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

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., $\chi^2 \subseteq \mathrm{Gam}$ and $\mathrm{N} \subseteq \mathrm{S}\alpha\mathrm{S}$), 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.

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

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

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

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

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

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

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,

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

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

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

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 $\mathbb{E}[N_m]$. ■

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 $|t| < h$, $\lim_{n\to\infty} \mathbb{M}_{X_n}(t) = \exp\left\{\lambda\left(e^t - 1\right)\right\} = \mathbb{M}_P(t)$, 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

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

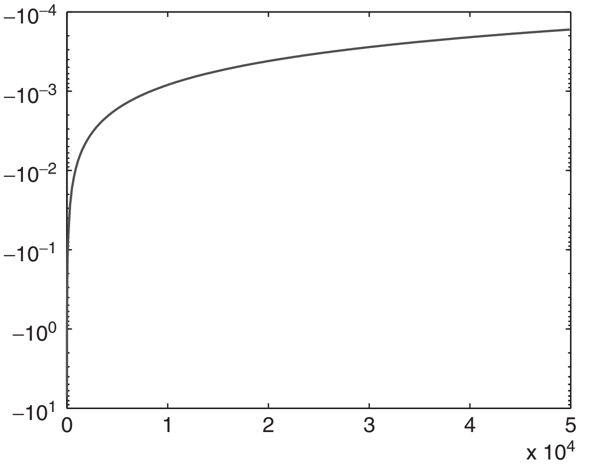

Figure 1.1 The relative percentage error of (1.19) as a function of λ

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 $a \to \infty$, $Y_a \overset{d}{\to} \text{N}(0, 1)$, 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 $S_n \sim \text{Gam}(n, 1)$ for $n \in \mathbb{N}$, so that, for large n, $S_n \overset{\text{app}}{\sim} \text{N}(n, n)$. 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,

$$\frac{\partial^k \mathbb{M}_{\mathbf{X}}(\mathbf{t})}{\partial t_1^{k_1} \partial t_2^{k_2} \cdots \partial t_n^{k_n}} = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} x_1^{k_1} x_2^{k_2} \cdots x_n^{k_n} \exp\{t_1x_1 + t_2x_2 + \cdots + t_nx_n\} f_{\mathbf{X}}(\mathbf{x})\, \mathrm{d}\mathbf{x},$$

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

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