Intermediate Probability

A Computational Approach

Marc S. Paolella

Swiss Banking Institute, University of Zurich, Switzerland

John Wiley & Sons, Ltd

Intermediate Probability

Intermediate Probability

A Computational Approach

Marc S. Paolella

Swiss Banking Institute, University of Zurich, Switzerland

BICENTENNIAL 1807 WILEY 2007 BICENTENNIAL

John Wiley & Sons, Ltd

Telephone (+44) 1243 779777

Email (for orders and customer service enquiries): cs-books@wiley.co.uk
Visit our Home Page on www.wileyeurope.com or www.wiley.com

Other Wiley Editorial Offices

John Wiley & Sons Inc., 111 River Street, Hoboken, NJ 07030, USA

Jossey-Bass, 989 Market Street, San Francisco, CA 94103-1741, USA

Wiley-VCH Verlag GmbH, Boschstr. 12, D-69469 Weinheim, Germany

John Wiley & Sons Australia Ltd, 42 McDougall Street, Milton, Queensland 4064, Australia

John Wiley & Sons (Asia) Pte Ltd, 2 Clementi Loop #02-01, Jin Xing Distripark, Singapore 129809

John Wiley & Sons Canada Ltd, 6045 Freemont Blvd, Mississauga, Ontario, L5R 4J3, Canada

Wiley also publishes its books in a variety of electronic formats. Some content that appears
in print may not be available in electronic books.

Anniversary Logo Design: Richard J. Pacifico

Library of Congress Cataloging-in-Publication Data

Paolella, Marc S.
Intermediate probability : a computational approach / Marc S. Paolella.
p. cm.
ISBN 978-0-470-02637-3 (cloth)
1. Distribution (Probability theory)–Mathematical models. 2. Probabilities. I. Title.
QA273.6.P36 2007
519.2 – dc22
2007020127

British Library Cataloguing in Publication Data

A catalogue record for this book is available from the British Library

ISBN-13: 978-0-470-02637-3

Typeset in 10/12 Times by Laserwords Private Limited, Chennai, India
Printed and bound in Great Britain by Antony Rowe Ltd, Chippenham, Wiltshire
This book is printed on acid-free paper responsibly manufactured from sustainable forestry
in which at least two trees are planted for each one used for paper production.

Chapter Listing

Contents

Preface

This book is a sequel to Volume I, *Fundamental Probability: A Computational Approach* (2006), `http://www.wiley.com/WileyCDA/WileyTitle/productCd-0470025948.html`, which covered the topics typically associated with a first course in probability at an undergraduate level. This volume is particularly suited to beginning graduate students in statistics, finance and econometrics, and can be used independently of Volume I, although references are made to it. For example, the third equation of Chapter 2 in Volume I is referred to as (I.2.3), whereas (2.3) means the third equation of Chapter 2 of the present book. Similarly, a reference to Section I.2.3 means Section 3 of Chapter 2 in Volume I.

The presentation style is the same as that in Volume I. In particular, computational aspects are incorporated throughout. Programs in Matlab are given for all computations in the text, and the book's website will provide these programs, as well as translations in the R language. Also, as in Volume I, emphasis is placed on solving more practical and challenging problems than is often done in such a course. As a case in point, Chapter 1 emphasizes the use of characteristic functions for calculating the density and distribution of random variables by way of (i) numerically computing the integrals involved in the inversion formulae, and (ii) the use of the fast Fourier transform. As many students may not be comfortable with the required mathematical machinery, a stand-alone introduction to complex numbers, Fourier series and the discrete Fourier transform are given as well.

The remaining chapters, in brief, are as follows.

Chapter 2 uses the tools developed in Chapter 1 to calculate the distribution of sums of random variables. I start with the usual, algebraically trivial examples using the moment generating function (m.g.f.) of independent and identically distributed (i.i.d) random variables (r.v.s), such as gamma and Bernoulli. More interesting and useful, but less commonly discussed, is the question of how to compute the distribution of a sum of independent r.v.s when the resulting m.g.f. is not 'recognizable', e.g., a sum of independent gamma r.v.s with different scale parameters, or the sum of binomial r.v.s with differing values of p, or the sum of independent normal and Laplace r.v.s.

Chapter 3 presents the multivariate normal distribution. Along with numerous examples and detailed coverage of the standard topics, computational methods for calculating the c.d.f. of the bivariate case are discussed, as well as partial correlation, which is required for understanding the partial autocorrelation function in time series analysis.

Chapter 4 is on asymptotics. As some of this material is mathematically more challenging, the emphasis is on providing careful and highly detailed proofs of basic results and as much intuition as possible.

Chapter 5 gives a basic introduction to univariate and multivariate saddlepoint approximations, which allow us to quickly and accurately invert the m.g.f. of sums of independent random variables without requiring the numerical integration (and occasional numeric problems) associated with the inversion formulae. The methods complement those developed in Chapters 1 and 2, and will be used extensively in Chapter 10. The beauty, simplicity, and accuracy of this method are reason enough to discuss it, but its applicability to such a wide range of topics is what should make this methodology as much of a standard topic as is the central limit theorem. Much of the section on multivariate saddlepoint methods was written by my graduate student and fellow researcher, Simon Broda.

Chapter 6 deals with order statistics. The presentation is quite detailed, with numerous examples, as well as some results which are not often seen in textbooks, including a brief discussion of order statistics in the non-i.i.d. case.

Chapter 7 is somewhat unique and provides an overview on how to help 'classify' some of the hundreds of distributions available. Of course, not all methods can be covered, but the ideas of nesting, generalizing, and asymmetric extensions are introduced. Mixture distributions are also discussed in detail, leading up to derivation of the variance–gamma distribution.

Chapter 8 is about the stable Paretian distribution, with emphasis on its computation, basic properties, and uses. With the unprecedented growth of it in applications (due primarily to its computational complexity having been overcome), this should prove to be a useful and timely topic well worth covering. Sections 8.3.2 and 8.3.3 were written together with my graduate student and fellow researcher, Yianna Tchopourian.

Chapter 9 is dedicated to the (generalized) inverse Gaussian and (generalized) hyperbolic distributions, and their connections. In addition to being mathematically intriguing, they are well suited for modelling a wide variety of phenomena. The author of this chapter, and all its problems and solutions, is my academic colleague Walther Paravicini.

Chapter 10 provides a quite detailed account of the singly and doubly noncentral F, t and beta distributions. For each, several methods for the exact calculation of the distribution are provided, as well as discussion of approximate methods, most notably the saddlepoint approximation.

The Appendix contains a list of tables, including those for abbreviations, special functions, general notation, generating functions and inversion formulae, distribution naming conventions, distributional subsets (e.g., $\chi^2 \subseteq \text{Gam}$ and $\text{N} \subseteq \text{S}\alpha\text{S}$), Student's t generalizations, noncentral distributions, relationships among major distributions, and mixture relationships.

As in Volume I, the examples are marked with symbols to designate their relative importance, with ⊖, ⊚ and ⊛ indicating low, medium and high importance, respectively. Also as in Volume I, there are many exercises, and they are furnished with stars to indicate their difficulty and/or amount of time required for solution. Solutions to all exercises, in full detail, are available for instructors, as are lecture notes for beamer presentation. As discussed in the Preface to Volume I, *not everything in the text is supposed to be (or could be) covered in the classroom.* I prefer to use lecture time for discussing the major results and letting students work on some problems (algebraically and with a computer), leaving some derivations and examples for reading outside of the classroom.

The companion website for the book is `http://www.wiley.com/go/intermediate`.

ACKNOWLEDGEMENTS

I am indebted to Ronald Butler for teaching and working with me on several saddlepoint approximation projects, including work on the doubly noncentral F distribution, the results of which appear in Chapter 10. The results on the saddlepoint approximation for the doubly noncentral *t* distribution represent joint work with Simon Broda. As mentioned above, Simon also contributed greatly to the section on multivariate saddlepoint methods. He has also devised some advanced exercises in Chapters 1 and 10, programmed Pan's (1968) method for calculating the distribution of a weighted sum of independent, central χ^2 r.v.s (see Section 10.1.4), and has proofread numerous sections of the book. Besides helping to write the technical sections in Chapter 8, Yianna Tchopourian has proofread Chapter 4 and singlehandedly tracked down the sources of all the quotes I used in this book. This book project has been significantly improved because of their input and I am extremely greatful for their help.

It is through my time as a student of, and my later joint work and common research ideas with, Stefan Mittnik and Svetlozar (Zari) Rachev that I became aware of the usefulness and numeric tractability via the fast Fourier transform of the stable Paretian distribution (and numerous other fields of knowledge in finance and statistics). I wish to thank them for their generosity, friendship and guidance over the last decade.

As already mentioned, Chapter 9 was written by Walther Paravicini, and he deserves all the credit for the well-organized presentation of this interesting and nontrivial subject matter. Furthermore, Walther has proofread the entire book and made substantial suggestions and corrections for Chapter 1, as well as several hundred comments and corrections in the remaining chapters. I am highly indebted to Walther for his substantial contribution to this book project.

One of my goals with this project was to extend the computing platform from Matlab to the R language, so that students and instructors have the choice of which to use. I wish to thank Sergey Goriatchev, who has admirably done the job of translating all the Matlab programs appearing in Volume I into the R language; those for the present volume are in the works. The Matlab and R code for both books will appear on the books' web pages.

Finally, I thank the editorial team Susan Barclay, Kelly Board, Richard Leigh, Simon Lightfoot, and Kathryn Sharples at John Wiley & Sons, Ltd for making this project go as smoothly and pleasantly as possible. A special thank-you goes to my copy editor, Richard Leigh, for his in-depth proofreading and numerous suggestions for improvement, not to mention the masterful final appearance of the book.

PART I

SUMS OF RANDOM VARIABLES

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.

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

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

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 $X \sim \mathrm{DUnif}(\theta_1, \theta_2)$ 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 $U \sim \mathrm{Unif}(0, 1)$. Then,

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

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

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

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

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

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

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

For $X \sim \text{Unif}(a, b)$, 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

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

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

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

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 $X \sim \text{Poi}(\lambda)$,

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

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,

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

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

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

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 $\mathbb{E}[N_m]$. ■

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 $|t| < h$, $\lim_{n\to\infty} \mathbb{M}_{X_n}(t) = \exp\left\{\lambda\left(\mathrm{e}^t - 1\right)\right\} = \mathbb{M}_P(t)$, where $P \sim \mathrm{Poi}(\lambda)$. That is, $X_n \xrightarrow{d} \mathrm{Poi}(\lambda)$. 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 $P_\lambda \sim \mathrm{Poi}(\lambda)$, $\lambda \in \mathbb{R}_{>0}$, and $Y_\lambda = (P_\lambda - \lambda)/\sqrt{\lambda}$. From (1.5),

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

Writing

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

we see that

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

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, $Y_\lambda \xrightarrow{d} \mathrm{N}(0,1)$ as $\lambda \to \infty$. This should not be too surprising in light of the skewness and kurtosis results of Example 1.4.

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

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

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

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

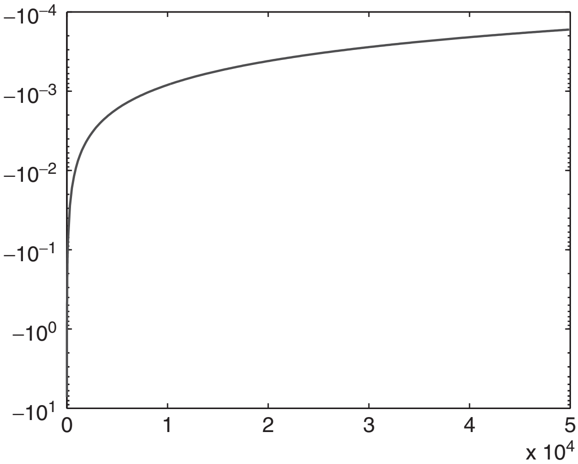

Figure 1.1 The relative percentage error of (1.19) as a function of λ

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

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

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,

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

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

and

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

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

$$\begin{aligned}
&\frac{1}{\Gamma(k)}\int_{-\infty}^{0}(-t_1)^{k-1}\left[\frac{\partial^k}{\partial t_2^k}\mathbb{E}\left[\mathrm{e}^{t_1X_1}\mathrm{e}^{t_2X_2}\right]\right]_{t_2=0}\mathrm{d}t_1\\
&=\frac{1}{\Gamma(k)}\mathbb{E}\left[\left[\frac{\partial^k}{\partial t_2^k}\mathrm{e}^{t_2X_2}\right]_{t_2=0}\int_{-\infty}^{0}(-t_1)^{k-1}\mathrm{e}^{t_1X_1}\,\mathrm{d}t_1\right]\\
&=\frac{1}{\Gamma(k)}\mathbb{E}\left[X_2^k\int_0^\infty u^{k-1}\mathrm{e}^{-uX_1}\,\mathrm{d}u\right]=\mathbb{E}\left[\left(\frac{X_2}{X_1}\right)^k\right].
\end{aligned}$$

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),

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

***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 $\mathbb{E}\left[e^{itX}\right]$, 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 $z \cdot \overline{z} = (a + bi)(a - bi) = a^2 - b^2 i^2 = a^2 + b^2$ 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

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

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

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

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

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

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),

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

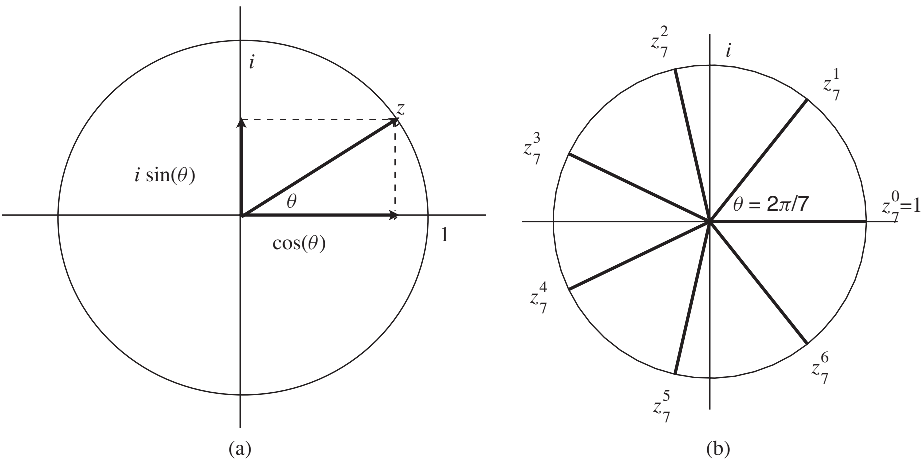

Figure 1.2 (a) Geometric representation of complex number $z = \cos(\theta) + i\sin(\theta)$ in the complex plane. (b) Plot of powers of $z_n = \exp(2\pi i/n)$ for $n = 7$, demonstrating that $\sum_{j=0}^{n-1} z_n^j = 0$

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

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

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

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

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

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

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),

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

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),

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

***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

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

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),

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

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 $Y \sim \mathrm{Unif}(0, 1)$. ■