the time (in seconds) it takes for the function to repeat, and its *frequency* is the number of times the wave repeats in a given time unit, say one second. Because the function in Figure 1.6 (apparently) repeats every one second, its *fundamental period* is $T = 1$, and its *fundamental frequency* is $f = 1/T$. The trigonometric functions sin and cos are natural period functions, and form the basis into which many periodic functions can be decomposed.

Let θ denote the angle in the usual depiction of the unit circle (similar to that in Figure 1.2), measured in radians (the arc length), of which the total length around the circle is 2π. Let θ be the function of time t given by $\theta(t) = \omega t$, so that ω is the *angular velocity*. If we take $\omega = 2\pi/T$ and let t move from 0 to T, then $\theta(t) = \omega t$ will move from zero to 2π, and thus $\cos(\omega t)$ will make one complete cycle, and similarly for sin. Now with $n \in \mathbb{N}$, $\cos(n\omega t)$ will make n complete cycles, where the angular velocity is $n\omega$. This (hopefully) motivates one to consider using $\cos(n\omega t)$ and $\sin(n\omega t)$, for $n = 0, 1, \ldots,$ as a 'basis', in some sense, for decomposing periodic functions such as that shown in Figure 1.6. That is, we wish to express a periodic function g with fundamental period T as

$$g(t) = \sum_{n=0}^{\infty} \left[a_n \cos(n\omega t) + b_n \sin(n\omega t) \right], \quad \omega = \frac{2\pi}{T}, \tag{1.75}$$

where the a_i and b_i represent the *amplitudes* of the components. Notice that, for $n = 0$, $\cos(n\omega t) = 1$ and $\sin(n\omega t) = 0$, so that a_0 is the constant term of the function, and there is no b_0 term. The function shown in Figure 1.6 is the weighted sum of four sin terms. Notice that if n is not an integer, then $\cos(n\omega t)$ will not coincide at $t = 0$ and $t = T$, and so will not be of value for decomposing functions with fundamental period T.

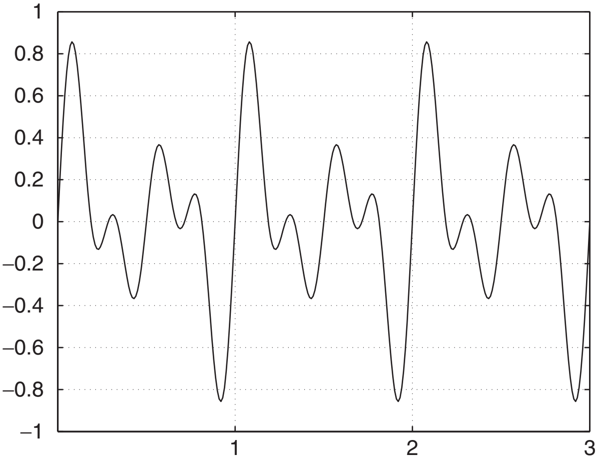

Figure 1.6 The periodic function (1.75) as a function of t, with all $a_i = 0$, $b_1 = 0.1$, $b_2 = 0.4$, $b_3 = 0.2$, $b_4 = 0.3$, and fundamental period $T = 1$

Recall from linear algebra the concepts of basis, span, dimension, orthogonality, linear independence, and orthonormal basis. In particular, for a given set S of vectors in $\mathbb{R}^n$, an orthonormal basis B can be constructed (e.g., via the Gramm–Schmidt process), such that any vector v which belongs to the span of S can be uniquely represented by the elements in B. The natural orthonormal basis for $\mathbb{R}^n$ is the set of unit vectors

$\mathbf{e}_1 = (1, 0, \ldots, 0)'$, $\mathbf{e}_2 = (0, 1, 0, \ldots, 0)'$, $\ldots$, $\mathbf{e}_n = (0, \ldots, 0, 1)'$, which are orthonormal because the inner product (in this case, the dot product) of $\mathbf{e}_i$ with itself is one, i.e., $\langle \mathbf{e}_i, \mathbf{e}_i \rangle = 1$, and that of $\mathbf{e}_i$ and $\mathbf{e}_j$, $i \neq j$, is zero. If $\mathbf{v} \in \mathbb{R}^n$, then it is trivial to see that, with the $\mathbf{e}_i$ so defined,

$$\mathbf{v} = \sum_{i=1}^{n} \langle \mathbf{v}, \mathbf{e}_i \rangle \, \mathbf{e}_i . \tag{1.76}$$

However, (1.76) holds for any orthonormal basis of $\mathbb{R}^n$. To illustrate, take $n = 3$ and the nonorthogonal basis $\mathbf{a}_1 = (1, 0, 0)'$, $\mathbf{a}_2 = (1, 1, 0)'$, $\mathbf{a}_3 = (1, 1, 1)'$, and use Gramm–Schmidt to orthonormalize this; in Matlab, this is done with

```
A=[1,0,0; 1,1,0; 1,1,1]', oA=orth(A)
```

Then type

```
v=[1,5,-3]'; a=zeros(3,1);
for i=1:3, a = a + ( v' * oA(:,i) ) * oA(:,i) ; end, a
```

to verify the claim.

In the Fourier series setting, it is helpful and common to start with taking $T = 2\pi$, so that $\omega = 1$, and assuming that f is defined over $[-\pi, \pi]$, and periodic, such that $g(x) = g(x + T) = g(x + 2\pi)$, $x \in \mathbb{R}$. (This range is easily extended to $[-r, r]$ for any $r > 0$, or to $[0, T]$, for any $T > 0$, as seen below.) Then we are interested in the sequence of functions

$$\{2^{-1/2}, \cos(n\cdot), \sin(n\cdot),\ n \in \mathbb{N}\},$$

because it forms an *infinite orthonormal basis*[9] for the set of (piecewise) continuous functions on $[-\pi, \pi]$ with respect to the inner product

$$\langle f, g \rangle := \frac{1}{\pi} \int_{-\pi}^{\pi} f(t)\overline{g(t)} \, \mathrm{d}t, \tag{1.77}$$

where $\overline{g}$ is the complex conjugate of function g. It is straightforward to verify orthonormality of these functions, the easiest of which is

$$\left\langle \frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}} \right\rangle = \frac{1}{\pi} \int_{-\pi}^{\pi} \frac{1}{2} \, \mathrm{d}t = 1.$$

The others require the use of basic trigonometric identities. For example, as $\cos^2(x) + \sin^2(x) = 1$ and, from (I.A.23b), $\cos(x + y) = \cos x \cos y - \sin x \sin y$, we have $\cos(2x) = \cos^2(x) - \sin^2(x) = 1 - 2\sin^2(x)$, implying

$$\big\langle \sin(nt), \sin(nt) \big\rangle = \frac{1}{\pi} \int_{-\pi}^{\pi} \sin^2(nt) \, \mathrm{d}t = \frac{1}{2\pi} \int_{-\pi}^{\pi} (1 - \cos(2nt)) \, \mathrm{d}t = 1.$$

[9] The concept of an infinite orthonormal basis is the cornerstone of the theory of so-called Hilbert spaces, and usually discussed in the framework of functional analysis. This is a fundamentally important branch of mathematics, with numerous real applications, including many in probability and statistics. Very accessible introductions are given by Light (1990) and Ha (2006).

Again using (I.A.23b), and the fact that $\cos(t) = \cos(-t)$,

$$\cos(m+n) = \cos(m)\cos(n) - \sin(m)\sin(n),$$
$$\cos(m-n) = \cos(m)\cos(n) + \sin(m)\sin(n),$$

and summing,

$$\cos(m+n) + \cos(m-n) = 2\cos(m)\cos(n),$$

we see, for $m \neq n$, that

$$\begin{aligned}\left\langle \cos(mt), \cos(nt) \right\rangle &= \frac{1}{2\pi}\int_{-\pi}^{\pi}\left[\cos((m+n)t) + \cos((m-n)t)\right]\mathrm{d}t \\ &= \frac{1}{2\pi}\left.\frac{\sin((m+n)t)}{m+n} + \frac{\sin((m-n)t)}{m-n}\right|_{-\pi}^{\pi} = 0.\end{aligned}$$

The reader is encouraged to verify similarly that $\langle\cos(nt), \cos(nt)\rangle = 1$, $\langle 1, \cos(nt)\rangle = \langle 1, \sin(nt)\rangle = 0$, $\langle\sin(mt), \cos(nt)\rangle = 0$, and $\langle\cos(mt), \sin(nt)\rangle = 0$. Thus, based on (1.76), it would appear that

$$a_0 = \left\langle g, \frac{1}{\sqrt{2}}\right\rangle = \frac{1}{\pi}\int_{-\pi}^{\pi}\frac{1}{\sqrt{2}}\,g(x)\,\mathrm{d}x,$$

$$a_n = \frac{1}{\pi}\int_{-\pi}^{\pi} g(t)\cos(nt)\,\mathrm{d}t, \quad \text{and} \quad b_n = \frac{1}{\pi}\int_{-\pi}^{\pi} g(t)\sin(nt)\,\mathrm{d}t, \quad n = 1, 2, \ldots,$$

and the Fourier series of g is $g(t) = a_0/\sqrt{2} + \sum_{n=1}^{\infty}\left[a_n\cos(nt) + b_n\sin(nt)\right]$. Equivalently, and often seen, one writes

$$g(t) = \frac{a_0}{2} + \sum_{n=1}^{\infty}\left[a_n\cos(nt) + b_n\sin(nt)\right], \tag{1.78}$$

with

$$a_n = \frac{1}{\pi}\int_{-\pi}^{\pi} g(t)\cos(nt)\,\mathrm{d}t \quad \text{and} \quad b_n = \frac{1}{\pi}\int_{-\pi}^{\pi} g(t)\sin(nt)\,\mathrm{d}t, \quad n = 0, 1, 2, \ldots, \tag{1.79}$$

so that the expression for a_n includes $n = 0$ (and, as $\sin(0) = 0$, $b_0 = 0$). The values of the a_i and b_i can be informally verified by multiplying the expression in (1.78) by, say, $\cos(nt)$, and integrating termwise.

***Remarks*:**
(a) The rigorous justification of termwise integration in (1.78) requires more mathematical sophistication, one requirement of which is that $\lim_{i\to\infty}\langle\mathbf{v}, \mathbf{e}_i\rangle = 0$, where the

$\mathbf{e}_i$ form an infinite orthonormal basis for some space V, and $\mathbf{v} \in V$. As a practical matter, the decomposition (1.75) will involve only a finite number of terms, with all the higher-frequency components being implicitly set to zero.
(b) Recall that an even function is symmetric about zero, i.e., $g : D \mapsto \mathbb{R}$ is even if $g(-x) = g(x)$ for all $x \in D$, and g is odd if $g(-x) = -g(x)$ (see the remarks after Example I.A.27. Thus, cos is even, and sin is odd, which implies that, if g is an even function and it has a Fourier expansion (1.75), then it will have no sin terms, i.e., all the $b_i = 0$, and likewise, if g is odd, then $a_i = 0$, $i \geq 1$. ■

⊖ ***Example 1.26*** Let g be the periodic function with $g(t) = t^2$ for $-r \leq t \leq r$, $r > 0$, and such that $g(t) = g(t + 2r)$. To deal with the fact that the range is relaxed such that r need not be π, observe that if $-r \leq t \leq r$, then $-\pi < \pi t/r < \pi$, so that

$$s(t) := \sin\left(\frac{\pi}{r}t\right) = \sin\left(\frac{\pi}{r}t + 2\pi\right) = \sin\left(\frac{\pi}{r}(t + 2r)\right) = s(t + 2r),$$

i.e., the function s has period $2r$, and likewise for $c(t) := \cos(\pi t/r)$. Thus,

$$\begin{aligned} g(t) &= \frac{a_0}{2} + \sum_{n=1}^{\infty} [a_n s(nt) + b_n c(nt)] \\ &= \frac{a_0}{2} + \sum_{n=1}^{\infty} \left[a_n \sin\left(\frac{n\pi t}{r}\right) + b_n \cos\left(\frac{n\pi t}{r}\right) \right], \end{aligned} \tag{1.80}$$

where

$$a_n = \frac{1}{r} \int_{-r}^{r} g(t) \cos\left(\frac{n\pi t}{r}\right) \mathrm{d}t, \quad b_n = \frac{1}{r} \int_{-r}^{r} g(t) \sin\left(\frac{n\pi t}{r}\right) \mathrm{d}t, \tag{1.81}$$

which, in this case, are

$$\begin{aligned} a_0 &= \frac{1}{r} \int_{-r}^{r} t^2 \, \mathrm{d}t = \frac{2}{3} r^2, \\ a_n &= \frac{1}{r} \int_{-r}^{r} t^2 \cos\left(\frac{n\pi t}{r}\right) \mathrm{d}t = \frac{4r^2 \cos(\pi n)}{\pi^2 n^2} = \frac{4r^2 (-1)^n}{\pi^2 n^2}, \quad n = 1, 2, \ldots, \\ b_n &= 0, \end{aligned}$$

having used repeated integration by parts for a_n, and where the b_n are zero because f is an even function. Thus,

$$g(t) = \frac{r^2}{3} + \frac{4r^2}{\pi^2} \sum_{n=1}^{\infty} \frac{(-1)^n}{n^2} \cos\left(\frac{n\pi t}{r}\right). \tag{1.82}$$

Figure 1.7 plots (1.82) for $r = 1$, truncating the infinite sum to have N terms. Series (1.82) with $t = 0$ and $r = \pi$ implies $0^2 = \pi^2/3 + 4 \sum_{n=1}^{\infty} (-1)^n n^{-2}$, i.e.,

$$\frac{\pi^2}{12} = \frac{1}{1^2} - \frac{1}{2^2} + \frac{1}{3^2} - \cdots,$$

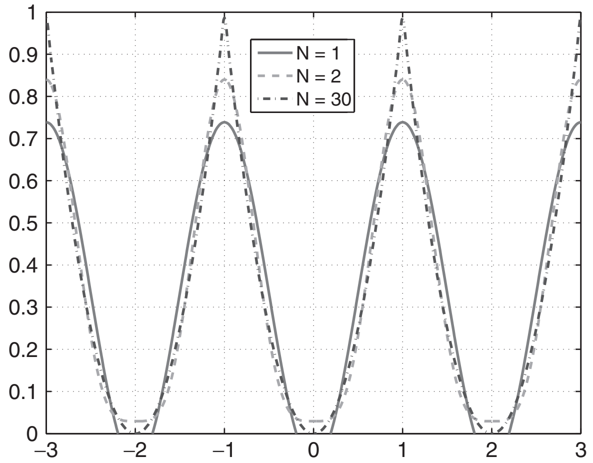

Figure 1.7 Fourier series (1.82) for $r = 1$, truncating the infinite sum to have N terms

and, with $t = \pi = r$, $\pi^2 = \pi^2/3 + 4\sum_{n=1}^{\infty} n^{-2}$, i.e.,

$$\boxed{\frac{\pi^2}{6} = \sum_{n=1}^{\infty} \frac{1}{n^2},}$$

which was stated without proof in (I.A.88). ■

Example 1.27 Assume that $g(t) = t^2 - t$ for $t \in [0, T]$, and $g(t) = g(t + T)$. This is plotted in Figure 1.8 for $T = 2$, using the Matlab code

```
g=[]; T=2;
for t=0:0.01:4*T
  u=t; while u>T, u=u-T; end,  g=[g u^2-u];
end
h=plot(0:0.01:4*T,g,'r.'), set(h,'MarkerSize',3)
```

Similar to Example 1.26, it is straightforward to verify that, for a periodic function defined on $t \in [0, T]$ and such that $g(t) = g(t + T)$,

$$g(t) = \frac{a_0}{2} + \sum_{n=1}^{\infty} \left[a_n \cos(n\omega t) + b_n \sin(n\omega t)\right], \quad \omega = \frac{2\pi}{T}, \tag{1.83}$$

with, for $n = 0, 1, 2, \ldots,$

$$a_n = \frac{2}{T}\int_0^T g(t)\cos(n\omega t)\,\mathrm{d}t, \quad b_n = \frac{2}{T}\int_0^T g(t)\sin(n\omega t)\,\mathrm{d}t. \tag{1.84}$$

In this case, a bit of straightforward work yields

$$a_0 = \frac{T(2T-3)}{3}, \quad a_n = \frac{T^2}{\pi^2 n^2}, \quad b_n = \frac{(1-T)T}{n\pi}, \quad n = 1, 2, \ldots, \tag{1.85}$$

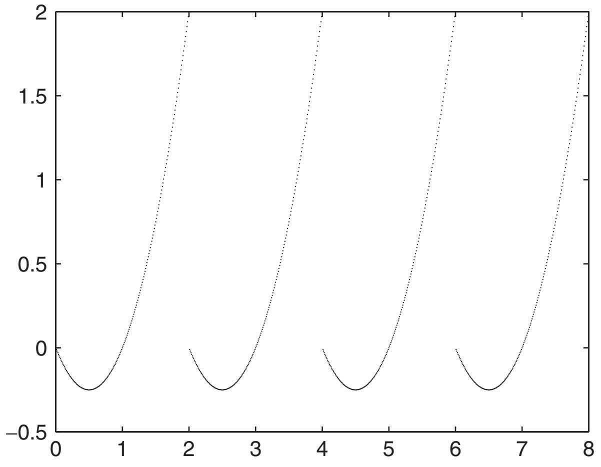

Figure 1.8 The periodic function $g(t) = t^2 - t$ for $t \in [0, T]$, $T = 2$, with $g(t) = f(t + T)$

and Figure 1.9 plots (1.83) with (1.85), for several values of N, the truncation value of the infinite sum. We see that, compared to the function in Example 1.26, N has to be quite large to obtain reasonable accuracy. ■

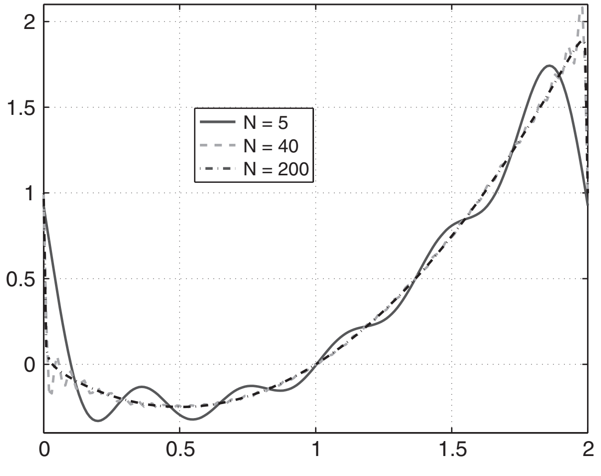

Figure 1.9 Approximation (1.83) and (1.85) to the function shown in Figure 1.8, truncating the infinite sum to have N terms

From (1.32) and (1.83),

$$\begin{aligned} g(t) &= \frac{a_0}{2} + \sum_{n=1}^{\infty} \left[\frac{a_n}{2} \left(e^{in\omega t} + e^{-in\omega t} \right) + \frac{b_n}{2i} \left(e^{in\omega t} - e^{-in\omega t} \right) \right] \\ &= \frac{a_0}{2} + \frac{1}{2} \sum_{n=1}^{\infty} \left[(a_n - ib_n) e^{in\omega t} + (a_n + ib_n) e^{-in\omega t} \right], \end{aligned}$$

and from (1.84), for $n \geq 1$,

$$a_n = \frac{2}{T}\int_0^T g(t)\cos(n\omega t)\,\mathrm{d}t = \frac{1}{T}\int_0^T g(t)\,\mathrm{e}^{in\omega t}\,\mathrm{d}t + \frac{1}{T}\int_0^T g(t)\,\mathrm{e}^{-in\omega t}\,\mathrm{d}t,$$

$$ib_n = \frac{2i}{T}\int_0^T g(t)\sin(n\omega t)\,\mathrm{d}t = \frac{1}{T}\int_0^T g(t)\,\mathrm{e}^{in\omega t}\,\mathrm{d}t - \frac{1}{T}\int_0^T g(t)\,\mathrm{e}^{-in\omega t}\,\mathrm{d}t,$$

so that

$$a_n - ib_n = \frac{2}{T}\int_0^T g(t)\,\mathrm{e}^{-in\omega t}\,\mathrm{d}t, \quad a_n + ib_n = \frac{2}{T}\int_0^T g(t)\,\mathrm{e}^{in\omega t}\,\mathrm{d}t,$$

yielding

$$g(t) = \frac{a_0}{2} + \sum_{n=1}^{\infty}\left[A_n\mathrm{e}^{in\omega t} + B_n\mathrm{e}^{-in\omega t}\right],$$

where

$$A_n = \frac{a_n - ib_n}{2} = \frac{1}{T}\int_0^T g(t)\,\mathrm{e}^{-in\omega t}\,\mathrm{d}t, \quad B_n = \frac{a_n + ib_n}{2} = \frac{1}{T}\int_0^T g(t)\,\mathrm{e}^{in\omega t}\,\mathrm{d}t.$$

The two terms in the sum are easily combined to give the *complex Fourier series expansion*

$$\boxed{g(t) = \sum_{n=-\infty}^{\infty} C_n\mathrm{e}^{in\omega t}, \quad C_n = \frac{1}{T}\int_0^T g(t)\,\mathrm{e}^{-in\omega t}\,\mathrm{d}t, \quad \omega = \frac{2\pi}{T}.} \tag{1.86}$$

(Note that the term with $n = 0$ is correct.) As a check, note that, for $k = (n - m) \neq 0$,

$$\int_0^T \mathrm{e}^{i(n-m)\omega t}\,\mathrm{d}t = \int_0^T \exp\left(\frac{2\pi ikt}{T}\right)\mathrm{d}t = \int_0^T \cos\left(\frac{2\pi kt}{T}\right)\mathrm{d}t + i\int_0^T \sin\left(\frac{2\pi kt}{T}\right)\mathrm{d}t$$
$$= \frac{T}{2\pi k}\sin(2\pi k) + i\frac{T}{2\pi k}(1 - \cos 2\pi k) = 0,$$

so that

$$\int_0^T \mathrm{e}^{i(n-m)\omega t}\,\mathrm{d}t = \begin{cases} T, & \text{if } n = m, \\ 0, & \text{if } n \neq m, \end{cases}$$

i.e., the expression for C_m from (1.86) is (assuming the validity of exchanging sum and integral)

$$\frac{1}{T}\int_0^T g(t)\,\mathrm{e}^{-im\omega t}\,\mathrm{d}t = \frac{1}{T}\int_0^T \left(\sum_{n=-\infty}^{\infty} C_n\mathrm{e}^{in\omega t}\right)\mathrm{e}^{-im\omega t}\,\mathrm{d}t$$
$$= \frac{1}{T}\sum_{n=-\infty}^{\infty} C_n\int_0^T \mathrm{e}^{i(n-m)\omega t}\,\mathrm{d}t = C_m,$$

as expected. It should be clear from the form of (1.86) that, if we started with g defined on $[-r, r]$ for $r > 0$, periodic with period $T = 2r$ such that $g(x) = g(x + 2r)$, then

$$g(t) = \sum_{n=-\infty}^{\infty} C_n e^{in\omega t}, \quad C_n = \frac{1}{2r}\int_{-r}^{r} g(t)\, e^{-in\omega t}\, dt, \quad \omega = \frac{2\pi}{T} = \frac{\pi}{r}, \tag{1.87}$$

which the reader is encouraged to check using (1.80) and (1.81).

1.3.2 Discrete and fast Fourier transforms

Based on (1.86), it would (informally) seem natural to construct a discrete version of the C_n by taking

$$G_n = \frac{1}{T}\sum_{t=0}^{T-1} g_t e^{-2\pi int/T}, \quad n = 0, \ldots, T-1, \tag{1.88}$$

where $g_t = g(t)$, with the subscript serving as a reminder that g is evaluated at discrete points. For vector $\mathbf{g} = (g_0, \ldots, g_{T-1})$, we will write $\mathbf{G} = \mathcal{F}(\mathbf{g})$, where $\mathbf{G} = (G_0, \ldots, G_{T-1})$. Expression (1.88) is the definition of the *(forward, complex) discrete Fourier transform*, or DFT. The T values of g_t can be recovered from the *inverse discrete Fourier transform*, defined as

$$g_t = \sum_{n=0}^{T-1} G_n e^{2\pi int/T}, \quad t = 0, \ldots, T-1, \tag{1.89}$$

and we write $\mathbf{g} = \mathcal{F}^{-1}(\mathbf{G})$. The programs in Listings 1.4 and 1.5 compute (1.88) and (1.89), respectively.

```
function G=dft(g)
  T=length(g); z=exp(-2*pi*i/T); G=zeros(T,1); t=0:(T-1);
  for n=0:(T-1),  w=z.^(t*n); G(n+1)=sum(g.*w'); end, G=G/T;
```

Program Listing 1.4 The forward discrete Fourier transform (1.88)

```
function g=idft(G)
  T=length(G); z=exp(2*pi*i/T); g=zeros(T,1); t=0:(T-1);
  for n=0:(T-1),  w=z.^(t*n); g(n+1)=sum(G.*w'); end
```

Program Listing 1.5 The inverse discrete Fourier transform (1.89)

Function $\mathcal{F}^{-1}$ is the inverse of $\mathcal{F}$ in the sense that, for a set of complex numbers $\mathbf{g} = (g_0, \ldots, g_{T-1})$, $\mathbf{g} = \mathcal{F}^{-1}(\mathcal{F}(\mathbf{g}))$. To see this, let $\mathbf{G} = \mathcal{F}(\mathbf{g})$ and define $z_T := e^{2\pi i/T}$. From (1.89) and (1.88), the tth element of $\mathcal{F}^{-1}(\mathbf{G})$, for $t \in \{0, 1, \ldots, T-1\}$, is

$$\sum_{n=0}^{T-1} G_n e^{2\pi int/T} = \sum_{n=0}^{T-1} G_n z_T^{nt} = \sum_{n=0}^{T-1}\left(\frac{1}{T}\sum_{s=0}^{T-1} g_s z_T^{-ns}\right) z_T^{nt} = \frac{1}{T}\sum_{s=0}^{T-1} g_s \left(\sum_{n=0}^{T-1}\left(z_T^{t-s}\right)^n\right).$$

Now recalling (1.38), we see that, if $s = t$, then the inner sum is T, and zero otherwise, so that the tth element of $\mathcal{F}^{-1}(\mathbf{G})$ is g_t. This holds for all $t \in \{0, 1, \ldots, T-1\}$, so that $\mathcal{F}^{-1}(\mathcal{F}(\mathbf{g})) = \mathbf{g}$. This is numerically illustrated in the program in Listing 1.6.

```
T=2^8; g=randn(T,1)+i*randn(T,1); G=dft(g); gg=idft(G); max(abs(g-gg))
```

Program Listing 1.6 Illustrates numerically that $\mathbf{g} = \mathcal{F}^{-1}(\mathcal{F}(\mathbf{g}))$

The *fast Fourier transform*, or FFT, is the DFT, but calculated in a judicious manner which capitalizes on redundancies. Numerous algorithms exist which are optimized for various situations, the simplest being when $T = 2^p$ is a power of 2, which we illustrate. As in Černý (2004, Section 7.4.1), splitting the DFT (1.88) into even and odd indices gives, with $0 \le n \le T-1$ and $z_T := \mathrm{e}^{-2\pi i/T}$,

$$\begin{aligned} G_n = \frac{1}{T}\sum_{t=0}^{T-1} g_t \left(z_T^n\right)^t &= \frac{1}{T}\sum_{t=0}^{(T/2)-1} g_{2t}\left(z_T^n\right)^{2t} + \frac{1}{T}\sum_{t=0}^{(T/2)-1} g_{2t+1}\left(z_T^n\right)^{2t+1} \\ &= \frac{1}{T}\sum_{t=0}^{(T/2)-1} g_{2t}\left(z_T^{2n}\right)^{t} + \frac{1}{T}z_T^n\sum_{t=0}^{(T/2)-1} g_{2t+1}\left(z_T^{2n}\right)^{t} \\ &=: S_1 + S_2. \end{aligned}$$

As $z_T^{T/2} = \left(\mathrm{e}^{-2\pi i/T}\right)^{T/2} = \mathrm{e}^{-\pi i} = -1$ and $z_T^T = \mathrm{e}^{-2\pi i} = 1$, the DFT of the $(n + T/2)$th element is

$$\begin{aligned} G_{n+T/2} &= \frac{1}{T}\sum_{t=0}^{(T/2)-1} g_{2t}\left(z_T^{n+T/2}\right)^{2t} + \frac{1}{T}\sum_{t=0}^{(T/2)-1} g_{2t+1}\left(z_T^{n+T/2}\right)^{2t+1} \\ &= \frac{1}{T}\sum_{t=0}^{(T/2)-1} g_{2t}\left(z_T^{2n}z_T^{T}\right)^{t} + \frac{1}{T}\left(z_T^n z_T^{T/2}\right)\sum_{t=0}^{(T/2)-1} g_{2t+1}\left(z_T^{2n}z_T^{T}\right)^{t} \\ &= \frac{1}{T}\sum_{t=0}^{(T/2)-1} g_t\left(z_T^{2n}\right)^{t} - \frac{1}{T}z_T^n\sum_{t=0}^{(T/2)-1} g_t\left(z_T^{2n}\right)^{t} \\ &= S_1 - S_2, \end{aligned}$$

so that the sums S_1 and S_2 only need to be computed for $n = 0, 1, \ldots, (T/2) - 1$. Recalling that T is a power of 2, this procedure can be applied to S_1 and S_2, and so on, p times. It is easy to show that this recursive method will require, approximately, of the order of $T(\log_2(T) + 1)$ complex-number multiplications, while the DFT requires of the order of T^2. For $T = 2^6 = 64$ data points, the ratio $T/\log_2(T)$ is about 10, i.e., the FFT is 10 times faster, while with $T = 2^{10} = 1024$ data points, the FFT is about 100 times faster.

In Matlab (version 7.1), the command `fft` is implemented in such a way that it computes (what we, and most books, call) the inverse Fourier transform (1.89), while the command `ifft` computes the DFT (1.88).[10] This is illustrated in the code in Listing 1.7.

```
T=2^10; g=randn(T,1)+i*randn(T,1);
s=cputime; G1=dft(g);  t1=cputime-s % how long does the direct DFT take?
s=cputime; for r=1:1000, G2=ifft(g); end, t2=(cputime-s)/1000
timeratio=t1/t2  % compare to the FFT
thedifference = max(abs(G1-G2)) % is zero, up to roundoff error
```

Program Listing 1.7 The DFT (1.88) is computed via the FFT by calling Matlab's `ifft` command

1.3.3 Applying the FFT to c.f. inversion

Similar to Černý (2004, Section 7.5.3), let X be a continuous random variable with p.d.f. f_X and c.f. φ_X, so that, for $\ell, u \in \mathbb{R}$ and $T \in \mathbb{N}$,

$$\varphi_X(s) = \int_{-\infty}^{\infty} \mathrm{e}^{isx} f_X(x)\,\mathrm{d}x \approx \int_{\ell}^{u} \mathrm{e}^{isx} f_X(x)\,\mathrm{d}x \approx \sum_{n=0}^{T-1} \mathrm{e}^{isx_n} P_n\ , \tag{1.90}$$

with

$$P_n := f_X(x_n)\,\Delta x, \quad x_n := \ell + n(\Delta x), \quad \Delta x := \frac{u-\ell}{T},$$

whereby the second approximation is valid (in the sense that it can be made arbitrarily accurate by choosing T large enough) from the definition of the Riemann integral, and the first approximation is valid (in the sense that it can be made arbitrarily accurate by choosing $-\ell$ and u large enough) because, as an improper integral, φ_X exists, and we *assume* that f is such that $\lim_{x\to\infty} f_X(x) = \lim_{x\to-\infty} f_X(x) = 0$. (While virtually all p.d.f.s in practice will satisfy this assumption, it is not true that if f_X is a proper continuous p.d.f., then $\lim_{x\to\infty} f_X(x) = \lim_{x\to-\infty} f_X(x) = 0$. See Problem 1.20.)

We assume that $\varphi_X(s)$ is numerically available for any s, and it is the P_n which we wish to recover. This will be done using the DFT (via the FFT). With the inverse DFT (1.89) in mind, dividing both sides of (1.90) by $\mathrm{e}^{is\ell}$ and equating $sn(\Delta x)$ with $2\pi nt/T$ gives

$$\varphi_X(s)\,\mathrm{e}^{-is\ell} \approx \sum_{n=0}^{T-1} \mathrm{e}^{isn(\Delta x)} P_n = \sum_{n=0}^{T-1} \mathrm{e}^{2\pi int/T} P_n =: g_t, \tag{1.91}$$

yielding $\mathbf{g} = \mathcal{F}^{-1}(\mathbf{P})$, with $\mathbf{g} = (g_0, \ldots, g_{T-1})$ and $\mathbf{P} = (P_0, \ldots, P_{T-1})$. As $s(\Delta x) = 2\pi t/T$, it makes sense to take a T-length grid of s-values as, say,

$$s_t = \frac{2\pi t}{T(\Delta x)}, \quad t = -\frac{T}{2}, -\frac{T}{2}+1, \ldots, \frac{T}{2}-1,$$

[10] To make matters more confusing, according to the Matlab's help file on its FFT, it defines the DFT and its inverse as we do (except for where the factor $1/T$ is placed), but this is not what they compute.

so $g_t \approx \varphi_X(s_t)\,\mathrm{e}^{-is_t\ell}$ from (1.91). The relation between $\mathcal{F}$ and $\mathcal{F}^{-1}$ implies $\mathbf{P} = \mathcal{F}(\mathbf{g})$, so applying the DFT (via the FFT) to $\mathbf{g}$ yields $\mathbf{P}$, the nth element of which is $P_n = f_X(x_n)\,\Delta x$. Finally, dividing this $\mathbf{P}$ by Δx yields the p.d.f. of X at the T-length grid of x-values $\ell,\ \ell + (\Delta x),\ \ldots, \ell + (T-1)(\Delta x) \approx u$. Clearly, choosing (Δx) small and T large (and ℓ and u appropriately) will yield greater accuracy, though the higher T is, the longer the FFT calculation. Typically, T is chosen as a power of 2, so that the most efficient FFT algorithms are possible. The code in Listing 1.8 implements this for inverting the normal and χ^2 characteristic functions, and plots the RPE of the computed p.d.f. versus the true p.d.f.

```
T=2^15; dx=0.002; % tuning parameters which influence quality

% ell is the lower limit in finite integral approximation to the c.f.
ell=-3; % use this when inverting the standard normal.
ell=0; % use 0 for the chi^2.

t=-(T/2):(T/2)-1; s = 2*pi*t/T/dx;

%%%%%%%%%%%%% Choose the c.f.  %%%%%%%%%%%%%%%%%%%%%%%%%%%%
phi = exp(-(s.^2)/2); % the standard normal c.f.
% df=4; phi = (1-2*i*s).^(-df/2); % the chi^2(df) c.f.
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

g = phi.*exp(-i*s*ell);
P=ifft(g); % Remember: in Matlab, ifft computes the DFT
pdf = P/dx;

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
if max(abs(imag(pdf)))>1e-7, 'problem!', end
%% There should be no imaginary part, so check for this
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
f=abs(pdf); f=f(end:-1:1); % this is necessary.
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

x=ell+(dx/1):dx:T*dx; % next 2 lines in case x is off by one
if length(x)<length(f), x=[x (T+1)*dx]; end
if length(x)>length(f), x=x(1:end-1); end

xup=4; % set to around 20 for the chi^2
loc=find(diff(x>xup)); % just up to x=xup
x=x(10:loc); f=f(10:loc);

true = normpdf(x); plot(x,100*(f-true)./true)
%true = chi2pdf(x,df); plot(x,100*(f-true)./true)
```

Program Listing 1.8 Illustrates the use of the FFT for inverting the standard normal c.f., or the χ^2, at an equally spaced grid of points. The program in Listing 1.9 is similar, but is more advanced and is recommended in practice

The implementation in Listing 1.8 can be improved in several ways. One is to dynamically choose the step size Δx and the exponent p in the grid size $T = 2^p$ as a function of where the density is to be calculated. This requires having an upper bound on p based on available computer memory required for the FFT. In addition, for given values of p and Δx (i.e., for the same computational cost), higher accuracy can be obtained by using a variation on the above method, as detailed in Mittnik, Doganoglu and Chenyao (1999) and the references therein. We implement this in the program in Listing 1.9; it should be used in practice instead of the simpler code in Listing 1.8.

Moreover, the program in Listing 1.8 only delivers the p.d.f. at an equally spaced grid of points. This is fine for plotting purposes, but if the likelihood is being calculated, the data points are almost certainly not equally spaced. If the density of X is desired at a particular set of points $\mathbf{z} = (z_1, \ldots, z_k)$, then, provided that $\ell \leq z_1$ and $z_k \leq u$, some form of interpolation can be used to get $f_X(\mathbf{z})$. This could easily be accomplished by using the built-in Matlab function `interp1`, but this is problematic when a massive number of data points are involved (such as $T = 2^{14}$ or more). As such, the program in Listing 1.9 calls the function `wintp1`, given in Listing 1.10, which sets up a moving window for the (linear) interpolation, resulting in a sizeable speed increase.

```
function pdf=fftnoncentralchi2(z,df,noncen)

% set the tuning parameters for the FFT
pmax=18; step=0.01; p=14;
maxz=round(max(abs(z)))+5;
while((maxz/step+1)>2^(p-1))
  p=p+1;
end;
if p > pmax, p=pmax; end
if maxz/step+1 >2^(p-1)
  step=(maxz+1)*1.001/(2^(p-1));
end;

% compute the pdf on a grid of points
zs=sort(z); [xgrd,bigpdf]=runthefft(p,step, df,noncen);

% Now use linear interpolation to get the pdf at the desired values.
pdf=wintp1(xgrd,bigpdf,zs);


function [x,p]=runthefft(n,h, df,noncen)
n=2^n; x=(0:n-1)'*h-n*h/2; s=1/(h*n);
t=2*pi*s*((0:n-1)'-n/2);
sgn=ones(n,1); sgn(2:2:n)=-1*ones(n/2,1);
%%%%%%%%%% This is now the c.f. %%%%%%%%%%
CF = (1-2*i*t).^(-df/2) .* exp((i*t*noncen)./(1-2*i*t));
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
phi=sgn.*CF; phi(n/2+1)=sgn(n/2+1); p=s.*abs(fft(phi));
```

Program Listing 1.9 Uses the FFT and linear interpolation to invert the specified c.f. and compute the p.d.f. at the vector of points `z`. In this case, the c.f. corresponds to a noncentral χ^2(`df,noncen`) random variable, as in (1.73). The code for function `wintp1` is in Listing 1.10

```
function r=wintp1(x,y,z,w)
% interpolate with a window of size w on z.
if nargin<4, w=length(z); end;
m=length(z); n=floor(m/w); r=[];
for i=1:n
 dz=z((i-1)*w+1:i*w); k=interval(x,dz);
 t=intpol(x(k),y(k),dz); r=[r;t];
end;
if (rem(m,w)~=0),
    e=rem(m,w); g=[n*w+1:n*w+e]; dz=z(g);k=interval(x,dz);
    t=intpol(x(k),y(k),dz); r=[r;t];
end;

function i=interval(x,y)
% find the x: min(x)<min(y) and max(x)> max(y)
i1=find(x<min(y)); i2=find(x>max(y)); i=i1(size(i1,1)):i2(1);
j=i(1); while (x(j)>y(1)), j=j-1; end;
k=max(i); while(x(k)<max(y)),   k=k+1; end;
i=[(j:i(1)-1)';i;(max(i)+1:k)'];

function r=intpol(x,y,z);
% linear interp of z to f(x,y), needs min(x)< z < max(x)
n=length(z); r=[]; j=1; i=1;
while i<=n, flag=i;
  while (flag==i)
    if (z(i)>=x(j) & z(i)<x(j+1))
      r=[r;y(j+1)-(x(j+1)-z(i))*(y(j+1)-y(j))/(x(j+1)-x(j))];
        i=i+1;
    else, j=j+1; end;
  end;
end;
```

Program Listing 1.10 Interpolation program called by the function in Listing 1.9. This was written by Toker Doganoglu and used in the implementation detailed in Mittnik, Doganoglu and Chenyao (1999)

To illustrate, we use the c.f. in (1.73), corresponding to the noncentral χ^2 distribution. For $\theta = 0$, this reduces to the c.f. of the usual χ^2 r.v., so that we can easily check the results, while methods of computing the p.d.f. of the noncentral χ^2 distribution to machine precision are discussed in detail in Chapter 10. Figure 1.10 shows the p.d.f. as computed via program `fftnoncentralchi2` using 5 degrees of freedom and several values of the noncentrality parameter θ. What is remarkable is the massive cut in computation time compared to the use of the p.d.f. inversion formula in Section 1.2.5, which would have to be evaluated for each ordinate in Figure 1.10.

1.4 Multivariate case

This section extends some of the ideas for univariate continuous r.v.s to the multivariate case. Paralleling the vector m.g.f., the c.f. of the d-dimensional continuous random vector $\mathbf{X} = (X_1, \ldots, X_d)$ is given by

$$\varphi_{\mathbf{X}}(\mathbf{t}) = \mathbb{E}\left[e^{i\mathbf{t}'\mathbf{X}}\right], \tag{1.92}$$

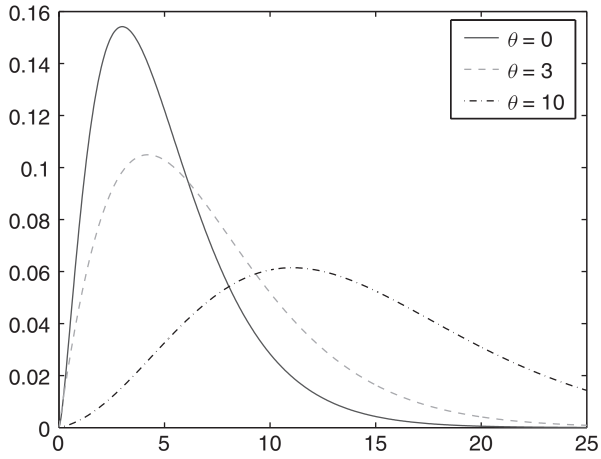

Figure 1.10 The noncentral $\chi^2(5, \theta)$ p.d.f. computed via the FFT

where $\mathbf{t} = (t_1, \ldots, t_d)$. The p.d.f. of $\mathbf{X}$ can be recovered by

$$\boxed{f_{\mathbf{X}}(\mathbf{x}) = \frac{1}{(2\pi)^d} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \mathrm{e}^{-i\mathbf{t}'\mathbf{x}} \varphi_{\mathbf{X}}(\mathbf{t}) \, \mathrm{d}t_1 \cdots \mathrm{d}t_d,}$$

which reduces to (1.54) when $d = 1$.

Consider the case when $\mathbb{M}_X$ exists, and $\varphi_X(t) = \mathbb{M}_X(it)$. In terms of the c.g.f. $\mathbb{K}_{\mathbf{X}}$, the multivariate survivor function

$$\overline{F}_{\mathbf{X}}(\mathbf{x}) = \Pr(X_1 > x_1, \ldots, X_d > x_d)$$

can be computed as

$$\boxed{\overline{F}_{\mathbf{X}}(\mathbf{x}) = \frac{1}{(2\pi i)^d} \int_{c_1 - i\infty}^{c_1 + i\infty} \cdots \int_{c_d - i\infty}^{c_d + i\infty} \exp\left(\mathbb{K}_{\mathbf{X}}(\mathbf{t}) - \mathbf{t}'\mathbf{x}\right) \frac{\mathrm{d}\mathbf{t}}{\prod_{j=1}^{d} t_j},} \tag{1.93}$$

where $c_k \in \mathbb{R}_{>0}$, $k = 1, 2, \ldots, d$, such that they are in the convergence region of $\mathbb{K}_{\mathbf{X}}$. See Kolassa (2003, p. 276) for proof. Expression (1.93) reduces to (1.100), given below in Problem 1.12 for $d = 1$.

In the case where the c.g.f. is not available, or, more likely, the multivariate c.d.f. $F_{\mathbf{X}}$ is desired (instead of the survivor function), Shephard (1991) has generalized Gil-Peleaz's univariate result to the d-dimensional case when the mean of $\mathbf{X}$ exists. He shows that, for $d \geq 2$,

$$\begin{aligned} u(\mathbf{x}) := {} & 2^d F_{\mathbf{X}}(x_1, \ldots, x_d) \\ & - 2^{d-1}[F_{\mathbf{X}}(x_2, x_3, \ldots, x_d) + \cdots + F_{\mathbf{X}}(x_1, \ldots, x_{d-2}, x_{d-1})] \end{aligned}$$

$$+ 2^{d-2}[F_{\mathbf{X}}(x_3, x_4, \ldots, x_d) + \cdots + F_{\mathbf{X}}(x_1, \ldots, x_{d-3}, x_{d-2})]$$
$$+ \cdots + (-1)^d \tag{1.94}$$

is given by

$$u(\mathbf{x}) = \begin{cases} \dfrac{2(-2)^d i^{d-1}}{(2\pi)^d} \displaystyle\int_0^\infty \cdots \int_0^\infty \underset{t_2}{\Delta} \cdots \underset{t_d}{\Delta} \operatorname{Im}\left[\frac{\mathrm{e}^{-i\mathbf{t}'\mathbf{x}}\varphi_{\mathbf{X}}(\mathbf{t})}{\prod_{j=1}^d t_j}\right] \mathrm{d}\mathbf{t}, & \text{if } d \text{ is odd,} \\ \dfrac{2(-2)^d i^{d}}{(2\pi)^d} \displaystyle\int_0^\infty \cdots \int_0^\infty \underset{t_2}{\Delta} \cdots \underset{t_d}{\Delta} \operatorname{Re}\left[\frac{\mathrm{e}^{-i\mathbf{t}'\mathbf{x}}\varphi_{\mathbf{X}}(\mathbf{t})}{\prod_{j=1}^d t_j}\right] \mathrm{d}\mathbf{t}, & \text{if } d \text{ is even,} \end{cases} \tag{1.95}$$

where $\underset{t}{\Delta}\, \eta(t) = \eta(t) + \eta(-t)$. (A similar, but slightly more complicated expression which is also valid for $d = 1$ is also given by Shephard).

For the bivariate case $d = 2$, use of (1.94) and (1.95) yields the expression

$$F_{\mathbf{X}}(x_1, x_2) = -\frac{1}{4} + \frac{1}{2}\big[F_X(x_1) + F_X(x_2)\big] - \frac{1}{2\pi^2}\int_0^\infty \int_0^\infty g(x_1, x_2, t_1, t_2)\, \mathrm{d}t_1\, \mathrm{d}t_2, \tag{1.96}$$

where

$$g(x_1, x_2, t_1, t_2) = \operatorname{Re}\left[\frac{\exp\{-it_1x_1 - it_2x_2\}\varphi(t_1, t_2)}{t_1t_2} - \frac{\exp\{-it_1x_1 + it_2x_2\}\varphi(t_1, -t_2)}{t_1t_2}\right].$$

Note that, given only $\varphi_{\mathbf{X}}(t_1, t_2)$, the univariate c.d.f.s $F_{X_1}(x_1)$ and $F_{X_2}(x_2)$ can be evaluated by univariate inversion of $\varphi_{X_1}(t_1) \equiv \varphi_{\mathbf{X}}(t_1, 0)$ and $\varphi_{X_2}(t_2) \equiv \varphi_{\mathbf{X}}(0, t_2)$, respectively. See Problem 1.21 for an illustration.

1.5 Problems

Opportunity is missed by most people because it is dressed in overalls and looks like work. (Thomas Edison)

It is common sense to take a method and try it; if it fails, admit it frankly and try another. But above all, try something. (Franklin D. Roosevelt)

1.1. ★ Derive the results in (1.9).

1.2. ★ Let $Y \sim \text{Lap}\,(0, 1)$. The raw integer moments of Y were shown in Chapter I.7 to be $\mathbb{E}[Y^r] = r!$, $r = 0, 2, 4, \ldots$, and $\mathbb{E}[Y^r] = 0$, for odd r. Verify these for the first six integer moments of Y by using the m.g.f. Also show that the variance is 2 and the kurtosis is 6.

1.3. ★ Let $X \sim \text{Gam}\,(\alpha, \beta)$. Derive the m.g.f., expected value and variance of $Z = \ln X$. Also compute $\text{Cov}\,(Z, X)$.

1.4. ★ Calculate the cumulant generating function and, from this, the mean, variance, skewness, and kurtosis of the binomial distribution with parameters n and p. What happens to the skewness and kurtosis as $n \to \infty$?

1.5. ★ Let $X \sim \text{NBin}(r, p)$, with p.m.f. (I.4.19), i.e.,

$$f_X(x; r, p) = \binom{r+x-1}{x} p^r (1-p)^x \, \mathbb{I}_{\{0,1,\ldots\}}(x).$$

(a) Show that the m.g.f. of X is

$$\mathbb{M}_X(t) = \left(\frac{p}{1-(1-p)\,\mathrm{e}^t}\right)^r, \quad t < -\log(1-p). \tag{1.97}$$

(b) Similar to Problem 1.4, use the c.g.f. to calculate the mean, variance, skewness, and kurtosis of X. What happens to the skewness and kurtosis as $r \to \infty$?

(c) Using (1.97), calculate the c.g.f. of a geometric random variable Y with p.m.f. (I.4.18), i.e., when the (single) success is counted. Use it to calculate the mean and variance of Y. Extend this to the negative binomial case.

1.6. For $X \sim \text{Poi}(\lambda)$, calculate the mean and variance using just the m.g.f. (i.e., not the c.g.f.).

1.7. ★ Let $X_1, \ldots, X_{k+1}$ be independent random variables, each from a Poisson distribution with parameters $\lambda_1, \ldots, \lambda_{k+1}$.

(a) Derive the m.g.f. of $Y = \sum_{i=1}^{k+1} X_i$ and determine its distribution.

(b) Show that the conditional distribution of $X_1, \ldots, X_k$ given $X_1 + \cdots + X_{k+1} = n$ is multinomial with the parameters n, λ_1/λ, λ_2/λ, $\ldots$, λ_k/λ, where $\lambda = \lambda_1 + \cdots + \lambda_{k+1}$. Observe that this generalizes the result in (I.8.6) of Example I.8.4.

1.8. ★ For continuous positive random variable X with m.g.f. $\mathbb{M}_X(t)$,

$$\mathbb{E}\left[X^{-1}\right] = \int_0^\infty \mathbb{M}_X(-t)\,\mathrm{d}t \tag{1.98}$$

and, more generally,

$$\mathbb{E}\left[X^{-c}\right] = \frac{1}{\Gamma(c)} \int_0^\infty t^{c-1} \mathbb{M}_X(-t)\,\mathrm{d}t, \tag{1.99}$$

as shown by Cressie *et al.* (1981).

(a) Let $Z \sim \text{Gam}(\alpha, \beta)$. Calculate $\mathbb{E}\left[Z^{-c}\right]$ directly and using (1.99).

(b) Prove (1.99). Hint: First verify that $\int_0^\infty \mathrm{e}^{-tx} t^{c-1}\,\mathrm{d}t = \Gamma(c)\,x^{-c}$.

1.9. ★ ★ Show that the m.g.f. $\mathbb{M}_K(t) =: \mathbb{M}_N(t)$ for the r.v. K in Banach's matchbox problem (see Problem I.4.13) with N matches satisfies

$$\mathbb{M}_N(t) = \frac{e^{2t}}{2e^t - 1}\mathbb{M}_{N-1}(t) + \frac{e^t - 1}{2e^t - 1}\binom{N - \frac{1}{2}}{N}$$

Using this, compute $\mathbb{M}'_N(t)$ and $\mathbb{E}[K]$. (Contributed by Markus Haas)

1.10. Use (1.15) to compute $\mathbb{E}[N_m]$. (Contributed by Markus Haas)

1.11. ★ Let $X \sim \text{Unif}(0, 1)$, for which we certainly do not need inversion formulae to compute the p.d.f. and c.d.f., but their use is instructive.

(a) Compute the c.f. $\varphi_X(t)$.

(b) For $F_X = 0.5$, the integrand in (1.70) has to be zero. Verify this directly.

(c) Consider what happens to the integrand at the extremes of the range of integration in (1.70), i.e., with $z(t)$ and $g(t)$ as defined in (1.70), for $x \in (0, 1)$, compute $\lim_{t\to\infty} z(t)$, $\lim_{t\to\infty} g(t)$, $\lim_{t\to 0} z(t)$, and $\lim_{t\to 0} g(t)$.

(d) For this c.f., decide if the use of (1.70) or (1.71) is preferred. Using $x = 0.2$, plot the integrands.

1.12. ★ Recall inversion formula (1.54), i.e.,

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

for X continuous. (Contributed by Simon Broda)

(a) Derive the expression in (1.56), i.e.,

$$f_X(x) = \frac{1}{2\pi i}\int_{-i\infty}^{i\infty} \exp\{\mathbb{K}_X(s) - sx\}\,ds,$$

from (1.54).

(b) The integral in (1.56) can also be expressed as

$$f_X(x) = \frac{1}{2\pi i}\int_{c-i\infty}^{c+i\infty} \exp\{\mathbb{K}_X(s) - sx\}\,ds,$$

where c is an arbitrary real constant in the convergence strip of $\mathbb{K}_X$. This is called the Fourier–Mellin integral and is a standard result in the theory of Laplace transforms; see Schiff (1999, Ch. 4). Write out the double integral for the survivor function

$$\overline{F}_X(x) = 1 - F_X(x) = \int_x^{\infty} f_X(y)\,dy,$$

exchange the order of the integrals, and confirm that c needs to be positive for this to produce the valid integral expression

$$\overline{F}_X(x) = \frac{1}{2\pi i}\int_{c-i\infty}^{c+i\infty} \exp\{\mathbb{K}_X(s) - sx\}\,\frac{\mathrm{d}s}{s}, \quad c > 0. \tag{1.100}$$

Similarly, derive the expression for the c.d.f. given by

$$F_X(x) = -\frac{1}{2\pi i}\int_{c-i\infty}^{c+i\infty} \exp\{\mathbb{K}_X(s) - sx\}\,\frac{ds}{s}, \quad c < 0. \tag{1.101}$$

(c) Write a program to compute the c.d.f. based on the previous result, and compute it using the c.f. for the uniform distribution.

1.13. ★ As in Gut (2005, p. 442), a random variable X is said to have an *infinitely divisible distribution* iff, for each $n \in \mathbb{N}$, there exist i.i.d. r.v.s $X_{n,k}$, $1 \le k \le n$, such that

$$X \stackrel{d}{=} \sum_{k=1}^{n} X_{n,k},$$

recalling (1.17). Equivalently, X is has an infinitely divisible distribution iff

$$\varphi_X(t) = \left[\varphi_{X_{n,1}}(t)\right]^n, \quad \forall n \in \mathbb{N}, \tag{1.102}$$

For example, if $\varphi_X(t) = \mathrm{e}^{-|t|}$, then taking $\varphi_{X_{n,1}}(t)$ to be $\mathrm{e}^{-|t|/n}$ shows that

$$\varphi_X(t) = \left[\mathrm{e}^{-|t|/n}\right]^n = \mathrm{e}^{-|t|}.$$

(See also Example 1.22). This demonstrates that a Cauchy random variable is infinitely divisible and, if $X_i \sim \mathrm{Cau}(c_i)$, then $X = \sum_{i=1}^n X_i \sim \mathrm{Cau}\left(\sum_{i=1}^n c_i\right)$.

Let $X \sim \mathrm{Lap}(0, 1)$. First show that the characteristic function of X is given by

$$\varphi_X(t) = \left(1 + t^2\right)^{-1}.$$

Then show that a Laplace r.v. is infinitely divisible.

1.14. ★ Verify (1.68).

1.15. ★ Derive (1.54) from (1.65).

1.16. Use Euler's formula to show that $i^i = \mathrm{e}^{-\pi/2}$.

1.17. ★★ This problem proves some useful properties of the Laplace transform and is based on Schiff (1999).

(a) Let f be (piecewise) continuous on $[0,\infty)$ and let $F(s) = \int_0^\infty e^{-st} f(t)\,dt$ converge uniformly for all $s \in D \subset \mathbb{C}$, i.e., for all $s \in D$,

$$\text{for any } \epsilon > 0,\ \exists\, T > 0 \text{ s.t., for } \tau \ge T, \quad \left| \int_\tau^\infty e^{-st} f(t)\,dt \right| < \epsilon. \quad (1.103)$$

Prove that

$$\lim_{s \to s_0} \int_0^\infty e^{-st} f(t)\,dt = \int_0^\infty \lim_{s \to s_0} e^{-st} f(t)\,dt, \quad (1.104)$$

i.e., that $F(s_0) = \lim_{s \to s_0} F(s)$. Hint: This amounts to showing that

$$\left| \int_0^\infty e^{-st} f(t)\,dt - \int_0^\infty e^{-s_0 t} f(t)\,dt \right|$$

is arbitrarily small for s sufficiently close to s_0, and consider that

$$\left| \int_0^\infty e^{-st} f(t)\,dt - \int_0^\infty e^{-s_0 t} f(t)\,dt \right| \le \int_0^{t_0} \left| e^{-st} - e^{-s_0 t} \right| |f(t)|\,dt$$
$$+ \left| \int_{t_0}^\infty \left(e^{-st} - e^{-s_0 t} \right) |f(t)|\,dt \right|.$$

(b) Recall Fubini's theorem (I.A.159), which states that if $f : A \subset \mathbb{R}^2 \to \mathbb{R}$ is continuous and A is closed and bounded, then

$$\int_a^b \int_0^\tau f(x,t)\,dt\,dx = \int_0^\tau \int_a^b f(x,t)\,dx\,dt. \quad (1.105)$$

To extend this to

$$\int_a^b \int_0^\infty f(x,t)\,dt\,dx = \int_0^\infty \int_a^b f(t,x)\,dx\,dt, \quad (1.106)$$

assume $f : A \subset \mathbb{R}^2 \to \mathbb{R}$ is continuous on each rectangle $a \le x \le b$, $0 \le t \le T$, $T > 0$, and $\int_0^\infty f(x,t)\,dt$ converges uniformly for all $x \in [a,b]$, i.e.,

$$\text{for any } \epsilon > 0,\ \exists\, T > 0 \text{ s.t., for } \tau \ge T, \quad \left| \int_\tau^\infty f(x,t)\,dt \right| < \frac{\epsilon}{b-a}. \quad (1.107)$$

Then taking $\lim_{\tau \to \infty}$ of both sides of (1.105) yields

$$\lim_{\tau \to \infty} \int_a^b \int_0^\tau f(x,t)\,dt\,dx = \int_0^\infty \int_a^b f(t,x)\,dx\,dt,$$

and the l.h.s. converges if the 'tail area' $\lim_{\tau \to \infty} \int_a^b \int_\tau^\infty f(x,t)\,dt\,dx = 0$. But this is true because, for $\tau \ge T$, (1.107) implies $\int_a^b \int_\tau^\infty f(x,t)\,dt\,dx < \epsilon$.

Let $f(x,t)$ and $D_1 f = (\partial/\partial x) f$ be continuous on each rectangle $a \leq x \leq b$, $0 \leq t \leq T, T > 0$, assume $F(x) = \int_0^\infty f(x,t)\,\mathrm{d}t$ exists and $\int_0^\infty D_1 f(x,t)\,\mathrm{d}t$ converges uniformly. Use (1.104) and (1.106) to show that

$$\frac{\mathrm{d}}{\mathrm{d}x} F(x) = \int_0^\infty D_1 f(x,t)\,\mathrm{d}t, \quad a < x < b. \tag{1.108}$$

(c) Let f be (piecewise) continuous on $[0,\infty)$ of exponential order α, and let $F(s) = \mathcal{L}\{f\}$. Show that $(-1)^n t^n f(t)$ is of exponential order, and then use (1.108) to show that, for real s,

$$\frac{\mathrm{d}}{\mathrm{d}s} F(s) = \mathcal{L}\{-tf(t)\}, \quad s > \alpha.$$

Induction then shows that

$$\frac{\mathrm{d}^n}{\mathrm{d}s^n} F(s) = \mathcal{L}\left\{(-1)^n t^n f(t)\right\}, \quad n \in \mathbb{N}, \quad s > \alpha. \tag{1.109}$$

(d) Let f be (piecewise) continuous on $[0,\infty)$ of exponential order α, and let $F(s) = \mathcal{L}\{f\}$, such that $\lim_{t\downarrow 0} f(t)/t$ exists. Use (1.106) to show that, for real s,

$$\int_s^\infty F(x)\,\mathrm{d}x = \mathcal{L}\left\{\frac{f(t)}{t}\right\}, \quad s > \alpha. \tag{1.110}$$

Use this to simplify the integral $\int_0^\infty \mathrm{e}^{-st} t^{-1} \sin(t)\,\mathrm{d}t$.

(e) Let f and f' be continuous functions on $[0,\infty)$, and f is of exponential order α. Show that

$$\mathcal{L}\left\{f'(t)\right\} = s\mathcal{L}\{f(t)\} - f(0). \tag{1.111}$$

(f) The *convolution* of two functions, f and g, is given by

$$(f * g)(t) = \int_0^t f(x)\, g(t-x)\,\mathrm{d}x, \tag{1.112}$$

if the integral exists, and is a central concept when studying the sum of random variables (Chapter 2). Let f and g be (piecewise) continuous on $[0,\infty)$ and of exponential order α. Show that

$$\mathcal{L}\{f * g\} = \mathcal{L}\{f\}\mathcal{L}\{g\}, \quad \mathrm{Re}(s) > \alpha. \tag{1.113}$$

Hint: It will be necessary to reverse the order of two improper integrals, the justification for which can be shown as follows. As in Schiff (1999, p. 93), let $h(x,y)$ be a function such that

$$\int_0^\infty \int_0^\infty |h(x,y)|\,\mathrm{d}x\,\mathrm{d}y < \infty,$$

i.e., h is absolutely convergent. Define

$$a_{mn} = \int_n^{n+1} \int_m^{m+1} |h(x, y)| \, dx \, dy \quad \text{and} \quad b_{mn} = \int_n^{n+1} \int_m^{m+1} h(x, y) \, dx \, dy,$$

and use (I.A.97) to show that

$$\int_0^\infty \int_0^\infty h(x, y) \, dx \, dy = \int_0^\infty \int_0^\infty h(x, y) \, dy \, dx.$$

1.18. ★ Similar to (1.112), for $\mathbf{g} = (g_0, \ldots, g_{T-1})$ and $\mathbf{h} = (h_0, \ldots, h_{T-1})$ vectors of complex numbers, their (circular) convolution is $\mathbf{c} = (c_0, \ldots, c_{T-1}) = \mathbf{g} * \mathbf{h}$, where

$$c_t = \sum_{j=0}^{T-1} g_{t-j} h_j, \quad \text{with } g_{t-j} = g_{T+t-j}. \tag{1.114}$$

Show algebraically that $\mathcal{F}(\mathbf{g} * \mathbf{h}) = T \mathcal{F}(\mathbf{g}) \mathcal{F}(\mathbf{h})$, where the product of the two DFTs means the elementwise product of their components. Also write a Matlab program to numerically verify this.

1.19. ★ ★ The *Fourier transform* is the result of considering what happens to the Fourier series (1.87) when letting the fundamental period T of function g tend towards infinity, i.e., letting the fundamental frequency $\Delta f := 1/T$ go to zero. The C_n coefficients measure the contribution (amplitude) of particular discrete frequencies, and we are interested in the limiting case for which the C_n becomes a continuum, so that the expression for $g(t)$ can be replaced by an integral. To this end, let $f_n = n\omega/(2\pi) = n/T$, so that, with $r = T/2$, taking the limit of the expressions in (1.87) in such a way that, as n and T approach ∞, $f_n = n/T$ stays finite gives

$$\begin{aligned} g(t) &= \lim_{T\to\infty} \sum_{n=-\infty}^{\infty} C_n e^{2\pi i f_n t} = \lim_{T\to\infty} \sum_{n=-\infty}^{\infty} \left[\int_{-T/2}^{T/2} g(t) \, e^{-2\pi i f_n t} \, dt \right] e^{2\pi i f_n t} \Delta f \\ &= \int_{-\infty}^{\infty} \left[\int_{-\infty}^{\infty} g(t) \, e^{-2\pi i f t} \, dt \right] e^{2\pi i f t} \, df, \end{aligned}$$

where f now denotes the continuous limiting function of f_n. Let $\omega = 2\pi f$, with $d\omega = 2\pi \, df$, and set $G(\omega)$ to be the inner integral, so that we obtain the pair

$$\boxed{G(\omega) = \int_{-\infty}^{\infty} g(t) \, e^{-i\omega t} \, dt \qquad g(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} G(\omega) \, e^{i\omega t} \, d\omega,} \tag{1.115}$$

where the former is the *Fourier transform*, and the latter is the *inverse Fourier transform*.

For $T > 0$ and $a \in \mathbb{R}$, let $g(t) = a\mathbb{I}(|t| \le T)$. Compute the Fourier transform $G(\omega)$ of $g(t)$ and verify it via the inverse Fourier transform. See the footnote for a hint (and remove one star from the difficulty level).[11]

1.20. ★ Try the following question on the mathematician of your choice. If f is a probability density function, is it true that $f(x) \to 0$ when $x \to \pm\infty$? The answer is no. We can construct a counter-example as follows (leaving aside some technical details concerning convergence).

(Contributed by Walther Paravicini)

(a) Assume that f_n is a p.d.f. for each natural number n. Provided that we can give a meaning to the weighted sum $f(x) := \sum_{n=1}^{\infty} 2^{-n} f_n(x)$, $x \in \mathbb{R}$, convince yourself that f is again a p.d.f.

(b) For all natural numbers n, let $\sigma_n > 0$ and $\mu_n \in \mathbb{R}$. Let f_n be the p.d.f. of an r.v. distributed as $\mathrm{N}(\mu_n, \sigma_n^2)$. For a given n, calculate the maximum of $f_n(x)$, $x \in \mathbb{R}$, in terms of σ_n and μ_n.

(c) Find a choice of σ_n and μ_n such that the maximum of $2^{-n} f_n(x)$, $x \in \mathbb{R}$, does not converge to 0 if $n \to \infty$ and such that the argument where the maximum is attained converges to ∞.

(d) Write a Matlab program which calculates the density of the so-constructed p.d.f. $f(x) = \sum_{n=1}^{\infty} 2^{-n} f_n(x)$, $x \in \mathbb{R}$. Produce a plot of the density which shows the behaviour of $f(x)$ for large x. For a suitable choice of σ_n and μ_n, the plot should show a series of ever-increasing spikes which get steeper and steeper.

1.21. ★ This problem illustrates the numeric implementation of (1.94) and (1.95) in a nontrivial case, but one for which we can straightforwardly calculate the correct answer to an arbitrary precision for comparison purposes, viz., the bivariate normal distribution, which was first seen in Example I.8.5 and will be more formally introduced and detailed in Chapter 3. At this point, interest centres purely on the mechanics of inverting a bivariate c.f. (Contributed by Simon Broda)

Let $\mathbf{X} = (X_1, X_2)'$ follow a bivariate normal distribution, i.e.,

$$\begin{bmatrix} X_1 \\ X_2 \end{bmatrix} \sim \mathrm{N}\left(\begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}, \begin{bmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix} \right), \tag{1.116}$$

with density

$$f_{X_1,X_2}(x, y) = K \exp\left\{ -\frac{X^2 - 2\rho XY + Y^2}{2\left(1-\rho^2\right)} \right\},$$

where

$$K^{-1} = 2\pi\sigma_1\sigma_2\left(1-\rho^2\right)^{1/2}, \qquad X = \frac{x-\mu_1}{\sigma_1}, \qquad Y = \frac{y-\mu_2}{\sigma_2},$$

for $\mu_1, \mu_2 \in \mathbb{R}$, $\sigma_1, \sigma_2 \in \mathbb{R}_{>0}$, and $|\rho| < 1$.

[11] Hint: For the Fourier transform $G(\omega)$ of $g(t)$, use (1.32), and for the inverse Fourier transform, first show that $\sin(a-b) + \sin(a+b) = 2\sin(a)\sin(b)$, and use this with the results in Example I.A.21 and a result mentioned in Example I.A.30.

Preferable methods for evaluation of $F_{X_1,X_2}(x_1, x_2)$ are discussed in Section 3.4. The joint c.f. of (X_1, X_2) is given by $\varphi(t_1, t_2) = \mathbb{M}_{\mathbf{X}}(i\mathbf{t})$ from (3.16), i.e.,

$$\varphi(t_1, t_2) = \exp\left[i(t_1\mu_1 + t_2\mu_2) - \frac{1}{2}\left(\sigma_1^2 t_1^2 + 2\rho\sigma_1\sigma_2 t_1 t_2 + \sigma_2^2 t_2^2\right)\right]. \tag{1.117}$$

Write a Matlab program to accomplish this.

1.22. ★ ★ This problem informally derives inversion formulae for the ratio of two random variables. (Contributed by Simon Broda)

(a) Let X_1 and X_2 be continuous r.v.s with c.f. φ_{X_1,X_2} and such that $\mathbb{E}[X_1]$ and $\mathbb{E}[X_2]$ exist. Show that an inversion formula for the density of $R := X_1/X_2$ is given by

$$f_R(r) = \frac{1}{\pi^2}\int_0^\infty \int_{-\infty}^\infty \mathrm{Re}\left[\frac{\varphi_2(t_1, -t_2 - rt_1)}{t_2}\right] \mathrm{d}t_1 \, \mathrm{d}t_2, \tag{1.118}$$

where

$$\varphi_2(t_1, t_2) = \frac{\partial}{\partial t_2}\varphi_{X_1,X_2}(t_1, t_2).$$

Hint: First show that, with $X_3 := X_1 - rX_2$,

$$\Pr\left(\frac{X_1}{X_2} < r\right) = \Pr(X_3 < 0) + \Pr(X_2 < 0) - 2\Pr(X_3 < 0 \wedge X_2 < 0), \tag{1.119}$$

and use this in conjunction with (1.96) and the fact that

$$\varphi_{X_3,X_2}(t_1, t_2) = \varphi_{X_1,X_2}(t_1, t_2 - rt_1)$$

to derive an inversion formula for the c.d.f. of R in terms of the joint c.f. of (X_1, X_2). Differentiate with respect to r and use the fact that $\varphi_{X_1,X_2}(t_1, t_2) = \overline{\varphi}_{X_1,X_2}(-t_1, -t_2)$.

(b) For the case where $F_{X_2}(0) = 0$, Geary's result (1.62) can be used, i.e.,

$$f_R(r) = \frac{1}{2\pi i}\int_{-\infty}^\infty \left[\frac{\partial \varphi_{X_1,X_2}(s, t)}{\partial t}\right]_{t=-rs} \mathrm{d}s.$$

In the event that $F_{X_2}(0)$ is positive but 'sufficiently small', (1.62) can be used to approximate (1.118). Consider $R_1 = X_1/X_2$, where (X_1, X_2) has a bivariate normal distribution as in (1.116), with joint c.f. given in (1.117). Write a Matlab program that evaluates the density of R by means of (1.62) and (1.118). Use your program to evaluate the density of R and $R_2 = X_2/X_1$, where $\mu_1 = 2$, $\mu_2 = 0.1$, $\sigma_1 = \sigma_2 = 1$, $\rho = 0$, and demonstrate that the relative error incurred by using (1.62) over (1.118) is larger for R.

See also Example 2.18 on how to compute the p.d.f. of R_1 and R_2.

1.23. Prove the second part of (1.38) by using the closed-form solution of a finite geometric sum for complex numbers. It is interesting to see this optically: for $n = 6$, write a program which plots the $(z_n^k)^j$ on a circle for $j = 0, \ldots, n-1$, but each with a slight perturbation so that duplicates can be recognized.