ASYMPTOTIC AND NUMERICAL METHODS FOR APPROXIMATING DISTRIBUTIONS

by

Min Tsao

M.Sc., Simon Fraser University, Canada, 1990

A THESIS SUBMITTED IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY
in the Department of Mathematics & Statistics

© Min Tsao 1995
SIMON FRASER UNIVERSITY
August 1995

APPROVAL

Name: Min Tsao

Degree: Doctor of Philosophy

Title of thesis: Asymptotic and Numerical Methods for Approximating Distributions

Examining Committee:

Chairman: Dr. G. A. C. Graham

Dr. R. D. Routledge
Senior Supervisor

Dr. C. B. Dean

Dr. R. A. Lockhart

Dr. C. J. Schwarz

Dr. B. Jørgensen
External Examiner
University of British Columbia

Date Approved: June 30, 1995

PARTIAL COPYRIGHT LICENSE

I hereby grant to Simon Fraser University the right to lend my thesis, project or extended essay (the title of which is shown below) to users of the Simon Fraser University Library, and to make partial or single copies only for such users or in response to a request from the library of any other university, or other educational institution, on its own behalf or for one of its users. I further agree that permission for multiple copying of this work for scholarly purposes may be granted by me or the Dean of Graduate Studies. It is understood that copying or publication of this work for financial gain shall not be allowed without my written permission.

Title of Thesis/Project/Extended Essay

Asymptotic and Numerical Methods for

Approximating Distributions.

Author:_____
    (signature)

        MIN TSAO
(name)

        May 11, 1995
(date)

Abstract

The problem of computing the distribution of a sum or an average of independently, identically distributed observations arises in many statistical applications. In most cases, asymptotic methods or numerical methods have to be used.

Daniels' (1954) saddlepoint expansion for the density function and Lugannani and Rice's (1980) expansion for the cumulative distribution function are amongst the most widely used asymptotic methods. Although the integrated saddlepoint expansion is generally considered to be as accurate as Lugannani and Rice's expansion for approximating the cumulative distribution function, the theoretical relationship between these two expansions remains largely unknown. We show that Lugannani and Rice's expansion may be differentiated to obtain the saddlepoint expansion, and give the exact relationship between their coefficients. We then discuss two applications of this result, and study the uniform validity of the saddlepoint expansion, a problem which arises naturally in our investigation of the relationship.

We also study asymptotic expansions for general distributions that are asymptotically normal. We discuss a formal method for deriving expansions and a family of expansions to which it leads. Through this family, we examine and compare known expansions, such as the Edgeworth expansion and the saddlepoint expansion, and discuss generalizations of these. We also examine the accuracy of these generalizations to problems where exact solutions are available and demonstrate that they are indeed accurate.

A numerical method for computing distributions whose moment generating functions are known is also discussed. The method makes use of the fact that the moment generating function is an integral transformation of the density function, and computes the density function by solving the integral equation that defines the moment generating function. The advantages of this method compared to the asymptotic methods will be discussed.

Dedication

To my parents

Acknowledgments

Contents

Chapter 1

Introduction

In many statistical applications, distributions of various statistics, frequently some functions of a sample mean, play a central role. The exact distributions, however, are often difficult to obtain. The problem of approximating these distributions has been studied by generations of statisticians, and continues to provide a fertile ground for new research. This thesis consists of work centered on three subjects concerning asymptotic and numerical methods for approximating these distributions.

Efforts for obtaining accurate approximations for the distribution of a sample mean have not only resulted in reliable numerical approximations, but have also led to major theoretical breakthroughs in asymptotic theory related to statistics and probability. Assume that $X_1, \ldots, X_n$ are independently identically distributed (i.i.d.) with an underlying density $f(x)$. Let $\bar{X}$ be their arithmetic mean. Then $\bar{X}$ is usually asymptotically normally distributed in the sense that its standardized version is asymptotically standard normal. Although the exact distribution of $\bar{X}$ may be difficult to calculate, its moment generating function often is not. Methods exist that allow one to approximate the distribution based on the first few moments of $\bar{X}$, e.g., fitting a Pearson curve, or using the first few terms of the Edgeworth expansion. These methods usually provide satisfactory approximation in the center of the distribution, but they are often not accurate in the tail. The Pearson curve method, though typically accurate out as far as 5th or 95th percentiles, is often inaccurate beyond that range and is of limited value as a theoretical tool. The Edgeworth expansion

is perhaps more important as a theoretical tool than as a practical approximation. It too is only accurate in the center of the distribution, and can even be negative in the far tails.

In his pioneering paper in 1954, H. E. Daniels introduced the method of saddlepoint approximation into statistics and derived an asymptotic series expansion for the density function. This asymptotic expansion is obtained by first expressing the density as the Fourier inversion of its moment generating function, and then expanding the inversion formula using the method of steepest descents. Unlike Edgeworth expansion, which is in powers of $1/\sqrt{n}$, Daniels' expansion is in powers of $1/n$. It provides accurate approximation to the density, even for very small sample sizes. Perhaps the most important feature of this expansion is that its accuracy usually does not deteriorate in the tail. The leading term of Daniels' series is called the saddlepoint approximation.

The corresponding asymptotic expansion for the cumulative distribution function was obtained by Lugannani and Rice in 1980. It was also obtained by first expressing the tail area of the distribution using a Fourier inversion formula, and then expanding this formula while taking into consideration a pole in the integrand. Like Daniels' expansion for the density function, it is in powers of $1/n$ while the Edgeworth expansion for the cumulative distribution function is in powers of $1/\sqrt{n}$. It is very accurate over the entire domain of $\bar{x}$, even for small sample sizes. Although the numerically integrated saddlepoint approximation has also been known as an accurate approximation to the cumulative distribution function, Lugannani and Rice's approximation has the advantage that it does not involve numerical integration, and is thus easy to compute.

The first subject of this thesis is the relationship between Daniels' expansion for the density and Lugannani and Rice's expansion for the corresponding cumulative distribution function. Our interest in this relationship was initially raised by an interesting conjecture by Lugannani and Rice. In their 1980 paper, they compared their series with the integrated Daniels' series numerically, and observed that they both have errors of comparable size. They therefore conjectured that their series and the integrated Daniels' series should always be "in errors by the same order of magnitude". Due to the complexity of the expressions for the errors of these two approximations, a direct algebraic comparison seems intractable. We shall therefore turn our attention to the mathematical relationship between the two series, and try to answer the conjecture by first understanding this relationship. Since the

cumulative distribution function may be differentiated to obtain the corresponding density function, our investigation focuses on whether or not a similar relationship exists between these two expansions. We shall prove that such a relationship indeed exists. This relationship not only leads to an answer to Lugannani and Rice's conjecture, but also establishes the derivative of any truncated Lugannani and Rice's series as an asymptotic approximation to the density function.

A problem closely related to our proof for this relationship is that of the uniform validity of Daniels' expansion. We shall discuss general conditions under which it is uniformly valid on compact subsets, and show that these conditions are met by commonly used continuous density functions for which the saddlepoint approximation can be derived.

The second subject centers on general expansions for densities of distributions that are asymptotically normal but not necessarily that of a sample mean or standardized mean. We introduce a formal method for deriving expansions for these densities and derive a family of formal asymptotic expansions, which includes the saddlepoint expansion and the Edgeworth expansion as special cases. This family of expansions is defined with respect to sequences derived from the cumulant generating functions. These sequences are usually asymptotic for distributions that are asymptotically normal. The derivation of this family of expansions also hints that such sequences are the most natural sequences with respect to which the asymptotic expansions of the densities be defined. The sequences simplify to essentially $\{1/(\sqrt{n})^r\}$ and $\{1/n^r\}$ for the cases of standardized mean and sample mean. The validity of these expansions, however, needs to be established for each case, and this in general can be difficult. We shall consider Jørgensen's exponential dispersion models for which the validity can be established. We shall also study the numerical accuracy of the saddlepoint approximation in this broader context through numerical examples.

The third subject is concerned with a numerical method for computing distributions whose moment generating functions are known. We look at the equation that defines (defining equation for) a moment generating function from the integral equation point of view, and compute the distribution by solving the Fredholm integral equation of the first kind given by the defining equation. Fredholm integral equations of the first kind, however, are usually difficult to compute, and existing methods do not seem to work well for our purpose of computing the defining equation. We shall first discuss a way to refine the quadrature method

for Fredholm equations of the first kind, a simple method that can be easily implemented. This leads to a general method which may be used to solve a variety of equations, including the defining equation for the moment generating function. This general method will be presented as a method for Fredholm equations of the first kind. Its relevance to the theme of the thesis, i.e., approximating distributions, is discussed separately. The method is easy to carry out for distributions with bounded domains. When the domains are unbounded, however, the method is difficult to implement. For these cases, we shall discuss the use of Bellman *et al.*'s (1966) quadrature method for computing distributions. These numerical methods have certain advantages over asymptotic methods, e.g., the moment generating functions need be evaluated only at a small number of points. This and other advantages will be discussed and illustrated with examples.

Following is an overview of the thesis:

In Chapter 2 we first review various properties of complex moment generating functions. These properties lead to the inversion integrals for both the density function and the cumulative distribution function, from which asymptotic expansions were found. We then review how Daniels' expansion for the density function of the mean of a sample of i.i.d. observations and Lugannani and Rice's expansion for the corresponding cumulative distribution function were derived.

In Chapter 3, we first formally differentiate Lugannani and Rice's expansion for the cumulative distribution function and obtain an asymptotic series that resembles Daniels' series for the density function. We then prove that this formal differentiation is valid under the condition that Daniels' series is uniformly valid, and the resulting series is indeed that of Daniels'. The uniform validity of Daniels' series, together with applications of the relationship between the two series will also be discussed in this chapter.

Chapter 4 begins with a discussion on Charlier differential series and the formal method for deriving asymptotic expansions. The family of expansions is then derived through the method. We then study this family for the sample mean and the standardized sample mean in detail. We also discuss Jørgensen's saddlepoint approximation for exponential dispersion models in the context of this family, and some numerical examples which illustrate the accuracy of the saddlepoint approximation.

In Chapter 5, we discuss general quadrature methods and our method for solving Fredholm integral equations of the first kind. We then discuss the use of our method and Bellman *et al.*'s (1966) method for computing distributions and compare them with asymptotic methods.

We finish this introduction with a brief review of work which we shall cite most frequently. Daniels' (1954) pioneering work derived the saddlepoint expansion for the density of the mean of a sample of i.i.d. observations. He also discussed the existence of the saddlepoint, as well as the uniform validity of the saddlepoint approximation for four important classes of densities. Barndorff-Nielsen and Cox (1979) brought the importance of saddlepoint approximation to light, and showed how it may be used in a variety of important applications. They also further discussed the uniform validity of the saddlepoint expansion. Lugannani and Rice (1980) derived the corresponding expansion for the tail probability of the sample mean. This expansion is uniformly valid, and compares very favorably to other asymptotic approximations for the tail probability. Daniels (1987) compared Lugannani and Rice's approximation to other approximations, and extended this approximation to the tail probability for the mean of i.i.d. lattice random variables. He also reproduced Lugannani and Rice's derivation in more convenient notation. Jensen (1988, 1991) gave important results on the uniform validity of Daniels' saddlepoint expansion and an approximation for the tail probability. Reid (1988) provided a comprehensive review of papers on saddlepoint approximation and its applications. Finally, we have found the two books by Field and Ronchetti (1990) and Barndorff-Nielsen and Cox (1989) very helpful in preparing this thesis. Both books provide a detailed account of the development of the method of saddlepoint approximation, as well as some of its more important applications. Barndorff-Nielsen and Cox (1989) also discussed a broad range of other asymptotic techniques in statistics.

Terminology and notation: We use the terms 'asymptotic expansion', 'asymptotic series' and 'asymptotic approximation', interchangeably. We say that $f(x)$ is *asymptotically equivalent or equal to* $g(x)$ under the limit $x \to x_0$ and write $f(x) \sim g(x)$ if $f(x)/g(x) \to 1$ as $x \to x_0$. We use R to denote the set of all real numbers, and R_+ the set of all positive real numbers. Unless specified, a density function is assumed to be continuous (in its *domain*) and defined on R. By the *domain* of a density function, $f(x)$, we mean the shortest interval in which $f(x) \not\equiv 0$. Finally, by $f(x) \in L^p(R)$, we mean that $\int_R |f(x)|^p dx < \infty$.

Chapter 2

Two Expansions for the

Distribution of the Sample Mean

In this chapter, we show how Daniels' saddlepoint expansion for the density function and Lugannani and Rice's expansion for the tail probability of the sample mean may be derived. The basic steps that led to both expansions are: (i) expressing the function to be expanded as the inversion of the characteristic function, and (ii) expanding the inversion integral where the sample size plays the role of a large parameter at the saddlepoint. We shall present step (i) in a slightly different way, i.e., we shall express the function to be expanded as the inversion of the complex moment generating function. A definition of a complex moment generating function is given below. Its important properties, including those we shall use for step (i), are discussed in the first section. We choose to work with the complex moment generating function instead of the characteristic function because it leads directly, without any additional complex analysis argument, to the inversion formulas needed for step (ii). More importantly, we need to use it extensively in later chapters. Step (ii) is discussed in the second section.

2.1 Complex moment generating functions

To be consistent with conventional notation used in complex analysis, in this section we use $z = x + iy$ where $x, y \in R$ to denote a complex variable. For this section only, we use T for a random variable with cumulative distribution function $F(t)$ and continuous probability density function $f(t) = F'(t)$. We call $M(z)$ given by the following integral transformation of $f(t)$ the complex moment generating function of T:

$$M(z) = \int_{-\infty}^{\infty} e^{zt} f(t) dt. \tag{2.1.1}$$

When $y = 0$, the complex moment generating function is the ordinary real moment generating function $M(x)$. When $x = 0$, it is the characteristic function, $M(iy)$. In this thesis, we only consider complex moment generating functions that are analytic in some *open vertical strip* containing the imaginary axis. The corresponding characteristic functions are thus also analytic. We call $K(z) = \log M(z)$ the complex cumulant generating function.

If the kernel in (2.1.1) is e^{-zt} rather than e^{zt}, the resulting integral transformation of $f(t)$ is the well-known bilateral Laplace-Stieltjes transformation. Denote this transformation by $L(z)$, then

$$M(z) = L(-z). \tag{2.1.2}$$

Bilateral Laplace-Stieltjes transformations have been shown to possess many important properties, see e.g., Kawata (1972) and LePage (1961). These properties can be transferred directly to $M(z)$ through (2.1.2). The only reason for working with $M(z)$ rather than $L(z)$ is that $M(z)$ is the direct extension of the moment generating function, which is commonly used in statistics. Although we are mainly interested in proper distributions with continuous densities, most of the results included in this section actually hold for integral transformations (2.1.1) of a much broader class of real functions.

• *Region of existence and elementary properties*

The collection of points at which the real moment generating function $M(x)$ exists forms an interval I where $0 \in I \subseteq R$. The complex moment generating function $M(z)$ exists in a vertical strip G in the complex plane (Z-plane) that contains the imaginary axis. Furthermore, the interval I is the intersection of G and the real axis. $M(z)$ has the following

two important properties: (a) it is analytic in G; and (b) $M(0) = 1$ and $|M(x+iy)| \leq M(x)$, with equality if and only if $y = 0$.

Proofs for the region of existence and the analyticity of $M(z)$ follow from those for the Bilateral Laplace-Stieltjes transformation, which may be found in Kawata (1972) and LePage (1961). A proof for the inequality in (b) may be found in Daniels (1954).

● *Calculation of the complex moment generating function*

The complex moment generating function may be simply obtained by replacing the x in $M(x)$ by z. To see this is true, $M(x) = M(x + i0)$ is an analytic function of x in I. Upon replacing x by z, we obtain the analytic continuation of $M(x)$, which is analytic and coincides with $M(z)$ in G. As a special case, upon replacing x in $M(x)$ by iy, one obtains the characteristic function of T.

● *Asymptotic properties*

Lemma 2.1: *Let $Re(z) = x \in I$ where I is the real interval in which $M(x)$ exists, then*

$$|M(z)| = \left| \int_{-\infty}^{\infty} e^{zt} f(t) dt \right| \to 0 \ \ as \ \ |y| \to \infty,$$

where $z = x + iy$.

Proof: Let $h(t) = e^{xt} f(t)$. Since $x \in I$, $h(t)$ is integrable over R. $M(z)$ can thus be regarded as the Fourier transform of $h(t)$. The Riemann-Lebesgue lemma then implies that $|M(z)| \to 0$ as $|y| \to \infty$. #

Furthermore, assume that $f(t)$ is differentiable in its domain. Then $h(t)$ is also differentiable in the domain. If $h(t)$ vanishes at the extremes of the domain, and $h'(t)$ is integrable, then $|M(z)| = o(|y|^{-1})$ as $|y| \to \infty$. This result may be proved by using the arguments in the proof for Lemma 4 on page 487 in Feller (1966). Note that there is an error in this Lemma. Although it deals with characteristic functions only, a condition equivalent to $h(t)$ vanishing at the extremes of the domain is needed but is not included. If $h(t)$ is bounded but is not zero at at least one of the extremes and $h'(t)$ is integrable, then $|M(z)| = O(|y|^{-1})$. For example, for the exponential distribution the complex moment generating function is $M(z) = 1/(1 - z)$ and $h(t)$ is e^{xt-t}. It is easy to see that as t approaches the lower extreme of the domain, $t = 0$, $h(t) \to 1$ and that $|M(z)| = O(|y|^{-1})$ as $|y| \to \infty$.

Theorem 2.1: *Let $f(t)$ be a continuous density function, and $[x_1, x_2]$ be a bounded subinterval of I. Then $|M(z)| \to 0$ uniformly with respect to $x \in [x_1, x_2]$ as $|y| \to \infty$.*

Proof: When $f(t)$ vanishes outside a finite interval, the above theorem may be proved by using a result on the uniform convergence of the Laplace transform of an almost piecewise continuous function[1] on a bounded domain (see, e.g., Theorem 12-1 in LePage, 1961).

We now show that for other density functions, for any $\epsilon > 0$, there exists a $y(\epsilon) > 0$ such that $|M(z)| < \epsilon$ when $|y| > y(\epsilon)$ for all $x \in [x_1, x_2]$. For any $C > 0$,

$$
\begin{aligned}
|M(z)| &= \left| \int_{-\infty}^{\infty} e^{zt} f(t) dt \right| \\
&\leq \int_{-\infty}^{-C} e^{xt} f(t) dt + \left| \int_{-C}^{C} e^{zt} f(t) dt \right| + \int_{C}^{\infty} e^{xt} f(t) dt \\
&\leq \int_{-\infty}^{-C} e^{x_1 t} f(t) dt + \left| \int_{-C}^{C} e^{zt} f(t) dt \right| + \int_{C}^{\infty} e^{x_2 t} f(t) dt.
\end{aligned}
\tag{2.1.3}
$$

Since $M(x_1)$ and $M(x_2)$ exist, there exists $C > 0$ such that the first and third terms in (2.1.3) are each less than $\epsilon/4$. For the integral in the middle, we may treat $f(t)$ as if it vanishes outside $(-C, C)$. Thus there exists a $y(\epsilon) > 0$ such that it is bounded by $\epsilon/2$ when $|y| > y(\epsilon)$ uniformly for $t \in [x_1, x_2]$. It follows that when $|y| > y(\epsilon)$, $|M(z)|$ is uniformly bounded by ϵ. #

• *Inversion formula*

$F(t)$ and $f(t)$ may be obtained by inverting the complex moment generating function. To derive the inversion formula, we need the following result for the bilateral Laplace-Stieltjes transformation. More general versions of this result and their proofs may be found on pages 267 and 268 in Kawata (1972). Denote the region of existence for $L(z)$ by G_L, we have $G_L = \{z : -Re(z) \in I\}$.

Theorem 2.2: *Let $F(t)$ be a proper distribution with density function, $f(t) = F'(t)$, which is continuous in its domain. Let $z = c + iy \in G_L$. Then*

$$
(i) \quad \frac{1}{2\pi i} \int_{c-i\infty}^{c+i\infty} \frac{L(z)}{z} e^{zt} dz = \begin{cases} F(t) & \text{if } c > 0, \\ F(t) - 1 & \text{if } c < 0, \end{cases}
\tag{2.1.4}
$$

$$
and \quad (ii) \quad \frac{1}{2\pi i} \int_{c-i\infty}^{c+i\infty} L(z) e^{zt} dz = f(t).
\tag{2.1.5}
$$

---

[1]See page 240 in LePage (1961) for a definition.

We now prove the following inversion formula of $M(z)$ by using the above result.

Theorem 2.3: *Let $F(t)$ be a proper distribution with density function, $f(t) = F'(t)$, which is continuous in its domain. Let $z = b + iy \in G$. Then*

$$(i) \quad \frac{1}{2\pi i} \int_{b-i\infty}^{b+i\infty} \frac{M(z)}{z} e^{-zt} dz = \begin{cases} -F(t) & \text{if } b < 0, \\ 1 - F(t) & \text{if } b > 0. \end{cases} \quad (2.1.6)$$

$$and \quad (ii) \quad \frac{1}{2\pi i} \int_{b-i\infty}^{b+i\infty} M(z) e^{-zt} dz = f(t). \quad (2.1.7)$$

Proof:

(i) When $b < 0$, let $c = -b > 0$, then $c + iy \in G_L$ and (2.1.4) gives

$$\begin{aligned} F(t) &= \frac{1}{2\pi i} \int_{c-i\infty}^{c+i\infty} \frac{L(z)}{z} e^{zt} dz \\ &= \frac{1}{2\pi i} \int_{c-i\infty}^{c+i\infty} \frac{M(-z)}{z} e^{zt} dz \\ &= -\frac{1}{2\pi i} \int_{b-i\infty}^{b+i\infty} \frac{M(z)}{z} e^{-zt} dz. \end{aligned}$$

Similarly, we can use (2.1.4) to show that when $b > 0$,

$$F(t) - 1 = -\frac{1}{2\pi i} \int_{b-i\infty}^{b+i\infty} \frac{M(z)}{z} e^{-zt} dz.$$

(ii) may be obtained similarly by using (2.1.5). #

• *Complex moment generating function of a convolution*

Let T_1 and T_2 be two independent random variables with distribution functions $F_1(t)$ and $F_2(t)$. Let $M_1(z)$ and $M_2(z)$ be the complex moment generating functions of these two random variables. Then the convolution $F_1 * F_2$ has complex moment generating function $M_1(z) \times M_2(z)$.

This property, together with the inversion formula of the complex moment generating function, enable us to express the density function and distribution function of the mean of n independently, identically distributed observations as the inverse transformations of the nth power of the complex moment generating function of a single observation. The inversion integrals are then expanded to obtain the saddlepoint expansion to the density, and Lugannani and Rice's expansion to the distribution function.

2.2 The two expansions

To prepare for our investigation on the relationship between Daniels' expansion for the density and Lugannani and Rice's expansion for the tail probability (Chapter 3), it is helpful to see first how they are derived. There are different ways for deriving these expansions. Derivations shown in this section are based on that given by Daniels (1954 and 1987).

For consistency with Daniels' notation, we now use $T = \tau + iy$ for a complex variable instead of $z = x + iy$. Also, we use $M(T)$ to denote the complex moment generating function of a single observation X_i, and $\bar{X}$ to denote the mean of a sample of size n.

- *The integrability of $|M(T)|$*

The condition that for each $\tau \in I$, there exists a $\nu > 0$ such that

$$\int_{\tau-i\infty}^{\tau+i\infty} |M(T)|^\nu dy < \infty \qquad (2.2.8)$$

is important to the derivation of Daniels' expansion. First of all, it ensures that $\bar{X}$ will actually *have* a density function, which can be expressed as a contour integral involving $M(T)$. Secondly, it is indispensable in proving that Daniels' expansion for the integral is indeed an asymptotic expansion of the density function. For brevity, we write $|M(T)| \in L^\nu(R)$ if $M(T)$ satisfies (2.2.8). We now show that it implies that $|M(T)| \in L^\gamma(R)$ for any $\gamma \geq \nu$, and that the density function of $\bar{X}$, $f_n(\bar{x})$, exists and is continuous for $n \geq \nu$.

By the elementary properties of $M(T)$, $|M(T)/M(\tau)| \leq 1$. Thus for any $\gamma > \nu$,

$$
\begin{aligned}
\int_{\tau-i\infty}^{\tau+i\infty} |M(T)|^\gamma dy &= M^\gamma(\tau) \int_{\tau-i\infty}^{\tau+i\infty} \left|\frac{M(T)}{M(\tau)}\right|^\gamma dy \\
&\leq M^{\gamma-\nu}(\tau) \int_{\tau-i\infty}^{\tau+i\infty} |M(T)|^\nu dy < \infty, \qquad (2.2.9)
\end{aligned}
$$

i.e., $|M(T)| \in L^\gamma(R)$, or equivalently, $|M(T)|^\gamma \in L^1(R)$. To see that it ensures the existence and continuity of $f_n(\bar{x})$, assume that $\nu = \nu_0$ at $\tau = 0$. Then for $n \geq \nu_0$ the characteristic function of $\sum X_i$, $M^n(iy)$, is in $L^1(R)$. It follows from the Fourier inversion theorem, e.g., Theorem 3 on page 482 in Feller (1966), that the density function for $\sum X_i$ and thus $f_n(\bar{x})$ are bounded and continuous. *For the rest of this chapter, we assume that (2.2.8) is satisfied.* Although we only need it at $\tau = 0$ for the existence and continuity of $f_n(\bar{x})$, we shall need it at other values of τ later on for demonstrating the asymptotic nature of Daniels' expansion.

It is not difficult to see that (2.2.8) is satisfied when $|M(T)|$ is $O(|y|^{-1})$ or $o(|y|^{-1})$ as $|y| \to \infty$. See Section 2.1 regarding this and other asymptotic properties of $M(T)$. Later in Section 3.4, we shall prove that it is satisfied by complex moment generating functions of *all* commonly used continuous distributions.

• *Inversion formulas for the density and tail probability of* $\bar{X}$

Using the continuity of $f_n(\bar{x})$, the convolution property in Section 2.1, and the inversion formula in Theorem 2.3, one can show that $f_n(\bar{x})$ can be expressed in terms of $M(T)$ as

$$f_n(\bar{x}) = \frac{n}{2\pi i} \int_{\tau - i\infty}^{\tau + i\infty} M^n(T) e^{-nT\bar{x}} dT, \qquad (2.2.10)$$

Recall that $K(T) = \log M(T)$, thus (2.2.10) may be written as

$$f_n(\bar{x}) = \frac{n}{2\pi i} \int_{\tau - i\infty}^{\tau + i\infty} e^{n[K(T) - T\bar{x}]} dT. \qquad (2.2.11)$$

Similarly, the tail probability $Q_n(\bar{x}) = P(\bar{X} > \bar{x})$ can be expressed as

$$Q_n(\bar{x}) = \frac{1}{2\pi i} \int_{c - i\infty}^{c + i\infty} e^{n[K(T) - T\bar{x}]} \frac{dT}{T}, \qquad (2.2.12)$$

where $c > 0$ and $c \in I$.

• *Watson's lemma*

The asymptotic nature of the expansions will be established with the aid of the following version of Watson's lemma. Discussions concerning this lemma may be found in Copson (1965) or Jeffreys and Jeffreys (1950).

Watson's Lemma: *If* $\psi(z)$ *is analytic in a neighborhood of* $z = 0$ *and bounded for real* $z = w$ *in an interval* $-A \leq w \leq B$ *with* $A > 0$ *and* $B > 0$, *then*

$$\sqrt{\frac{n}{2\pi}} \int_{-A}^{B} e^{-nw^2/2} \psi(w) dw \sim \psi(0) + \frac{1}{2n} \psi''(0) + \cdots + \frac{1}{(2n)^r} \psi^{(2r)}(0) + \cdots$$

is an asymptotic expansion in powers of n^{-1}.

• *The saddlepoint*

The saddlepoint corresponding to $\bar{x}$, $\hat{T}$, is the real number in the region of analyticity of $M(T)$ that satisfies the equation

$$K'(T) - \bar{x} = 0. \qquad (2.2.13)$$

To see that $\hat{T}$ is indeed a saddlepoint, recall from the elementary properties that $|M(\hat{T} + iy)| \leq |M(\hat{T})|$. Thus $|\exp\{K(\hat{T} + iy)\}| \leq |\exp\{K(\hat{T})\}|$, or equivalently $|\exp\{K(\hat{T} + iy) - (\hat{T} + iy)\bar{x}\}| \leq |\exp\{K(\hat{T}) - \hat{T}\bar{x}\}|$. It follows that $Re\{K(\hat{T} + iy) - (\hat{T} + iy)\bar{x}\} \leq K(\hat{T}) - \hat{T}\bar{x}$ for any y, and thus $\frac{\partial}{\partial y}Re\{K(T) - T\bar{x}\} = 0$ at $T = \hat{T}$. This and (2.2.13) imply that $\nabla Re\{K(\hat{T}) - \hat{T}\bar{x}\} = 0$. By the analyticity of $K(T) - T\bar{x}$ and the Cauchy-Riemann equations, $\nabla Im\{K(\hat{T}) - \hat{T}\bar{x}\} = 0$. Thus $\hat{T}$ is either a maximum or a saddlepoint. By the maximum modulus principle, $\hat{T}$ must be a saddlepoint.

In general, the saddlepoint has to be computed numerically. Since $K''(\hat{T}) > 0$, $\hat{T}$ is an increasing function of $\bar{x}$. Also, $K'(0) = \mu \ (= E(X_i))$. Thus the saddlepoint corresponding to μ is 0. For more discussion concerning these properties and the existence and uniqueness of the saddlepoint, see Daniels (1954).

- *The two expansions*

For convenience of our presentation, we now give the two expansions here. Daniels' expansion for $f_n(\bar{x})$ and Lugannani and Rice's expansion for $Q_n(\bar{x})$ are

$$f_n(\bar{x}) \sim g_n(\bar{x}) \sum_{r=0}^{\infty} \frac{a_r}{n^r}, \tag{2.2.14}$$

and

$$Q_n(\bar{x}) \sim 1 - \Phi(\hat{W}n^{\frac{1}{2}}) + \phi(\hat{W}n^{\frac{1}{2}}) \sum_{r=0}^{\infty} \frac{b_r}{n^{r+\frac{1}{2}}}, \tag{2.2.15}$$

where $g_n(\bar{x})$ is the saddlepoint approximation given by

$$g_n(\bar{x}) = \left[\frac{n}{2\pi K''(\hat{T})}\right]^{\frac{1}{2}} e^{n[K(\hat{T}) - \hat{T}\bar{x}]}, \tag{2.2.16}$$

$\hat{W} = sgn(\hat{T})\{2[\hat{T}K'(\hat{T}) - K(\hat{T})]\}^{\frac{1}{2}}$, Φ and ϕ are the cumulative distribution function and probability density function of the standard normal distribution, respectively. The a_r's and the b_r's are coefficients in (3.3) in Daniels (1954) and (4.5) in Daniels (1987), respectively. In general they are functions of $\bar{x}$. For brevity we may write $K^{(r)}(\hat{T})$ as $\hat{K}^{(r)}$. We may also write $K^{(r)}(\hat{T})/[K''(\hat{T})]^{r/2}$ as $\lambda_r(\hat{T})$ or λ_r. Later in this section, we shall give expressions for a_0, a_1, b_0 and b_1 in terms of $\hat{T}$, the $\hat{K}^{(r)}$'s and the λ_r's.

- *Derivation of Daniels' expansion*

We use the method of steepest descents. We first choose a proper contour, ξ, for the inversion integral (2.2.11), which contains a small section of the curve of steepest descent that passes through the saddlepoint. Using elementary properties of $M(T)$ and the integrability of $|M(T)|$, we show that the contribution to the inversion integral from the part of the contour outside a small neighborhood of the saddlepoint is essentially the product of $g_n(\bar{x})$ and a quantity which converges to zero faster than any power of $1/n$. We then use Watson's Lemma to show that the series in (2.2.14) is an asymptotic expansion for the contribution from the part inside the neighborhood. It follows that the series is an asymptotic expansion for the inversion integral. Consequently, it is an asymptotic expansion for $f_n(\bar{x})$.

The contour is obtained by deforming a small section of the line $T = \hat{T} + iy$ near the saddlepoint as described below. First set τ in (2.2.11) to $\hat{T}$ and then replace the section of $T = \hat{T} + iy$ inside the neighborhood defined by $|T - T_0| < \delta$ by the curve of steepest descent which is that branch of $Im\{K(T) - T\bar{x}\} = 0$ touching $T = \hat{T} + iy$ at $\hat{T}$. We denote this part of the deformed contour by ω. Outside this neighborhood, the contour is continued along curves orthogonal to ω on which $Re\{K(T) - T\bar{x}\}$ remains constant. These curves can be shown to meet $T = \hat{T} + iy$ provided δ is small enough. From the points where they meet, the contour resumes the original $T = \hat{T} + iy$. The value of the inversion integral under this deformed contour is the same as that under the original $T = \hat{T} + iy$ so long as δ is small enough that the deformed contour is inside G. Also, ω can be chosen symmetric with respect to the real axis since $Im\{K(T) - T\bar{x}\}$ is an odd function of y. For more discussion concerning the contour, see Daniels (1954). Field and Ronchetti (1990, p. 28) contains a graph of the contour.

On the above contour of integration, $|e^{K(T)-T\bar{x}}|$ reaches its absolute maximum at the saddlepoint $\hat{T}$. To see this true, we first note that starting from the saddlepoint the value of $Re\{K(T) - T\bar{x}\}$, and hence that of $|e^{K(T)-T\bar{x}}|$, decreases as T moves along ω away from the saddlepoint until the contour reaches the outside of the neighborhood $|T - T_0| < \delta$. It then remains constant until the contour meets the original path $T = \hat{T} + iy$. On the original path, according to the elementary properties of $M(T)$, $|M(T)| < |M(\hat{T})|$. Thus $|M(T)e^{-T\bar{x}}| < |M(\hat{T})e^{-\hat{T}\bar{x}}|$ and, equivalently, $|e^{K(T)-T\bar{x}}| < |e^{K(\hat{T})-\hat{T}\bar{x}}|$.

Furthermore, by the asymptotic properties of $M(T)$, it approaches 0 as $|y|$ approaches infinity. Thus on the contour outside the neighborhood $|T - T_0| < \delta$, ξ/ω, it cannot be

arbitrarily close to $|M(\hat{T})|$; i.e., there exists some $\rho \in (0, 1)$ such that,

$$\frac{\sup\{|M(T)e^{-T\bar{x}}| : T \in \xi/\omega\}}{|M(\hat{T})e^{-\hat{T}\bar{x}}|} \leq \rho.$$

We now rewrite the inversion integral (2.2.11) as

$$f_n(\bar{x}) = \frac{n}{2\pi i} e^{n[K(\hat{T})-\hat{T}\bar{x}]} \int_\xi e^{n[K(T)-T\bar{x}-K(\hat{T})+\hat{T}\bar{x}]} dT. \qquad (2.2.17)$$

A bound on the contribution to the integral in (2.2.17) from ξ/ω is given by

$$\left| \int_{\xi/\omega} e^{n[K(T)-T\bar{x}-K(\hat{T})+\hat{T}\bar{x}]} dT \right| = \left| \int_{\xi/\omega} \left[\frac{M(T)e^{-T\bar{x}}}{M(\hat{T})e^{-\hat{T}\bar{x}}} \right]^n dT \right|$$

$$\leq \frac{\rho^{n-\nu}}{|M(\hat{T})|^\nu} \int_{\xi/\omega} \left| M(T)e^{-T\bar{x}+\hat{T}\bar{x}} \right|^\nu |dT|. \qquad (2.2.18)$$

The integral in (2.2.18) is finite since $|M(T)|^\nu$ is integrable over ξ/ω and $|e^{-T\bar{x}+\hat{T}\bar{x}}|$ is 1 except on a small finite section of ξ/ω just outside the neighborhood. Thus the contribution to the integral in (2.2.17) from ξ/ω is $O(\rho^n)$.

On ω, $K(T) - T\hat{x} - K(\hat{T}) + \hat{T}\bar{x}$ is real and analytic, and at $\hat{T}$, its first derivative is 0. In order to utilize Watson's lemma, we introduce a new variable w such that

$$\begin{aligned}
\frac{-w^2}{2} &= K(T) - T\hat{x} - K(\hat{T}) + \hat{T}\bar{x} \\
&= \frac{1}{2} K''(\hat{T})(T-\hat{T})^2 + \frac{1}{6} K'''(\hat{T})(T-\hat{T})^3 + \cdots \\
&= \frac{1}{2} z^2 + \frac{1}{6} \lambda_3 z^3 + \frac{1}{24} \lambda_4 z^4 + \cdots,
\end{aligned}$$

where $z = (T - \hat{T})[K''(\hat{T})]^{1/2}$, and w is chosen to have the same sign as $Im(z)$ on the contour. Inversion of the series yields an expansion

$$z = iw + \frac{1}{6} \lambda_3 w^2 + \left[\frac{1}{24} \lambda_4 - \frac{5}{72} \lambda_3^2 \right] iw^3 + \cdots$$

convergent in some neighborhood of $w = 0$. This new variable w is a continuous one-to-one function of z and has range $(-A, B)$ where A and B are positive. Thus

$$\int_\omega e^{n[K(T)-T\hat{x}-K(\hat{T})+\hat{T}\bar{x}]} dT = \frac{1}{K''(\hat{T})^{1/2}} \int_{-A}^B e^{-nw^2/2} \frac{dz}{dw} dw. \qquad (2.2.19)$$

By noting that

$$\frac{dz}{dw} = i + \frac{1}{3}\lambda_3 w + i\left[\frac{1}{8}\lambda_4 - \frac{5}{24}\lambda_3^2\right]w^2 + \cdots, \tag{2.2.20}$$

and applying Watson's lemma to the integral in the right-hand side of (2.2.19), we obtain

$$\int_\omega e^{n[K(T) - T\hat{x} - K(\hat{T}) + \hat{T}\hat{x}]}dT \sim i\left[\frac{2\pi}{nK''(\hat{T})}\right]^{1/2}\left\{a_0 + \frac{a_1}{n} + \frac{a_2}{n^2} + \cdots\right\}, \tag{2.2.21}$$

where, by (2.2.20), $a_0 = 1$ and $a_1 = \frac{1}{8}\lambda_4 - \frac{5}{24}\lambda_3^2$. For a general expression for the a_i's, see Daniels (1954). It follows from (2.2.21) and (2.2.18) that

$$\int_\xi e^{n[K(T) - T\hat{x} - K(\hat{T}) + \hat{T}\hat{x}]}dT \sim i\left[\frac{2\pi}{nK''(\hat{T})}\right]^{1/2}\left\{a_0 + \frac{a_1}{n} + \frac{a_2}{n^2} + \cdots\right\}. \tag{2.2.22}$$

Inversion formula (2.2.17) then implies that

$$\begin{aligned}
f_n(\bar{x}) &\sim \left[\frac{n}{2\pi K''(\hat{T})}\right]^{\frac{1}{2}} e^{n[K(\hat{T}) - \hat{T}\bar{x}]}\left\{a_0 + \frac{a_1}{n} + \frac{a_2}{n^2} + \cdots\right\} \\
&= g_n(\bar{x})\left\{a_0 + \frac{a_1}{n} + \frac{a_2}{n^2} + \cdots\right\}.
\end{aligned} \tag{2.2.23}$$

We have obtained Daniels' expansion for the density function (2.2.23).

• *Derivation of Lugannani and Rice's expansion*

It is natural to try a routine application of the method of steepest descents to the inversion integral of the tail probability (2.2.12). Indeed, an asymptotic expansion for the tail probability may be obtained in this way (see Daniels, 1987) at any $\bar{x}$ value except μ. At μ, the saddlepoint is 0 and is thus a singularity of the integrand. The method of steepest descents does not apply. Another problem with this approach is that even though an asymptotic expansion can be obtained for any $\bar{x} \neq \mu$, the expansion does not provide an accurate approximation for $Q_n(\bar{x})$ near μ, even for large n values, due to the presence of the pole.

Lugannani and Rice (1980) used a method developed for obtaining asymptotic expansions for contour integrals where the integrand has a simple pole near a saddlepoint, and derived an asymptotic expansion for the tail probability which is uniformly valid over the entire range of $\bar{x}$. Daniels (1987) gave the following concise account of their derivation.

The basic idea is to evaluate the contributions to the inversion integral from the pole and the saddlepoint separately. This is achieved by introducing a new variable which enables one to write the inversion integral as the sum of two integrals. The value of the first integral represents the contribution from the pole. The integrand of this integral has a pole at the origin, but may be integrated exactly. The value of the second integral represents the contribution from the saddlepoint. The integrand has a saddlepoint but no singularity at or near μ. Watson's lemma is then used to obtain an asymptotic expansion for this integral. The new variable, W, is defined by the relation

$$\frac{1}{2}W^2 - \hat{W}W = K(T) - T\bar{x},$$

where $\hat{W}$ is as in (2.2.15). The inversion integral (2.2.12) is now expressed as

$$Q_n(\bar{x}) = \frac{1}{2\pi i} \int_{c-i\infty}^{c+i\infty} e^{n[\frac{1}{2}W^2 - \hat{W}W]} \left(\frac{1}{T}\frac{dT}{dW}\right) dW \quad (c > 0). \qquad (2.2.24)$$

Equation (2.2.24) may be rearranged as

$$\begin{aligned}
Q_n(\bar{x}) &= \frac{1}{2\pi i} \int_{c-i\infty}^{c+i\infty} e^{n[\frac{1}{2}W^2 - \hat{W}W]} \frac{dW}{W} \\
&\quad + \frac{1}{2\pi i} e^{-\frac{1}{2}n\hat{W}^2} \int_{c-i\infty}^{c+i\infty} e^{\frac{1}{2}n(W-\hat{W})^2} \left(\frac{1}{T}\frac{dT}{dW} - \frac{1}{W}\right) dW.
\end{aligned} \qquad (2.2.25)$$

The first integral in (2.2.25) has a singularity at $W = 0$ (or $\bar{x} = \mu$), and has value $1 - \Phi(\hat{W}n^{\frac{1}{2}})$. To see that the integrand of the second integral has no singularity at or near μ, we examine its asymptotic behavior as $\bar{x}$ approaches μ, or equivalently, as T approaches 0. When $|T|$ is small, $W \sim AT$ where

$$A = \frac{K'(\hat{T}) - K'(0)}{\hat{W}} = \frac{\bar{x} - \mu}{\hat{W}}$$

when $T \neq 0$, and $A = \{K''(0)\}^{-1/2}$ when $T = 0$. Hence $T^{-1}dT/dW \sim W^{-1}$. Since dT/dW is analytic in a neighborhood of $W = 0$, so is $T^{-1}dT/dW - W^{-1}$. It follows that the path of integration for the second integral may be set to $W = \hat{W} + iy$ where $\hat{W}$ is the saddlepoint, even when $\hat{W} = 0$. Watson's lemma is then applied to expand this integral, and the resulting expansion is uniformly valid across μ. Detailed calculations lead to Lugannani and Rice's expansion for the tail probability,

$$Q_n(\bar{x}) = 1 - \Phi(\hat{W}n^{\frac{1}{2}}) + \phi(\hat{W}n^{\frac{1}{2}}) \left\{ \frac{b_0}{n^{\frac{1}{2}}} + \frac{b_1}{n^{\frac{3}{2}}} + \cdots + \frac{b_k}{n^{k+\frac{1}{2}}} + O(n^{-k-\frac{3}{2}}) \right\}, \qquad (2.2.26)$$

where, by letting $\hat{U} = \hat{T}(\hat{K}'')^{\frac{1}{2}}$,

$$b_0 = \frac{1}{\hat{U}} - \frac{1}{\hat{W}}, \quad \text{and} \quad b_1 = \frac{1}{\hat{U}}(\frac{1}{8}\lambda_4 - \frac{5}{24}\lambda_3^2) - \frac{\lambda_3}{2\hat{U}^2} - \frac{1}{\hat{U}^3} + \frac{1}{\hat{W}^3}. \qquad (2.2.27)$$

Formulas for further coefficients may be found in Lugannani and Rice (1980).

When $\hat{U}$, and hence $\hat{W}$, are small, the individual terms making up b_0 and b_1 are seen to be large. However, by expanding $\hat{W}$ in powers of $\hat{U}$ as

$$\hat{W} = \hat{U} - \frac{1}{6}\lambda_3\hat{U}^2 + \frac{1}{24}(\lambda_4 - \frac{1}{3}\lambda_3^2)\hat{U}^3 - \frac{1}{24}(\frac{1}{5}\lambda_5 - \frac{1}{6}\lambda_3\lambda_4 + \frac{1}{18}\lambda_3^2)\hat{U}^4 + \cdots,$$

they reduce to

$$b_0 = -\frac{1}{6}\lambda_3 + \frac{1}{24}(\lambda_4 - \lambda_3^2)\hat{U} + O(\hat{U}^2), \quad b_1 = \frac{1}{40}\lambda_5 - \frac{5}{48}\lambda_3\lambda_4 + \frac{35}{432}\lambda_3^2 + O(\hat{U}). \qquad (2.2.28)$$

Every b_r can be expanded in this way as a series in powers of $\hat{U}$ whose leading term is $O(1)$. Thus the b_r's remain finite as $\hat{U}$ crosses the origin. In practice, to approximate the tail probability one usually uses the leading term (b_0) of the expansion and writes this approximation as follows:

$$Q_n(\bar{x}) = 1 - \Phi(\hat{\zeta}) + \phi(\hat{\zeta})\left\{\frac{1}{\hat{z}} - \frac{1}{\hat{\zeta}} + O(n^{-\frac{3}{2}})\right\}, \qquad (2.2.29)$$

where $\hat{z} = \hat{T}(n\hat{K}'')^{\frac{1}{2}}$, and $\hat{\zeta} = \{2n(\hat{T}\bar{x} - \hat{K})\}^{\frac{1}{2}}sgn(\hat{T})$. At $\bar{x} = \mu$, $\hat{T} = 0$, $\hat{\zeta} = 0$ and $\hat{U} = 0$. By (2.2.28), $b_0 = \lambda_3(0)$. Thus (2.2.29) reduces to

$$Q_n(\mu) = \frac{1}{2} - \frac{1}{6}\frac{\lambda_3(0)}{(2\pi n)^{\frac{1}{2}}} + O(n^{-\frac{3}{2}}). \qquad (2.2.30)$$

The expansion for the tail probability (2.2.26) leads to the following expansion for the cumulative distribution function $F_n(\bar{x})$

$$F_n(\bar{x}) = \Phi(\hat{W}n^{\frac{1}{2}}) - \phi(\hat{W}n^{\frac{1}{2}})\left\{\frac{b_0}{n^{\frac{1}{2}}} + \frac{b_1}{n^{\frac{3}{2}}} + \cdots + \frac{b_k}{n^{k+\frac{1}{2}}} + O(n^{-k-\frac{3}{2}})\right\}. \qquad (2.2.31)$$

We shall refer to this expansion as Lugannani and Rice's expansion for $F_n(\bar{x})$.

Note that the uniform validity of (2.2.26) and (2.2.31) means that the $O(n^{-k-\frac{3}{2}})$ terms in these equations are independent of $\bar{x}$. Discussions on uniform validity of expansions for contour integrals may be found in Wasow (1965) and Bleistein and Handelsman (1975). For

a definition of this uniform validity, see Appendix A.

- *The orders of the relative errors of the expansions*

The absolute error of the saddlepoint approximation (2.2.16) is $g_n(\bar{x})O(1/n)$, and that of the Lugannani and Rice's approximation (2.2.29) is $\phi(\hat{\zeta})O(n^{-3/2})$ or $\phi(\hat{W}n^{1/2})O(n^{-3/2})$. Since both the density and the tail probability tend to be small at the far tail, these absolute errors do not tell us much about the accuracy of the approximations. It is thus important to examine their relative errors.

The relative error of the saddlepoint approximation is given by

$$\left| \frac{f_n(\bar{x}) - g_n(\bar{x})}{f_n(\bar{x})} \right| = \left| \frac{g_n(\bar{x})}{f_n(\bar{x})} - 1 \right| = O(1/n).$$

To examine the relative error of Lugannani and Rice's approximation (2.2.29), and in general that of (2.2.31), we first calculate the asymptotic order of Mill's ratio, $R(\hat{W}n^{1/2})$. Using the following identity from Kendall and Stuart (1969, Vol. 1, p. 137)

$$R(x) = \frac{|x|}{2\sqrt{\pi}} \int_0^\infty \frac{e^{-t}t^{-1/2}}{t + x^2/2} dt,$$

it is easy to see that $R(\hat{W}n^{1/2}) = O(1/n^{1/2})$. It follows that

$$
\begin{aligned}
\frac{Q_n(\bar{x})}{1 - \Phi(\hat{W}n^{1/2}) + \phi(\hat{W}n^{1/2})b_0/n^{1/2}} &= 1 - \frac{\phi(\hat{W}n^{1/2})O(1/n^{3/2})}{1 - \Phi(\hat{W}n^{1/2}) + \phi(\hat{W}n^{1/2})b_0/n^{1/2}} \\
&= 1 - \frac{O(1/n^{3/2})}{R(\hat{W}n^{1/2}) + b_0/n^{1/2}} \\
&= 1 - O(1/n).
\end{aligned}
\tag{2.2.32}
$$

It is not difficult to show using (2.2.32) that the relative error of Lugannani and Rice's approximation (2.2.29) is also $O(1/n)$. Furthermore, it is uniformly $O(1/n)$ since all the $O(\cdot)$'s in (2.2.32) are uniformly valid. The absolute and relative errors of other truncated versions of the two expansions may be discussed analogously.

- *Renormalization*

It is necessary to check whether an approximation for a density function is (1) non-negative, and (2) integrates to 1. The saddlepoint approximation $g_n(\bar{x})$ satisfies the first

condition, but does not in general integrate to 1. This problem is usually solved by multiplying $g_n(\bar{x})$ by a constant c, which is the reciprocal of the number that $g_n(\bar{x})$ integrates to. This strategy is known as *renormalization*, and $cg_n(\bar{x})$ is referred to as the *renormalized* saddlepoint approximation. Renormalization is usually carried out numerically. It can be time consuming since at each $\bar{x}$ the saddlepoint has to be computed using Newton's method. To reduce the amount of computation, Daniels (1987) suggested using $\hat{T}$ as the variable of integration. This can be achieved by a simple change of variable and eliminates the need to compute saddlepoints. The renormalized saddlepoint approximation is often more accurate than the original saddlepoint approximation. Another benefit of renormalization is that one may store the values of $g_n(\bar{x})$ used for renormalization, and use them for computing an integrated saddlepoint approximation to approximate the distribution function.

It is also important to know whether Lugannani and Rice's approximation (2.2.29) is non-negative, non-decreasing and approaches 1/0 as $\bar{x}$ approaches the lower/upper end of its domain. Due to the complexity of the formula, these conditions are usually difficult to verify. We have not encountered any numerical example where they are not met. We shall further discuss these conditions, in particular, the last condition, in the next chapter.

Chapter 3

The Relationship Between the Two Expansions

3.1 Introduction

A natural alternative to Lugannani and Rice's approximation for the tail probability is that of numerically integrated saddlepoint approximation. Which one of the two approximations is more accurate? In their 1980 paper, Lugannani and Rice compared the two approximations through a numerical example and wrote, "this example suggests the conjecture that the integration of Daniels' series and our series for $Q_n(\bar{x})$ both give approximations to $Q_n(\bar{x})$ that are in error by the same order of magnitude". This conjecture was subsequently supported by numerical evidence provided by various authors, e.g., Daniels (1983), Field and Ronchetti (1990). Nevertheless, from a theoretical point of view the conjecture remains largely unanswered. The fact that the error terms for both approximations do not have simple expressions, and no simple and practical bounds are available made this conjecture difficult to pursue. See Lugannani and Rice (1980) for further discussion.

Our interest in the mathematical relationship between the two expansions was raised by the possibility that such a relationship, should it exist, may hold the key to the conjecture. More specifically, if a relationship between the coefficients of the two expansions can be established, then we may be able to express the integrated saddlepoint approximation in terms of the coefficients of Lugannani and Rice's expansion. A direct comparison of the two approximations may then be possible. There is reason to believe some kind relationship indeed exists since they are both obtained by expanding the inversion integrals at the saddlepoint. This unknown relationship itself became an interesting question, and subsequently the focus of our investigation. Since the cumulative distribution function may be differentiated to obtain the corresponding density function, we asked whether the relationship between their asymptotic expansions is similar.

In this chapter, we investigate this relationship by focusing on the derivatives of the entire and truncated Lugannani and Rice series for $F_n(\bar{x})$. Our investigation starts with the derivation of a formal relationship which suggests that the derivative of Lugannani and Rice's series is Daniels' series. This formal relationship is then rigorously established by using asymptotic arguments. Applications of this result are then discussed. The following is an outline for this chapter. In Section 2 we formally differentiate Lugannani and Rice's series for $F_n(\bar{x})$ and show that the resulting series resembles Daniels' series for $f_n(\bar{x})$. In Section 3 we prove, under a uniform validity assumption concerning Daniels' series, that we can indeed differentiate Lugannani and Rice's series for $F_n(\bar{x})$ to obtain Daniels' series. This result is the main result of this chapter. In Section 4 we show that the uniform validity condition required to prove the relationship is in general true. We then apply the main result in Section 5 to study the relationship between the truncated versions of the two series, which establishes the derivative of a truncated Lugannani and Rice series as an alternative asymptotic approximation to the density function and provides an answer to Lugannani and Rice's conjecture. We also derive a saddlepoints approximation for the standardized mean, a result related to the uniform validity of the saddlepoint expansion for the mean.

Throughout this chapter, we assume that X_i has a continuous density function $f(x)$ whose domain is an interval on the real line. As the existence of the two series depends on the existence of the saddlepoint, we shall only be concerned with those $\bar{x}$ values that

have saddlepoints. These $\bar{x}$ values form an interval. For this chapter, we shall refer to this interval as the domain of $\bar{x}$.

3.2 Formal differentiation

The derivative of $\hat{T}$ with respect to $\bar{x}$ and the derivative of $\hat{W}$ with respect to $\hat{T}$ are used repeatedly in the differentiation. It is not difficult to show that they are

$$\frac{d\hat{T}}{d\bar{x}} = \frac{1}{K''(\hat{T})}, \quad \text{and} \quad \frac{d\hat{W}}{d\hat{T}} = \frac{\hat{T}K''(\hat{T})}{\hat{W}}.$$

We now formally differentiate Lugannani and Rice's series for $F_n(\bar{x})$ (2.2.31).

$$
\begin{aligned}
f_n(\bar{x}) &= \frac{dF_n(\bar{x})}{d\bar{x}} \\
&\sim \frac{d\hat{T}}{d\bar{x}} \frac{d}{d\hat{T}} \left[\Phi(\hat{W}n^{\frac{1}{2}}) - \phi(\hat{W}n^{\frac{1}{2}}) \sum_{r=0}^{\infty} \frac{b_r}{n^{r+\frac{1}{2}}} \right] \\
&= \frac{\phi(\hat{W}n^{\frac{1}{2}})}{K''(\hat{T})} \left[\frac{d(\hat{W}n^{\frac{1}{2}})}{d\hat{T}} + \hat{W}n^{\frac{1}{2}} \frac{d(\hat{W}n^{\frac{1}{2}})}{d\hat{T}} \sum_{r=0}^{\infty} \frac{b_r}{n^{r+\frac{1}{2}}} - \sum_{r=0}^{\infty} \frac{db_r}{d\hat{T}} \frac{1}{n^{r+\frac{1}{2}}} \right] \\
&= g_n(\bar{x}) \left[\frac{\hat{T}(\hat{K}'')^{\frac{1}{2}}}{\hat{W}} + \hat{T}(\hat{K}'')^{\frac{1}{2}} \sum_{r=0}^{\infty} \frac{b_r}{n^r} - \sum_{r=0}^{\infty} (\hat{K}'')^{-\frac{1}{2}} \frac{db_r}{d\hat{T}} \frac{1}{n^{r+1}} \right] \\
&= g_n(\bar{x}) \left\{ \frac{\hat{T}(\hat{K}'')^{\frac{1}{2}}}{\hat{W}} + \hat{T}(\hat{K}'')^{\frac{1}{2}}b_0 + \sum_{r=1}^{\infty} \frac{1}{n^r} \left[\hat{T}(\hat{K}'')^{\frac{1}{2}}b_r - (\hat{K}'')^{-\frac{1}{2}} \frac{db_{r-1}}{d\hat{T}} \right] \right\} \quad (3.2.1)
\end{aligned}
$$

Recalling from (2.2.27) that

$$b_0 = \frac{1}{\hat{T}(\hat{K}'')^{\frac{1}{2}}} - \frac{1}{\hat{W}}, \tag{3.2.2}$$

it is easy to see that

$$\frac{\hat{T}(\hat{K}'')^{\frac{1}{2}}}{\hat{W}} + \hat{T}(\hat{K}'')^{\frac{1}{2}}b_0 = 1. \tag{3.2.3}$$

Thus (3.2.1) may be written as

$$f_n(\bar{x}) \sim g_n(\bar{x}) \left\{ 1 + \sum_{r=1}^{\infty} \frac{1}{n^r} \left[\hat{T}(\hat{K}'')^{\frac{1}{2}}b_r - (\hat{K}'')^{-\frac{1}{2}} \frac{db_{r-1}}{d\hat{T}} \right] \right\}. \tag{3.2.4}$$

With $b_{-1} = -\hat{W}$, (3.2.4) may be written in a more compact form

$$f_n(\bar{x}) \sim g_n(\bar{x}) \sum_{r=0}^{\infty} \frac{1}{n^r} \left[\hat{T}(\hat{K}'')^{\frac{1}{2}} b_r - (\hat{K}'')^{-\frac{1}{2}} \frac{db_{r-1}}{d\hat{T}} \right]$$

$$= g_n(\bar{x}) \sum_{r=0}^{\infty} \frac{c_r}{n^r}, \qquad (3.2.5)$$

where

$$c_r = \hat{T}(\hat{K}'')^{\frac{1}{2}} b_r - (\hat{K}'')^{-\frac{1}{2}} \frac{db_{r-1}}{d\hat{T}}. \qquad (3.2.6)$$

The formal relation (3.2.5) suggests that the asymptotic series in the right-hand side of (3.2.5) is an asymptotic expansion of the density function $f_n(\bar{x})$. Furthermore, since Daniels' series (2.2.23) is an asymptotic expansion of $f_n(\bar{x})$ and the asymptotic series in the right-hand side of (3.2.5) is in the form of Daniels' series, it suggests that the two series should coincide with one another, that is $c_r = a_r$ ($r = 0, 1, \ldots$). From (3.2.4) and the argument below we see that $c_0 = 1$ and $c_1 = \frac{1}{8}\lambda_4 - \frac{5}{24}\lambda_3^2$, which indeed match a_0 and a_1, respectively. If this formal relation can be rigorously established, then by the uniqueness of the asymptotic expansion with respect to the asymptotic sequence $\{1/n^r\}$, the entire series (3.2.5) coincides with Daniels' series (2.2.23), and thus $c_r = a_r$ ($r = 0, 1, \ldots$).

We conclude this section by showing that $a_1 = c_1$, i.e.,

$$a_1 = \hat{T}(\hat{K}'')^{\frac{1}{2}} b_1 - (\hat{K}'')^{-\frac{1}{2}} \frac{db_0}{d\hat{T}}, \qquad (3.2.7)$$

which is also a useful identity for the proof of Theorem 3.1 in the next section. Using the expressions for b_0 and b_1 shown in (2.2.27), we can show that

$$(\hat{K}'')^{-\frac{1}{2}} \frac{db_0}{d\hat{T}} = -\frac{\hat{K}^{(3)}}{2\hat{T}(\hat{K}'')^2} - \frac{1}{T^2 \hat{K}''} + \frac{\hat{T}(\hat{K}'')^{\frac{1}{2}}}{\hat{W}^3}, \qquad (3.2.8)$$

and that

$$\hat{T}(\hat{K}'')^{\frac{1}{2}} b_1 = \left(\frac{1}{8}\lambda_4 - \frac{5}{24}\lambda_3^2 \right) - \frac{\hat{K}^{(3)}}{2\hat{T}(\hat{K}'')^2} - \frac{1}{T^2 \hat{K}''} + \frac{\hat{T}(\hat{K}'')^{\frac{1}{2}}}{\hat{W}^3}. \qquad (3.2.9)$$

Equation (3.2.7) then follows from (3.2.8) and (3.2.9).

3.3 The main theorem

When a function of two variables $f(s,t)$ has an asymptotic series expansion in variable s, it is not always true that formally differentiating this series with respect to t will result in an asymptotic series for the partial derivative $f_t(s,t)$. Wasow (1965, pp. 43-48) discussed conditions under which this is true. But these conditions are in general not satisfied here. In this section, we state and prove the main result of this chapter, i.e., the following theorem, which presents a sufficient condition under which the formal relation (3.2.5) derived by differentiating Lugannani and Rice's series for $F_n(\bar{x})$ is valid.

Theorem 3.1. *Let $D_{\bar{x}}$ be a bounded closed interval in the domain of $\bar{x}$. If Daniels' series (2.2.23) is uniformly valid in $D_{\bar{x}}$, then the series in (3.2.5) obtained by differentiating Lugannani and Rice's series (2.2.31) coincides with Daniels' series (2.2.23) in $D_{\bar{x}}$.*

To prove the theorem, we first establish conditions, through Lemma 3.1 and 3.2 below, under which a *power* series expansion of $f(s,t)$ in s may be differentiated with respect to t to obtain a power series expansion of $f_t(s,t)$. Lemma 3.1 gives a necessary and sufficient condition for uniform validity of an asymptotic series. Lemma 3.2 uses the condition given by Lemma 3.1 to show that when asymptotic expansions for both $f_t(s,t)$ and $f(s,t)$ with respect to the asymptotic sequence $\{s^r\}$ exist, the expansion for $f_t(s,t)$ is the derivative of that for $f(s,t)$, provided it is uniformly valid. Lemma 3.1 and its proof are taken, essentially, from Chapter 3 in *Asymptotic Expansions for Ordinary Differential Equations* by Wasow (1965). Theorem 3.1 and the lemmas in this section all involve the notion of uniform validity of asymptotic expansions for functions of two variables, which we had come across a few times toward the end of Chapter 2. For discussion concerning this uniform validity, as well as generalizations of Lemma 3.1, see Appendix A.

Lemma 3.1. (Wasow, 1965) *Let $f(s,t)$ be bounded in $D_s \times D_t$ where $0 \in D_s$, and $h_r(t)$ ($r = 0, 1, \ldots$) be bounded in D_t. Then*

$$f(s,t) \sim \sum_{r=0}^{\infty} h_r(t)s^r \quad as \quad s \to 0 \qquad (3.3.10)$$

uniformly for $t \in D_t$ iff for every m the function $E_m(s,t)$ defined by the relation

$$f(s,t) = \sum_{r=0}^{m} h_r(t)s^r + E_m(s,t)s^{m+1} \tag{3.3.11}$$

is bounded in $D_s \times D_t$.

Proof:

(a) If the expansion (3.3.10) is uniformly valid for $t \in D_t$ then there exists, for every $\varepsilon > 0$, a $\delta_m(\varepsilon) > 0$, independent of t, such that in $D_s \times D_t$

$$\left| s^{-m} \left[f(s,t) - \sum_{r=0}^{m} h_r(t)s^r \right] \right| < \varepsilon,$$

whenever $|s| \le \delta_m(\varepsilon)$. Applying the inequality with $m+1$ instead of m we find immediately that

$$\left| s^{-(m+1)} \left[f(s,t) - \sum_{r=0}^{m} h_r(t)s^r \right] \right| < \varepsilon + |h_{m+1}(t)|, \quad |s| \le \delta_m(\varepsilon),$$

which implies the boundedness of $E_m(s,t)$ for the part of $D_s \times D_t$ in which $|s| \le \delta_m(\varepsilon)$. For $|s| > \delta_m(\varepsilon)$, $(s,t) \in D_s \times D_t$, the boundedness of

$$E_m(s,t) = s^{-(m+1)} \left[f(s,t) - \sum_{r=0}^{m} h_r(t)s^r \right] \tag{3.3.12}$$

follows from the boundedness of $f(s,t)$ and of the $h_r(t)$'s.

(b) If $E_m(s,t)$, as defined by formula (3.3.12), is bounded we have

$$s^{-m} \left[f(s,t) - \sum_{r=0}^{m} h_r(t)s^r \right] = E_m(s,t)s,$$

which tends uniformly in D_t to zero, as $s \to 0$ in D_s. #

Lemma 3.2. *Assume $\frac{\partial f(s,t)}{\partial t}$ is integrable in t and bounded in $D_s \times D_t$, where $0 \in D_s$ and D_t is a bounded closed interval on the real line. If*

$$f(s,t) \sim \sum_{r=0}^{\infty} h_r(t)s^r \quad as \quad s \to 0, \tag{3.3.13}$$

and

$$\frac{\partial f(s,t)}{\partial t} \sim \sum_{r=0}^{\infty} l_r(t)s^r \quad as \quad s \to 0, \tag{3.3.14}$$

where (3.3.14) is uniformly valid in D_t and $l_r(t)$'s are continuous functions of t, then the $h_r(t)$'s are differentiable in D_t and

$$l_r(t) = \frac{dh_r(t)}{dt} \quad for \quad r = 0, 1, \ldots$$

Proof:

Since $\frac{\partial f(s,t)}{\partial t}$ and the $l_r(t)$'s are bounded, (3.3.14) and Lemma 3.1 imply that for every m the function $E_m(s,t)$ defined by the relation

$$\frac{\partial f(s,t)}{\partial t} = \sum_{r=0}^{m} l_r(t)s^r + E_m(s,t)s^{m+1}$$

is bounded in $D_s \times D_t$. For t and t' in D_t

$$\int_t^{t'} \frac{\partial f(s,v)}{\partial v} dv = \sum_{r=0}^{m} \left[\int_t^{t'} l_r(v)dv \right] s^r + \left[\int_t^{t'} E_m(s,v)dv \right] s^{m+1}.$$

Since $\int_t^{t'} l_r(v)dv$ $(r = 0, 1, \ldots, m)$ are bounded in D_t, and $\int_t^{t'} \frac{\partial f(s,v)}{\partial v} dv$ and $\int_t^{t'} E_m(s,v)dv$ $(m = 0, 1, \ldots)$ are bounded in $D_s \times D_t$, by Lemma 3.1

$$\int_t^{t'} \frac{\partial f(s,v)}{\partial v} dv \sim \sum_{r=0}^{\infty} \left[\int_t^{t'} l_r(v)dv \right] s^r$$

as $s \to 0$ uniformly in D_t. That is:

$$f(s,t') - f(s,t) \sim \sum_{r=0}^{\infty} \left[\int_t^{t'} l_r(v)dv \right] s^r$$

as $s \to 0$ uniformly in D_t. On the other hand, (3.3.13) implies that

$$f(s,t') - f(s,t) \sim \sum_{r=0}^{\infty} \left[h_r(t') - h_r(t) \right] s^r$$

as $s \to 0$. By the uniqueness of asymptotic expansion with respect to a given asymptotic sequence,

$$h_r(t') - h_r(t) = \int_t^{t'} l_r(v)dv \quad for \quad r = 0, 1, \ldots \tag{3.3.15}$$

Since l_r is continuous and (3.3.15) holds for any $t' \in D_t$, (3.3.15) implies that $h'_r(t)$ exists and equals $l_r(t)$ $(r = 0, 1, \ldots)$. #

Nonetheless, since $f_n(\bar{x})$ in (2.2.23) is not bounded when n approaches infinity, and Lugannani and Rice's series (2.2.31) is not a standard power series, Lemma 3.2 cannot be

directly applied to (2.2.31) and (2.2.23) to prove the theorem. We now focus on a new function, $I(n, \bar{x})$, defined below, and show that this function and its partial derivative have asymptotic power series expansions. Lemma 3.2 is then used to establish the relationship between these two power series, which leads to the theorem.

Lemma 3.3. *Let D_n be the set of positive integers, $D_{\bar{x}}$ be a bounded closed interval in the domain of $\bar{x}$, and $I(n, \bar{x})$ be defined by the relation*

$$F_n(\bar{x}) = \Phi(\hat{W} n^{\frac{1}{2}}) - \frac{\phi(\hat{W} n^{\frac{1}{2}}) b_0}{n^{\frac{1}{2}}} - \phi(\hat{W} n^{\frac{1}{2}}) I(n, \bar{x}). \tag{3.3.16}$$

Then

$$(i) \quad I(n, \bar{x}) \sim \sum_{r=1}^{\infty} \frac{b_r}{n^{r+\frac{1}{2}}} \text{ as } n \to \infty \text{ uniformly for } \bar{x} \in D_{\bar{x}}. \tag{3.3.17}$$

Furthermore, if (2.2.23) is uniformly valid in $D_{\bar{x}}$, then

$$(ii) \quad \frac{\partial I(n, \bar{x})}{\partial \bar{x}} \text{ is continuous in } \bar{x} \text{ and bounded in } D_n \times D_{\bar{x}}, \text{ and}$$

$$(iii) \quad \frac{\partial I(n, \bar{x})}{\partial \bar{x}} \sim \sum_{r=1}^{\infty} h_r(\bar{x}) \frac{1}{n^{r+\frac{1}{2}}} \text{ uniformly with respect to } \bar{x} \in D_{\bar{x}},$$

where $h_r(\bar{x}) = \frac{db_r}{d\bar{x}}$ for $r = 1, 2, \ldots$

Proof:

(i) is readily obtained upon substituting Lugannani and Rice's series for $F_n(\bar{x})$ (2.2.31) in (3.3.16). To show (ii) is true, differentiate both sides of (3.3.16) with respect to $\bar{x}$. We obtain

$$f_n(\bar{x}) = g_n(\bar{x}) \left[1 - \frac{(\hat{K}'')^{-\frac{1}{2}} \frac{db_0}{dT}}{n} + n^{\frac{1}{2}} \hat{T} (\hat{K}'')^{\frac{1}{2}} I(n, \bar{x}) - n^{-\frac{1}{2}} (\hat{K}'')^{\frac{1}{2}} \frac{\partial I(n, \bar{x})}{\partial \bar{x}} \right]. \tag{3.3.18}$$

Define $D_1(n, \bar{x})$ and $L_1(n, \bar{x})$ by the following relations

$$\frac{f_n(\bar{x})}{g_n(\bar{x})} = 1 + \frac{a_1}{n} + \frac{D_1(n, \bar{x})}{n^2}, \tag{3.3.19}$$

$$I(n, \bar{x}) = \frac{b_1}{n^{1+\frac{1}{2}}} + \frac{L_1(n, \bar{x})}{n^{2+\frac{1}{2}}}. \tag{3.3.20}$$

For any finite n, $f_n(\bar{x})/g_n(\bar{x})$ and $I(n, \bar{x})$ are bounded in $D_{\bar{x}}$. When n approaches infinity, by the uniform validity assumption on (2.2.23) and (i), they approach the leading terms in

the right-hand side of (3.3.19) and (3.3.20) uniformly, respectively. Thus $f_n(\bar{x})/g_n(\bar{x})$ and $I(n,\bar{x})$ are bounded in $D_n \times D_{\bar{x}}$. Lemma 3.1 then implies that $D_1(n,\bar{x})$ and $L_1(n,\bar{x})$ are both bounded in $D_n \times D_{\bar{x}}$. They are also continuous in $\bar{x}$ since $f_n(\bar{x})/g_n(\bar{x})$ and $I(n,\bar{x})$, and a_1 and b_1 are all continuous in $\bar{x}$.

Substitute (3.3.19), (3.3.20) into (3.3.18), and use the identity (3.2.7), we obtain

$$\frac{\partial I(n,\bar{x})}{\partial \bar{x}} = \hat{T}\left[\frac{L_1(n,\bar{x})}{n^{1+\frac{1}{2}}}\right] - (\hat{K}'')^{-\frac{1}{2}}\left[\frac{D_1(n,\bar{x})}{n^{1+\frac{1}{2}}}\right].$$

On $D_{\bar{x}}$, $\hat{T}$ and $(\hat{K}'')^{-\frac{1}{2}}$ are both continuous and bounded. Thus it follows from the boundedness and continuity of $D_1(n,\bar{x})$ and $L_1(n,\bar{x})$, $\frac{\partial I(n,\bar{x})}{\partial \bar{x}}$ is continuous in $\bar{x}$ and bounded in $D_n \times D_{\bar{x}}$.

To show (iii), rewrite (3.3.18) as

$$\frac{\partial I(n,\bar{x})}{\partial \bar{x}} = n^{\frac{1}{2}}(\hat{K}'')^{-\frac{1}{2}}\left[-\frac{f_n(\bar{x})}{g_n(\bar{x})} + 1 - \frac{(\hat{K}'')^{-\frac{1}{2}}\frac{db_0}{d\hat{T}}}{n} + n^{\frac{1}{2}}\hat{T}(\hat{K}'')^{\frac{1}{2}}I(n,\bar{x})\right].$$

By substituting (2.2.23) for $f_n(\bar{x})$ and the asymptotic series in (i) for $I(n,\bar{x})$ into the above equation, we obtain

$$\frac{\partial I(n,\bar{x})}{\partial \bar{x}} \sim n^{\frac{1}{2}}(\hat{K}'')^{-\frac{1}{2}}\left[\sum_{r=1}^{\infty}\frac{\hat{T}(\hat{K}'')^{\frac{1}{2}}b_r}{n^r} - \sum_{r=1}^{\infty}\frac{a_r}{n^r} - \frac{(\hat{K}'')^{-\frac{1}{2}}\frac{db_0}{d\hat{T}}}{n}\right] \quad \text{as } n \to \infty.$$

By using the identity $a_1 = \hat{T}(\hat{K}'')^{\frac{1}{2}}b_1 - (\hat{K}'')^{-\frac{1}{2}}\frac{db_0}{d\hat{T}}$, and collecting terms according to the powers of n we have

$$\begin{aligned}\frac{\partial I(n,\bar{x})}{\partial \bar{x}} &\sim \sum_{r=1}^{\infty}\left[\hat{T}b_{r+1} - (\hat{K}'')^{-\frac{1}{2}}a_{r+1}\right]\frac{1}{n^{r+\frac{1}{2}}} \\ &= \sum_{r=1}^{\infty}h_r(\bar{x})\frac{1}{n^{r+\frac{1}{2}}}, \end{aligned} \tag{3.3.21}$$

where

$$h_r(\bar{x}) = \hat{T}b_{r+1} - (\hat{K}'')^{-\frac{1}{2}}a_{r+1}. \tag{3.3.22}$$

The asymptotic expansion in the right-hand side of (3.3.21) is uniformly valid due to the uniform validity assumption on Daniels' series (2.2.23) and the uniform validity of the expansion in (i). Since the h_r's are functions of a_r's, b_r's, $\hat{K}''$, and $\hat{T}$, they are continuous in

$\bar{x}$. (ii) and Lemma 2 then imply that $h_r(\bar{x}) = \frac{db_r}{d\bar{x}}$ for $r = 1, 2, \ldots$ #

Proof of Theorem 3.1:

By substituting the asymptotic expansions of $I(n, \bar{x})$ and $\frac{\partial I(n,\bar{x})}{\partial \bar{x}}$ in Lemma 3.3 into (3.3.18) and collecting terms according to the powers of n, we obtain (3.2.5). Thus (3.2.5) is indeed a uniformly valid asymptotic expansion of $f_n(\bar{x})$ for $\bar{x}$ in $D_{\bar{x}}$. By the uniqueness of asymptotic expansion with respect to the asymptotic sequence $\{1/n^r\}$, this uniformly valid asymptotic series coincides with that of Daniels. #

The above proof centers on proving the asymptotic validity of the series in the right-hand side of (3.2.5). The general relationship between the coefficients of Daniels' series and Lugannani and Rice's series, i.e., c_r in (3.2.6) equals a_r, is a consequence of this validity and the uniqueness of asymptotic expansion with respect to a given asymptotic sequence. One may also prove the theorem without explicitly using the uniqueness property by showing that (3.3.22) implies $c_r = a_r$ $(r = 0, 1, \ldots)$.

To conclude, we note that the uniform validity condition required by the theorem is not as restrictive as it appears to be. We show in the following section that this condition is rather easily satisfied.

3.4 Uniform validity of Daniels' expansion

Theory announced to date on the uniform validity of Daniels' expansion falls into two categories, i.e., results on uniform validity of the saddlepoint approximation as $\bar{x}$ approaches the end points of its domain, and results on uniform validity of the entire series in some given compact subset of the domain. Daniels (1954) studied four classes of densities for uniform validity in the tail. Jensen (1988) further studied these four classes. These two papers showed that for the four classes of continuous densities, the saddlepoint approximation is uniformly valid as $\bar{x}$ approaches the end points. The results in the latter category were presented by Barndorff-Nielsen and Cox (1979), in which they gave a condition and outlined a proof for the uniform validity of the Edgeworth expansion for a parametric family of densities in any compact subset of the parameter space. This result led to the uniform validity of Daniels' series in any compact subset in the parameter space. These results,

however, appear to be valid only for the exponential families considered in their paper. The outlined proof in their paper and that in Jensen (1988) for the uniform validity of the saddlepoint approximation both use Feller's (1966) method for verifying the uniform validity of the Edgeworth expansion. Jensen (1991) gave a set of three conditions under which Daniels' series is uniformly valid in a subset of the parameter space. Note that the parameter that we have been referring to is the saddlepoint, $\hat{T}$. Uniformity results in terms of this parameter may be translated into results in terms of $\bar{x}$ using the relation $\bar{x} = K'(\hat{T})$.

In this section, we prove two theorems concerning the uniform validity of Daniels' series on compact sets. The first theorem (Theorem 3.2) identifies two conditions under which the series is uniformly valid. This theorem is similar to the univariate version of the corollary in the appendix in Barndorff-Nielsen and Cox (1979), but it uses stronger conditions than the corollary and its validity is not restricted to a certain type of densities. The proof is also based on Feller's method for verifying the uniform validity of Edgeworth expansion, and the theorem resembles the fundamental lemma (1) in Jensen (1991). Nevertheless, we were unaware of Jensen's (1991) result at the time we proved the theorem, and our proof is also somewhat different from that outlined in Jensen (1991, 1988). There has been no detailed proof in the literature so far. We thus include our proof here for the completeness of our discussion. The second theorem (Theorem 3.3) shows that Daniels' series is, for all practical purposes, always uniformly valid.

Theorem 3.2. *Let $\bar{X}$ be the mean of n i.i.d. observations with underlying density function $f(x)$, and $D_{\bar{x}} = [x_1, x_2]$ be a bounded closed interval in the domain of $\bar{x}$. If (i) there exist $\gamma > 0$ and $\kappa > 0$ such that for $\bar{x} \in D_{\bar{x}}$,*

$$\int_{-\infty}^{\infty} |\phi_{\bar{x}}(y)|^{\gamma} dy < \kappa < \infty,$$

where

$$\phi_{\bar{x}}(y) = \frac{e^{[K(\hat{T}+iy)-(\hat{T}+iy)\bar{x}]}}{e^{[K(\hat{T})-\hat{T}\bar{x}]}}, \qquad (3.4.23)$$

and (ii) for any $y_0 > 0$, there exists a $\rho_0 \in (0,1)$ such that for all $\bar{x} \in D_{\bar{x}}$, $|\phi_{\bar{x}}(y)| < \rho_0$ when $|y| > y_0$, then Daniels' expansion (2.2.23) is uniformly valid in $D_{\bar{x}}$.

Function $\phi_{\bar{x}}(y)$ is the characteristic function corresponding to the conjugate density

$$h_{\bar{x}}(u) = c(\bar{x})e^{\hat{T}u}f(u+\bar{x}),$$

where $c(\bar{x}) = e^{-K(\hat{T})+\hat{T}\bar{x}}$. It arises naturally in Daniels' second derivation of the saddlepoint expansion (2.2.23) (Section 4 in Daniels, 1954) which establishes the asymptotic property of the expansion through that of the Edgeworth expansion. Using the relationship $K'(\hat{T}) = \bar{x}$, one can show that the mean and variance of a random variable, $U_{\bar{x}}$, whose density is $h_{\bar{x}}(u)$, are zero and $K''(\hat{T})$, respectively. To prove Theorem 3.2, we need the following lemma.

Lemma 3.4. *For any given r, the rth absolute moment of the random variable $U_{\bar{x}}$, $M_r(\bar{x}) = E(|U_{\bar{x}}|^r)$, is bounded in $D_{\bar{x}}$.*

Proof:

Let (l, u) be the largest open interval in which the moment generating function $M(t)$ exists. At a saddlepoint $\hat{T}$, $K''(\hat{T}) > 0$ (Daniels, 1954). Thus $\bar{x} = K'(\hat{T})$ is a strictly increasing function of $\hat{T}$. Since $K'(\hat{T})$ is also continuous, it has a continuous and strictly increasing inverse function. Hence, the set of saddlepoints that correspond to the $\bar{x}$'s in $[x_1, x_2]$ form an interval $[\hat{T}_1, \hat{T}_2] \subset (l, u)$, where $\hat{T}_1 = \hat{T}(x_1)$, and $\hat{T}_2 = \hat{T}(x_2)$. For each $r \in \{0, 1, \dots\}$

$$
\begin{aligned}
M_r(\bar{x}) &= \int_{-\infty}^{\infty} |u|^r c(\bar{x}) e^{\hat{T}u} f(u + \bar{x}) du \\
&= c(\bar{x}) e^{-\hat{T}\bar{x}} \int_{-\infty}^{\infty} |y - \bar{x}|^r e^{\hat{T}y} f(y) dy.
\end{aligned}
$$

Choose $\delta > 0$ such that $\hat{T}_1 - \delta$ and $\hat{T}_2 + \delta$ are in (l, u), and choose $C > 0$ such that for all $\bar{x}$ in $[x_1, x_2]$, $|y - \bar{x}|^r e^{\hat{T}y} < e^{(\hat{T}_2+\delta)y}$ when $y > C$, and $|y - \bar{x}|^r e^{\hat{T}y} < e^{(\hat{T}_1-\delta)y}$ when $y < -C$. Then

$$
\begin{aligned}
\int_{-\infty}^{\infty} |y - \bar{x}|^r e^{\hat{T}y} f(y) dy &\leq \int_{-C}^{C} |y - \bar{x}|^r e^{\hat{T}y} f(y) dy \\
&+ \int_{-\infty}^{-C} e^{(\hat{T}_1-\delta)y} f(y) dy + \int_{C}^{\infty} e^{(\hat{T}_2+\delta)y} f(y) dy. \quad (3.4.24)
\end{aligned}
$$

Note that the last two terms in (3.4.24) are bounded by $M(\hat{T}_1 - \delta)$ and $M(\hat{T}_2 + \delta)$ respectively, which are both finite constants since $\hat{T}_1 - \delta$ and $\hat{T}_2 + \delta$ are in (l, u). Furthermore, $c(\bar{x}) e^{-\hat{T}\bar{x}}$ and $|y - \bar{x}|^r e^{\hat{T}y}$ are continuous and bounded in $\{(\bar{x}, y) : \bar{x} \in [x_1, x_2], \text{ and } |y| < C\}$. Thus $M_r(\bar{x})$ is bounded in $D_{\bar{x}}$ for any given r. #

The above proof may be easily modified to show that $M_r(\bar{x})$ is actually continuous in $D_{\bar{x}}$. We also need the following inequality concerning the Taylor expansion of a characteristic function. Assuming that $M_{(r+1)}$ for some distribution with characteristic function $\varphi(t)$ is

finite, then the following inequality is valid

$$\left| \varphi(t) - \varphi(0) - \frac{t}{1!}\varphi'(0) - \cdots - \frac{t^r}{r!}\varphi^{(r)}(0) \right| < M_{(r+1)} \frac{|t|^{(r+1)}}{(r+1)!}. \qquad (3.4.25)$$

See Feller (1966, p. 487) for a proof.

Proof of Theorem 3.2:

The density function $f_n(\bar{x})$ may be expressed in terms of $\phi_{\bar{x}}$ as (Section 4 in Daniels, 1954)

$$\begin{aligned} f_n(\bar{x}) &= e^{n[K(\hat{T}) - \hat{T}\bar{x}]} \frac{n}{2\pi} \int_{-\infty}^{\infty} \phi_{\bar{x}}^n(y) dy \\ &= g_n(\bar{x}) \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \phi_{\bar{x}}^n \left(\frac{t}{\sigma(\bar{x})\sqrt{n}} \right) dt, \end{aligned} \qquad (3.4.26)$$

where $\sigma(\bar{x}) = \sqrt{K''(\hat{T})}$. The basic idea is to write the integral in (3.4.26) as the sum of two integrals, and then to show that one of these two integrals has an asymptotic approximation in powers of $1/n$ that is uniformly valid with respect to all $\bar{x} \in D_{\bar{x}}$ while the other integral converges to zero uniformly faster than any power of $1/n$ as n goes to infinity. The asymptotic approximation multiplied by $g_n(\bar{x})$ then forms a uniformly valid expansion for the density $f_n(\bar{x})$. This approximation must coincide with Daniels' expansion due to the uniqueness of asymptotic expansions with respect to a given asymptotic sequence. We can thus conclude that Daniels' expansion is uniformly valid.

Define function $\psi_{\bar{x}}(t)$ by the following relation

$$\psi_{\bar{x}}(t) = \log \phi_{\bar{x}}(t) + \frac{1}{2}\sigma^2(\bar{x})t^2. \qquad (3.4.27)$$

To obtain the Taylor approximation for $\psi_{\bar{x}}(t)$ up to and including the term of degree r ($r \geq 3$) near $t = 0$, we note that, by (3.4.25) and the fact that $\phi'_{\bar{x}}(0) = E(U_{\bar{x}}) = 0$, $\phi_{\bar{x}}(t)$ has Taylor approximation

$$\phi_{\bar{x}}(t) = 1 - \frac{1}{2}\sigma^2(\bar{x})t^2 + \frac{\phi_{\bar{x}}^{(3)}(0)}{3!}t^3 + \cdots + \frac{\phi_{\bar{x}}^{(r)}(0)}{r!}t^r + O(|t|^{(r+1)}),$$

where coefficients for terms of even powers are real. By Lemma 3.4 the absolute moments of the family of distributions indexed by $\bar{x}$, $U_{\bar{x}}$, where $\bar{x} \in D_{\bar{x}}$ are bounded. Thus (3.4.25) implies that the $O(|t|^{(r+1)})$ term is uniformly valid for all $\phi_{\bar{x}}(t)$. For $|z| < 1$,

$$\log(1 + z) = z - \frac{z^2}{2} + \frac{z^3}{3} - \cdots + (-1)^{(p-1)}\frac{z^p}{p} + O(|z|^{(p+1)}). \qquad (3.4.28)$$

Let

$$z(\bar{x}) = -\frac{1}{2}\sigma^2(\bar{x})t^2 + \frac{\phi_{\bar{x}}^{(3)}(0)}{3!}t^3 + \cdots + \frac{\phi_{\bar{x}}^{(r)}(0)}{r!}t^r + O(|t|^{(r+1)}). \tag{3.4.29}$$

Since $\left|\phi_{\bar{x}}^{(k)}(0)\right| \le M_k(\bar{x})$, by Lemma 3.4, there exists a constant $C < \infty$ such that $\left|\phi_{\bar{x}}^{(k)}(0)\right| < C$ for $k = 1, 2, \ldots, r$. Thus $z(\bar{x}) < 1$ for all $\bar{x}$ if $|t|$ is small. Equations $(3.4.27), (3.4.28)$ and $(3.4.29)$ can then be used to obtain a power expansion for $\psi_{\bar{x}}(t)$ that is valid near $t = 0$ for all $\bar{x}$. This expansion has only terms with power 3 or higher, and coefficients for terms of even powers are real. More specifically, let p in $(3.4.28)$ be sufficiently large, say $p > r$. Then there exists a polynomial of degree $r - 2$, $\psi_{\bar{x}}^{(r)}(t)$, such that $\psi_{\bar{x}}^{(r)}(0) = 0$ and

$$\psi_{\bar{x}}(t) = \log \phi_{\bar{x}}(t) + \frac{1}{2}\sigma^2(\bar{x})t^2 = t^2\psi_{\bar{x}}^{(r)}(t) + O(|t|^{(r+1)}) \quad \text{as } t \to 0, \tag{3.4.30}$$

where the term $O(|t|^{(r+1)})$ is uniform with respect to $\bar{x}$. Note that only the first term, z, in the right-hand side of $(3.4.28)$ will contribute a t^3 to $\psi_{\bar{x}}(t)$, and, from $(3.4.29)$, the coefficient of this term is $\phi_{\bar{x}}^{(3)}(0)/6$. Consequently, the coefficient of t in $\psi_{\bar{x}}^{(r)}(t)$ is also $\phi_{\bar{x}}^{(3)}(0)/6$. We shall use this fact later on in this proof. We now put

$$p_{\bar{x}}(t) = \sum_{k=1}^{r-2} \frac{1}{k!}\left[\left(\frac{t}{\sigma(\bar{x})}\right)^2 \psi_{\bar{x}}^{(r)}\left(\frac{t}{\sigma(\bar{x})\sqrt{n}}\right)\right]^k.$$

Then $p_{\bar{x}}(t)$ is a polynomial in t whose coefficients depend on n and the first r moments of $U_{\bar{x}}$. The inversion formula $(3.4.26)$ can now be expressed as

$$\begin{aligned}
f_n(\bar{x}) &= g_n(\bar{x})\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{-\frac{1}{2}t^2}\exp\left\{n\psi_{\bar{x}}\left(\frac{t}{\sigma(\bar{x})\sqrt{n}}\right)\right\}dt \\
&= g_n(\bar{x})\frac{1}{\sqrt{2\pi}}\left\{\int_{-\infty}^{\infty} e^{-\frac{1}{2}t^2}[1 + p_{\bar{x}}(t)]\,dt \right. \\
&\quad \left. + \int_{-\infty}^{\infty} e^{-\frac{1}{2}t^2}\left[\exp\left\{n\psi_{\bar{x}}\left(\frac{t}{\sigma(\bar{x})\sqrt{n}}\right)\right\} - 1 - p_{\bar{x}}(t)\right]dt\right\}. \tag{3.4.31}
\end{aligned}$$

For $p_{\bar{x}}(t)$, coefficients for terms with odd powers have $1/\sqrt{n}$ to some odd powers as factors, and coefficients for terms with even powers are real and have $1/\sqrt{n}$ to some even powers as factors. If m is an odd integer, then $t^m \exp\{-\frac{1}{2}t^2\}$ is an odd function and integrates to zero over $(-\infty, \infty)$. Thus the result of the first integral in $(3.4.31)$ may be written as a polynomial in $1/n$ whose coefficients are all real and can be expressed in terms of the first r moments of $U_{\bar{x}}$. The order of this polynomial in $1/n$ is $(r - 2)^2/2$ if r is an even number, and is the biggest integer smaller than $(r - 2)^2/2$ if r is an odd number.

Note that since $r \geq 3$, $(r-2)^2/2 \geq r/2 - 1$. If the second integral is $o(1/n^{r/2-1})$, then the sum of those terms of the polynomial with degree less than or equal to $r/2 - 1$ may be used as an asymptotic approximation for the integral in (3.4.26). We now show that the second integral is indeed $o(1/n^{r/2-1})$ uniformly in $D_{\bar{x}}$. Specifically, we show that there is a suitably small $\delta > 0$ such that the contribution to the integral from the interval $|t| \leq \delta\sigma(\bar{x})\sqrt{n}$ is uniformly $o(1/n^{r/2-1})$, and that from $|t| > \delta\sigma(\bar{x})\sqrt{n}$ is uniformly $O(\rho^n)$ for some $\rho < 1$.

To evaluate the contribution from the interval, $|t| \leq \delta\sigma(\bar{x})\sqrt{n}$, we use the inequality,

$$\left| e^\alpha - 1 - \sum_1^{r-2} \frac{\beta^k}{k!} \right| \leq \left| e^\alpha - e^\beta \right| + \left| e^\beta - 1 - \sum_1^{r-2} \frac{\beta^k}{k!} \right|$$

$$\leq e^\gamma \left\{ |\alpha - \beta| + \frac{1}{(r-1)!} |\beta|^{r-1} \right\}, \tag{3.4.32}$$

where $\gamma > \max\{|\alpha|, |\beta|\}$. In the present discussion, α and β are

$$\alpha = n\psi_{\bar{x}}\left(\frac{t}{\sigma(\bar{x})\sqrt{n}} \right) \quad \text{and} \quad \beta = n \left(\frac{t}{\sigma(\bar{x})\sqrt{n}} \right)^2 \psi_{\bar{x}}^{(r)}\left(\frac{t}{\sigma(\bar{x})\sqrt{n}} \right),$$

respectively. Since $\sigma(\bar{x})$ is positive and bounded away from zero for $\bar{x} \in D_{\bar{x}}$, for any $\epsilon > 0$, by (3.4.30), there exists $\delta > 0$ such that for $|t| < \delta$,

$$\left| \psi_{\bar{x}}(t) - t^2 \psi_{\bar{x}}^{(r)}(t) \right| < \epsilon \sigma^r(\bar{x}) |t|^r, \tag{3.4.33}$$

uniformly in $\bar{x}$. Thus for $|t| \leq \delta\sigma(\bar{x})\sqrt{n}$ or $|t|/(\sigma(\bar{x})\sqrt{n}) < \delta$, we have $|\alpha - \beta| < \epsilon|t|^r/n^{r/2-1}$ uniformly. The coefficient of t in $\psi_{\bar{x}}^{(r)}(t)$ being $\phi_{\bar{x}}^{(3)}(0)/6$ and bounded, we can suppose that for $|t| < \delta$,

$$\left| \psi_{\bar{x}}^{(r)}(t) \right| < a|t| < \frac{1}{4}\sigma^2(\bar{x}) \tag{3.4.34}$$

uniformly, provided $a > 1 + |\phi_{\bar{x}}^{(3)}(0)|$ for all $\bar{x}$. Thus $|\beta| < t^2/4$. Also, we require that for $|t| < \delta$,

$$|\psi_{\bar{x}}(t)| < \frac{1}{4}\sigma^2(\bar{x})t^2, \tag{3.4.35}$$

so that $|\alpha| < t^2/4$. It then follows from (3.4.32) that for $|t| < \delta\sigma(\bar{x})\sqrt{n}$, the absolute value of the integrand for the second integral in (3.4.31) is *less* than

$$B_{\bar{x}}(t) = e^{-\frac{1}{4}t^2} \left[\frac{\epsilon|t|^r}{n^{r/2-1}} + \frac{a^{r-1}}{(r-1)!} \frac{|t|^{3(r-1)}}{(\sigma(\bar{x})\sqrt{n})^{r-1}} \right], \tag{3.4.36}$$

where to obtain the last term in the square brackets, we have used the bound $a|t|$ for $|\psi_{\bar{x}}^{(r)}(t)|$ given by (3.4.34). Since $\sigma(\bar{x})$ is bounded away from zero, the integral of $B_{\bar{x}}(t)$ over $(-\infty, \infty)$ is uniformly of order $O(\epsilon n^{-r/2+1})$ for all $\bar{x}$ in $D_{\bar{x}}$. Furthermore, since ϵ is arbitrary, we can write this order as $o(n^{-r/2+1})$. Thus the contribution to the second integral in (3.4.31) from the interval $|t| \leq \delta\sigma(\bar{x})\sqrt{n}$ is indeed uniformly $o(n^{-r/2+1})$ with respect to $\bar{x} \in D_{\bar{x}}$.

The contribution from the interval $|t| > \delta\sigma(\bar{x})\sqrt{n}$ may be evaluated using conditions (i) and (ii) stated in the theorem. By (ii) for $y > \delta$ there exists a $\rho_\delta > 0$ such that $|\phi_{\bar{x}}(y)| < \rho_\delta < 1$ for $\bar{x} \in D_{\bar{x}}$. Thus this contribution is bounded by

$$\rho_\delta^{n-\gamma} \int_{-\infty}^{\infty} \left|\phi_{\bar{x}}\left(\frac{t}{\sigma(\bar{x})\sqrt{n}}\right)\right|^\gamma dt + \int_{|t|>\delta\sigma(\bar{x})\sqrt{n}} e^{-\frac{1}{2}t^2}\left[1 + p_{\bar{x}}(t)\right] dt. \tag{3.4.37}$$

By condition (i) the first term in (3.4.37) is bounded by $\rho_\delta^{n-\gamma}\kappa$, which tends to zero more rapidly than any power of $1/n$. Using integration by parts, one can show that the second term is dominated by $\exp\{-\frac{1}{2}\delta^2\sigma^2(\bar{x})n\}$, which approaches zero uniformly in $\bar{x}$, faster than any power of $1/n$.

It follows that the second integral in (3.4.31) is uniformly $o(n^{-r/2+1})$. Let k be the largest integer smaller than or equal to $r/2 - 1$, and let the c_l's be the coefficients of the polynomial in $1/n$ resulting from the first integral. Then we can write (3.4.31) as

$$f_n(\bar{x}) = g_n(\bar{x})\left\{1 + \frac{c_1}{n} + \cdots + \frac{c_k}{n^k} + o\left(\frac{1}{n^k}\right)\right\}. \tag{3.4.38}$$

Note that the term $o(1/n^k)$ does not just represent the value of the second integral. It represents the sum of this value and terms of the polynomial that do not appear in (3.4.38). The c_l's are functions of $\bar{x}$. Specifically, they are fractions with $\sigma(\bar{x})$ in their denominators and $\phi_{\bar{x}}^{(i)}(0)$ $(i = 1, \ldots, r)$ in their numerators. Thus they are bounded for $\bar{x} \in D_{\bar{x}}$. It follows that the contribution to the error term $o(1/n^k)$ from the polynomial, i.e., $c_{k+1}/n^{k+1} + c_{k+2}/n^{k+2} + \cdots$, is uniformly $o(1/n^k)$ for $\bar{x} \in D_{\bar{x}}$. This and the fact that the second integral is uniformly $o(n^{-r/2+1})$ imply that the $o(1/n^k)$ in (3.4.38) is uniformly valid in $D_{\bar{x}}$.

By the uniqueness of asymptotic expansion with respect to the asymptotic sequence $\{1/n^l\}$, The asymptotic approximation (3.4.38) is that of Daniels' expansion (2.2.23) truncated at k. This implies that the truncated Daniels' expansion is uniformly valid in $D_{\bar{x}}$. By letting $r = 3, 4, \ldots,$ we conclude that the entire Daniels' expansion is uniformly valid. #

We now briefly comment on the method used in the above proof. The method offers an alternative to the method of steepest descents for showing the pointwise validity of the saddlepoint expansion. When $D_{\bar{x}}$ degenerates into a single point, the second condition of the theorem is always true, and the first condition is equivalent to the integrability of the complex moment generating function discussed in Section 2.2. We shall further comment on this condition at the end of this section. The advantage of this method compared to the method of steepest descents is that it may be used to prove the uniform validity of the saddlepoint expansion on bounded closed intervals as we have shown. Note that we did not give expressions for the coefficients c_i $(i = 1, 2, \ldots)$ in the proof. Since (3.4.38) is the same as Daniels' series (2.2.23), the c_i's are the same as the a_i's in (2.2.23).

Also, under a stronger set of conditions, the method may be used to prove uniform validity in the entire domain of $\bar{x}$. Specifically, if (1) for any given r, $M_r(\bar{x})$ is bounded, (2) $\phi''_{\bar{x}}(0) = \sigma^2(\bar{x})$ is bounded away from zero, and (3) the two conditions stated in the theorem are valid, throughout the domain, then the above proof may be adapted to show that the saddlepoint expansion is uniformly valid throughout the domain. This is so because under these conditions, the coefficients of the polynomial in $1/n$ are all bounded throughout the domain and the second integral is uniformly of order $o(n^{-r/2+1})$. Note that the set of sufficient conditions for uniform validity in Jensen (1991) includes only (1) and (3) but not (2). We do not have details of his proof. In order to use the above proof without condition (2), one would have to look into the asymptotic behavior of the $\phi^{(r)}_{\bar{x}}(0)$'s, e.g., the relative speed at which $\phi''_{\bar{x}}(0)$ and $\phi^{(r)}_{\bar{x}}(0)$ $(r > 2)$ converge to zero, to establish the boundedness of the coefficients and to eliminate the $\sigma(\bar{x})$ term in (3.4.36). Even when these can be accomplished, the above proof only implies the uniform validity for those $\bar{x}$ such that $\sigma(\bar{x}) = O(1/n^\theta)$ for some $\theta < 1/2$. This last point is clear from the discussion on (3.4.37).

We now prove the following theorem which shows that Daniels' series is in general uniformly valid in a bounded closed interval.

Theorem 3.3. *Let $f(x)$ be a continuous density function defined on an interval with end points e_1 and e_2 $(e_1 < e_2)$. If $\lim_{x \to e_i} f(x) = 0$ when e_i is infinity, and $f(x) = O(|e_i - x|^{-\alpha})$ for some $\alpha < 1$ when e_i is finite for $i = 1, 2$, then Daniels' series (2.2.23) is uniformly valid in any bounded closed interval, $D_{\bar{x}} = [x_1, x_2]$, inside the interval.*

The conditions in this theorem imply that: (1) when the domain of $f(x)$ is an unbounded interval, for any given $\beta > 0$ $f(x)$ is bounded by β outside some bounded interval, and (2) for any $p > 1$ such that $\alpha p < 1$, $f(x) \in L^p(R)$. In the case of the Gamma distribution with density $f(x) = x^{-1/2}e^{-x}/\Gamma(1/2)$, for example, $e_1 = 0$, $e_2 = \infty$, $\alpha = 1/2$ and $\lim_{x \to e_2} f(x) = 0$. We shall need the following inequality in our proof of the theorem:

Hausdorff-Young Inequality: *Let $1 < p \leq 2$ and φ be the Fourier transformation (characteristic function) of f. If $f \in L^p(R)$, then*

$$\|\varphi\|_q \leq (2\pi)^{\frac{1}{q}}\|f\|_p,$$

where $1/p + 1/q = 1$.

A proof of this inequality may be found in Sogge (1993) or Titchmarsh (1948).

Proof of Theorem 3.3:

We use the notation introduced in the proof for Lemma 3.4. We show that under conditions stated in Theorem 3.3, conditions (i) and (ii) in Theorem 3.2 are true.

(i) Since $[\hat{T}_1, \hat{T}_2] \subset (l, u)$, there exists a $p \in (1, 2]$ such that $\alpha p < 1$ and $[p\hat{T}_1, p\hat{T}_2] \subset (l, u)$. For a constant $\beta > 0$, let $L > 0$ be such that $f(y) < \beta$ when $|y| > L$. Then

$$
\begin{aligned}
\int_{-\infty}^{\infty} h_{\bar{x}}^p(u)du &= \int_{-\infty}^{\infty} \left[c(\bar{x})e^{\hat{T}u}f(u+\bar{x})\right]^p du \\
&= \left[c(\bar{x})e^{-\hat{T}\bar{x}}\right]^p \int_{-\infty}^{\infty} e^{p\hat{T}y}f(y)^p dy \\
&\leq \left[c(\bar{x})e^{-\hat{T}\bar{x}}\right]^p \left[\int_{-L}^{L} e^{p\hat{T}y}f(y)^p dy + \beta^{p-1}\int_{|y|>L} e^{p\hat{T}y}f(y)dy\right].
\end{aligned}
$$

Choose $B_1, B_2 > 0$ such that $\left[c(\bar{x})e^{-\hat{T}\bar{x}}\right]^p < B_1$, and $e^{p\hat{T}y} < B_2$ for $(\bar{x}, y) \in \{(\bar{x}, y) : \bar{x} \in [x_1, x_2], \text{ and } |y| < L\}$, then

$$\int_{-\infty}^{\infty} h_{\bar{x}}^p(u)du < B_1\left[B_2\int_{-L}^{L} f(y)^p dy + \beta^{p-1}M(p\hat{T})\right]. \tag{3.4.39}$$

$M(p\hat{T})$ is continuous and thus bounded in $p\hat{T} \in [p\hat{T}_1, p\hat{T}_2]$. Since $\alpha p < 1$, $f(x)$ is in $L^p(R)$, thus $\int_{-L}^{L} f(y)^p dy$ is bounded. It follows that the left-hand side of (3.4.39) is bounded by some constant, say $B > 0$, that is independent of $\bar{x}$.

By the Hausdorff-Young inequality, the characteristic function corresponding to density $h_{\bar{x}}(u)$, $\phi_{\bar{x}}(u)$, satisfies

$$\int_{-\infty}^{\infty} |\phi_{\bar{x}}(u)|^{\gamma} \, du \leq 2\pi B^{\frac{\gamma}{p}},$$

where γ is given by $1/p + 1/\gamma = 1$. Letting κ be $2\pi B^{\frac{\gamma}{p}}$ and noting that $\gamma > 0$, we obtain condition (i) in Theorem 3.2.

(ii) $\phi_{\bar{x}}(y)$ may be rewritten as follows:

$$\phi_{\bar{x}}(y) = \frac{M(\hat{T} + iy)}{M(\hat{T})} e^{i\bar{x}y}. \tag{3.4.40}$$

Thus to prove (ii) it suffices to show that for any $y_0 > 0$ there exist a $\rho_0 \in (0, 1)$ such that the ratio $r(y) = |M(\hat{T} + iy)/M(\hat{T})| < \rho_0$ when $|y| > y_0$ for all $\bar{x} \in [x_1, x_2]$.

Under the conditions stated in Theorem 3.3, the density function $f(x)$ is continuous and thus Theorem 2.1 gives the asymptotic behavior of $M(z)$. Let ρ_1 be in $(0, 1)$ and $\delta = \min_{\hat{T} \in [\hat{T}(x_1), \hat{T}(x_2)]} \{M(\hat{T})\}$. Since $\delta > 0$ and by Theorem 2.1 $|M(\hat{T} + iy)| \to 0$ uniformly as $|y| \to \infty$ with respect to $\hat{T} \in [\hat{T}(x_1), \hat{T}(x_2)]$, there exists $y_1 > 0$ such that for $|y| > y_1$, $|M(\hat{T}+iy)/\delta| < \rho_1$. Without loss of generality, we assume y_0 is in $(0, y_1)$. On the union of the two bounded, closed, rectangular regions given by $\{z : \hat{T} \in [\hat{T}(x_1), \hat{T}(x_2)]$ and $|y| \in [y_0, y_1]\}$ in the complex plane, $|M(\hat{T} + iy)| < M(\hat{T})$, and $r(y)$ is a continuous function. Thus $r(y)$ is bounded by a constant $\rho_2 < 1$.

Let $\rho_0 = \max\{\rho_1, \rho_2\}$, then $r(y) = |M(\hat{T} + iy)/M(\hat{T})| < \rho_0$ when $|y| > y_0$ for all $\hat{T} \in [\hat{T}(x_1), \hat{T}(x_2)]$. (ii) is thus proven. #

Theorem 3.3 may be generalized to densities with discontinuities. Specifically, the conditions may be replaced by that $f(x)$ be almost piecewise continuous, bounded outside some finite interval, and be in $L^p(R)$ for some $p > 1$ without affecting the conclusion.

The conditions stated in Theorem 3.3 are satisfied by commonly used continuous density functions with moment generating functions, including the four classes of density functions studied by Daniels (1954) and Jensen (1988) and all densities with closed bounded interval domains. This implies that the uniform validity of Daniels' expansion in any bounded closed interval is in general true. Note that to establish the relationship between Daniels' expansion and Lugannani and Rice's expansion (Theorem 3.1) at a certain point $\bar{x}$ in the interior of

the domain, we only need the uniform validity of Daniels' expansion in a small bounded closed interval containing the $\bar{x}$. Since Theorem 3.3 holds for *any* bounded closed interval in the interior, together with Theorem 3.1, they imply that Daniels' series is the derivative of Lugannani and Rice's series everywhere in the interior of the domain.

To conclude this section, we recall that to derive the saddlepoint expansion (Section 2.2), we used the condition that $|M(T)|^{\nu}$ is integrable for some $\nu > 0$. This condition is equivalent to $|\phi_{\bar{x}}(y)|^{\nu}$ be integrable for some $\nu > 0$. From the proof of Theorem 3.3, we see that such a condition is satisfied when conditions in Theorem 3.3 are met.

3.5 Applications

We discuss two applications of Theorem 3.1 and saddlepoints approximation for the density of standardized mean in this section. The first application establishes the derivative of a truncated Lugannani and Rice series as an asymptotic approximation to the density function. The second gives an answer to Lugannani and Rice's conjecture that we described at the beginning of this chapter. The saddlepoints approximation for the standardized mean is based on the saddlepoint approximation for the density of the mean, and its relative error is also of order $1/n$. We include this result in this section since it is an application of the uniform validity results discussed in this chapter.

1. *The derivative of a truncated Lugannani and Rice series as an asymptotic approximation to the density function.*

We consider the two expansions for $\bar{x}$ in some interval $D_{\bar{x}}$ where conditions stated in Theorem 3.1 are satisfied. Denote the sum of the first $m+2$ terms, including $\Phi(\hat{W}n^{\frac{1}{2}})$, of (2.2.31) by $F_n^{(m)}(\bar{x})$, and the sum of the first $m+1$ terms in (2.2.23) by $f_n^{(m)}(\bar{x})$, i.e.,

$$F_n^{(m)}(\bar{x}) = \Phi(\hat{W}n^{\frac{1}{2}}) - \phi(\hat{W}n^{\frac{1}{2}}) \sum_{r=0}^{m} \frac{b_r}{n^{r+\frac{1}{2}}},$$

and

$$f_n^{(m)}(\bar{x}) = g_n(\bar{x}) \sum_{r=0}^{m} \frac{a_r}{n^r}.$$

By differentiating $F_n^{(m)}(\bar{x})$, and then applying (3.2.6), we obtain

$$\frac{dF_n^{(m)}(\bar{x})}{d\bar{x}} = g_n(\bar{x}) \left[\sum_{r=0}^{m} \frac{a_r}{n^r} - \frac{(\hat{K}'')^{\frac{1}{2}}}{n^{m+1}} \frac{db_m}{d\bar{x}} \right] . \qquad (3.5.41)$$

Since $f_n^{(m)}(\bar{x})$ satisfies

$$f_n(\bar{x}) = g_n(\bar{x}) \left[\sum_{r=0}^{m} \frac{a_r}{n^r} + O\left(\frac{1}{n^{m+1}} \right) \right] , \qquad (3.5.42)$$

it follows that

$$f_n(\bar{x}) = g_n(\bar{x}) \left[\sum_{r=0}^{m} \frac{a_r}{n^r} - \frac{(\hat{K}'')^{\frac{1}{2}}}{n^{m+1}} \frac{db_m}{d\bar{x}} + O\left(\frac{1}{n^{m+1}} \right) \right] . \qquad (3.5.43)$$

Equations (3.5.41) and (3.5.43) imply that the derivative of $F_n^{(m)}(\bar{x})$ is an asymptotic approximation for $f_n(\bar{x})$, and that its error is of the same order as that of $f_n^{(m)}(\bar{x})$.

We give prominence to the derivative of $F_n^{(0)}(\bar{x})$ given below.

$$\begin{aligned}
\frac{dF_n^{(0)}(\bar{x})}{d\bar{x}} &= g_n(\bar{x}) \left[1 - \frac{1}{n} (\hat{K}'')^{\frac{1}{2}} \frac{db_0}{d\bar{x}} \right] \\
&= g_n(\bar{x}) \left[1 - \frac{1}{n} \left(\frac{|\hat{T}|(\hat{K}'')^{\frac{1}{2}}}{[2(\hat{T}\hat{K}' - \hat{K})]^{\frac{3}{2}}} - \frac{1}{\hat{T}^2 \hat{K}''} - \frac{\hat{K}^{(3)}}{2\hat{T}(\hat{K}'')^2} \right) \right] . \qquad (3.5.44)
\end{aligned}$$

We shall refer to (3.5.44) as the *adjusted saddlepoint approximation*, and the second term in the square brackets as the *adjustment term*.

The advantage of the adjusted saddlepoint approximation is that it in general does not need to be numerically renormalized since $F_n^{(0)}(\bar{x})$ generally approaches 0/1 when $\bar{x}$ approaches the lower/upper end of its domain. However, it also raises the following concerns: (1) it could be negative when the adjustment term is greater than 1, and (2) the adjustment term may compromise the accuracy of the original saddlepoint approximation. Nevertheless, our experience with the adjusted saddlepoint approximation has yet to validate these concerns. The first problem can only emerge when $F_n^{(0)}(\bar{x})$ is a decreasing function of $\bar{x}$. We have not found any example where this happens. Based on examples that we looked at, the adjusted saddlepoint approximation is actually more accurate than the original, and is often substantially more accurate near the mean. We now further illustrate this point with the following two examples. The distribution in the first example has unbounded domain, and that in the second example is bounded.

Example 3.1: Gamma(α, β)

$$f(x) = \frac{1}{\beta^\alpha \Gamma(\alpha)} x^{\alpha-1} e^{-x/\beta}, \quad x \geq 0.$$

$$K(T) = -\alpha \log(1 - \beta T), \quad K'(\hat{T}) = \frac{\alpha\beta}{1 - \beta\hat{T}} = \bar{x}, \quad K''(\hat{T}) = \frac{\bar{x}^2}{\alpha}.$$

From (2.2.16), the saddlepoint approximation for the mean of a sample of size n is

$$g_n(\bar{x}) = \left(\frac{n\alpha}{2\pi}\right)^{\frac{1}{2}} e^{n\alpha} \left(\frac{1}{\alpha\beta}\right)^{n\alpha} \bar{x}^{n\alpha-1} e^{-n\bar{x}/\beta}.$$

The exact result is

$$f_n(\bar{x}) = \frac{n^{n\alpha}}{\beta^{n\alpha}\Gamma(n\alpha)} \bar{x}^{n\alpha-1} e^{-n\bar{x}/\beta}.$$

Example 3.2: Uniform$[-1, 1]$

$$f(x) = \frac{1}{2}, \quad -1 \leq x \leq 1.$$

The density function for the mean of n independent observations from this distribution is given by (Seal, 1951),

$$f_n(\bar{x}) = \frac{n^n}{2^n(n-1)!} \sum_{k=0}^{n} (-1)^k \binom{n}{k} \left\langle 1 - \bar{x} - \frac{2k}{n} \right\rangle^{n-1}.$$

where $\langle z \rangle = z$ for $z \geq 0$, and $\langle z \rangle = 0$ for $z < 0$. The cumulant generating function and its first two derivatives are

$$K(T) = \log\left(\frac{\sinh(T)}{T}\right), \quad K'(\hat{T}) = \coth(\hat{T}) - \frac{1}{T} = \bar{x}, \quad K''(\hat{T}) = \frac{1}{\hat{T}^2} - \operatorname{csch}^2(\hat{T}).$$

The saddlepoint approximation is given by

$$g_n(\bar{x}) = \left(\frac{n}{2\pi}\right)^{\frac{1}{2}} \left[\frac{1}{\hat{T}^2} - \operatorname{csch}^2(\hat{T})\right]^{-\frac{1}{2}} \left(\frac{\sinh(\hat{T})}{\hat{T}}\right)^n e^{-\hat{T}\bar{x}n}.$$

The following tables contain, along with the exact values, the values of the saddlepoint approximation (spa) and of the adjusted saddlepoint approximation (aspa). Table 1 is for the case where the underlying distribution is gamma with both parameters equal to 2, and a sample size of 3. Table 2 is for the case where the underlying distribution is uniform$[-1,1]$, and a sample size of 5. Renormalization, though it can be quite involved, will in general

Table 3.1: Approximations to the density function for the mean of 3
independent observations from a gamma(2,2) distribution

$\bar{x}$	0.5	1.5	2.5	3.5	4.5
spa	0.6133652	0.3694524	0.0117770	0.0001570	1.3673e-06
aspa	0.6049278	0.3644115	0.0116238	0.0001551	1.3510e-06
exact	0.6049129	0.3643613	0.0116147	0.0001548	1.3484e-06

Table 3.2: Approximations to the density function for the mean of 5
independent observations from a uniform$[-1, 1]$ distribution

$\bar{x}$	0.1	0.3	0.5	0.7	0.9
spa	1.4461734	0.8411568	0.2628890	0.0340165	0.0004137
aspa	1.4022376	0.8128860	0.2521337	0.0323275	0.0004034
exact	1.4021810	0.8121745	0.2522786	0.0329590	0.0004069

improve the accuracy of the saddlepoint approximation. The renormalized spa values for
the uniform case may be found in Field and Ronchetti (1990), and are indeed more accurate
than the unrenormalized ones. However, even compared with these renormalized values,
aspa values are still more accurate.

When $\bar{x}$ is near the mean μ, the individual terms making up the adjustment term are
seen to be large. The adjustment term, as given in (3.5.44), is undefined at $\bar{x} = \mu$ where
$\hat{T} = 0$. One may thus be concerned with the accuracy of adjusted saddlepoint approximation
near μ. We now show that as $\bar{x}$ approaches μ the adjustment term has a finite limit, and
that this limit equals $-a_1/n$, the second term in Daniels' expansion (2.2.23). Using identity
(3.2.7), the adjustment term may be expressed in terms of a_1 as

$$\frac{1}{n}(\hat{K}'')^{-\frac{1}{2}}\frac{db_0}{d\hat{T}} = \frac{1}{n}\left[\hat{T}(\hat{K}'')^{\frac{1}{2}}b_1 - a_1\right].\tag{3.5.45}$$

As $\bar{x}$ approaches μ, $\hat{T}$ approaches 0 and, by (2.2.28), b_1 has a finite limit. Thus the right-
hand side of (3.5.45) approaches $-a_1/n$. It follows that at $\bar{x} = \mu$ the adjusted saddlepoint
approximation reduces to

$$\frac{dF_n^{(0)}(\bar{x})}{d\bar{x}} = g_n(\bar{x})\left[1 + \frac{a_1}{n}\right],$$

and in the neighborhood of μ where $\hat{T}(\hat{K}'')^{\frac{1}{2}}b_1 = O(1/n)$, it satisfies

$$\begin{aligned}
\frac{dF_n^{(0)}(\bar{x})}{d\bar{x}} &= g_n(\bar{x})\left\{1 + \frac{1}{n}\left[\hat{T}(\hat{K}'')^{\frac{1}{2}}b_1 - a_1\right]\right\} \\
&= g_n(\bar{x})\left\{1 + \frac{a_1}{n} + O\left(\frac{1}{n^2}\right)\right\}.
\end{aligned}$$

This means that near the mean the adjusted saddlepoint approximation is a second order approximation, and agrees with the examples where it is seen to be substantially more accurate than the original saddlepoint approximation.

The above derivation may be generalized to show that, for any m, the derivative of $F_n^{(m)}(\bar{x})$ is of smaller asymptotic error near the mean when compared with $f_n^{(m)}(\bar{x})$, and that at the mean it equals $f_n^{(m+1)}(\bar{x})$. This observation, however, does not constitute a reason for using the derivative of $F_n^{(m)}(\bar{x})$ to approximate the density instead of $f_n^{(m)}(\bar{x})$. The approximation with a smaller asymptotic error is not necessarily more accurate for small and moderate sample sizes. Also, with asymptotic approximations of any order available, other considerations, such as the simplicity of the formula and the actual accuracy observed through numerical examples, usually take precedence over the order of asymptotic error.

Nevertheless, for computing numerical approximations of the density, we recommend the adjusted saddlepoint approximation due to the advantage and accuracy we discussed earlier. When asymptotic approximations to both $F_n(\bar{x})$ and $f_n(\bar{x})$ are sought, the adjusted saddlepoint approximation is a particularly attractive alternative to the saddlepoint approximation since the corresponding approximation to $F_n(\bar{x})$, $F_n^{(0)}(\bar{x})$, is easily available.

2. *Lugannani and Rice's conjecture.*

Lugannani and Rice's conjecture states: "the integration of Daniels' series and our series for $Q_n(\bar{x})$ both give approximations to $Q_n(\bar{x})$ that are in error by the same order of magnitude". We now examine this conjecture in a simple setting where there is a saddlepoint for every possible $\bar{x}$ value, and $F_n^{(m)}(\bar{x})$ approaches 0/1 when $\bar{x}$ approaches the lower/upper end point of its domain. Since, by (2.2.31),

$$F_n(\bar{x}) - F_n^{(m)}(\bar{x}) = \phi(\hat{W}n^{\frac{1}{2}})O(1/n^{m+\frac{3}{2}}), \qquad (3.5.46)$$

where $O(1/n^{m+\frac{3}{2}})$ is uniform over the domain, the asymptotic behavior of $F_n^{(m)}(\bar{x})$ with respect to $\bar{x}$ may be determined through (3.5.46) by examining that of $\hat{W}$. When $|\hat{W}(\bar{x})|$ approaches infinity at the lower/upper end point, $\phi(\hat{W}n^{\frac{1}{2}})$ approaches 0. Thus there exists an n_0 such that when $n > n_0$ the right-hand side of (3.5.46) approaches 0. Since $F_n(\bar{x})$ approaches 0/1, (3.5.46) implies that $F_n^{(m)}(\bar{x})$ approaches 0/1.

In light of (3.5.46), the conjecture may be formulated in terms of $F_n(\bar{x})$ as

$$F_n(\bar{x}) - \int_{s_l}^{\bar{x}} f_n^{(m)}(y)dy = \phi(\hat{W}n^{\frac{1}{2}})O(1/n^{m+\frac{3}{2}}), \qquad (3.5.47)$$

where s_l is the lower end of the domain, and the constants associated with the symbol O in (3.5.46) and that in (3.5.47) may be different. We note that Lugannani and Rice (1980) are not specific about which truncated versions of their expansion for $Q_n(\bar{x})$ and Daniels' expansion for $f_n(\bar{x})$ the conjecture refers to. The numerical example prior to the conjecture in that paper compares $Q_n^{(1)}(\bar{x})$ $(= 1 - F_n^{(1)}(\bar{x}))$ with integrated $f_n^{(2)}(\bar{x})$. The formulation (3.5.47) implies that $F_n^{(m)}(\bar{x})$ is to be compared with integrated $f_n^{(m)}(\bar{x})$. We choose to do so because it is most meaningful to compare approximations formed with the same number of terms (m) from the asymptotic components of the two expansions in terms of their orders of asymptotic error. This formulation is somewhat subjective but the subjectivity is not critical to our discussion. Our main purpose is to demonstrate the use of the theorem in answering this conjecture. Should the conjecture be formulated differently, the theorem is still applicable, although the conclusion may be different.

In order to answer the conjecture, we need to first express the difference between $F_n(\bar{x})$ and integrated $f_n^{(m)}(\bar{x})$ in terms of quantities whose asymptotic orders can be evaluated. To see how this may be achieved, we rewrite the derivative of $F_n^{(m)}(y)$, given by (3.5.41), at a point y in the domain as

$$\frac{dF_n^{(m)}(y)}{dy} = f_n^{(m)}(y) - \frac{1}{n^{m+\frac{1}{2}}}\phi(Wn^{\frac{1}{2}})\frac{db_m}{dT}\frac{dT}{dy}, \qquad (3.5.48)$$

where $T = T(y)$ is the saddlepoint corresponding to y and $W = sgn(T)\{2[TK'(T) - K(T)]\}^{\frac{1}{2}}$. By integrating, from s_l to $\bar{x}$, both sides of (3.5.48), we obtain

$$F_n^{(m)}(\bar{x}) = \int_{s_l}^{\bar{x}} f_n^{(m)}(y)dy - \frac{1}{n^{m+\frac{1}{2}}}R_n^{(m)}(\bar{x}), \qquad (3.5.49)$$

where

$$R_n^{(m)}(\bar{x}) = \int_{T(s_l)}^{\hat{T}} \phi(Wn^{\frac{1}{2}})\frac{db_m}{dT}dT. \qquad (3.5.50)$$

Equations (3.5.46) and (3.5.49) lead to the difference between $F_n(\bar{x})$ and integrated $f_n^{(m)}(\bar{x})$

$$F_n(\bar{x}) - \int_{s_l}^{\bar{x}} f_n^{(m)}(y)dy = \phi(\hat{W}n^{\frac{1}{2}})O(1/n^{m+\frac{3}{2}}) - \frac{1}{n^{m+\frac{1}{2}}}R_n^{(m)}(\bar{x}). \qquad (3.5.51)$$

Laplace's method for integrals with large parameters (see, e.g., Murray, 1974) may now be used to obtain the asymptotic order of $R_n^{(m)}(\bar{x})$. The derivative of $TK'(T) - K(T)$ is $TK''(T)$, where $K''(T)$ is positive. Thus $TK'(T) - K(T)$ is monotonely decreasing for $T < 0$ and monotonely increasing for $T > 0$. At $T = 0$, it has its minimum 0. It follows that function $-W^2(T)$ has its global maximum at $T = 0$, and is monotonely increasing for $T < 0$ and monotonely decreasing for $T > 0$. The integral in the right-hand side of (3.5.50) can thus be expanded by using Laplace's method at $T = 0$ when $\hat{T} \geq 0$ (or $\bar{x} \geq \mu$), and at $\hat{T}$ when $\hat{T} < 0$ (or $\bar{x} < \mu$). This gives

$$R_n^{(m)}(\bar{x}) = \begin{cases} \phi(\hat{W}n^{\frac{1}{2}})O(1/n) & \text{if } \bar{x} < \mu, \\ O(1/n^{\frac{1}{2}}) & \text{if } \bar{x} \geq \mu. \end{cases} \tag{3.5.52}$$

Equations (3.5.51) and (3.5.52) then imply

$$F_n(\bar{x}) - \int_{s_l}^{\bar{x}} f_n^{(m)}(y)dy = \begin{cases} \phi(\hat{W}n^{\frac{1}{2}})O(1/n^{m+\frac{3}{2}}) & \text{if } \bar{x} < \mu, \\ O(1/n^{m+1}) & \text{if } \bar{x} \geq \mu. \end{cases} \tag{3.5.53}$$

For $\bar{x} > \mu$, $F_n(\bar{x})$ may be more accurately approximated by subtracting from 1 the integral of $f_n^{(m)}(y)$ over $(\bar{x}, s_u)$, where s_u is the upper end of the domain. By essentially repeating the above procedure, we obtain

$$F_n(\bar{x}) - \left(1 - \int_{\bar{x}}^{s_u} f_n^{(m)}(y)dy\right) = \phi(\hat{W}n^{\frac{1}{2}})O(1/n^{m+\frac{3}{2}}) \text{ if } \bar{x} > \mu. \tag{3.5.54}$$

It follows from (3.5.53) and (3.5.54) that for $\bar{x} \neq \mu$, the smallest error achievable by using integrated truncated Daniels' series $f_n^{(m)}(y)$ to approximate $F_n(\bar{x})$ is $\phi(\hat{W}n^{\frac{1}{2}})O(1/n^{m+\frac{3}{2}})$, the same as that of truncated Lugannani and Rice's $F_n^{(m)}(y)$. At the mean, it is $O(1/n^{m+1})$, but that of Lugannani and Rice's is $O(1/n^{m+\frac{3}{2}})$. Thus Lugannani and Rice's conjecture, as formulated by (3.5.47), is correct everywhere, except at the mean.

3. *Saddlepoints expansion for the density of a standardized mean.*

Consider the standardized mean, Z_n, given by

$$Z_n = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}}, \tag{3.5.55}$$

where μ and σ are the mean and standard deviation of X, respectively. For simplicity of presentation, we shall write a realization of Z_n as z rather than z_n. The density, $f_{Z_n}(z)$, has

Edgeworth expansion, which shows that the relative error of approximating this density with that of the standard normal, $\phi(z)$, is in general of order $1/\sqrt{n}$. We now use the uniform validity of the saddlepoint expansion in a small neighborhood of the mean to develop a saddlepoints expansion for $f_{Z_n}(z)$ whose relative error is of order $1/n$. The reason for using the term "saddlepoint*s* expansion" will become apparent later.

Using (3.5.55) we can express $\bar{x}$ as a function of z and n. We have

$$\bar{x}(z,n) = \frac{\sigma}{\sqrt{n}}z + \mu, \quad \text{and} \quad \frac{\partial \bar{x}(z,n)}{\partial z} = \frac{\sigma}{\sqrt{n}}.$$

Thus the $f_{Z_n}(z)$ may be expressed in terms of $f_n(\bar{x})$ as

$$f_{Z_n}(z) = \frac{\sigma}{\sqrt{n}}f_n(\bar{x}(z,n)). \tag{3.5.56}$$

Since $f_{Z_n}(z)$ and $f_n(\bar{x}(z,n))$ only differ by a known factor $\sigma/\sqrt{n}$, an approximation to $f_n(\bar{x}(z,n))$ multiplied by the factor may be used as an approximation to $f_{Z_n}(z)$. Furthermore, the relative errors of these two approximations are equal.

For any fixed z, $\bar{x}(z,n)$ changes with n and is increasingly close to μ as n gets larger. If (2.2.23) is uniformly valid in some neighborhood of μ, i.e., the relative error for using the sum of the first $r + 1$ terms of the expansion to approximate $f_n(\bar{x})$ is uniformly of order $O\left(1/n^{r+1}\right)$ in that neighborhood, then $f_{Z_n}(z)$ has expansion

$$f_{Z_n}(z) = \frac{\sigma}{\sqrt{n}}g_n(\bar{x}(z,n))\left\{1 + \frac{a_1(\bar{x}(z,n))}{n} + \cdots + \frac{a_r(\bar{x}(z,n))}{n^r} + O\left(\frac{1}{n^{r+1}}\right)\right\}, \tag{3.5.57}$$

where, to recognize their dependence on z and n, we write the a_r's as $a_r(\bar{x}(z,n))$. It should be noted that expansion (3.5.57) is not a power series expansion in Poincaré's sense in that the a_r's are not constants and will change with the large parameter n. Nevertheless, it is a meaningful expansion since the $O\left(1/n^{r+1}\right)$ term in the expansion is indeed valid. In terms of Poincaré's general definition of asymptotic expansion, it is an expansion with respect to the asymptotic sequence $\{\phi_r(n)\}$ where $\phi_r(n) = a_r(\bar{x}(z,n))/n^r$ and the coefficients of the $\phi_r(n)$'s are all 1. The expansion may thus be written in terms of $\{\phi_r(n)\}$ as

$$f_{Z_n}(z) = \frac{\sigma}{\sqrt{n}}g_n(\bar{x}(z,n))\left\{1 + \frac{a_1(\bar{x}(z,n))}{n} + \cdots + \frac{a_r(\bar{x}(z,n))}{n^r} + O\left(\frac{a_{r+1}(\bar{x}(z,n))}{n^{r+1}}\right)\right\}. \tag{3.5.58}$$

Furthermore, the condition that (2.2.23) is uniformly valid in some neighborhood of μ implies not only pointwise validity of (3.5.57), but also its uniform validity in any compact subset

on the real line. If (2.2.23) is uniformly valid throughout the domain of $\bar{X}$, then (3.5.57) is also uniformly valid in the domain of Z_n, although the domain expands with n when that of $\bar{X}$ is not the entire real line.

It should also be noted that (3.5.56) may not be used the other way around to obtain an expansion for $f_n(\bar{x})$ through the Edgeworth expansion for $f_{Z_n}(z)$ unless $\bar{x} = \mu$. This is because for any fixed $\bar{x}$ ($\neq \mu$), the value of z goes to infinity as n increases. Since the Edgeworth expansion for $f_{Z_n}(z)$ is not uniformly valid on the entire real line, the above argument will not work. At $\bar{x} = \mu$, the Edgeworth expansion for $f_{Z_n}(z)$ leads to an expansion for $f_n(\bar{x})$, which is the saddlepoint expansion for $f_n(\bar{x})$ at $\bar{x} = \mu$.

We shall denote the leading term of (3.5.57) by $h_n(z)$ and call it the saddlepoint*s* approximation to $f_{Z_n}(z)$ as it involves a sequence of saddlepoints which converge to zero as n increases. Let $\hat{T}_n$ be the saddlepoint corresponding to $\bar{x}(z, n)$, the saddlepoints approximation is given by

$$h_n(z) = \frac{\sigma}{\sqrt{2\pi K''(\hat{T}_n)}} e^{n\{K(\hat{T}_n) - \bar{x}\hat{T}_n\}}. \tag{3.5.59}$$

Since $\hat{T}_n \to 0$, it is not difficult to show that $K''(\hat{T}_n) \to \sigma^2$ and $n\{K(\hat{T}_n) - \bar{x}\hat{T}_n\} \to -z^2/2$. Thus $h_n(z)$ approaches $\phi(z)$. This agrees with the Edgeworth expansion that $f_{Z_n}(z)$ approaches $\phi(z)$ as n approaches infinity.

Table 3.3: Relative errors of one and three-term Edgeworth expansions and saddlepoints approximation

n	z	$\phi(z)$	$f_e(z)$	$h_n(z)$
	0.50	0.164e-01	0.611e-03	0.313e-01
n=5	1.50	0.567e-01	0.544e-02	0.366e-01
	2.50	0.563e-01	0.817e-02	0.419e-01
	3.50	0.865e+01	0.973e+00	0.167e-01
	0.50	0.816e-02	0.293e-03	0.154e-01
n=10	2.00	0.266e-01	0.229e-02	0.176e-01
	3.50	0.718e+00	0.346e-01	0.198e-01
	5.00	0.104e+04	*0.145e+04	0.836e-02

To demonstrate the accuracy of the saddlepoints approximation, consider the approximation of the density of the standardized mean of n uniform$[-1, 1]$ observations. The exact density may be found in Field and Ronchetti (p. 17, 1990). We compare the saddlepoints

approximation with one-term and three-term Edgeworth expansions. The one-term Edgeworth expansion is the standard normal density, which for this particular case also has relative error of order $1/n$. The three-term Edgeworth expansion, $f_e(z)$, has relative error of order $1/n^{3/2}$ and is given by

$$f_e(z) = \phi(z)\left\{1 - \frac{1}{20n}(z^4 - 6t^2 + 3)\right\}. \qquad (3.5.60)$$

Table 3.3 contains the relative error of the three approximations at some equally spaced points for $n = 5$ and $n = 10$. $f_e(z)$ can produce negative approximations at the tail and an asterisk in the $f_e(z)$ column indicates that the corresponding approximation is negative.

Even when the sample size is as small as 5, the saddlepoints approximation's relative error is less than 5% throughout the domain, while the Edgeworth expansions are good in the middle but unacceptable at the tail. When sample size reaches 10, the three-term Edgeworth expansion starts to produce negative approximations at the tail. Further computation can show that when sample size is greater than 20, the relative error of the saddlepoints approximation is uniformly less than 1% while that for the Edgeworth expansions can be as large as 10^{15} at the tail.

Finally, Lugannani and Rice's expansion for $F_n(\bar{x})$ may be used to derive a saddlepoints approximation for the cumulative distribution function of Z_n since it is uniformly valid.

Chapter 4

General Asymptotic Expansions

4.1 Introduction

In this chapter, we introduce a family of formal expansions for density functions of one-dimensional continuous distributions that are asymptotically normal. The expansions are derived through a formal method which is analogous to the formal argument Daniels (1954) used to demonstrate the asymptotic validity of the saddlepoint expansion. This family (1) provides a systematic framework for important expansions for densities of the sample mean and the standardized mean, such as the saddlepoint expansion and the Edgeworth expansion, and (2) leads to natural generalizations of these expansions for distributions that are asymptotically normal but are not necessarily related to a sample mean. Much of the discussion in this chapter will be centered on these two points. The foundation of this method is the inversion formula given by Theorem 2.3:

$$f_n(x) = \frac{1}{2\pi i} \int_{\tau - i\infty}^{\tau + i\infty} e^{K_n(T) - Tx} dT, \tag{4.1.1}$$

where $K_n(T)$ is the cumulant generating function for X_n with density $f_n(x)$, $T = \tau + iy$, and $\tau \in (l, u)$, which is the largest interval in which $K_n(\tau)$ exists. Inversion formula (4.1.1)

expresses the density function as a contour integral for which an asymptotic expansion often exists. We assume X_n is asymptotically normal as n approaches a certain limit, usually infinity, where n is not necessarily a sample size but merely an index or parameter that may assume values other than integers. Also, we assume that the interval (l, u) does not change with n. For simplicity of presentation, we shall use x to denote a realization of X_n.

To highlight the difference between our approach and Charlier's (1906) approach for obtaining expansions for densities, we shall discuss the Charlier differential series and the Edgeworth expansion in Section 4.2. In Section 4.3, we introduce the formal method and derive the family of formal expansions for $f_n(x)$. In Section 4.4, we give a systematic account of known expansions related to the sample mean and the standardized mean in the context of this family of expansions. In Section 4.5, we study the validity of the family, in particular the validity of the saddlepoint approximation, for Jørgensen's exponential dispersion models. We conclude with some numerical examples in Section 4.6.

4.2   Charlier differential series and Edgeworth expansion

The Charlier differential series represents perhaps the first attempt to study expansions for density functions in a systematic manner. It first appeared in Charlier (1906). In a review paper, Wallace (1958) discussed this series and gave some related historical notes. Discussions concerning this series may also be found in Kendall and Stuart (1969). To facilitate comparison between this series and other expansions, we give the following derivation which is slightly different from that shown by Wallace (1958). We use the inversion formula (4.1.1) as the starting point. Let $f(x)$ be the density to be expanded and $g(x)$ be the developing function with respect to which $f(x)$ is to be expanded. Then let $K_f(t)$ and $K_g(t)$ be the cumulant generating functions of the two distributions. Let κ_i and κ_i' $(i = 1, 2, \ldots)$ be the cumulants of these two distributions, respectively. Then

$$
\begin{aligned}
f(x) &= \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{K_f(it) - itx} dt \\
&= \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{K_f(it) - K_g(it)} e^{K_g(it)} e^{-itx} dt \\
&= \frac{1}{2\pi} \int_{-\infty}^{\infty} \exp\left\{ \sum_{r=1}^{\infty} (\kappa_r - \kappa_r') \frac{(it)^r}{r!} \right\} e^{K_g(it)} e^{-itx} dt. \qquad (4.2.2)
\end{aligned}
$$

Assume all derivatives of $g(x)$ vanish at the extremes of its domain. Then by the definition of $K_g(it)$ and integration by parts, we have

$$e^{K_g(it)} = \frac{(-1)^r}{(it)^r} \int_{-\infty}^{\infty} e^{itx} g^{(r)}(x) dx, \quad \text{or} \quad (it)^r e^{K_g(it)} = \int_{-\infty}^{\infty} e^{itx} (-1)^r g^{(r)}(x) dx.$$

Using Fourier inversion, we obtain

$$(-1)^r g^{(r)}(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} (it)^r e^{K_g(it)} e^{-itx} dt.$$

Thus from (4.2.2) we have formally

$$f(x) = \exp\left\{ \sum_{r=1}^{\infty} (\kappa_r - \kappa_r') \frac{(-D)^r}{r!} \right\} g(x), \tag{4.2.3}$$

where D denotes the differential operator, d/dx. More generally, if $K_f(x)$ and $K_g(x)$ exist in some interval I, then $f(x)$ may be formally expressed as

$$f(x) = \frac{1}{2\pi i} \int_{\tau - i\infty}^{\tau + i\infty} \exp\left\{ \sum_{r=0}^{\infty} \left[K_f^{(r)}(\tau) - K_g^{(r)}(\tau) \right] \frac{(T - \tau)^r}{r!} \right\} e^{K_g(T) - Tx} dT, \tag{4.2.4}$$

where $\tau \in I$. By

$$(-1)^r g^{(r)}(x) = \frac{1}{2\pi i} \int_{\tau - i\infty}^{\tau + i\infty} T^r e^{K_g(T) - Tx} dT,$$

we have

$$f(x) = \exp\left\{ \sum_{r=0}^{\infty} \left[K_f^{(r)}(\tau) - K_g^{(r)}(\tau) \right] \frac{(-D - \tau)^r}{r!} \right\} g(x). \tag{4.2.5}$$

An important special case arises when $f(x)$ $(= f_n(x))$ is the density of the standardized mean of n i.i.d. observations, and $g(x)$ is the density of the standard normal distribution, $\phi(x)$. In this case (4.2.3) becomes

$$f_n(x) = \exp\left\{ \sum_{r=3}^{\infty} \kappa_r \frac{(-D)^r}{r!} \right\} \phi(x). \tag{4.2.6}$$

Let λ_r $(r = 1, 2, \ldots)$ be the standardized cumulants for the underlying distribution of the standardized mean, then $\kappa_r = \lambda_r / n^{r/2-1}$. Substituting $\lambda_r / n^{r/2-1}$ for κ_r in (4.2.6) and rearranging terms according to the powers of $1/\sqrt{n}$, we obtain the Edgeworth expansion

$$f_n(x) = \phi(x) + \sum_{r=3}^{\infty} \frac{P_r(-\phi(t))}{n^{r/2-1}}, \tag{4.2.7}$$

where $P_r(\cdot)$ is a polynomial of degree $3(r-2)$ with coefficients depending on $\lambda_3, \lambda_4, \ldots, \lambda_r$. The powers of the polynomial should be interpreted as derivatives, e.g., $\phi^2(x) = D^2\phi(x)$. See Section 4.4 for further details. The expansion (4.2.7) may be written as

$$f_n(x) = \phi(x) + \frac{1}{\sqrt{n}}\frac{\lambda_3}{6}H_3(x)\phi(x) + \frac{1}{n}\left[\frac{\lambda_4}{24}H_4(x) + \frac{\lambda_3^2}{72}H_6(x)\right]\phi(x) + \cdots, \qquad (4.2.8)$$

where $H_r(x)$ is the Hermite polynomial of order r. Discussions concerning the validity of the Edgeworth expansion (4.2.8) may be found in Cramér (1962) and Feller (1966). Here we present a result which may be found on page 506 in Feller (1966). Note that this result does not require the existence of all cumulants of the underlying distribution.

Theorem 4.1: *Suppose the third moment of the underlying distribution exists and its characteristic function ψ satisfies that $|\psi|^\nu$ is integrable for some $\nu \geq 1$, then $f_n(x)$ exists for $n \geq \nu$ and as $n \to \infty$*

$$f_n(x) = \phi(x) + \frac{1}{\sqrt{n}}\frac{\lambda_3}{6}H_3(x)\phi(x) + o(1/\sqrt{n}) \qquad (4.2.9)$$

uniformly in x.

Note that equation (4.2.9) is equivalent to

$$f_n(x) = \phi(x)\left\{1 + \frac{1}{\sqrt{n}}\frac{\lambda_3}{6}H_3(x) + o(1/\sqrt{n})\right\} \qquad (4.2.10)$$

at any fixed x. The $o(1/\sqrt{n})$ term in (4.2.10) denotes the relative error of the expansion and can be obtained from the error term in (4.2.9) by dividing by $\phi(x)$. A careful examination of Feller's proof can show that it is uniformly valid in any bounded subset of R, and is not uniformly valid in any unbounded subset. At the mean, $x = 0$ and $H_r(x) = 0$ if r is odd, and thus the Edgeworth expansion becomes an expansion in integer powers of $1/n$. Because of this, as approximations to $f_n(x)$, truncated versions of the Edgeworth expansion with even number of terms are asymptotically more accurate at the mean than anywhere else.

4.3 General expansions for densities

We now introduce the formal method and derive the family of expansions for density functions. Following the development of the Edgeworth expansion, the method consists of two

steps, i.e., (1) expanding density functions in terms of quantities related to the cumulant generating functions, and (2) examining the asymptotic properties of these quantities, rearranging terms in the expansions according to their asymptotic orders, and thus obtaining formal asymptotic expansions. We shall illustrate the method through examples and highlight differences between it and the derivation of Charlier differential series.

- *The expansion at the saddlepoint*

We first define the generalized saddlepoint approximation and briefly discuss the saddlepoint. Let $\hat{T}_n \in (l, u)$ be the saddlepoint corresponding to x, i.e., the solution to equation $K'_n(T) - x = 0$. We define $g_n(x)$, the generalized saddlepoint approximation for $f_n(x)$ as

$$g_n(x) = \sqrt{\frac{1}{2\pi K''_n(\hat{T}_n)}} e^{K_n(\hat{T}_n) - \hat{T}_n x}. \qquad (4.3.11)$$

It should be noted that in general the saddlepoint corresponding to x may vary with n. The existence of a saddlepoint has been discussed by Daniels (1954). Since qualitative features of $f_n(x)$ such as whether it has a finite domain are usually available, Theorems 6.1 and 6.2 concerning the existence of the saddlepoint in Daniels (1954) may be used to show the existence of a saddlepoint for each x value when $f_n(x)$ has finite domain.

To obtain the expansion at the saddlepoint, set τ in (4.1.1) to $\hat{T}_n$. On the contour near $\hat{T}_n$, $K_n(T) - Tx$ has Taylor expansion

$$K_n(T) - Tx = K_n(\hat{T}_n) - \hat{T}_n x - \frac{1}{2}K''_n(\hat{T}_n)y^2 - \frac{1}{6}K_n^{(3)}(\hat{T}_n)iy^3 + \frac{1}{24}K_n^{(4)}(\hat{T}_n)y^4 + \cdots \quad (4.3.12)$$

Equation (4.1.1) is then formally rewritten as

$$f_n(x) = \frac{1}{2\pi}e^{K_n(\hat{T}_n) - \hat{T}_n x}\int_{-\infty}^{\infty} e^{-\frac{1}{2}K''_n(\hat{T}_n)y^2 - \frac{1}{6}K_n^{(3)}(\hat{T}_n)iy^3 + \cdots}dy. \qquad (4.3.13)$$

Letting $v = [K''_n(\hat{T}_n)]^{1/2}y$ and $\lambda_{(n,j)}(\hat{T}_n) = K_n^{(j)}(\hat{T}_n)/[K''_n(\hat{T}_n)]^{j/2}$ for $j = 3, 4, \ldots$, then (4.3.13) becomes

$$
\begin{aligned}
f_n(x) &= \frac{1}{2\pi[K''_n(\hat{T}_n)]^{1/2}}e^{K_n(\hat{T}_n) - \hat{T}_n x}\int_{-\infty}^{\infty} e^{-\frac{1}{2}v^2 - \frac{1}{6}\lambda_{(n,3)}(\hat{T}_n)iv^3 + \frac{1}{24}\lambda_{(n,4)}(\hat{T}_n)v^4 + \cdots}dv \\
&= g_n(x)\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{-\frac{1}{2}v^2 - \frac{1}{6}\lambda_{(n,3)}(\hat{T}_n)iv^3 + \frac{1}{24}\lambda_{(n,4)}(\hat{T}_n)v^4 + \cdots}dv \\
&= g_n(x)\int_{-\infty}^{\infty} \phi(v)e^{-\frac{1}{6}\lambda_{(n,3)}(\hat{T}_n)iv^3 + \frac{1}{24}\lambda_{(n,4)}(\hat{T}_n)v^4 + \cdots}dv. \qquad (4.3.14)
\end{aligned}
$$

Expanding the function exp $(\cdot)$ in the integrand, we get

$$
\begin{aligned}
f_n(x) = \; & g_n(x) \int_{-\infty}^{\infty} \phi(v) \left\{ 1 - \frac{1}{6}\lambda_{(n,3)}(\hat{T}_n)iv^3 + \right. \\
& \left. \frac{1}{24}\lambda_{(n,4)}(\hat{T}_n)v^4 + \cdots - \frac{1}{72}\lambda_{(n,3)}^2(\hat{T}_n)v^6 + \cdots \right\} dv.
\end{aligned}
\tag{4.3.15}
$$

Integrating (4.3.15) term-by-term and noting that odd moments of the standard normal distribution are zero, we obtain

$$
f_n(x) \sim g_n(x) \left\{ 1 + \frac{1}{8}\lambda_{(n,4)}(\hat{T}_n) + \frac{1}{2 \times 4!}\lambda_{(n,6)}(\hat{T}_n) + \cdots - \frac{5}{24}\lambda_{(n,3)}^2(\hat{T}_n) + \cdots \right\}.
\tag{4.3.16}
$$

This is the expansion at the saddlepoint.

● *The expansion at the origin*

Another special point for $K_n(T) - Tx$ is the origin at which it is zero. Set τ in (4.1.1) to 0. On the contour near the origin,

$$
K_n(T) - Tx = (K_n'(0) - x)iy - \frac{1}{2}K_n''(0)y^2 + \frac{1}{3!}K_n^{(3)}(0)(iy)^3 + \frac{1}{4!}K_n^{(4)}(0)(iy)^4 + \cdots
\tag{4.3.17}
$$

Since $K_n'(0) = \mu_n$ and $K_n''(0) = \sigma_n^2$, where μ_n and σ_n^2 are the mean and variance of X_n, (4.3.17) may be written as

$$
\begin{aligned}
K_n(T) - Tx = \; & (\mu_n - x)iy - \frac{1}{2}\sigma_n^2 y^2 + \frac{1}{3!}K_n^{(3)}(0)(iy)^3 + \frac{1}{4!}K_n^{(4)}(0)(iy)^4 + \cdots \\
= \; & -\frac{\sigma_n^2}{2}\left[y - \frac{(\mu_n - x)i}{\sigma_n^2} \right]^2 - \frac{(\mu_n - x)^2}{2\sigma_n^2} + \frac{1}{3!}K_n^{(3)}(0)(iy)^3 + \cdots
\end{aligned}
\tag{4.3.18}
$$

With $c = (\mu_n - x)/\sigma_n^2$, equation (4.1.1) may be rewritten as

$$
f_n(x) = \frac{1}{2\pi}e^{-(\mu_n - x)^2/2\sigma_n^2} \int_{-\infty}^{\infty} e^{-\frac{\sigma_n^2}{2}(y - ic)^2 + \frac{1}{3!}K_n^{(3)}(0)(iy)^3 + \frac{1}{4!}K_n^{(4)}(0)(iy)^4 + \cdots}dy.
\tag{4.3.19}
$$

Letting $v = \sigma_n(y - ic)$ and $\lambda_{(n,j)}(0) = K_n^{(j)}(0)/\sigma_n^j$ for $j \geq 3$, we may write (4.3.19) as

$$
\begin{aligned}
f_n(x) = \; & \frac{1}{2\pi\sigma_n}e^{-(\mu_n - x)^2/2\sigma_n^2} \int_{-\infty - ic\sigma_n}^{\infty - ic\sigma_n} e^{-\frac{1}{2}v^2 + \frac{1}{3!}\lambda_{(n,3)}[i(v + ic\sigma_n)]^3 + \frac{1}{4!}\lambda_{(n,4)}[i(v + ic\sigma_n)]^4 + \cdots}dv \\
= \; & \phi_n(x) \int_{-\infty - ic\sigma_n}^{\infty - ic\sigma_n} \phi(v)e^{\frac{1}{3!}\lambda_{(n,3)}[i(v + ic\sigma_n)]^3 + \frac{1}{4!}\lambda_{(n,4)}[i(v + ic\sigma_n)]^4 + \cdots}dv,
\end{aligned}
\tag{4.3.20}
$$

where ϕ_n is the density of $N(\mu_n, \sigma_n^2)$, and for brevity we have written $\lambda_{(n,j)}(0)$ as $\lambda_{(n,j)}$. Expanding the function $\exp(\cdot)$ in the integrand, we get

$$
\begin{aligned}
f_n(x) \;=\; & \phi_n(x) \int_{-\infty-ic\sigma_n}^{\infty-ic\sigma_n} \phi(v) \left\{ 1 + \frac{1}{3!} \lambda_{(n,3)} [i(v + ic\sigma_n)]^3 \right. \\
& \left. + \; \frac{1}{4!} \lambda_{(n,4)} [i(v + ic\sigma_n)]^4 + \cdots + \frac{1}{2!(3!)^2} \lambda_{(n,3)}^2 [i(v + ic\sigma_n)]^6 + \cdots \right\} dv \quad (4.3.21)
\end{aligned}
$$

Since $[i(v+ic\sigma_n)]^3 = -iv^3 + 3v^2c\sigma_n + i3v(c\sigma_n)^2 - (c\sigma_n)^3$, which has terms involving v^2 and v^0, thus unlike in (4.3.16) $\lambda_{(n,3)}$ will not disappear after term-by-term integration. This term-by-term integration may be easily carried out by noting that $\phi(v)v^r$ is an entire function, and thus the contour of integration in (4.3.21) may be deformed from $Im(v) = -c\sigma_n$ to $Im(v) = 0$. The following identity may be used to simplify this integration

$$
\int_{-\infty}^{\infty} \phi(v)(iv - x)^k dv = (-1)^k H_k(x), \qquad (4.3.22)
$$

for $k = 0, 1, \ldots$, where $H_k(x)$ is the Hermite polynomial of degree k (see Appendix B for a proof). This identity and (4.3.21) lead to

$$
\begin{aligned}
f_n(x) \;\sim\; & \phi_n(x) \left\{ 1 + \frac{1}{3!} \lambda_{(n,3)} (-1)^3 H_3(c\sigma_n) \right. \\
& \left. + \; \frac{1}{4!} \lambda_{(n,4)} (-1)^4 H_4(c\sigma_n) + \cdots + \frac{1}{2!(3!)^2} \lambda_{(n,3)}^2 (-1)^6 H_6(c\sigma_n) + \cdots \right\}, \quad (4.3.23)
\end{aligned}
$$

or equivalently

$$
\begin{aligned}
f_n(x) \;\sim\; & \phi_n(x) \left\{ 1 - \frac{1}{3!} \lambda_{(n,3)} H_3\left(\frac{\mu_n - x}{\sigma_n}\right) \right. \\
& \left. + \; \frac{1}{4!} \lambda_{(n,4)} H_4\left(\frac{\mu_n - x}{\sigma_n}\right) + \cdots + \frac{1}{2!(3!)^2} \lambda_{(n,3)}^2 H_6\left(\frac{\mu_n - x}{\sigma_n}\right) + \cdots \right\}. \quad (4.3.24)
\end{aligned}
$$

This is the expansion for $f_n(x)$ at the origin.

• *Expansions at other points*

At any fixed $\tau \in (l, u)$ where $K''(\tau) > 0$, the argument demonstrated above may be used to derive an expansion for $f_n(x)$. The leading term in the Taylor expansion of $K_n(T) - Tx$ at $T = \tau$ is in general a nonzero constant $K_n(\tau) - \tau x$. Thus the resulting expansion is

$$
\begin{aligned}
f_n(x) \;\sim\; & e^{K_n(\tau) - \tau x} \phi_{(n,\tau)}(x) \left\{ 1 - \frac{\lambda_{(n,3)}(\tau)}{3!} H_3\left(\frac{\mu_n - x}{\sigma_n}\right) + \right. \\
& \left. \frac{\lambda_{(n,4)}(\tau)}{4!} H_4\left(\frac{\mu_n - x}{\sigma_n}\right) + \cdots + \frac{\lambda_{(n,3)}^2(\tau)}{2!(3!)^2} H_6\left(\frac{\mu_n - x}{\sigma_n}\right) + \cdots \right\} \quad (4.3.25)
\end{aligned}
$$

where $\phi_{(n,\tau)}(x)$ is the density of the normal distribution with mean $\mu_n = K'_n(\tau)$ and variance $\sigma_n^2 = K''_n(\tau)$, and $\lambda_{(n,r)}(\tau) = K_n^{(r)}(\tau)/[K''_n(\tau)]^{r/2}$. This gives a family of expansions indexed by τ, which includes the expansions at the saddlepoint and the origin. This completes the first step of the method.

Compared with Charlier differential series, the family of expansions (4.3.25) is obtained without an explicit developing function $g(x)$. The key difference between the method which led to the family (4.3.25) and that which led to Charlier series is that the method completes a square using the first and second order terms of the Taylor expansion of $K_n(T) - Tx$ and expands only $\exp(R_3)$ before integration, where R_3 is the sum of terms in the Taylor expansion with orders three and higher. An important consequence of this is that the family of expansions obtained are expansions based on the $\lambda_{(n,j)}(\tau)$'s, which usually have certain asymptotic properties with respect to n if the X_n's are asymptotically normal. On the other hand, that for Charlier series calculates the Taylor expansion for $K_f(T) - K_g(T)$ and expands $\exp\{$the entire expansion$\}$ before integration. Explicit expressions of the terms of the expansion are usually difficult to obtain. This makes it difficult to examine the asymptotic properties of these terms.

• *Formal expansions based on the sequence* $\{\lambda_{(n,r)}(\cdot)\}$

We now proceed with the second step of the method stated at the beginning of this section. Expansions (4.3.16) and (4.3.25) are not particularly useful from an asymptotic expansion point of view in that they, like Charlier differential series, do not use information concerning the asymptotic properties of the distribution being expanded, and are not asymptotic expansions. To transform them into asymptotic expansions, we generalize the idea of Edgeworth, i.e., making use of the asymptotic properties of individual terms in the expansions. This involves examining the asymptotic orders of the $\lambda_{(n,r)}(\cdot)$'s, and then rearranging terms in the curly brackets in (4.3.16) and (4.3.25) in ascending order according to the rates at which the $\lambda_{(n,r)}(\cdot)$'s approach zero. The notation for the sequence, $\{\lambda_{(n,r)}(\cdot)\}$, is different from notation commonly used for asymptotic sequences. It may help to reduce the confusion created by this difference to keep in mind that r is the index of the sequence, and that $\{\lambda_{(n,r)}(\cdot)\}$, when written in commonly used notation, is essentially $\{\lambda_r(n)\}$.

Examination of the asymptotic orders of the $\lambda_{(n,r)}(\cdot)$'s, though it involves calculating the

$\lambda_{(n,r)}(\cdot)$'s and may be time consuming, is straightforward and will not be further commented on. In the following, we shall focus on rearranging terms in (4.3.16) and (4.3.25) in ascending order according to the rates at which the $\lambda_{(n,r)}(\cdot)$'s approach zero. In order to illustrate how to do it, we shall make some assumptions concerning the asymptotic orders of the $\lambda_{(n,r)}(\cdot)$'s. We first consider the expansion at the saddlepoint (4.3.16) under certain assumptions, which are true when X_n is a sample mean. We assume that $\{\lambda_{(n,i)}(\hat{T}_n)\}_{i=2}^{\infty}$ form an asymptotic sequence with respect to n; i.e., $\lambda_{(n,i+1)}(\hat{T}_n) = o(\lambda_{(n,i)}(\hat{T}_n))$ as n approaches a certain limit, usually infinity, for $i \geq 2$. In particular, to make it easy to compare this expansion at the saddlepoint with the saddlepoint expansion for the sample mean, we assume that $\lambda_{(n,4)}(\hat{T}_n) = O(\lambda_{(n,3)}^2(\hat{T}_n))$. By arranging the terms in the curly brackets in (4.3.16) in ascending order according to the rates at which they approach zero, we have

$$f_n(x) = g_n(x) \left\{ 1 + \left[\frac{1}{8} \lambda_{(n,4)}(\hat{T}_n) - \frac{5}{24} \lambda_{(n,3)}^2(\hat{T}_n) \right] + o\left(\lambda_{(n,4)}(\hat{T}_n) \right) \right\}. \qquad (4.3.26)$$

This is the *generalized saddlepoint expansion* for $f_n(x)$. With more assumptions concerning the rates at which the $\lambda_{(n,i)}(\hat{T}_n)$'s approach zero, further terms in (4.3.26) may be given.

To discuss the asymptotic properties of other members of the family (4.3.25), we also need to consider the Hermite polynomials that appear in the expansion. If the absolute value of their common argument, $(\mu_n - x)/\sigma_n$, goes to infinity when n goes to infinity, then since H_k is a polynomial of order k, the reciprocals of these polynomials, i.e., $1/H_k$ ($k = 3, 4, \ldots$), will form an asymptotic sequence with respect to n. Thus (4.3.25) contains ratios of terms in two asymptotic sequences, and its asymptotic properties need careful examination. To avoid this complication, we focus on cases for which $(\mu_n - x)/\sigma_n$ is bounded. In these cases, the relative rate at which terms in the curly bracket, such as $\lambda_{(n,3)}(\tau)H_3$ and $\lambda_{(n,4)}(\tau)H_4$, approach zero is determined by that of the $\lambda_{(n,i)}(\tau)$'s. Now assume that $\lambda_{(n,i+1)}(\tau) = o(\lambda_{(n,i)}(\tau))$ for $i \geq 2$. Then we may rewrite (4.3.25) as

$$f_n(x) = e^{K_n(\tau) - \tau x} \phi_{(n,r)}(x) \left\{ 1 - \frac{1}{6} \lambda_{(n,3)}(\tau) H_3 \left(\frac{\mu_n - x}{\sigma_n} \right) + o(\lambda_{(n,3)}(\tau)) \right\}, \qquad (4.3.27)$$

where $\mu_n = K_n'(\tau)$ and $\sigma_n = K_n''(\tau)$. In particular, at $\tau = 0$ we have

$$f_n(x) = \phi_n(x) \left\{ 1 + \frac{1}{6} \lambda_{(n,3)} \left[\left(\frac{x - \mu_n}{\sigma_n} \right)^3 - 3 \left(\frac{x - \mu_n}{\sigma_n} \right) \right] + o(\lambda_{(n,3)}) \right\}. \qquad (4.3.28)$$

We shall refer to (4.3.27) as the *general expansion* for $f_n(x)$, and (4.3.28) as the *generalized Edgeworth expansion*. The assumption that $(x - \mu_n)/\sigma_n$ approaches a finite limit is usually

satisfied when the X_n's are standardized variables, e.g., the standardized mean of n i.i.d. observations. Note that all the assumptions made above are for illustrative purposes only. A $\{X_n\}$ may not satisfy these assumptions. In this case terms in the curly brackets of (4.3.16) and (4.3.25) should be rearranged accordingly, and the general expansions, the generalized saddlepoint and Edgeworth expansions will be different from those shown above.

It is important to keep in mind that the generalized saddlepoint approximation and the general expansions we obtained this way are only *formal* asymptotic expansions for $f_n(x)$. Their validity still needs to be verified. Also, the generalized Edgeworth expansion can be valid for cases where only the first few cumulants exist, e.g., Theorem 4.1 on the Edgeworth expansion in Section 4.2.

A common feature of the expansions discussed above is that the asymptotic sequences involved all have the same first term, 1. Thus the second terms of these sequences measure the asymptotic relative error for approximating $f_n(x)$ with the leading term of the expansion.

● *The generalized saddlepoint and Edgeworth expansions as special cases*

From the perspective of the family of expansions (4.3.27), these two expansions are special cases with the generalized Edgeworth expansion being the one derived at $\tau = 0$ whose leading term is a normal density, and the generalized saddlepoint expansion being the one derived at the saddlepoint for which the leading term simplifies to $\exp\{K_n(\hat{T}_n) - \hat{T}_n x\}$. The latter usually has the smallest relative error for approximating $f_n(x)$ due to the disappearance of $\lambda_{(n,3)}(\hat{T}_n)$. The generalized saddlepoint expansion may represent a collection of expansions in that it will assume a different member of the family for different values of n should the saddlepoint vary with n.

Loosely speaking, the generalized Edgeworth expansion tells us the asymptotic distance between $f_n(x)$ and the normal densities, while the generalized saddlepoint expansion tells us the smallest asymptotic distance between $f_n(x)$ and the family. From this point of view, the generalized saddlepoint expansion is more suitable for approximating $f_n(x)$.

4.4 Expansions for the densities of the sample mean and standardized mean

The general expansions are derived without assumptions on the X_n's beyond that they are asymptotically normal (under which the assumptions on the $\lambda_{(n,r)}$'s made above are usually true), and include expansions at different τ values. It is in this sense that we call them the general expansions. However, such general expansions are meaningful only if they simplify to known results for special cases. We now show for the special cases of the sample mean and the standardized sample mean that the general expansions indeed coincide with known expansions.

● *Expansions for the density of the sample mean*

It may be readily verified that when $X_n = \bar{W}$, the average of n independent copies of a random variable W, the generalized saddlepoint approximation (4.3.11) is the same as the saddlepoint approximation given by Daniels (1954). See (2.2.16). In this case, let $K(T)$ be the cumulant generating function of W, then $K_n(T) = nK(T/n)$. Let $\hat{T}$ be the solution of $K'(\hat{T}) = x$, then $\hat{T}_n = n\hat{T}$. Furthermore, $K_n(\hat{T}_n) = nK(\hat{T})$ and $K_n''(\hat{T}_n) = K''(\hat{T})/n$. Thus the generalized saddlepoint approximation (4.3.11) is the same as Daniels' saddlepoint approximation (2.2.16).

To examine the asymptotic property of the generalized saddlepoint expansion (4.3.26), we first note that $K_n^{(r)}(\hat{T}_n) = K^{(r)}(\hat{T})/n^{r-1}$ for any $r \in N$. It follows that $\lambda_{(n,r)}(\hat{T}_n) = K^{(r)}(\hat{T})/[n^{r/2-1}(K''(\hat{T}))^{r/2}]$ for $r \geq 2$. Denote $K^{(r)}(\hat{T})/(K''(\hat{T}))^{r/2}$ by λ_r. It is not difficult to show that

$$\frac{1}{8}\lambda_{(n,4)}(\hat{T}_n) - \frac{5}{24}\lambda_{(n,3)}^2(\hat{T}_n) = \frac{1}{n}\left(\frac{1}{8}\lambda_4(\hat{T}) - \frac{5}{24}\lambda_3^2(\hat{T})\right)$$

which is a_1 in Daniels' expansion (2.2.23). Further terms in expansion (4.3.26) may be constructed for this particular case and it can be shown that they are equal to the corresponding terms in (2.2.23). Thus the generalized saddlepoint expansion (4.3.26) coincides with Daniels' saddlepoint expansion (2.2.23). Hence it is a valid expansion for $f_n(\bar{x})$.

It may also be easily verified using the same argument demonstrated above that the general expansion of the form (4.3.27) coincides with the expansion Daniels (1954) derived through the Edgeworth expansion at τ. See (4.3) in Section 4 in Daniels (1954). We shall

refer to this (4.3) as D(4.3) [1]. Daniels (1954) stated that the family of expansions given by D(4.3) are asymptotic expansions, and in particular at $\tau = 0$ it reduces to the Edgeworth series for $f_n(\bar{x})$. This, however, is not accurate. The reason is that the Edgeworth expansion for the standardized variable $Z_u = (\bar{U} - E(U))/(\sigma_u/\sqrt{n})$ may not be used to obtain an asymptotic expansion for the density of $\bar{U}$ at anywhere except for $E(U)$ (see the second paragraph on page 48 for a related discussion). The distribution of the random variable U described before D(4.3) has mean $E(U) = K'(\tau) - \bar{x}$, but D(4.3) was derived through the Edgeworth expansion for Z_u at $\bar{U} = 0$, or $Z_u = -[K'(\tau) - \bar{x}][n/K''(\tau)]^{1/2}$. Thus when $K'(\tau) - \bar{x} \neq 0$, expansions given by D(4.3) are not valid. More specifically, the coefficient A_1 in D(4.3), for example, is in general $O(n^{3/2})$. Thus the second term in D(4.3), $A_1/\sqrt{n}$, is in general $O(n)$. Hence D(4.3) cannot even be an asymptotic expansion in a formal sense. This illustrates the importance of the assumption that $(\mu_n - x)/\sigma_n$ approaches a finite limit as n approaches infinity that we made in arriving at (4.3.27).

To conclude, for the case of sample mean only one member of the family, i.e., the generalized saddlepoint expansion, is a valid asymptotic expansion for the density.

● *Expansions for the density of the standardized mean*

(i) The Edgeworth expansion and the generalized Edgeworth expansion

We now show that (4.3.28) coincides with the Edgeworth expansion (4.2.10). Let

$$X_n = \frac{\bar{W} - \mu}{\sigma/\sqrt{n}}, \tag{4.4.29}$$

where μ and σ are the mean and standard deviation of W. The cumulant generating function of X_n, K_n, is given by

$$K_n(T) = -\frac{\sqrt{n}\mu}{\sigma}T + nK\left(\frac{T}{\sigma\sqrt{n}}\right). \tag{4.4.30}$$

Denote the cumulants of W by κ_i ($i = 1, 2, \ldots$). Then for any finite T, $K_n(T)$ has a Taylor expansion, provided that n is sufficiently large, i.e.,

$$K_n(T) = \frac{T^2}{2!} + \frac{T^3}{3!}\frac{\lambda_3(W)}{\sqrt{n}} + \frac{T^4}{4!}\frac{\lambda_4(W)}{n} + \cdots, \tag{4.4.31}$$

---

[1]There is an error in the expressions for A_1 and A_2 in D(4.3). There should be a negative sign in front of each $[K'(\tau) - \bar{x}]$.

where $\lambda_r(W) = \kappa_r/\sigma^r$ are the standardized cumulants for W. It is not difficult to see from
(4.4.29) and (4.4.31) that $\mu_n = 0$, $\sigma_n = 1$, $K_n^{(3)}(0) = \lambda_3(W)/\sqrt{n} = \mu_3(W)/(\sigma^3\sqrt{n})$, and
$K_n^{(i+1)}(0)/K_n^{(i)}(0) = O(1/\sqrt{n})$ for $i > 2$, where $\mu_3(W)$ is the third moment and cumulant of
W. It follows that $\lambda_{(n,r)} = K_n^{(r)}(0)$ for $(r \geq 3)$, and (4.3.28) may be written as

$$f_n(x) = \phi(x)\left\{1 + \frac{1}{\sqrt{n}}\frac{\mu_3(W)}{6\sigma^3}(x^3 - 3x) + o(1/\sqrt{n})\right\},$$

which coincides with the two-term Edgeworth expansion for $f_n(x)$.

The full Edgeworth expansion for the standardized mean is obtained by using the follow-
ing version of the relation (4.4.31) which expresses the characteristic function of X_n, $\psi_n(t)$,
as

$$\psi_n(t) = \exp\left\{\frac{(it)^3}{3!}\frac{\lambda_3(W)}{\sqrt{n}} + \frac{(it)^4}{4!}\frac{\lambda_4(W)}{n} + \frac{(it)^5}{5!}\frac{\lambda_5(W)}{n^{3/2}} + \cdots\right\}e^{-t^2/2}. \qquad (4.4.32)$$

Expanding the $\exp(\cdot)$ term in (4.4.32) we obtain

$$\psi_n(t) = \left\{1 + \frac{(it)^3}{3!}\frac{\lambda_3(W)}{\sqrt{n}} + \frac{(it)^4}{4!}\frac{\lambda_4(W)}{n} + \cdots + \frac{1}{2!}\frac{(it)^6}{(3!)^2}\frac{(\lambda_3(W))^2}{n} + \cdots\right\}e^{-t^2/2}. \qquad (4.4.33)$$

Since $(it)^k e^{-t^2/2}$ has Fourier inversion $H_k(x)\phi(x)$ (see Appendix B for further discussion on
$H_k(x)$ and $\phi(x)$), then Fourier inversion of (4.4.33) gives

$$\begin{aligned}
f_n(x) &= \phi(x)\left\{1 + \frac{1}{3!}\frac{\lambda_3(W)}{\sqrt{n}}H_3(x) + \right. \\
&\qquad \left. \frac{1}{4!}\frac{\lambda_4(W)}{n}H_4(x) + \cdots + \frac{1}{2!(3!)^2}\frac{(\lambda_3(W))^2}{n}H_6(x) + \cdots\right\}. \qquad (4.4.34)
\end{aligned}$$

The full Edgeworth expansion is then obtained by rearranging terms in (4.4.34) in ascending
powers of $1/\sqrt{n}$.

The right-hand side of (4.3.28) may also be further expanded with respect to the $\lambda_{(n,k)}$'s.
It is not difficult to show that the coefficient for, say $\lambda_{(n,k)}^2$ is $H_{2k}/[2!(k!)^2]$, and so on. Since
$\lambda_{(n,k)} = \lambda_k(W)/(\sqrt{n})^{(k-2)}$ for $k = 3, 4, \ldots$, (4.3.24) coincides with (4.4.34). The further
expanded (4.3.28) is obtained by rearranging terms in (4.3.24) in ascending powers of $1/\sqrt{n}$,
and is thus the same as the Edgeworth expansion.

(ii) The generalized saddlepoint expansion and saddlepoints expansion

Recall from Chapter 3 that the saddlepoints expansion for the density of a standardized mean

was derived through the saddlepoint approximation for the mean, and its relative error is of the same order as that of the saddlepoint expansion provided the saddlepoint expansion is uniformly valid in some neighborhood of the mean. We now show that the saddlepoints expansion is actually the generalized saddlepoint expansion for the standardized mean.

For a fixed x, $\hat{T}_n$ is the solution of equation $K'_n(T) = x$, i.e., from (4.4.30),

$$-\frac{\sqrt{n}\mu}{\sigma} + \frac{\sqrt{n}}{\sigma}K'\left(\frac{T}{\sigma\sqrt{n}}\right) = x.$$

Since $x = (\bar{w} - \mu)\sqrt{n}/\sigma$, the above equation is equivalent to

$$K'\left(\frac{T}{\sigma\sqrt{n}}\right) = \bar{w}. \tag{4.4.35}$$

Let t_n be the solution to $K'(T) = \bar{w}$, then by (4.4.35), $t_n = \hat{T}_n/(\sigma\sqrt{n})$, and $\hat{T}_n = t_n\sigma\sqrt{n}$. Thus from (4.4.30) in terms of t_n, $K_n(\hat{T}_n)$ is

$$K_n(\hat{T}_n) = -\frac{\sqrt{n}\mu}{\sigma}t_n\sigma\sqrt{n} + nK(t_n) = nK(t_n) - n\mu t_n. \tag{4.4.36}$$

Furthermore,

$$K'_n(\hat{T}_n)\hat{T}_n = \left\{-\frac{\sqrt{n}\mu}{\sigma} + \frac{\sqrt{n}}{\sigma}K'(t_n)\right\}t_n\sigma\sqrt{n} = -n\mu t_n + nt_n K'(t_n), \tag{4.4.37}$$

and

$$K''_n(\hat{T}_n) = \frac{1}{\sigma^2}K''(t_n). \tag{4.4.38}$$

It follows from (4.4.36) and (4.4.37) that

$$K_n(\hat{T}_n) - K'_n(\hat{T}_n)\hat{T}_n = n[K(t_n) - t_n K'(t_n)]. \tag{4.4.39}$$

Equations (4.4.38) and (4.4.39) imply that the generalized saddlepoint approximation given by (4.3.11) is the same as the saddlepoints approximation (3.5.59), for which $\hat{T}_n$ and $\bar{x}(z, n)$ correspond to t_n and $\bar{w}$ in the present discussion.

By repeating the procedure we followed to show the equivalence of the generalized saddlepoint expansion and the saddlepoint expansion, one can show that the entire generalized saddlepoint expansion (4.3.26) coincides with the saddlepoints expansion (3.5.57).

(iii) Validity of other members in the family

Recall that for the sample mean only the generalized saddlepoint approximation is valid. For the standardized mean, however, besides the generalized saddlepoint approximation and the generalized Edgeworth expansion, the following one-term expansion given by (4.3.27),

$$f_n(x) = e^{K_n(\tau) - \tau x} \phi_{(n,\tau)}(x) \left\{ 1 + O(\lambda_{(n,3)}(\tau)) \right\},$$ (4.4.40)

is also valid. Although in this case the $O(\lambda_{(n,3)}(\tau))$ term in (4.4.40) and $O(\lambda_{(n,4)}(\tau))$ term in (4.3.27) may be easily further expanded, verification of the asymptotic validity of expansions with more terms than that in (4.4.40) is difficult. That is why we only consider (4.4.40) for which the validity of the family can be established through that of the Edgeworth expansion.

We now prove the validity of (4.4.40). The following equation will be used repeatedly in the proof:

$$\left[1 + O\left(\frac{1}{n^p}\right) \right]^q = 1 + O\left(\frac{1}{n^p}\right),$$ (4.4.41)

where $p, q \in R$ and $p > 0$. This follows directly from the fact that for $|x| < 1$, the function $(1 + x)^q$ has Taylor expansion

$$1 + c_1 x + c_2(x) x^2,$$

where c_1 is a constant and $c_2(x)$ is a function bounded in absolute value. By (4.4.31) we obtain the following asymptotic approximations for the derivatives of $K_n(T)$: (1) $K_n(T) = T^2/2 + O(n^{-1/2})$, (2) $K_n'(T) = T + O(n^{-1/2})$, (3) $K_n''(T) = 1 + O(n^{-1/2})$, and (4) $K_n^{(r)}(T) = O(1/n^{r/2-1})$ for $r \geq 3$. Denote the leading term of the expansion (4.4.40) by $l_{(n,\tau)}(x)$. We have

$$l_{(n,\tau)}(x) = \frac{1}{\sqrt{2\pi K_n''(\tau)}} \exp\{K_n(\tau) - \tau x - \frac{(x - K_n'(\tau))^2}{2K_n''(\tau)}\}.$$

Then (1), (2), (3) and (4.4.41) lead to

$$
\begin{aligned}
l_{(n,\tau)}(x) &= \frac{\exp\left\{ \frac{1}{2}\tau^2 + O(n^{-1/2}) - \tau x - \frac{[x - \tau + O(n^{-1/2})]^2}{2[1 + O(n^{-1/2})]} \right\}}{\sqrt{2\pi[1 + O(n^{-1/2})]}} \\
&= \frac{1 + O(n^{-1/2})}{\sqrt{2\pi}} \exp\left\{ \frac{-x^2/2 + O(n^{-1/2})}{1 + O(n^{-1/2})} \right\} \\
&= \phi(x) e^{O(n^{-1/2})} \left[1 + O(n^{-1/2}) \right] \\
&= \phi(x) \left[1 + O(n^{-1/2}) \right].
\end{aligned}
$$ (4.4.42)

Also, (3) and (4) imply that $\lambda_{(n,3)}(\tau) = O(n^{-1/2})$. Thus (4.4.42) may be written as

$$l_{(n,\tau)}(x) = \phi(x) \left\{ 1 + O(\lambda_{(n,3)}(\tau)) \right\}.$$ (4.4.43)

By the Edgeworth expansion, $f_n(x) = \phi(x)\left[1 + O(n^{-1/2})\right]$. Thus

$$
\begin{aligned}
\frac{f_n(x)}{l_{(n,\tau)}(x)} &= \frac{\phi(x)\left[1 + O(n^{-1/2})\right]}{\phi(x)\left[1 + O(n^{-1/2})\right]} \\
&= 1 + O(n^{-1/2}) \text{ or } 1 + O(\lambda_{(n,3)}(\tau)).
\end{aligned} \tag{4.4.44}
$$

This proves the validity of (4.4.40).

The relative accuracy of $l_{(n,\tau)}(x)$ and one-term Edgeworth expansion, i.e., $\phi(x)$, is not clear from (4.4.40). But we expect that when τ is close to the saddlepoint, $l_{(n,\tau)}(x)$ will be more accurate.

● *Some remarks*

To conclude this section, we note that the seemingly very different expansions, such as the Edgeworth expansion for the standardized mean and the saddlepoint expansion for the mean are in fact special cases of the same general expansion (4.3.27). This is, perhaps, not unexpected since the general expansion was derived at an arbitrary τ without using specific information on the form of the cumulant generating function. The general expansion (4.3.27) would not be of much interest to us if its only use is as a unified treatment for the known expansions. However, it also identifies the basic asymptotic quantity, i.e., $\{\lambda_{(n,r)}(\tau)\}$, based on which asymptotic expansions for distributions may be found, and has thus generalized the Edgeworth and saddlepoint expansions for which $\{\lambda_{(n,r)}(\tau)\}$ happen to be, essentially, power series $\{n^{-1/2}\}$ and $\{n^{-1}\}$, respectively. Such generalization provides formal expansions (4.3.27) to be rigorously examined which may turn out to be valid asymptotic expansions. The saddlepoints expansion and the expansions given by (4.4.40), for example, were first derived through the general expansion formula as formal expansions. They both turned out indeed to be valid expansions, although neither is a power series expansion.

4.5 Saddlepoint expansion for Jørgensen's exponential dispersion model

In this section we examine another case, i.e., Jørgensen's exponential dispersion model, for which the generalized saddlepoint expansion may be expressed as a power series. The asymptotics for such a model are with respect to a large parameter, which is in general not a sample size and is continuous. To be consistent with notation commonly used for this model, λ will be used, for this section only, to denote the large parameter instead of n. Also, τ and σ will be used for different purposes and this will be made clear in the discussion.

Jørgensen's (1987) exponential dispersion model for a one-dimensional continuous random variable Y is defined by the probability density function,

$$p(y; \theta, \lambda) = a(\lambda, y) \exp[\lambda\{y\theta - \kappa(\theta)\}], \quad y \in R \qquad (4.5.45)$$

for suitable functions a and κ, where $\lambda \in \Lambda \subseteq R_+$, $1 \in \Lambda$ and $\theta \in \Theta \subseteq R$. This model is an important contribution to the theory of exponential families and has wide application in generalized linear regression analysis. Its moment generating function is

$$M(t; \theta, \lambda) = \exp[\lambda\{\kappa(\theta + t/\lambda) - \kappa(\theta)\}].$$

The cumulant generating function, $K(t; \theta, \lambda)$, is thus $\lambda\{\kappa(\theta + t/\lambda) - \kappa(\theta)\}$. By repeatedly differentiating $K(t; \theta, \lambda)$, one can obtain the $K^{(r)}(t; \theta, \lambda)$'s. In particular, the ith cumulant, $\kappa_i(\theta, \lambda)$, for Y is

$$\kappa_i(\theta, \lambda) = \kappa^{(i)}(\theta)\lambda^{1-i} \quad i = 1, 2, \ldots$$

Define $\tau(\theta) = \kappa'(\theta)$ and $V(\mu) = \kappa''(\tau^{-1}(\mu))$. Then the expectation of Y, μ, is given by

$$\mu = \tau(\theta),$$

and the variance is given by

$$\sigma^2 V(\mu),$$

where V is called the variance function for the model and $\sigma^2 = 1/\lambda$ is the dispersion parameter. The set $\Omega = \tau(\text{int}\Theta)$ is called the mean domain. The smallest closed convex subset of R with measure 1 is called the convex support. When the mean domain coincides

with the convex support for all values of λ, the model is said to be *steep*. In this section, we only work with y values in the mean domain of the model. When the model is steep such a restriction is unnecessary. It is, however, essential when the model is nonsteep since a y value outside the mean domain does not have a corresponding saddlepoint. For details on steepness and other properties of the model, see Jørgensen (1992).

The model is closed with respect to λ under addition, and this leads to the asymptotic normality of the model in the sense that

$$\frac{Y - \mu}{\sigma} \xrightarrow{d} N(0, V(\mu)) \quad \text{as} \quad \sigma \to 0. \tag{4.5.46}$$

Because of its asymptotic normality (4.5.46), asymptotics concerning the behavior of the model as σ or $1/\lambda$ approaches zero fit readily into the framework of the general expansion. The deviance, $d(y, \mu)$, for a single observation from the model is

$$d(y, \mu) = 2[y\tilde{\theta} - \kappa(\tilde{\theta}) - \{y\tau^{-1}(\mu) - \kappa(\tau^{-1}(\mu))\}],$$

where $\tilde{\theta}$ is the value of θ that maximizes $y\theta - \kappa(\theta)$. Jørgensen (1992, Theorem 3.3.1) introduced the saddlepoint approximation for the density (4.5.45),

$$p(y; \theta, \lambda) \simeq \sqrt{\frac{1}{2\pi\sigma^2 V(y)}} \exp\{-d(y, \mu)/(2\sigma^2)\}, \tag{4.5.47}$$

and demonstrated that the approximation converges to $p(y; \theta, \lambda)$ as σ approaches zero.

In this section, we look at the entire saddlepoint expansion for the model. Specifically, we give the following extended version of Theorem 3.3.1 in Jørgensen (1992).

Theorem 4.2: *Let $g_{(\lambda,\theta)}(y)$ be the saddlepoint approximation given by (4.5.47), and $\tilde{\theta}$ be such that $\tau(\tilde{\theta}) = y$. Then for a fixed θ,*

$$p(y; \theta, \lambda) = g_{(\lambda,\theta)}(y) \left\{ 1 + \frac{a_1}{\lambda} + \cdots + \frac{a_r}{\lambda^r} + O\left(\frac{1}{\lambda^{r+1}}\right) \right\}, \tag{4.5.48}$$

where the a_i's are functions of y for any $r \in N$. The expansion (4.5.48) is uniform in y in any compact subset if $p(y; \theta, 1)$ satisfies the following conditions: (1) it is piecewise continuous in any finite interval, (2) it is bounded outside a finite interval, and (3) it belongs to $L^q(R)$ for some $q > 1$.

The three conditions for the uniform validity are the generalized versions of conditions in

Theorem 3.3, which we had commented on near the end of Section 3.4. We now give a simple proof for the above theorem. By inverting the moment generating function we obtain

$$
\begin{aligned}
p(y; \theta, \lambda) &= \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{\lambda\{\kappa(\theta+it/\lambda)-\kappa(\theta)\}-ity} dt \\
&= \frac{\lambda}{2\pi} \int_{-\infty}^{\infty} e^{\lambda[\kappa(\theta+is)-\kappa(\theta)-isy]} ds \\
&= \frac{\lambda}{2\pi} \int_{-\infty}^{\infty} e^{\lambda[K(s;\theta,1)-isy]} ds,
\end{aligned}
\tag{4.5.49}
$$

where we have used a change of variable, $s = t/\lambda$.

We first consider the special case where λ only assumes integer values. In this case, (4.5.49) implies that the density $p(y; \theta, \lambda)$ is that of the mean of λ i.i.d. observations from the model $p(y; \theta, 1)$. For this case, the pointwise validity and uniform validity of the saddlepoint expansion as defined by Daniels (1954) have been proved in Chapters 2 and 3. Thus the theorem will be proved upon showing that the saddlepoint approximation given by (4.5.47) coincides with that given by Daniels. Written in the present notation, Daniels' saddlepoint approximation is

$$
\sqrt{\frac{\lambda}{2\pi K''(s_0, \theta, 1)}} e^{\lambda[K(s_0,\theta,1)-s_0 y]},
\tag{4.5.50}
$$

where s_0 is the saddlepoint satisfying: $K'(s_0; \theta, 1) = y$. To see (4.5.47) is indeed (4.5.50), note that $K(s; \theta, 1) = \kappa(\theta+s) - \kappa(\theta)$. Thus $K'(s; \theta, 1) = \kappa'(\theta+s)$, and $K''(s; \theta, 1) = \kappa''(\theta+s)$. The saddlepoint s_0 satisfies $\kappa'(\theta + s) = y$, and thus the equation $\tilde{\theta} = \theta + s_0$. It follows that

$$
\begin{aligned}
K(s_0, \theta, 1) - s_0 y &= \kappa(\theta + s_0) - \kappa(\theta) - s_0 y \\
&= \kappa(\tilde{\theta}) - \kappa(\theta) - (\tilde{\theta} - \theta)y \\
&= -d(y, \mu)/2,
\end{aligned}
$$

and $K''(s_0, \theta, 1) = \kappa''(\theta + s_0) = \kappa''(\tilde{\theta}) = V(y)$. The theorem has thus been proved for integer values of λ.

Note that (4.5.47) and (4.5.50) are identical whether or not λ is an integer. In general, λ is not necessarily an integer. The theorem follows from the equivalence of (4.5.47) and (4.5.50) and the observation that the method of steepest descents, which was used by Daniels (1954) to show the pointwise validity of the saddlepoint expansion, does not use the information that n is an integer and may be applied to (4.5.49) to derive (4.5.48). The derivation is

literally identical to that shown by Daniels (1954) except that the large parameter in this case is λ instead of n. Consequently, the a_r's in (4.5.48) are the same as the a_r's in Daniels' saddlepoint expansion (2.2.23). As to the uniform validity of (4.5.48) on a compact set, again the proof for the uniform validity of Daniels' saddlepoint expansion that we demonstrated in Chapter 3 does not use the information that n is an integer, and is thus applicable to Theorem 4.2 as well.

Expansion (4.5.48) is the generalized saddlepoint expansion for the model. This can be readily confirmed using similar arguments to those shown in the previous section. For the standardized variable $(Y - \mu)/\sigma$, we also can show that the generalized Edgeworth expansion for both the density and the cumulative distribution function are valid. This indicates that the density of the standardized variable approaches that of the standard normal density with a relative error diminishing at the rate of $O(\sigma)$. Furthermore, the cumulative distribution function of the model has asymptotic expansions such as Lugannani and Rice's expansion and other expansions discussed in Daniels (1987). The general expansions given by (4.4.40) are all valid.

Jørgensen (1992, Theorem 3.3.3) also discussed saddlepoint approximation for a class of dispersion models. Here we consider the following modified version of this theorem, which gives the order of the relative error of the approximation.

Theorem 4.3: *Consider the dispersion model for the random variable Y of the form*

$$p(y; \mu, \sigma^2) = a(\sigma^{-2}) \exp\{t(y - \mu)/\sigma^2\}, \quad y \in R \qquad (4.5.51)$$

where function t satisfies (i) $t(y)$ has a global maximum at $y = 0$ in the sense that for any bounded neighborhood ω of 0, $\sup\{t(y) : y \notin \omega\} < t(0)$, and (ii) there exists a neighborhood of $y = 0$, δ, in which $t''(y)$ exists and is continuous. Then

$$p(y; \mu, \sigma^2) = g_{(\mu,\sigma)}(y) \left\{1 + O(\sigma^2)\right\} \quad as \quad \sigma \to 0, \qquad (4.5.52)$$

and

$$X = (Y - \mu)/\sigma \xrightarrow{d} N(0, V) \quad for \quad \sigma \to 0. \qquad (4.5.53)$$

The leading term, $g_{(\mu,\sigma)}(y)$, in (4.5.52) is the saddlepoint approximation,

$$g_{(\mu,\sigma)}(y) = (2\pi\sigma^2 V)^{-1/2} \exp\{-d(y, \mu)/(2\sigma^2)\}, \qquad (4.5.54)$$

where $d(y, \mu) = 2\{t(0) - t(y - \mu)\}$ is the deviance, and $V = -1/t''(0)$.

Proof: (a) We now use Laplace's method to prove (4.5.52). The key step is to obtain an asymptotic expansion for the quantity $a(\sigma^{-2})e^{t(0)/\sigma^2}$. Although function a is not specified, such an expansion is possible due to the fact that p is a density function. We have

$$\begin{aligned}
1 &= \int_{-\infty}^{\infty} p(y; 0, \lambda^{-1})dy \\
&= \int_{-\infty}^{\infty} a(\lambda)e^{\lambda t(y)}dy \\
&= a(\lambda)\left\{\int_{y \in \omega} e^{\lambda t(y)}dy + \int_{y \notin \omega} e^{\lambda t(y)}dy\right\},
\end{aligned} \qquad (4.5.55)$$

where ω is chosen to be a bounded neighborhood of 0. By condition (i) there exists a positive constant $\rho < 1$ such that $\exp\{t(y) - t(0)\} < \rho$ for $y \notin \omega$. The asymptotic order of the second integral in the curly brackets can thus be evaluated as the following:

$$\begin{aligned}
\int_{y \notin \omega} e^{\lambda t(y)}dy &= e^{\lambda t(0)}\left[e^{-t(0)}\int_{y \notin \omega} e^{(\lambda-1)(t(y)-t(0))}e^{t(y)}dy\right] \\
&\leq e^{\lambda t(0)}\left[e^{-t(0)}\rho^{\lambda-1}\int_{y \notin \omega} e^{t(y)}dy\right].
\end{aligned}$$

Owing to the existence of $\int e^{t(y)}dy$, this integral is $e^{\lambda t(0)}O(\rho^\lambda)$. By Laplace's method, the first integral is

$$\sqrt{\frac{2\pi}{-\lambda t''(0)}}e^{\lambda t(0)}\left\{1 + O\left(\frac{1}{\lambda}\right)\right\}.$$

Since ρ^λ converges to zero faster than any power of $1/\lambda$, the second integral is asymptotically negligible. Thus equation (4.5.55) becomes

$$1 = a(\lambda)\sqrt{\frac{2\pi}{-\lambda t''(0)}}e^{\lambda t(0)}\left\{1 + O\left(\frac{1}{\lambda}\right)\right\}.$$

It follows that

$$(2\pi\sigma^2 V)^{-1/2} = a(\sigma^{-2})e^{t(0)/\sigma^2}\left\{1 + O(\sigma^2)\right\},$$

or equivalently,

$$a(\sigma^{-2})e^{t(0)/\sigma^2} = (2\pi\sigma^2 V)^{-1/2}\left\{1 + O(\sigma^2)\right\}. \qquad (4.5.56)$$

The density function (4.5.51) is now expressed by using (4.5.56) as

$$p(y; \mu, \sigma^2) = a(\sigma^{-2})\exp\{t(y - \mu)/\sigma^2\}$$

$$= a(\sigma^{-2})e^{t(0)/\sigma^2}\exp\{-t(0)/\sigma^2 + t(y-\mu)/\sigma^2\}$$

$$= (2\pi\sigma^2 V)^{-1/2}\exp\{-d(y,\mu)/(2\sigma^2)\}\left\{1 + O(\sigma^2)\right\}, \qquad (4.5.57)$$

which proves (4.5.52).

(b) To prove the asymptotic normality as stated in (4.5.53), let the density of X be $q(x;\mu,\sigma^2)$. Then

$$q(x;\mu,\sigma^2) = p((\sigma x + \mu),\mu,\sigma^2)\frac{dy}{dx} = \sigma a(\sigma^{-2})e^{t(\sigma x)/\sigma^2}. \qquad (4.5.58)$$

The distribution function of X, $Q(x;\mu,\sigma^2)$, is thus given by

$$Q(x;\mu,\sigma^2) = \int_{-\infty}^{x}\sigma a(\sigma^{-2})e^{t(\sigma v)/\sigma^2}dv.$$

We now prove (4.5.53) by showing that $\lim_{\sigma\to 0} Q(x;\mu,\sigma^2) = \Phi(x/\sqrt{V})$. Using (4.5.56), we obtain the following expansion of $Q(x;\mu,\sigma^2)$

$$Q(x;\mu,\sigma^2) = (2\pi V)^{-1/2}\int_{-\infty}^{x}e^{[t(\sigma v)-t(0)]/\sigma^2}dv\left\{1 + O(\sigma^2)\right\}. \qquad (4.5.59)$$

Consider the contribution to the integral in (4.5.59) from the interval $(-\infty, b/\sigma)$, where $b < 0$ and $b \in \delta$, the neighborhood of zero in which $t(y)$ is twice continuously differentiable. We have

$$\int_{-\infty}^{b/\sigma}e^{[t(\sigma v)-t(0)]/\sigma^2}dv = \frac{1}{\sigma}\int_{-\infty}^{b}e^{[t(v)-t(0)]/\sigma^2}dv.$$

Choose b close enough to zero such that the function $t(v) - t(0)$ has a global maximum at $v = b$. Since it has a continuous second derivative in a neighborhood of b, by Laplace's method, we obtain

$$\int_{-\infty}^{b/\sigma}e^{[t(\sigma v)-t(0)]/\sigma^2}dv = \frac{1}{\sigma}e^{[t(b)-t(0)]/\sigma^2}\left\{\frac{\sigma^2}{t'(b)} + O(\sigma^4)\right\}. \qquad (4.5.60)$$

Thus for the chosen b the contribution from $(-\infty, b/\sigma)$ converges to zero faster than any power of σ. It follows from (4.5.59) and (4.5.60) that the asymptotic normality is established if we can show that

$$\lim_{\sigma\to 0}\int_{b/\sigma}^{x}e^{[t(\sigma v)-t(0)]/\sigma^2}dv = \int_{-\infty}^{x}e^{t''(0)v^2/2}dv. \qquad (4.5.61)$$

Let σ be small enough so that $b < \sigma x$ and $(b, \sigma x) \subset \delta$. Then the function $t(\sigma v) - t(0)$ has a second order Taylor expansion for all $\sigma v \in (b, \sigma x)$, or equivalently for all $v \in (b/\sigma, x)$. That is $t(\sigma v) - t(0) = \sigma^2 t''(\xi(v)) v^2/2$ where $|\xi(v)| < |\sigma v|$ for all $v \in (b/\sigma, x)$. Thus

$$\int_{b/\sigma}^x e^{[t(\sigma v) - t(0)]/\sigma^2} dv = \int_{b/\sigma}^x e^{t''(\xi(v)) v^2/2} dv.$$

Since $|\xi(v)| < |\sigma v| < \max\{|b|, |\sigma x|\}$, we are able to choose b close enough to zero and let σ be small enough that so that $|\xi(v)|$ is very close to zero for all v. Consequently, $t''(\xi(v))$ is then very close to $t''(0)$. We now use these observations to prove (4.5.61).

Let $u = \sup\{t''(y) : y \in \delta\}$. Since $t''(0) < 0$ and $t''(y)$ is continuous in δ, we may assume that $-\infty < u < 0$. For any given $\epsilon > 0$, let c be such that

$$\int_{-\infty}^c e^{u v^2/2} dv < \epsilon/4.$$

It follows that if $\xi(v) \in \delta$, then

$$\int_{b/\sigma}^c e^{t''(\xi(v)) v^2/2} dv < \int_{b/\sigma}^c e^{u v^2/2} dv < \epsilon/4, \quad \text{and} \quad \int_{-\infty}^c e^{t''(0) v^2/2} dv < \epsilon/4.$$

Due to the continuity of $t''(y)$ in δ and the fact that $|\xi(v)| < \max\{|b|, |\sigma x|\}$, we can choose b and σ_0 such that for $\sigma < \sigma_0$, not only $\xi(v) \in \delta$, but also to be so close to 0 that

$$\left| \int_c^x e^{t''(\xi(v)) v^2/2} dv - \int_c^x e^{t''(0) v^2/2} dv \right| < \epsilon/2.$$

It follows that

$$
\begin{aligned}
\left| \int_{b/\sigma}^x e^{[t(\sigma v) - t(0)]/\sigma^2} dv - \int_{-\infty}^x e^{t''(0) v^2/2} dv \right| &= \left| \int_{b/\sigma}^x e^{t''(\xi(v)) v^2/2} dv - \int_{-\infty}^x e^{t''(0) v^2/2} dv \right| \\
&\leq \left| \int_c^x e^{t''(\xi(v)) v^2/2} dv - \int_c^x e^{t''(0) v^2/2} dv \right| + \\
&\quad \left| \int_{b/\sigma}^c e^{t''(\xi(v)) v^2/2} dv - \int_{-\infty}^c e^{t''(0) v^2/2} dv \right| \\
&\leq \epsilon/2 + \epsilon/4 + \epsilon/4 = \epsilon. \qquad (4.5.62)
\end{aligned}
$$

This proves (4.5.61). It follows that $\lim_{\sigma \to 0} Q(x; \mu, \sigma) = \Phi(x/\sqrt{V})$. #

In a private communication, Jørgensen pointed out that it may be possible to prove the asymptotic normality by using the following result: if p_n and p are density functions and $p_n \to p$, then $F_n \to F$, where F_n and F are the distribution functions corresponding

to p_n and p, respectively. We now show that this result provides a simpler proof than the one given above. To utilize this result, we first derive an asymptotic expansion for the density of X. Recall that the saddlepoint expansion for the mean was used to obtain the saddlepoints expansion for the standardized mean in Chapter 3. Here we can use the saddlepoint expansion for the density of Y to obtain a saddlepoint expansion for that of X. Specifically, by (4.5.56) and (4.5.58) we have

$$q(x; \mu, \sigma^2) = (2\pi V)^{-1/2} e^{[t(\sigma x) - t(0)]/\sigma^2} \left\{ 1 + O(\sigma^2) \right\}. \tag{4.5.63}$$

Let σ be small enough so that $\sigma x \in \delta$. Then $t(\sigma x) - t(0)$ has a Taylor expansion

$$t(\sigma x) - t(0) = \sigma^2 t''(\xi(x)) x^2 / 2,$$

where $|\xi(x)| < \sigma x$. Since $\xi(x) \to 0$ and thus $t''(\xi(x)) \to t''(0)$ as $\sigma \to 0$, it follows from (4.5.63) that

$$q(x; \mu, \sigma^2) \to (2\pi V)^{-1/2} e^{-x^2/(2V)} \quad \text{as} \quad \sigma \to 0.$$

The result stated above then implies (4.5.53). To conclude this section, we note that the above two proofs for the asymptotic normality are rather qualitative in the sense that they do not give the order in which the distribution function of X converges to that of $N(0, V)$. Should the analytic expression of $t(y)$ be available, this order may be derived by looking into the order in which the left-hand side of (4.5.61) approaches its right-hand side.

4.6 General expansions for distribution functions

A method similar to that discussed in Section 4.2 may be devised to derive expansions for the cumulative distribution function of X_n, $F_n(x)$. The base of the method is again an inversion formula for the distribution,

$$Q_n(x) = \frac{1}{2\pi i} \int_{\tau - i\infty}^{\tau + i\infty} e^{K_n(T) - Tx} \frac{dT}{T}, \tag{4.6.64}$$

where $Q_n(x)$ is the tail probability and $\tau > 0$. The expansions, however, become quite complicated due to the extra factor $1/T$ in the integrand and the restriction that $\tau > 0$, and come in many different forms. Following the derivation of the expansion at the origin in Section 4.2, suppose that we have obtained the Taylor expansion of $K_n(T) - Tx$ and so on,

and are about to perform the term-by-term integration similar to that shown in (4.3.21). Two different families of expansions may be derived by expanding $1/T$ before the term-by-term integration and retaining it. Yet another family of expansions may be obtained by simply integrating the general expansions given by (4.3.27), which for the case of the sample mean includes Lugannani and Rice's expansion for the cumulative distribution function (see Chapter 3). The relationship amongst these families have yet to be studied in detail, but it can be shown that for the simple case of the sample mean, they are all different. Thus it appears that there does not exist a single formula such as (4.3.27) for expansions for densities that would include all known expansions for the cumulative distribution function as special cases. We shall therefore not discuss general expansions for the cumulative distribution function further. Instead, we look at the generalization of two frequently used formulas for the standardized mean and sample mean.

- *The generalized Edgeworth expansion*

Consider the case where $\mu_n = 0$ and $\sigma_n = 1$. By formally integrating (4.3.28), we obtain the generalized Edgeworth expansion for a cumulative distribution function

$$F_n(x) = \Phi(x) + \frac{1}{6}\lambda_{(n,3)}(1 - x^2)\phi(x) + o(\lambda_{(n,3)}). \qquad (4.6.65)$$

It may be easily verified that (4.6.65) is the same as the Edgeworth expansion for a cumulative distribution function when X_n is the standardized mean. Also, (4.6.65) can be valid when X_n is not a standardized mean. Here we briefly consider the expansion for the cumulative distribution function of a U-statistic of degree 2. This statistic is defined by

$$U_n = \frac{2}{n(n-1)} \sum_{1 \le i < j \le n} h(X_i, X_j), \qquad (4.6.66)$$

where the X_i's are i.i.d. and h is a symmetric function of two variables with $E[h(X_1, X_2)] = 0$ and $E[h^2(X_1, X_2)] < \infty$. The asymptotic normality of U_n was proved by Hoeffding in 1948. Let σ_n be the standard deviation of U_n and $F_n(x)$ be the cumulative distribution function of U_n/σ_n, then under certain conditions Bickel, Götze and van Zwet (1986) showed that

$$F_n(x) = \Phi(x) + \frac{1}{6}\frac{\omega_3}{\sqrt{n}}(1 - x^2)\phi(x) + o(n^{-1/2}), \qquad (4.6.67)$$

where $\omega_3/\sqrt{n}$ is an approximation with error $o(1/n)$ to the third cumulant of U_n/σ_n, $\lambda_{(n,3)}$. With $X_n = U_n/\sigma_n$, (4.6.67) then implies that (4.6.65) is indeed valid. Furthermore, it can

be shown that the fourth cumulant of U_n/σ_n, $\lambda_{(n,4)}$, satisfies $\lambda_{(n,4)} = O(1/n)$. The right-hand side of (4.6.67) and thus that of (4.6.65) can be further expanded. The expansion in (4.6.67) is simpler than that in (4.6.65) in that it is defined in terms of a simpler asymptotic sequence $\{1/\sqrt{n}\}$ while (4.6.65) is defined in terms of $\{\lambda_{(n,r)}\}$ ($r \geq 3$). From the present point of view, $\{\lambda_{(n,r)}\}$ is a more natural asymptotic sequence upon which to base asymptotic expansions. It is not clear, however, which one is more accurate for small and moderate sample sizes, and whether or not $\{\lambda_{(n,r)}\}$ are easy to calculate.

● *Lugannani and Rice's approximation*

Lugannani and Rice's approximation for the tail probability of a sample mean is given by (2.2.29). To approximate the tail probability for X_n using this formula, we may treat X_n as a mean of k i.i.d. random variables where $k = 1$. The formula then leads to

$$LR_n(x) = 1 - \Phi(\xi_n) + \phi(\xi_n)\left[\frac{1}{z_n} - \frac{1}{\xi_n}\right], \qquad (4.6.68)$$

where $\xi_n = \{2[\hat{T}_n x - K_n(\hat{T}_n)]\}^{\frac{1}{2}}$, and $z_n = \hat{T}_n(K_n''(\hat{T}_n))^{\frac{1}{2}}$. Note that since (4.6.68) is not derived through the formal method, we do not have a formal order for its relative error. When X_n is a sample mean, this approximation is very accurate, even for small sample sizes. We shall examine its accuracy for other cases in the next section.

4.7 Numerical examples

There is no definite relationship between the accuracy of an asymptotic expansion at a fixed value of the asymptotic factor and the order of its relative error, although a high order expansion is often more accurate. Thus an asymptotic expansion, regardless of its order, is considered useful only after a substantial amount of numerical evidence which indicates its accuracy is gathered. In this section, we examine the numerical accuracy of the generalized saddlepoint approximation for the density function (4.3.11) and Lugannani and Rice's approximation for the cumulative distribution function (4.6.68) for some examples. For brevity, we may also refer to (4.3.11) as the saddlepoint approximation.

● **Example 1:**

Consider $f_n(x)$ with the following structure

$$f_n(x) = (1 - \theta_n)\phi(x) + \theta_n \varepsilon(x), \qquad (4.7.69)$$

where $\theta_n \to 0$ as $n \to \infty$, and $\varepsilon(x)$ is a density function. By rewriting $f_n(x)$ as

$$f_n(x) = \phi(x) + \theta_n \left[\varepsilon(x) - \phi(x)\right], \qquad (4.7.70)$$

it is clear that $f_n(x)$ is asymptotically $\phi(x)$, and the error of using $\phi(x)$ to approximate $f_n(x)$ is $O(\theta_n)$. Furthermore, X_n converges to the standard normal in distribution.

Let $M_\varepsilon(T)$ be the moment generating functions for $\varepsilon(x)$, then

$$K_n(T) = \frac{1}{2}T^2 + \log\left[1 + \theta_n \left(e^{-\frac{1}{2}T^2}M_\varepsilon(T) - 1\right)\right].$$

Assume that $e^{-\frac{1}{2}T^2}M_\varepsilon(T)$ is bounded, then $K_n(T)$ converges uniformly to the cumulant generating function of the standard normal distribution, $T^2/2$. By Weierstrass' theorem on uniformly convergent sequences of analytic functions, the derivatives of $K_n(T)$ also converge to those of $T^2/2$ uniformly. It can be shown that $K_n''(T) = O(1)$, $K_n^{(r)}(T) = O(\theta_n)$ for $r \geq 3$ and that the generalized saddlepoint approximation (4.3.11) is indeed an asymptotic approximation to $f_n(x)$. Also, $\lambda_{(n,r)} = O(\theta_n)$. Thus by (4.3.26) the formal order of the generalized saddlepoint approximation is $O(\theta_n)$, the same as that of the normal approximation. Is it still more accurate than $\phi(x)$ in this case?

Consider $f_n(x)$ given by

$$f_n(x) = (1 - \theta_n)\phi(x) + \theta_n I(x)\frac{\cos(x)}{2}, \qquad (4.7.71)$$

where $I(x)$ is 1 if $x \in [-\pi/2, \pi/2]$, and is 0 otherwise. Table 4.1 contains the exact values of $f_n(x)$, the normal approximation, $\phi(x)$, and values of the saddlepoint approximation, $g_n(x)$, for $\theta_n = 0.5, 0.3$, and 0.1. $g_n(x)$ here is not renormalized. Renormalization will in general improve the accuracy of saddlepoint approximation at those x values where $f_n(x)$ is relatively large, but may actually make the approximation at the far tail less accurate. It is thus recommended for densities, such as those that have finite domain and are bounded away from zero, whose magnitudes are relatively homogeneous over their domain.

When the value of the parameter θ_n is large ($\theta_n = 0.5$), the normal approximation is poor. Its relative error is approximately 100% except at the mean ($x = 0$). The saddlepoint

Table 4.1: Comparison of approximations to $f_n(x)$ with $\varepsilon(x) = I(x)cos(x)/2$

θ_n	x	0	2	4	6	8
$\theta_1 = 0.5$	$g_1(x)$	3.98942e-01	3.16164e-02	6.60181e-05	3.03786e-09	2.52613e-15
	exact	4.49471e-01	2.69954e-02	6.69151e-05	3.03794e-09	2.52613e-15
	$\phi(x)$	3.98942e-01	5.39909e-02	1.33830e-04	6.07588e-09	5.05227e-15
$\theta_2 = 0.3$	$g_2(x)$	3.98942e-01	4.07347e-02	9.31173e-05	4.25307e-09	3.53659e-15
	exact	4.29259e-01	3.77936e-02	9.36811e-05	4.25311e-09	3.53659e-15
	$\phi(x)$	3.98942e-01	5.39909e-02	1.33830e-04	6.07588e-09	5.05227e-15
$\theta_3 = 0.1$	$g_3(x)$	3.98942e-01	4.96086e-02	1.20254e-04	5.46827e-09	4.54704e-15
	exact	4.09048e-01	4.85918e-02	1.20447e-04	5.46829e-09	4.54704e-15
	$\phi(x)$	3.98942e-01	5.39909e-02	1.33830e-04	6.07588e-09	5.05227e-15

approximation, on the other hand, is reasonably accurate near the mean, and is remarkably accurate at the tail. As θ_n decreases to 0.3 and 0.1, the normal approximation becomes more accurate, but is still consistently less accurate than the saddlepoint approximation.

Table 4.2: Comparison of approximations to $Q_n(x)$ with $\varepsilon(x) = I(x)cos(x)/2$

θ_n	x	0	2	4	6	8
$\theta_1 = 0.5$	$LR_1(x)$	5.00000e-01	1.18940e-02	1.56423e-05	4.93965e-10	3.36432e-16
	exact	5.00000e-01	1.13750e-02	1.58356e-05	4.93293e-10	3.33066e-16
	$\Phi(x)$	5.00000e-01	2.27501e-02	3.16712e-05	9.86587e-10	6.66133e-16
$\theta_2 = 0.3$	$LR_2(x)$	5.00000e-01	1.61569e-02	2.20608e-05	6.91107e-10	4.46532e-16
	exact	5.00000e-01	1.59250e-02	2.21698e-05	6.90611e-10	4.44089e-16
	$\Phi(x)$	5.00000e-01	2.27501e-02	3.16712e-05	9.86587e-10	6.66133e-16
$\theta_3 = 0.1$	$LR_3(x)$	5.00000e-01	2.05219e-02	2.84705e-05	8.88121e-10	5.56044e-16
	exact	5.00000e-01	2.04751e-02	2.85041e-05	8.87929e-10	5.55111e-16
	$\Phi(x)$	5.00000e-01	2.27501e-02	3.16712e-05	9.86587e-10	6.66133e-16

Table 4.2 contains values of the tail probability $Q_n(x)$. Since $f_n(x)$ is symmetric, both Lugannani and Rice's formula and the normal approximation are trivially exact at the mean. At the tail, however, Lugannani and Rice's formula is consistently more accurate. Also, we note that Lugannani and Rice's formula is quite accurate even for $\theta_n = 0.5$ while normal approximation is not acceptable until $\theta_n = 0.1$.

● **Example 2:**

We consider approximating the distribution of the normalized sum of independent uniform distributions discussed in Example (d) of Section 4 in Chapter VIII in Feller (1966).

Let X_k be uniformly distributed between $-a_k$ and a_k. Then $\sigma_k^2 = a_k^2/3$. The normalized sum of the X_k's, W_n, is given by

$$W_n = \frac{X_1 + X_2 + \cdots + X_n}{s_n} \qquad (4.7.72)$$

where $s_n^2 = \sigma_1^2 + \sigma_2^2 + \cdots + \sigma_n^2$. If $a_1^2 + a_2^2 + \cdots + a_n^2 \to \infty$ and the a_k's remain bounded, then the Lindeberg condition is satisfied, and the distribution of W_n approaches the standard normal. We now consider the special cases where $a_k = k$ for $k = 1, 2, \ldots, 5$. The exact densities of $\sum_1^3 X_i$, $\sum_1^4 X_i$, and $\sum_1^5 X_i$, $h_3(x)$, $h_4(x)$, and $h_5(x)$, shown below are obtained through repeated convolution. The domains of these three densities are: [-6, 6], [-10, 10] and [-15, 15], respectively.

$$
\begin{aligned}
h_3(x) &= \frac{1}{96} \left[(x+6)_+^2 - (x+4)_+^2 - (x+2)_+^2 + (x-2)_+^2 + (x-4)_+^2 - (x-6)_+^2 \right], \\
h_4(x) &= \frac{1}{2304} \left[(x+10)_+^3 - (x+8)_+^3 - (x+6)_+^3 + 2(x)_+^3 - (x-6)_+^3 - (x-8)_+^3 \right. \\
&\quad \left. + (x-10)_+^3 \right], \\
h_5(x) &= \frac{1}{92160} \left[(x+15)_+^4 - (x+13)_+^4 - (x+11)_+^4 + (x+5)_+^4 + (x+3)_+^4 \right. \\
&\quad + (x+1)_+^4 - (x-1)_+^4 - (x-3)_+^4 - (x-5)_+^4 + (x-11)_+^4 + (x-13)_+^4 \\
&\quad \left. - (x-15)_+^4 \right].
\end{aligned}
$$

The densities of W_3, W_4 and W_5, $f_3(x)$, $f_4(x)$ and $f_5(x)$, can be obtained by using the relation $f_n(x) = s_n h_n(s_n x)$, and have domains $[-2.78, 2.78], [-3.16, 3.16]$ and $[-3.50, 3.50]$, respectively. Note that for the present example, a_k cannot always be allowed to be k due to the fact that the a_k's need to be bounded for the Lindeberg condition to apply. Since we are only interested in the accuracy of the saddlepoint approximation for $k = 3, 4$ and 5, we shall not specify the a_k's for $k > 5$. Table 4.3 contains, along with the exact values of the density functions, the values of the renormalized saddlepoint approximation, and that of the normal approximation. The x values are approximately equally spaced points in the positive half of the domains.

The saddlepoint approximation for the average of the mean of five uniform$[-1, 1]$ observations has been computed by Daniels (1954) and Field and Ronchetti (1990) to illustrate its remarkable accuracy for approximating the density of the sample mean. The maximum relative error of the renormalized saddlepoint approximation for the density h_5 is about 3%, which is comparable to that reported by Field and Ronchetti (1990).

Table 4.3: Comparison of approximations to density function $f_n(x)$

$f_n(x)$	x	0.0000	0.6925	1.3850	2.0775	2.7700
$f_3(x)$	$g_3(x)$	3.69101e-01	3.11023e-01	1.76792e-01	4.95893e-02	5.55914e-06
	exact	3.60041e-01	3.09682e-01	1.80745e-01	5.14500e-02	5.84455e-06
	$\phi(x)$	3.98942e-01	3.13888e-01	1.52887e-01	4.61000e-02	8.60520e-03
$f_4(x)$	$g_4(x)$	3.75409e-01	2.94665e-01	1.31653e-01	2.10719e-02	7.71770e-08
	exact	3.73324e-01	2.92380e-01	1.34334e-01	2.20003e-02	8.03275e-08
	$\phi(x)$	3.98942e-01	2.92580e-01	1.15411e-01	2.44863e-02	2.79425e-03
$f_5(x)$	$g_5(x)$	3.79532e-01	2.77795e-01	9.80271e-02	9.12319e-03	4.41027e-09
	exact	3.77719e-01	2.76990e-01	9.94923e-02	9.42633e-03	4.55929e-09
	$\phi(x)$	3.98942e-01	2.73244e-01	8.77960e-02	1.32337e-02	9.35772e-04

Table 4.4: Comparison of approximations to cumulative distribution function $F_n(x)$

$F_n(x)$	x	0.0000	0.6925	1.3850	2.0775	2.7700
$F_1(x)$	$\hat{F}_3(x)$	0.5000000	0.7416284	0.9121156	0.9881504	1.0000000
	exact	0.5000000	0.7376693	0.9090287	0.9879917	1.0000000
	$\phi(x)$	0.5000000	0.7556883	0.9169738	0.9811223	0.9971972
$F_2(x)$	$\hat{F}_4(x)$	0.5000000	0.7734851	0.9416066	0.9956302	1.0000000
	exact	0.5000000	0.7712989	0.9402327	0.9955662	1.0000000
	$\phi(x)$	0.5000000	0.7845054	0.9423718	0.9909239	0.9991836
$F_3(x)$	$\hat{F}_5(x)$	0.5000000	0.7990668	0.9601499	0.9982908	1.0000000
	exact	0.5000000	0.7977302	0.9594114	0.9982675	1.0000000
	$\phi(x)$	0.5000000	0.8078498	0.9590705	0.9954729	0.9997493

Table 4.4 contains approximations to the cumulative distribution function, $\hat{F}$, obtained through Lugannani and Rice's formula, the exact values and the values of the normal approximation. Although the normal approximation is reasonably accurate in all three cases, again Lugannani and Rice's formula consistently gives more accurate approximations.

● Example 3:

When the cumulant generating function of a discrete distribution is known, and this distribution is asymptotically normal, saddlepoint approximation may still be used to obtain a smoothed version of its probability mass function, see e.g., Daniels (1958). Lugannani and Rice's formula for the tail probability can also be used to approximate its tail probability. Although these two approximations are now only formal approximations, numerical evidence

suggests that they remain accurate. We now consider the use of Lugannani and Rice's formula for approximating the significance level of a simple one-sample permutation test.

Let $x_1, x_2, \ldots, x_n$ be a random sample from a population with distribution function $F(x - \theta)$, where $F(y)$ is symmetric. The hypothesis of interest is: H_0: $\theta = \theta_0$. Robinson (1982) considered a randomization test conditional on $|a_1|, |a_2|, \ldots, |a_n|$, where

$$a_i = \frac{x_i - \theta_0}{[\sum_{i=1}^{n}(x_i - \theta_0)^2]^{\frac{1}{2}}}. \qquad (4.7.73)$$

The test statistic, W_n, is then

$$W_n = \sum_{i=1}^{n} B_i |a_i|, \qquad (4.7.74)$$

where the B_i's are independent random variables taking values 1 and -1, each with probability 1/2. The cumulant generating function of W_n is

$$K_n(T) = \sum_{i=1}^{n} \log[\cosh(a_i T)]. \qquad (4.7.75)$$

The observed value is $w_n = \sum_{i=1}^{n} a_i$ and the significance level for a one-sided test is $P(W_n \geq w_n)$. Note that the distribution of W_n is symmetric. Thus for a two-sided test, the significance level is $2P(W_n \geq w_n)$. The exact value of the significance level is usually difficult to obtain due to the huge amount of computing needed to calculate it. However, the cumulant generating function of W_n (4.7.75) is easily available, and under the null hypothesis, the distribution of W_n is asymptotically normal. See Lehmann (1988) and Robinson (1982) for discussion concerning the asymptotic normality of W_n. Lugannani and Rice's formula may thus be used to approximate the significance level. Robinson (1982) derived the following two approximations for $P(W_n \geq w_n)$,

$$A = \exp\{K_n(\hat{T}_n) - \hat{T}_n w_n + \frac{1}{2}\hat{T}_n^2 \sigma_n^2\}[1 - \Phi(\hat{T}_n \sigma_n)],$$

and

$$B = A\left[1 + \frac{1}{6}\lambda_3(\hat{T}_n)V(\hat{T}_n \sigma_n)\right],$$

where $\hat{T}_n$ is the saddlepoint corresponds to the observed value w_n, $\sigma_n = \sqrt{K_n''(\hat{T}_n)}$, $\lambda_3(\hat{T}_n) = K_n^{(3)}(\hat{T}_n)/[K_n''(\hat{T}_n)]^{\frac{3}{2}}$, and

$$V(y) = \frac{(y^2 - 1)\phi(y)}{1 - \Phi(y)} - y^3.$$

Robinson called A the first saddlepoint approximation, and B the second saddlepoint approximation. Since we have used the term "saddlepoint approximation" for the approximation given by (4.3.11), to avoid confusion we shall refer to A and B as Robinson's first and second approximations to the significance level, respectively. Robinson's approximations may be used in other situations where the distribution being approximated is asymptotically normal. Daniels (1987) used them for the tail area of the mean of n i.i.d. observations, where their errors are $O(1/\sqrt{n})$ and $O(1/n)$, respectively.

Table 4.5 compares Lugannani and Rice's formula (4.6.68) with Robinson's approximations for two examples. The first example, also used by Robinson (1982), is taken from Fisher (1935, Section 21). The second example involves a set of twenty random numbers from a normal distribution with mean 0.3 and standard deviation 1. The data for both examples are listed in the end of this example. The hypothesis in both cases is H_0: $\theta_0 = 0$. For the first example, the exact significance level is given by Robinson (1982). For the second example, we used a simulated significance level based on 10,000 simulated values of W_n, where the a_i's are calculated using the data set # 2 given below.

Table 4.5: Approximations to the significance level of
a one sample randomization test.

Data Set	n	exact/simulated	A	B	LR
#1	15	0.052	0.048	0.052	0.052
#2	20	0.508	0.493	0.504	0.502

The examples suggest that values given by Robinson's second approximation and Lugannani and Rice's formula differ very little. They both are very close to the true significance level. Robinson's first approximation, however, is slightly less accurate, and gives values that are smaller than the exact/simulated significance levels. Robinson (1982) also discussed the use of his approximations in approximating the significance level of a two-sample permutation test, as well as computing confidence intervals for θ. One may also use Lugannani and Rice's formula in place of his approximations to perform these tasks.

Data set #1 used in Example 3:
-67 -48 6 8 14 16 23 24 28 29 41 49 56 60 75

Data set #2 used in Example 3:
-1.0819158 0.2442755 0.9777535 0.2314501 -0.6371420 -1.0340618 1.4874597

-1.8075270 1.7112674 -0.2267291 0.1596133 -0.9088350 0.2676362 0.6653836
1.8406825 1.1424801 0.3416494 -1.4551200 -0.1336391 1.4804197

- *Saddlepoint approximation for L statistics*

Easton and Ronchetti (1986) discussed saddlepoint approximation in situations where exact moment generating functions are not available but Edgeworth expansions for densities up to and including the term of order $1/n$ are available. Their approach is to approximate the moment generating function through the Edgeworth expansion, and use this approximation of the moment generating function to replace the true unknown moment generating function for calculating the saddlepoint approximation. They seemed to have obtained a remarkable result that the error of the saddlepoint approximation calculated this way is of order $O(1/n)$ uniformly. See, also, Field and Ronchetti (1990). The authors were not specific about the type of error they discussed. Judging by the key formula (2.4) in Easton and Ronchetti (1986), this error should be the absolute error rather than the relative error due to the presence of the absolute error of the Edgeworth expansion, $Z_n(x)$. For brevity, a formula in this paper, say (2.4), will be referred to in the following as ER(2.4).

Neither Easton and Ronchetti (1986) nor Field and Ronchetti (1990) contain a proof for the result. In attempting to construct a proof based on discussion in Easton and Ronchetti (1986), we encountered the following problem. To obtain the saddlepoint approximation, the integral in ER(2.4) needs to be expanded at the saddlepoint. The cumulant generating function, $\tilde{R}_n(T)$, however, is a function of n. Thus the expansion of the integral may not be obtained through the classical method of Laplace or steepest descents. Chaganty and Sethuraman (1986, Theorem 2.1) discussed a saddlepoint approximation type of expansion for an integral (see (2.12) in this paper) similar to that in ER(2.4). The integral has, in the place of $\tilde{R}_n(T)$ in ER(2.4), a true cumulant generating function of complex variable. They used rather strong conditions to establish the validity of their expansion. We have not found a way to prove that Easton and Ronchetti's expansion to the integral in ER(2.4), i.e., ER(2.6), is valid. Even if it is, only the order of the relative error (for using the expansion to approximate the integral) is available. It is not clear to us how the order of this relative error and that of $Z_n(x)$ imply that the error of using ER(2.6) to approximate $f_n(x)$ is of order $O(1/n)$, uniform for all x values.

Easton and Ronchetti (1986) applied their saddlepoint approximation to approximate distributions of some complicated L statistics. Their numerical results suggest that the saddlepoint approximation compares favorably to the normal approximation and the Edgeworth expansion. These numerical results have not only added further evidence to the remarkable accuracy of the saddlepoint approximation, but have also demonstrated its robustness in that a small variation of the approximation formula results in little loss in accuracy.

● *Some comments based on the examples*

The examples that we examined suggest that the saddlepoint approximation and Lugannani and Rice's formula are reliable approximations, and are increasingly more accurate as the distributions being approximated get closer to the normal distribution.

These and other examples that we have examined indicate that the saddlepoint approximation and Lugannani and Rice's formula give accurate numerical approximations to normal-like distributions. Davison and Hinkley (1988) applied saddlepoint approximation in a variety of resampling methods where the exact cumulant generating functions are not available. They used the cumulant generating function of the empirical distribution to compute the approximation and still obtained very accurate results. Field (1993) applied Lugannani and Rice's formula to approximate the tail probability of weighted chi-squares. Although a general asymptotic relationship between the error and the weights is not available, he observed that the approximation is very accurate. All this evidence suggests that these two approximations are in general robust, and their excellent accuracies are intrinsic in that they are accurate so long as the distribution being approximated is normal-like, whatever the underlying cause. This calls for a wider use of these approximations in place of the normal approximation, (perhaps) even in situations where asymptotic analyses of their relative errors are not available.

4.8   Concluding remarks

Although in Poincaré's definition of an asymptotic expansion,

$$f(z) = \sum_{n=1}^{N} a_n \phi_n(z) + o(\phi_N(z)),$$

the asymptotic sequence $\{\phi_n\}$ needs not to be a power series, important developments in the theory of asymptotic analysis are mostly concerned with power series expansions for functions. The developments in asymptotic expansions for distributions reflect that of the theory of asymptotic analysis. In fact all known expansions for distributions are power series expansions, and even in cases where the power series expansions are not the most natural expansions, efforts had been concentrated on obtaining power series expansions. There may be three reasons for the popularity of power series expansions: (1) they are simple to interpret, (2) for the important special cases of sample mean and standardized mean for i.i.d. observations, the well-known expansions are all power expansions, and (3) their validity can often be readily demonstrated using existing asymptotic methods.

In this chapter, we have derived a family of general expansions for density functions with respect to sequences based on $\{\lambda_{(n,r)}(\cdot)\}$. Our main goal has been to point out the possibility of expanding distributions that are asymptotically normal with respect to these sequences. The usefulness of the general expansions is evident from examples discussed in Sections 4.4, 4.5 and 4.6. Having said this, we note, however, that we have been relying on the asymptotic properties of known power expansions to prove the validity of the general expansions. A typical example is that of the proof for the validity of the saddlepoints expansion in Chapter 3. More direct methods for verifying the validity of these expansions are needed. Work is continuing to connect the general expansions with results obtained by Skovgaard (1986) and Chaganty and Sethuraman (1986), and to establish general conditions under which some or all members of the family are valid asymptotic expansions for the density function.

The numerical accuracy of the saddlepoint approximation and Lugannani and Rice's formula for cases where orders of their errors are not clear has been the focus of Section 4.7. The remarkable accuracy and robustness of these approximations demonstrated in that section should provide some assurance for using these approximations in the absence of rigorous analyses on their asymptotic properties.

Chapter 5

Quadrature Methods for Computing Distributions

5.1 Introduction

Let X be the mean of a sample of i.i.d. observations with domain $[a, b]$. Assume X has a continuous density function $f(x)$. The moment generating function of X, $M(t)$, is given by

$$M(t) = \int_a^b e^{tx} f(x) dx. \tag{5.1.1}$$

The moment generating function may be easily determined by using that of the individual observations. In theory, $f(x)$ and the cumulative distribution function of X, $F(x)$, may be computed by numerically evaluating the convolution formulas or integrating the inversion formulas of (5.1.1), but in practice the first approach may be time consuming and inaccurate. The second also has its drawbacks. It involves integrating the characteristic function over an infinite path, and requires a detailed analysis concerning the asymptotic behavior of the characteristic function. See Lugannani and Rice (1980) for further discussion. While there is a lack of successful numerical methods for approximating distributions, there are accurate

asymptotic approximations, such as Daniels' saddlepoint approximation and Lugannani and Rice approximation that we discussed extensively in earlier chapters. These approximations have had remarkable success. Unlike normal approximation, they are also often accurate for small sample sizes. However, there are several difficulties in using these approximations: (1) Often $M(t)$ can only be determined through numerical means, i.e., its analytic expression is not available. See, e.g., Daniels (1983). In such cases, it is rather expensive if not impossible to compute these approximations. (2) There is a lack of a practical method for estimating the errors of these approximations. Lugannani and Rice (1980) discussed an error bound for their approximation, but it often requires integrating a complex function over an infinite path. (3) Various examples suggest that in the case of a finite domain, when the density function being approximated is far from a normal density, these asymptotic approximations are often poor. See, e.g., Chapter 7 in Field and Ronchetti (1990).

In the present chapter, we discuss quadrature methods for computing distributions, in particular that of the mean of a sample of i.i.d. observations. Distributions whose domains are finite intervals and those whose domains are unbounded are dealt with separately. For computing distributions with finite domains, a refined quadrature method, the optimal method, is introduced. This method is built upon a measure of accuracy that we devised for identifying accurate quadrature methods for Fredholm integral equations of the first kind, such as (5.1.1). For distributions with domains which are bounded below or above but not both, we discuss Bellman *et al.*'s (1966) method for inverting Laplace transformations and its use for computing distributions of this type. These methods may be used to compute both the cumulative distribution function and the density function of a distribution, and have certain advantages over the traditional asymptotic methods.

Since the measure of accuracy and the optimal method may be used to solve other types of Fredholm equations of the first kind, we shall present them not as tools devised solely for the purpose of computing distributions but tools for solving general Fredholm integral equations of the first kind. Thus we divide this chapter into two parts with the first part devoted to the presentation of the measure of accuracy and the optimal method, and the second part to their application and the use of Bellman *et al.*'s method in computing distributions. For the completeness of presentation of the measure of accuracy and the optimal method, the first part also contains an introduction.

Notation in this chapter is different from that in previous chapters. We shall use, for example, $f(x)$ to denote the density of a sample mean instead of $f_n(x)$. The reason for this is that the sample size, unlike in asymptotic methods where it is the all important large parameter, plays no role here.

5.2 Part One: A measure of accuracy for quadrature methods for Fredholm equations of the first kind

5.2.1 Introduction

We consider quadrature methods for Fredholm integral equations of the first kind,

$$\int_a^b K(t,x)f(x)dx = y(t), \qquad t \in I_t \tag{5.2.2}$$

where $f(x)$ is the unknown, and I_t is the domain of t. Equation (5.2.2) may be written in operator form: $Kf = y$. Let x_j and w_j ($j = 1, 2, \dots N$) be the abscissae and weights of an N-point quadrature rule, Q_N. A quadrature method replaces the integral in (5.2.2) with Q_N for $t = t_i$ ($i = 1, 2, \dots, N$) and leads to N simultaneous equations

$$\sum_{j=1}^N w_j K(t_i, x_j)\hat{f}(x_j) = y(t_i), \qquad i = 1, 2, \dots, N. \tag{5.2.3}$$

Let $K^{(N)} = [w_j K(t_i, x_j)]_{N \times N}$. Provided that $K^{(N)}$ is nonsingular, (5.2.3) has a unique solution $\hat{f}(x_j)$, which is an approximation of $f(x_j)$ ($j = 1, 2, \dots, N$). For brevity, a method that uses Q_N and a certain set of t_i's will be referred to as method (Q_N, T), where $T = \{t_1, t_2, \dots, t_N\}$.

Despite the fact that they are the simplest of all numerical methods for solving (5.2.2), quadrature methods are rarely used in practice. One reason is that they frequently lead to inaccurate numerical solutions. Yet little is known concerning conditions under which they may be accurate. This difficulty is further discussed in the next section. For a comprehensive discussion concerning quadrature methods and their drawbacks, see Baker (1977). Nonetheless, as we shall see from subsequent examples, there are equations for which accurate quadrature methods exist. The question is how to identify these accurate methods

from a virtually infinite number of methods available. In the present section, we introduce a simple measure of accuracy that may be used to identify such methods. This measure may also be used to establish error bounds for the solutions. In Section 5.2.2, we examine, through an example, the difficulties of determining the accuracy of quadrature methods, and introduce the measure of accuracy. In Section 5.2.3, we discuss how this measure may be applied to identify accurate methods and calculate approximate error bounds. We shall provide some general discussion on basis selection and error bounds in Section 5.2.4.

5.2.2 The order of accuracy of a quadrature method

The choice of each of the three components for a quadrature method, i.e., the type of quadrature rule, the number of abscissae N and the t_i's, affects its accuracy. Unfortunately, unless $K(t, x)$ is of a certain special form, e.g., example 5.35 in Baker (1977), it is generally difficult to determine theoretically how the accuracy depends on the choices of these components. Even increasing the number of abscissae, a strategy commonly used to improve the accuracy of numerical methods, may fail to bring more accurate solutions for quadrature methods. Consider the following example

Example 5.1:

$$\int_{-1}^{1} e^{tx} f(x) dx = \frac{\sinh(t)}{t}, \qquad t \in (-\infty, \infty), \tag{5.2.4}$$

which has exact solution $f(x) = 0.5$ for $x \in [-1, 1]$. Denote the error of the quadrature formula at $t = t_i$ by r_i, and $\hat{f}(x_j) - f(x_j)$ by e_j. Let $R = (r_1, \ldots, r_N)^T$ and $E = (e_1, \ldots, e_N)^T$. The following table, computed by methods using N-Point Gauss-Legendre rules and t_i's that are equally spaced points between -1 and 1, contains $\|R\|_\infty$, the estimated condition number of $K^{(N)}$, $\kappa_1(K^{(N)})$, and $\|E\|_\infty$.

It can be shown that $K^{(N)}E = R$. Thus in general $\|E\|_\infty$ is small when both $\kappa_1(K^{(N)})$ and $\|R\|_\infty$ are small. Initially as N increases, $\kappa_1(K^{(N)})$ increases, but $\|R\|_\infty$ decreases. The decrease in $\|R\|_\infty$ offsets the increase in $\kappa_1(K^{(N)})$, and the accuracy of the numerical solution improves. However, once $\|R\|_\infty$ drops to near machine epsilon at $N = 7$, it cannot drop further, but $\kappa_1(K^{(N)})$ still increases very rapidly. Consequently, $\|E\|_\infty$ becomes larger. The influence of the type of quadrature rule is even more difficult to determine.

Table 5.1: The effect of the number of abscissae on the error.

N	$\|R\|_\infty$	$\kappa_1(K^{(N)})$	$\|E\|_\infty$
3	3.27×10^{-5}	4.57×10^1	1.34×10^{-4}
5	4.12×10^{-10}	1.53×10^4	1.15×10^{-7}
7	6.66×10^{-16}	1.32×10^7	2.29×10^{-10}
9	4.44×10^{-16}	2.15×10^{10}	7.00×10^{-8}
11	2.22×10^{-16}	5.65×10^{13}	3.34×10^{-5}
13	4.44×10^{-16}	3.26×10^{18}	9.66×10^1

Although theoretical analyses on the errors of quadrature methods seem intractable, one can still judge their accuracy through test problems. We propose the following measure of accuracy to explore this idea. To simplify our presentation, we shall assume that (5.2.2) has a unique solution.

Let $P_h = \text{Span}\{1, x, x^2, \ldots, x^h\}$ be the set of polynomials of degree h or less. The set of $h + 1$ test equations, $K f_l = K x^l$, have solutions x^l ($l = 0, 1, \ldots, h$). To measure the accuracy of a certain quadrature method, (Q_N, T), we compare its numerical solution to $K f_l = K x^l$, $\hat{f}_l$, with x^l for $l = 0, 1, \ldots, h$. Let

$$e_l = \max_{j \in \{1, \ldots, N\}} \{|x_j^l - \hat{f}_l(x_j)|\}, \qquad (5.2.5)$$

and $emax(h) = \max\{e_0, e_1, \ldots, e_h\}$. Write $emax(h)$ in scientific notation,

$$emax(h) = \alpha \times 10^{-r}, \qquad (5.2.6)$$

where $1 \leq \alpha < 10$ and r is an integer. We call r *the order of accuracy of method* (Q_N, T) *with respect to* P_h, and write this order as $r(P_h)$. $r(P_h)$ is a nonincreasing function of h, i.e., $r(P_{h+1}) \leq r(P_h)$. The following table shows the orders of accuracy of 5-point, 7-point and 9-point methods as used to compute Table 5.1 with respect to P_3, P_4, P_5, and P_6. The 7-

Table 5.2: Example 5.1 cont'd. Orders of accuracy for methods using N-point Gauss-Legendre rules with $t_i = -1 + \frac{2(i-1)}{N-1}$.

N	$r(P_3)$	$r(P_4)$	$r(P_5)$	$r(P_6)$
5	4	3	3	2
7	8	5	4	2
9	3	2	0	-1

point method has consistently the highest order of accuracy with respect to all four different

polynomial spaces. This agrees with Table 5.1, where the 7-point method was shown to be the most accurate.

To see how the order of accuracy may be used to establish error bounds for numerical solutions of quadrature methods, we first note that if $\hat{f}$ and $\hat{g}$ are the numerical solutions for equations where the exact solutions are f and g, then the numerical solution for the equation with exact solution $\alpha f + \beta g$ is $\alpha \hat{f} + \beta \hat{g}$. This linearity of the numerical solution implies that if (5.2.2) has solution $f \in P_h$, i.e.,

$$f = c_0 + c_1 x + c_2 x^2 + \cdots + c_h x^h, \tag{5.2.7}$$

where the c_i's are constants, then the corresponding numerical solution $\hat{f}$ is

$$\hat{f} = c_0 \hat{f}_0 + c_1 \hat{f}_1 + \cdots + c_h \hat{f}_h. \tag{5.2.8}$$

A bound on the error is then given by

$$\|f - \hat{f}\|_\infty \leq \sum_{j=0}^{h} |c_j| \times e_j. \tag{5.2.9}$$

or, in terms of the order of accuracy,

$$\|f - \hat{f}\|_\infty \leq (\sum_{j=0}^{h} |c_j|) \times \alpha \times 10^{-r(P_h)}, \tag{5.2.10}$$

where α is the same as in (5.2.6). The c_j's are in general unknown, but may be approximated by the method of least-squares. Also, the above bounds are often useful even if $f(x) \notin P_h$, provided it can be well approximated by some $p(x) \in P_h$. We shall further illustrate these points through Example 5.2 in the next section.

5.2.3 The optimal method

Different choices of quadrature rules and t_i's can also greatly affect the orders of accuracy. The following table is computed by methods using N-point composite Simpson's rules and Gauss-Legendre (G-L) rules with t_i's that are equally spaced between -2 and 2. With the same choices of N and T, methods that use Gauss-Legendre rules have considerably higher orders of accuracy. Also, by comparing orders of accuracy of methods that use the

Table 5.3: Example 5.1 cont'd. Orders of accuracy for methods using N-point composite Simpson's rules and G-L rules with $t_i = -2 + \frac{4(i-1)}{N-1}$.

N	Simpson's Rule			G-L Rule		
	$r(P_3)$	$r(P_4)$	$r(P_5)$	$r(P_3)$	$r(P_4)$	$r(P_5)$
5	1	1	1	3	3	2
7	1	1	1	6	5	4
9	0	0	0	6	5	4

above Gauss-Legendre rules with that of methods (shown in Table 5.2) that use the same quadrature rules but different t_i's, we see that the order of accuracy is heavily influenced by the choice of T. In general, to obtain high order methods, the larger the N, the wider spread in the t_i's needs to be. For methods that use the 15-point Gauss-Legendre rule, for example, $r(P_5)$ is actually negative when the t_i's are equally spaced between -2 and 2. When they are equally spaced between -11 and 11, $r(P_5)$ is 7. There is a limit, however, on how large one can make N and still obtain a high order method. When N is too large, $K^{(N)}$ may be too ill-conditioned regardless of the choice of quadrature rule and T.

In general, to look for the method with the highest order of accuracy with respect to a certain P_h, one should consider methods with different types of quadrature rule and different choices of N and T, so that methods of possible high orders are not overlooked. One computes and then compares their orders of accuracy. The method with the highest order found by this process will be referred to as the *optimal method with respect to P_h*, or simply the *optimal method*. The optimal method is not necessarily the one with the highest order possible. Nevertheless, when used to solve (5.2.2), it generally gives the most accurate solutions among methods considered.

We now use some of the methods discussed above to compute the following example involving the operator in Example 5.1. These methods all use the Gauss-Legendre rule.

Example 5.2:

$$\int_{-1}^{1} e^{tx} f(x)dx = \frac{e^{(t+1)} - e^{-(t+1)}}{(t+1)(e^1 - e^{-1})} \qquad t \in (-\infty, \infty). \tag{5.2.11}$$

The exact solution is $f(x) = e^x/(e - e^{-1})$ for $x \in [-1, 1]$. Table 5.4 contains, in the middle column, the maximum absolute error of the numerical solutions of a 7-point, a 9-point and a 15-point method, where the t_i's are equally spaced points in $[-2, 2], [-5, 5],$

and $[-11, 11]$, respectively. These methods have orders of accuracy $r(p_5) = 2$, 4, and 7, respectively. The 15-point method is the optimal method. The estimated error bounds,

Table 5.4: Maximum absolute errors and error bounds for Example 5.2.

N	$\|E\|_\infty$	Error Bound
7	2.04×10^{-7}	3.04×10^{-7}
9	1.48×10^{-7}	2.64×10^{-6}
15	8.01×10^{-10}	3.17×10^{-8}

computed using (5.2.9), are in the right-hand column. These bounds are calculated using the coefficients of the third order polynomials that best approximate the numerical solutions. The method of least-squares is used to compute these coefficients. The decision to use third order polynomials is based on the observation that the coefficients for x^4 and x^5, should they be used, are much smaller than those of the lower order terms. Also, without these terms, the error sum of squares is already very small. Adding them to the least-squares regression reduces the error sum of squares very little.

Note that although the exact solution in this example is not a polynomial, it can be well approximated by a third order polynomial and the optimal method with respect to P_5 does give the most accurate solution. In practice, since the exact solution is unknown, one in general does not know the value of h such that the solution may be well approximated by a member of P_h. Thus optimal methods with respect to P_h where h is reasonably large, such as $h = 5$ in this example, are recommended.

5.2.4 Basis selection and error bounds

Let $L \in R^{(h+1) \times (h+1)}$ be nonsingular and $X = (1, x, \ldots, x^h)^T$. The elements of $\Phi = (\phi_o, \phi_1, \ldots, \phi_h)^T$ where $\Phi = LX$ form a basis for P_h. Under this basis, $f \in P_h$ in (5.2.7) may be expressed as

$$f = C^T X = C^T L^{-1} \Phi = V^T \Phi, \qquad (5.2.12)$$

where $C = (c_0, c_1, \ldots, c_h)^T$ and $V = C^T L^{-1} = (v_0, v_1, \ldots, v_h)^T$. Substituting $\phi_l(x_j)$ for x_j^l and $\hat{\phi}_l(x_j)$ for $\hat{f}_l(x_j)$ in (5.2.5) for $l = 0, 1, \ldots, h$, where $\hat{\phi}_l$ is the numerical solution to $K\phi = K\phi_l$, the resulting e_l's may be used to define the order of accuracy and optimal method under $\{\phi_o, \phi_1, \ldots, \phi_h\}$ by following procedures described in Sections 5.2.2 and 5.2.3.

For (5.2.2) where $f \in P_h$, they also lead to the following bound on the error of $\hat{f}$

$$\|f - \hat{f}\|_\infty \leq \sum_{j=0}^{h} |v_j| \times e_j. \tag{5.2.13}$$

Although the order of accuracy of a method may change with the basis, we observed that the optimal method under one basis remains optimal under another. Furthermore, it also remains optimal even when the dimension of P_h is changed; e.g., among the three methods shown in Table 5.2, the 7-point method is consistently optimal with respect to P_3, P_4, P_5 and P_6. This independence of the optimal method on the choice of the basis raises an interesting question: is there a strategy for finding a basis under which the error bound given by (5.2.13) for the optimal method is the smallest or closest to the real error?

Since the v_j's and the e_j's will both vary with the basis, even for examples where the exact solutions are known it is usually difficult to determine theoretically whether a certain basis will lead to a bound smaller than that under, say $\{1, x, \ldots, x^h\}$. One exception to this is that when $f = \phi_i$ for some i, then $v_i = 1$ and $v_j = 0$ for $j \neq i$. The equality in (5.2.13) holds and its right-hand side gives the minimum bound. In practice, however, the exact solution is not known, thus the minimum bound cannot be achieved. If we fit the numerical solution with a polynomial using the method of least-squares, and let ϕ_i be this polynomial, then $v_i \approx 1$ and $v_j \approx 0$ for $i \neq j$. The resulting bound may be close to the minimum bound, provided the numerical solution is accurate. This bound may not be useful in practice since its accuracy depends on that of the numerical solution. A practical strategy for looking for a small bound may be to use a certain set of bases, such as $\{1, x, \ldots, x^h\}$ and the shifted Chebyshev polynomials, at the same time routinely, and compare the resulting error bounds.

For $f \notin P_h$, write f as $f = p_h + \varepsilon$ where $p_h \in P_h$. Let $\hat{p}_h$, $\hat{\varepsilon}$ and $\hat{f}$ be solutions to $Kg = Kp_h$, $Kg = K\varepsilon$ and (5.2.2), respectively, given by the same method. Then $\|f - \hat{f}\|_\infty \leq \|p_h - \hat{p}_h\|_\infty + \|\varepsilon - \hat{\varepsilon}\|_\infty$. While we may assume $\|p_h - \hat{p}_h\|_\infty$ is small when the method is optimal with respect to P_h, we still need to know the error in solving $Kg = K\varepsilon$. This is an important and yet complicated problem. Here we make two suggestions that may be useful for dealing with it. If it is known that $\varepsilon \in \Lambda$ where Λ is a function space of a finite dimension, one may consider deriving and using an optimal method with respect to both P_h and Λ. Without this information, one may select a high h value so that the solution can be well approximated by some $p_h \in P_h$ and solve the problem as if f is in P_h. This approach

usually works well in practice, e.g., Example 5.2.

5.2.5 Concluding remarks

We have found that optimal methods defined in Section 5.2.3 work well for solving a variety of problems where the kernel is not too flat and I_t is $(-\infty, \infty)$. A typical example is that of inverting Laplace transformations of functions that vanish outside some finite intervals. Under these two conditions, optimal methods with high orders of accuracy can usually be found, and numerical solutions given by these methods are very accurate. The error bound given by (5.2.9) is also quite reliable. We also found, however, that for equations where the two conditions are not met, optimal methods are sometimes of low orders of accuracy ($r(P_5) \leq 3$), and their numerical solutions may be poor. Although the simple basis, $\{1, x, \ldots, x^h\}$, is usually satisfactory in practice, basis selection remains an important problem to be further studied.

Finally, the idea behind the above order of accuracy is the same as that behind the precision of quadrature rules, i.e., evaluating numerical methods according to their performance on test problems involving polynomials. Since continuous functions on finite intervals can be well approximated by polynomials, when (5.2.2) has a continuous solution, the P_h's are natural choices relative to which the order of accuracy and optimal methods are defined. Nevertheless, when exact solutions are known to be in some other spaces, one should consider using these spaces instead.

Examples in this part were computed using Fortran 77 on a SPARC station. The linear system solver used was DLSARG, least-squares estimates were computed by DRCURV, and condition numbers were estimated by DLFCRG, all from IMSL.

5.3 Part Two: The use of quadrature methods for computing distributions

In this part, we discuss the use of the optimal method and Bellman *et al.*'s (1966) method for computing density functions, and compare their accuracy with that of the saddlepoint approximation. In Section 5.3.1, we discuss issues related to the implementation of the optimal method for (5.1.1). In Section 5.3.2, we discuss Bellman *et al.*'s (1966) method for cases where b or a (not both) is infinity. We then compare the optimal method with the saddlepoint approximation in Section 5.3.3.

5.3.1 Computing densities with finite domain $[a, b]$.

The kernel for both examples that we considered in Part One is e^{tx}, which is that in the equation that defines the moment generating function (5.1.1). The optimal methods for these examples both involve the Gauss-Legendre rule. This is not a coincidence. Based on our experience with different types of quadrature rule, for equation (5.1.1), the optimal method always involves the Gauss-Legendre rule. The domain of $f(x)$ for these examples is $[-1, 1]$, on which the Gauss-Legendre rule can be directly used without any transformation.

When $[a, b]$ is not $[-1, 1]$, in order to utilize the Gauss-Legendre rule, it is most convenient to transform $[a, b]$ to $[-1, 1]$ with the following transformation:

$$y = \frac{2x - (a + b)}{b - a} \quad \text{where} \quad x \in [a, b]. \tag{5.3.14}$$

The transformed variable y satisfies $y \in [-1, 1]$ and

$$x = \frac{(b - a)y + (a + b)}{2}. \tag{5.3.15}$$

Thus (5.1.1) becomes

$$\int_{-1}^{1} e^{t[\frac{(b-a)y}{2}]} f\left[\frac{(b - a)y + (a + b)}{2}\right] dy = \frac{2M(t)}{b - a} e^{-t[\frac{(a+b)}{2}]}. \tag{5.3.16}$$

We shall refer to (5.3.16) as the standardized equation for equation (5.1.1). Applying the optimal method with abscissae y_i $(i = 1, 2, \ldots, N)$ to (5.3.16), we obtain approximations for $f(x)$ at $x_i = x(y_i)$ for $i = 1, 2, \ldots, N$.

The operator in (5.3.16) has kernel $e^{t(b-a)y/2}$, for which e^{ty} is the special case where $b = -a = 1$. Still the Gauss-Legendre rule provides the most accurate methods for this operator. Note that the operator is location invariant (or invariant under a shift of the domain of a distribution) in that it only depends on the length of the domain, but not its location. Thus the optimal method for a certain standardized equation (distribution) with $b - a = l$ may be stored and used for other equations (distributions) that have the same domain length l.

Furthermore, the standardized equation (5.3.16) corresponding to a distribution will remain the same when its domain is shifted. To see this is true, we need to show that the right-hand side of (5.3.16) does not vary with the location of the domain. We have

$$\frac{2M(t)}{b-a}e^{-t[\frac{(a+b)}{2}]} = \frac{e^{-tl/2}}{l} \int_0^l e^{ty} f_0(y) dy = 2M_0(t)e^{-tl/2}/l, \qquad (5.3.17)$$

where $f_0(y)$ and $M_0(t)$ are the density function and the moment generating function for the shifted distribution whose domain is $[0, l]$, respectively. Notice that both a and b are absent from this simplified expression. Thus the right-hand side of (5.3.16) does not depend on the location of the domain; i.e., it is simply a different expression of the function at the right-hand side of (5.3.17) whose value does not depend on a and b.

Example 5.3:

$$\int_a^b e^{tx} f(x) dx = \frac{e^{(t+1)b} - e^{(t+1)a}}{(t+1)(e^b - e^a)}. \qquad (5.3.18)$$

The exact solution is $f(x) = e^x/(e^b - e^a)$. Consider two equations for which $[a, b]$ is $[1, 5]$ and $[11, 15]$, respectively. The distributions associated with these two equations are identical except for the locations of their domains. To obtain approximations for their density functions, we compute their standardized equations, (5.3.16). Table 5.5 shows $r(6)$, the exact

Table 5.5: Orders of accuracy, errors and error bounds of methods
for the standardized equations of (5.3.18).

N	$[a,b] = [1,5]$			$[a,b] = [11,15]$		
	r(6)	Exact	Error Bound	r(6)	Exact	Error Bound
9	4	5.30×10^{-7}	3.13×10^{-6}	4	5.30×10^{-7}	3.13×10^{-6}
15	6	1.58×10^{-8}	3.64×10^{-7}	6	1.58×10^{-8}	3.64×10^{-7}
19	7	7.49×10^{-9}	1.24×10^{-7}	7	7.49×10^{-9}	1.24×10^{-7}

error and the error bound (5.2.9) given by the following three methods for the standardized

equations: $N = 9$ with T a set of 9 equally spaced points on $[-2, 2]$; $N = 15$ with T a set of 15 equally spaced points on $[-6, 6]$; and $N = 19$ with T a set of 21 equally spaced points on $[-9, 9]$. These methods all use Gauss-Legendre rules. The 19-point method is the optimal method, and is very accurate for both equations. Also, the order of accuracy, the exact error and the error bound did not vary with the location of the domain. This reflects the fact that the two equations are in theory the same. Numerically, since a difference in two expressions of a certain function may lead to a difference in values corresponding to the two expressions, there may be a difference in values of the right-hand sides of the equations. Such a difference, however, is usually negligible and does not affect the accuracy of the methods and numerical solutions as can be seen in the above example.

To compute the density function at points other than the abscissae of the quadrature rule involved, one may use the method of interpolation. The abscissae of the Gauss-Legendre rules are very evenly spread in $[-1, 1]$. With highly accurate numerical solutions at these abscissae, accurate approximations throughout the entire domain can be obtained.

The optimal method for (5.1.1) may also be used to compute the cumulative distribution function, $F(x)$, corresponding to $f(x)$. To see how, we rewrite (5.1.1) as

$$\int_a^b e^{tx} F'(x) dx = M(t). \tag{5.3.19}$$

Integrating the left-hand side by parts, we obtain

$$\int_a^b e^{tx} F(x) dx = \frac{e^{tb} - M(t)}{t}. \tag{5.3.20}$$

The above equation has the same kernel as equation (5.1.1). Thus the optimal method for (5.1.1) is also optimal for this equation, and can thus be used to compute $F(x)$. As in the case of computing $f(x)$, we recommend that the standardized equation for (5.3.20) be used instead of (5.3.20) itself. This standardized equation is:

$$\int_{-1}^1 e^{t[\frac{(b-a)y}{2}]} F\left[\frac{(b-a)y + (a+b)}{2}\right] dy = \frac{2M^*(t)}{b-a} e^{-t[\frac{(a+b)}{2}]}, \tag{5.3.21}$$

where $M^*(t) = [e^{tb} - M(t)]/t$.

5.3.2 Computing densities with domain $[a, +\infty)$

The optimal method is mainly intended for (5.1.1) whose domain is a finite interval. This is because polynomials form a dense subspace in the space of continuous functions on any finite interval domain, and can approximate any continuous density function defined on the domain well. Hence we can use the P_h's to define the optimal method and expect it to be accurate for computing (5.1.1). When the domain is unbounded, e.g., $[a, +\infty)$, the polynomials are not in the solution space of equation (5.1.1). This is clear from the fact that no polynomial integrates to one over $[a, +\infty)$. They cannot even approximate a density function well due to that they either remain a constant, or approach positive or negative infinity as x approaches infinity. Thus they can no longer be used to define the optimal method for computing (5.1.1). Consequently, the optimal method is difficult to define unless some information, e.g., a basis of a space to which the solution belongs, are known.

Nevertheless, there are quadrature methods for dealing with situations where such information is not available. Bellman *et al.* (1966) studied quadrature methods for inverting Laplace transformations of continuous functions on $[a, +\infty)$. In this section, we discuss the use of their methods to solve (5.1.1), and demonstrate through examples that they are capable of delivering accurate numerical solutions.

Without loss of generality, we may assume that $a = 0$. So (5.1.1) becomes

$$M(t) = \int_0^\infty e^{tx} f(x) dx. \tag{5.3.22}$$

Note that in general $M(t)$ is not defined on the entire real line, rather it is defined on $(-\infty, t_0)$ for some $t_0 \geq 0$. Bellman *et al.* (1966) recommended using the transformation $y = e^{-x}$ to transform the equation into one with a bounded domain. With this transformation, $[0, \infty)$ is transformed onto $(0, 1]$, and (5.3.22) becomes

$$M(t) = \int_0^1 y^{-t-1} f(-\log(y)) dy. \tag{5.3.23}$$

A quadrature method can now be applied to (5.3.23) to compute $f(x)$ on a set of points $x_i = -\log(y_i)$ for $i = 1, 2, \ldots, N$ where the y_i's are the abscissae of the quadrature rule involved. Since the domain of $f(x)$ is unbounded, it may not be enough to compute $f(x)$ only at some finite number of points. They suggested the following argument which makes

it possible to compute $f(x)$ at any point. By (5.3.22),

$$
\begin{aligned}
\frac{M(t/c)}{c} &= \frac{1}{c} \int_0^\infty e^{tx/c} f(x) dx \\
&= \int_0^1 y^{-t-1} f(-c \times \log(y)) dy,
\end{aligned} \tag{5.3.24}
$$

where $y = e^{-x/c}$ and $c > 0$. If we need to estimate $f(x)$ at, say $x = 5$, let y_i be an abscissa of a certain quadrature rule, and choose c such that $5 = -c \times \log(y_i)$. Applying a quadrature method that uses this quadrature rule to (5.3.24), we obtain 5 as one of the points at which $f(x)$ is computed.

As to what type of quadrature method is best for equations of the form of (5.3.23) or (5.3.24), they recommended methods involving Gauss rules on [0,1] whose abscissae are the roots of the shifted Legendre polynomials.

Example 5.4: Let X be the sum of two independently exponentially distributed random variables with parameter 1. Then its density function and distribution function satisfy

$$
\int_0^1 y^{-t-1} f(-\log(y)) dy = \frac{1}{(1-t)^2} \qquad t < 1, \tag{5.3.25}
$$

and

$$
\int_0^1 y^{-t-1} F(-\log(y)) dy = \frac{-1}{t(1-t)^2} \qquad t < 0. \tag{5.3.26}
$$

Using methods based on Gauss rules with $T = -(0, 1, \ldots, 17)$ and $T = -(1, 2 \ldots, 18)$, respectively, we computed (5.3.25) and (5.3.26). The exact and the numerical solutions to (5.3.25) are in Table 5.6 under $f(x)$ and $\hat{f}(x)$, respectively. Those to (5.3.26) are in Table 5.7 under $F(x)$ and $\hat{F}(x)$, respectively. The c values associated with the numerical solutions are also indicated in the tables.

The numerical results indicate that the magnitude of the maximum absolute error of the numerical solutions is about 10^{-6}. Thus the quadrature method is quite accurate for this example. The tables contain numerical solutions at x values as large as 21, for which the density is as small as 10^{-8}, and the cumulative distribution function is as close to one as $1 - 10^{-8}$. To estimate these functions farther out in the tail, one may use even larger c values. We caution however that when c is too large, e.g., in the above example when c is greater than 5, the numerical solution at the tail, though very close to zero, may be negative

Table 5.6: Approximations for the density in Example 5.4 ($c = 4$).

x	$f(x)$	$\hat{f}(x)$	abs. error
0.21430e+02	0.10566e-07	0.81429e-08	0.24239e-08
0.11261e+02	0.14479e-03	0.14475e-03	0.35949e-07
0.70605e+01	0.60602e-02	0.60601e-02	0.99036e-07
0.45033e+01	0.49862e-01	0.49861e-01	0.17334e-06
0.27725e+01	0.17328e+00	0.17328e+00	0.23793e-06
0.15685e+01	0.32680e+00	0.32680e+00	0.27652e-06
0.75093e+00	0.35438e+00	0.35438e+00	0.27875e-06
0.24699e+00	0.19293e+00	0.19293e+00	0.23705e-06
0.18893e-01	0.18539e-01	0.18539e-01	0.13232e-06

in the case of computing $f(x)$, and slightly greater than one in the case of computing $F(x)$. As $f(x)$ is expected to be very close to zero and $F(x)$ very close to one at the far tail, they should be treated as zero and one, respectively.

Table 5.7: Approximations for the c.d.f in Example 5.4 ($c = 4$).

x	$F(x)$	$\hat{F}(x)$	abs. error
0.21430e+02	0.99999e+00	0.10000e+01	0.52861e-07
0.11261e+02	0.99984e+00	0.99984e+00	0.10569e-06
0.70605e+01	0.99308e+00	0.99308e+00	0.13486e-06
0.45033e+01	0.93906e+00	0.93906e+00	0.14613e-06
0.27725e+01	0.76421e+00	0.76421e+00	0.14313e-06
0.15685e+01	0.46484e+00	0.46484e+00	0.13012e-06
0.75093e+00	0.17368e+00	0.17368e+00	0.10997e-06
0.24699e+00	0.25916e-01	0.25916e-01	0.83615e-07
0.18893e-01	0.17625e-03	0.17629e-03	0.44330e-07

Note that one may also use Bellman *et al.*'s (1966) methods to compute distributions with finite domain. However, they are less accurate than the optimal method, and give no error bounds on their numerical solutions.

5.3.3 Comparison with the saddlepoint approximation

We have shown that the two quadrature methods discussed above are promising methods for computing the distribution of the sample mean. Such numerical methods are intended just for that and are of little use as theoretical tools.

In terms of accuracy, these quadrature methods are complements to asymptotic methods for computing the distribution of the mean in that they are more accurate overall for very small sample sizes, but less so for larger sample sizes. This is because quadrature methods, like numerical methods in general, are accurate when the exact solutions are smooth and are less accurate when they are not. When the sample size is small, the density function and the cumulative distribution function for the mean are relatively smooth. When the sample size is large, the density function is highly aggregated in a small region in its domain near the mean. It has a sharp spike, and the cumulative distribution function takes a steep jump in that region. Consequently, they are not smooth when viewed over the entire domain, and the quadrature methods become less accurate. Tables 5.8 and 5.9 below further illustrate this point.

Table 5.8: Approximations to the density of the mean of
5 i.i.d. uniform$[-1, 1]$ observations.

x	$f(x)$	$\hat{f}_{OM}(x)$	$\hat{f}_{SPA}(x)$	OM Error	SPA Error
0.00000e+00	0.14974e+01	0.14978e+01	0.15451e+01	0.36147e-03	0.47701e-01
0.28802e+00	0.85284e+00	0.85309e+00	0.88310e+00	0.24184e-03	0.30251e-01
0.55162e+00	0.16436e+00	0.16446e+00	0.17197e+00	0.10075e-03	0.76103e-02
0.76844e+00	0.11699e-01	0.11716e-01	0.11946e-01	0.17433e-04	0.24747e-03
0.92010e+00	0.16584e-03	0.16895e-03	0.16862e-03	0.31138e-05	0.27835e-05
0.99375e+00	0.62002e-08	0.82791e-06	0.63043e-08	0.82171e-06	0.10406e-09

Table 5.9: Approximations to the density of the mean of
15 i.i.d. uniform$[-1, 1]$ observations.

x	$f(x)$	$\hat{f}_{OM}(x)$	$\hat{f}_{SPA}(x)$	OM Error	SPA Error
0.00000e+00	0.26493e+01	0.26484e+01	0.26762e+01	0.92946e-03	0.26892e-01
0.28802e+00	0.42186e+00	0.42127e+00	0.42654e+00	0.58821e-03	0.46757e-02
0.55162e+00	0.18175e-02	0.16354e-02	0.18413e-02	0.18215e-03	0.23751e-04
0.76844e+00	0.19533e-06	-0.42294e-04	0.19692e-06	0.42489e-04	0.15981e-08
0.92010e+00	0.66255e-13	-0.10954e-04	0.66624e-13	0.10954e-04	0.36905e-15
0.99375e+00	0.21170e-28	-0.32721e-05	0.21288e-28	0.32721e-05	0.11792e-30

In these tables, $f(x)$ is the exact solution, $\hat{f}_{OM}(x)$ and $\hat{f}_{SPA}(x)$ are approximations given by the optimal method and the method of saddlepoint approximation. 'OM Error' and 'SPA Error' are the absolute errors of these two approximations, respectively. The optimal method used to compute $\hat{f}_{OM}(x)$ has $r(6) = 7$, and involves the Gauss-Legendre

rule with $N = 21$, and a T consisting of 21 equally spaced points between -20 and 20. When the sample size is 5, the optimal method is more accurate overall. When the sample size is 15, $\hat{f}_{OM}(x)$ becomes negative at the tail and the saddlepoint approximation becomes more accurate overall. In general, the quadrature methods are more accurate in the part of the domain where $f(x)$ is not too small relative to its maximum. If $f(x)$ does not vary too much in magnitude throughout the domain, as is the case in Examples 5.1 and 5.2, then they are more accurate than the asymptotic methods everywhere.

Apart from the optimal method being more accurate for very small sample sizes, quadrature methods in general also have the following advantages: (1) They are applicable whenever the moment generating function is known. Consider the case where we need to estimate the density of the sum of several independent observations. If these observations are not identically distributed, then asymptotic methods may not be applicable but quadrature methods still are. (2) Computationally, they are often less time consuming than the asymptotic method which involve computing the saddlepoint since they require only that the moment generating function be evaluated at a small number of points. In cases such as estimating the density functions of statistics given by estimating equations, the moment generating functions of the statistics often can only be expressed as contour integrals and their evaluation may be time consuming. In these cases, this advantage is particularly important. (3) For densities with bounded domains, the optimal method also gives a reliable bound on the maximum absolute error of its numerical solution.

To conclude, we note that there are many other numerical methods, such as the eigenfunction expansion method and a varieties of iterative methods, for solving Fredholm equations of the first kind. We have discussed only quadrature methods here because they are the simplest of all methods for solving these equations, and may be easily used by statisticians to compute distributions. They are also very accurate for this purpose. The use of quadrature methods for computing distributions defined on the entire real line remains to be further studied. We have yet to find successful methods for computing these distributions. Examples in this paper were computed using Fortran 77 on a SPARC station. The linear system solver used was DLSARG, and least-square estimates were computed by DRCURV, both from IMSL.

Appendix A

Uniform Validity of an

Asymptotic Expansion

Here we give the definition for uniform validity, and discuss two generalizations of Lemma 3.1 (Wasow, 1965) under weaker conditions.

Definition. *Uniform Validity of An Asymptotic Series*

Let $f(s,t)$ be defined on $D_s \times D_t$ where $0 \in D_s$. We say that an asymptotic expansion of $f(s,t)$ with respect to s,

$$f(s,t) \sim \sum_{r=0}^{\infty} h_r(t)s^r \quad as \quad s \to 0,$$

is uniformly valid in D_t if for any given $m \in N$,

$$s^{-m}\left[f(s,t) - \sum_{r=0}^{m} h_r(t)s^r\right] \to 0 \quad as \quad s \to 0 \tag{A.0.1}$$

uniformly with respect to all $t \in D_t$.

An important lemma related to uniform validity is Lemma 3.1 given below (see Section 3.3 for a proof). This lemma shows that when the specified conditions are met, the error of the truncated expansion which includes only the first $m + 1$ terms of the expansion is the

product of a bounded function and s^{m+1}.

Lemma 3.1. *(Wasow, 1965) Let $f(s,t)$ be bounded in $D_s \times D_t$ where $0 \in D_s$, and $h_r(t)$ (r = 0, 1, ...) be bounded in D_t. Then*

$$f(s,t) \sim \sum_{r=0}^{\infty} h_r(t)s^r \quad as \quad s \to 0 \tag{A.0.2}$$

uniformly for $t \in D_t$ iff for every m the function $E_m(s,t)$ defined by the relation

$$f(s,t) = \sum_{r=0}^{m} h_r(t)s^r + E_m(s,t)s^{m+1} \tag{A.0.3}$$

is bounded in $D_s \times D_t$.

Under a weaker set of conditions, i.e., without the condition that $f(s,t)$ be bounded in $D_s \times D_t$, we can prove the following lemma:

Lemma 3.1a: *Expansion (A.0.2) is uniformly valid and its coefficients, $h_r(t)$ (r = 0, 1, ...), are each bounded for $t \in D_t$ if and only if $f(s,t)$ is bounded in $\Omega \times D_t$ for some Ω such that $0 \in \Omega \subseteq D_s$, and for each m, $E_m(s,t)$ is bounded in $\Omega \times D_t$.*

Proof: For any given $\epsilon > 0$, by the definition of the uniform validity there exists a Ω where $0 \in \Omega \subseteq D_s$ such that $|f(s,t) - h_0(t)| < \epsilon$ uniformly with respect to $t \in D_t$ for $s \in \Omega$. Thus $|f(s,t)| < |h_0(t)| + \epsilon$ for $(s,t) \in \Omega \times D_t$. Since $h_0(t)$ is bounded in D_t, $f(s,t)$ is bounded in $\Omega \times D_t$. It then follows from Lemma 3.1 that the $E_m(s,t)$'s are bounded in $\Omega \times D_t$.

On the hand, if $f(s,t)$ and the $E_m(s,t)$'s are all bounded in $\Omega \times D_t$, then $h_0(t) = f(s,t) - E_0(s,t)s$ is bounded. Furthermore, $h_{r+1}(t) = E_r(s,t) - E_{r+1}(s,t)s$ (r = 0, 1, ...). Thus they are bounded. Also, the boundedness of the $E_m(s,t)$'s imply the uniform validity of the expansion. This is clear from part (b) of Wasow's proof for Lemma 3.1. #

Lemma 3.1a may be used to prove a lemma similar to Lemma 3.2 but without the condition that $f_t(s,t)$ be bounded. Compared to Lemma 3.1, Lemma 3.1a makes use of the fact that the conditions that $f(s,t)$ be bounded and $h_0(t)$ be bounded are essentially redundant, and it is applicable when $f(s,t)$ is not bounded in the entire $D_s \times D_t$. In general, if we define $E_{-1}(s,t) = f(s,t)$, then the boundedness of $\{E_r(s,t)\}_{r \geq m_0}$, where m_0

is an integer, depends only on that of $\{h_r(t)\}_{r \geq m_0 + 1}$ and vise versa. This leads to the following lemma:

Lemma 3.1b: *The $h_m(t)$'s where $m \geq m_0 + 1$ are each bounded in D_t and (A.0.1) holds for each $m \geq m_0$ uniformly in D_t if and only if for each $m \geq m_0$, there exists a subset of D_s, Ω_m, where $0 \in \Omega_m$, such that $E_m(s,t)$ is bounded in $\Omega_m \times D_t$.*

Proof: From part (a) of Wasow's proof for Lemma 3.1, for any given $\epsilon > 0$, there exists a $\delta_m(\epsilon) > 0$ such that

$$|E_m(s,t)| = \left| s^{-(m+1)} \left[f(s,t) - \sum_{r=0}^{m} h_r(t) s^r \right] \right| < \varepsilon + |h_{m+1}(t)|, \quad \text{if } |s| \leq \delta_m(\varepsilon).$$

Let $\Omega_m = \{s : |s| < \delta_m(\epsilon)\}$. Then the boundedness of $h_{m+1}(t)$ in D_t implies that of $E_m(s,t)$ in $\Omega_m \times D_t$. On the other hand, $h_{m+1}(t) = E_m(s,t) - E_{m+1}(s,t)s$. Thus for $m \geq m_0$ the boundedness of the $E_m(s,t)$'s implies that of the $h_{m+1}(t)$'s. It also implies that (A.0.1) holds uniformly. This is again clear from part (b) of Wasow's proof for Lemma 3.1. #

A consequence of Lemma 3.1a is that a uniform expansion with bounded coefficients may be written using the O symbol as

$$f(s,t) = \sum_{r=0}^{m} h_r(t) s^r + O(s^{m+1}) \quad m = 0, 1, \dots,$$

where $O(s^{m+1})$ is uniformly valid for $t \in D_t$.

Appendix B

An Identity Concerning Hermite Polynomials

We now prove identity (4.3.22). The Hermite polynomials, $H_k(x)$ $(k = 0, 1, \ldots)$, can be derived by repeatedly differentiating the density of the standard normal distribution, $\phi(x)$. Specifically, they are given by the following relationship

$$\frac{d^k \phi(x)}{dx^k} = (-1)^k H_k(x)\phi(x), \qquad (B.0.1)$$

where, by convention, $H_0(x) = 1$. It can be shown that the $H_k(x)$'s form a set of orthogonal polynomials in R with respect to the weight function, $\phi(x)$. Also, $H_k(x)\phi(x)$ has the Fourier transform $(it)^k e^{-t^2/2}$. See, e.g., Feller (1966), for a proof. We now discuss two properties of these polynomials which we shall use to prove the identity. These two properties are both derived from the equation,

$$\exp\{tx - \frac{1}{2}t^2\} = \sum_{k=0}^{\infty} \frac{t^k}{k!} H_k(x), \qquad (B.0.2)$$

which may be proved by noting that

$$\phi(x - t) = \frac{1}{\sqrt{2\pi}} \exp\{-\frac{1}{2}x^2 + tx - \frac{1}{2}t^2\} = \phi(x)\exp\{tx - \frac{1}{2}t^2\},$$

106

and that, by Taylor's theorem,

$$\phi(x - t) = \sum_{k=0}^{\infty} \frac{(-1)^k}{k!} t^k \frac{d^k \phi(x)}{dx^k} = \sum_{k=0}^{\infty} \frac{t^k}{k!} H_k(x)\phi(x).$$

The first property we need is the following differential equation,

$$\frac{d}{dx} H_k(x) = k H_{k-1}(x), \quad \text{for} \quad k = 1, 2, \ldots \tag{B.0.3}$$

This may be proved by differentiating both sides of (B.0.2) with respect to x and identifying coefficients for t^k. See, e.g., Kendall and Stuart (1969).

The second property is the following relationship between the constant term of $H_k(x)$ and the kth moment of the standard normal distribution. By setting x to zero in (B.0.2) we obtain

$$\exp\{\frac{1}{2}(it)^2\} = \sum_{k=0}^{\infty} \frac{t^k}{k!} H_k(0) = \sum_{k=0}^{\infty} \frac{(-i^2 t)^k}{k!} H_k(0) = \sum_{k=0}^{\infty} \frac{(it)^k}{k!} (-i)^k H_k(0). \tag{B.0.4}$$

The left-hand side of (B.0.4) is the moment generating function of the standard normal distribution evaluated at it. It follows that

$$H_k(0) = (-i)^{-k} m_k, \tag{B.0.5}$$

for $k = 0, 1, \ldots$, where m_k is the kth moment of the standard normal distribution. Since $m_k = 0$ when k is odd, (B.0.5) may be written as

$$H_k(0) = (i)^k m_k. \tag{B.0.6}$$

To prove (4.3.22), we show that if P_k satisfies

$$\int_{-\infty}^{\infty} \phi(v)(iv - x)^k dv = (-1)^k P_k(x), \tag{B.0.7}$$

for $k = 0, 1, \ldots$, then $P_k(x) = H_k(x)$. We first note that

$$P_k(0) = (-1)^k \int_{-\infty}^{\infty} \phi(v)(iv)^k = (-1)^k (i)^k m_k. \tag{B.0.8}$$

Again, since $m_k = 0$ when k is odd, (B.0.8) may be written as

$$P_k(0) = (i)^k m_k. \tag{B.0.9}$$

Furthermore, by differentiating (B.0.7) with respect to x we obtain

$$(-1)^k \frac{dP_k(x)}{dx} = -k \int_{-\infty}^{\infty} \phi(v)(iv - x)^{k-1} dv = k(-1)^k P_{k-1}(x),$$

for $k = 1, 2, \ldots$ Thus

$$\frac{d}{dx} P_k(x) = k P_{k-1}(x), \tag{B.0.10}$$

for $k = 1, 2, \ldots$ It follows that $H_k(x)$ and $P_k(x)$ are the solutions of the same differential equation (B.0.3) or (B.0.10). Furthermore, $H_0(x) = P_0(x) = 1$ and $H_k(0) = P_k(0) = (i)^k m_k$, by induction $P_k(x) = H_k(x)$ for $k = 0, 1, \ldots$

Bibliography

[1] Baker, C. T. H. (1977). *The Numerical Treatment of Integral Equation.* Clarendon Press, Oxford.

[2] Barndorff-Nielsen, O. E. (1978). *Information and Exponential Families in Statistical Theory.* Chichester: Wiley.

[3] Barndorff-Nielsen, O. E. and Cox, D. R. (1979). Edgeworth and saddle-point approximation with statistical applications. *J. R. Statist. Soc.,* **41**, No. 3, pp. 279-312.

[4] Barndorff-Nielsen, O. E. and Cox, D. R. (1989). *Asymptotic Techniques for Use in Statistics.* New York: Chapman and Hall.

[5] Barndorff-Nielsen, O. E. and Kluppelberg, C. (1992). A note on the tail accuracy of the univariate saddlepoint approximation. *Annal. Fac. Sci.,* Toulouse Series 6, Vol. I, pp. 5-14.

[6] Bellman, R.E., Kalaba, R.E. and Lockett, J.A. (1966). *Numerical Inversion of the Laplace Transform: Applications to Biology, Economics, Engineering, and Physics.* The Rand Corporation.

[7] Bhattacharya, R. N. and Ghosh, J. K. (1976). *Normal Approximation and Asymptotic Expansions.* New York: Wiley.

[8] Bickel, P. J., Götze, F. and van Zwet, W. R. (1986). The Edgeworth expansion for U-statistic of degree two. *Ann. Statist.,* **14**, pp. 1463-1484.

[9] Bleistein, N. and Handelsman, R. A. (1975). *Asymptotic Expansions of Integrals.* New York: Holt, Rinehart and Winston.

[10] Chaganty, N. R. and Sethuraman, J. (1985). Large deviation local limit theorems for arbitrary sequences of random variables. *Ann. Probab.,* **13**, pp. 97-114.

[11] Charlier, C. V. L. (1906). Über die Darstellung Willkürlicher Funktionen. *Ark. Mat. Astr. och Fys.,* Vol 2. **No. 20.**

[12] Copson, E. T. (1965). *Asymptotic Expansions.* Cambridge University Press.

[13] Cramér, H. (1946). *Mathematical Methods of Statistics.* Princeton: Princeton University Press.

[14] Cramér, H. (1962). *Random Variables and Probability Distributions*. Cambridge Tracts. 2nd edition.

[15] Daniels, H. E. (1954). Saddlepoint approximations in statistics. *Ann. Math. Statist.*, **25**, pp. 631–649.

[16] Daniels, H. E. (1958). Discussion of paper by D. R. Cox. *J. R. Statist. Soc.*, **B20**, pp. 236–238.

[17] Daniels, H. E. (1983). Saddlepoint approximations for estimating equations. *Biometrika*. **70**, pp. 89–96.

[18] Daniels, H. E. (1987). Tail probability approximations. *Int. Statist. Review*. **55**, pp. 37–48.

[19] Davis, P. J. (1984). *Methods of Numerical Integration*. 2nd ed. Orlando, Fla. : Academic Press.

[20] Davison, A. C. and Hinkley, D. V. (1988). Saddlepoint approximations in resampling methods. *Biometrika*. **75**, pp. 417–431.

[21] Delves, L. M. and Mohamed, J. L. (1985) *Computational Methods for Integral Equations*. Cambridge: Cambridge University Press.

[22] Easton, G. S. and Ronchetti, E. (1986). General saddlepoint approximations with applications to L statistics. *JASA*. **86**, pp. 420–430.

[23] Edgeworth, F. Y. (1905). The law of error. *Cambridge Philos. Trans.*, Vol. 20, pp. 33–66 and 113–141.

[24] Feller, W. (1966). *An Introduction to Probability Theory and Its Applications*. Vol. 2. New York: Wiley.

[25] Field, C. A. and Ronchetti, E. (1990). Small sample asymptotics. *Inst. Math. Statist*. Lecture Notes-Monograph Series Vol. 13. IMS, Hayward, CA.

[26] Field, C. A. (1993). Tail areas of linear combinations of chi-squares and non-central chi-squares. *J. Statist. Comput. Simul.*, **45**, pp. 243–248.

[27] Fisher, R. A. (1935). *The Design of Experiments*. Edinburgh: Oliver and Boyd.

[28] Hoeffding, W. (1948). A class of statistics with asymptotically normal distributions. *Ann. Math. Statist.*, **19**, pp. 293–325.

[29] Jeffreys, H. and Jeffreys, B. S. (1950). *Methods of Mathematical Physics*. Cambridge University Press.

[30] Jensen, J. L. (1988). Uniform saddlepoint approximations. *Adv. Appl. Prob.*, **20**, pp. 622–634.

[31] Jensen, J. L. (1991). Uniform saddlepoint approximations and log-concave densities. *J. R. Statist. Soc.*, **B53**, pp. 157–172.

[32] Jørgensen, B. (1986). Some properties of exponential dispersion models. *Scand. J. Statist.*, **13**, pp. 187–198.

[33] Jørgensen, B. (1987). Exponential dispersion models (with discussion). *J. Roy. Statist. Soc.*, **B 49**, pp. 127–162.

[34] Jørgensen, B. (1992). *The Theory of Exponential Models and Analysis of Deviance.* Lecture notes. School of Linear Models, University of São Paulo. 2nd edition. Monografias de Matemática #51, IMPA, Rio de Janeiro.

[35] Kawata, T. (1972) *Fourier Analysis in Probability Theory.* New York: Academic Press.

[36] Kendall, M. G. and Stuart, A. (1969). *Advanced Theory of Statistics.* Vol. I (3rd edn.) High Wycombe: Griffin.

[37] Lehmann, E. L. (1986). *Testing Statistical Hypothesis.* New York: Wiley.

[38] LePage, W. R. (1961). *Complex Variables and the Laplace Transform for Engineers.* New York: McGraw-Hill.

[39] Lugannani, R. and Rice, S. O. (1980). Saddlepoint approximation for the distribution of the sum of independent random variables. *Adv. Appl. Prob.*, **12**, pp. 475–490.

[40] Markushevich, A. I. (1965). *Theory of Functions of a Complex Variable.* Rev. Eng. ed. / translated and edited by Richard A. Silverman. Vol. I & II. Englewood Cliffs, N.J. : Prentice-Hall.

[41] Murray, J. D. (1974). *Asymptotic Analysis.* Oxford : Clarendon Press.

[42] Olver, F. W. J. (1974). *Asymptotics and Special Functions.* New York: Academic Press.

[43] Reid, N. (1988). Saddlepoint methods and statistical inference. *Statistical Science*, Vol. 3, No. 2, pp. 213–238.

[44] Robinson, J. (1982). Saddlepoint approximations for permutation tests and confidence intervals. *J. R. Statist. Soc.*, **B 44**, pp. 91–101.

[45] Rudin, W. (1966). *Real and Complex Analysis.* New York: McGraw-Hill.

[46] Seal, H. L. (1951). Spot the prior reference. *J. Inst. of Actuaries, Students' Soc.*, **10**, pp. 255–258.

[47] Skovgaard, I. M. (1986). On multivariate Edgeworth expansions. *Int. Statist. Review*, **54**, pp. 169-186.

[48] Sogge, C. (1993). *Fourier Integral in Classical Analysis.* Cambridge University Press.

[49] Srivastava, M. S. and Yau, W. K. (1988). Tail probability approximations of general statistics. Technical Report No. 88-38. Center for Multivariate Analysis, University of Pittsburgh.

[50] Titchmarsh, E. C. (1948). *Introduction to the Theory of Fourier Integrals.* 2nd ed., Oxford University Press.

[51] Wallace, D. L. (1958). Asymptotic approximations to distributions. *Ann. Math. Statist.*, **29**. pp. 635–654.

[52] Wasow, W. (1965). *Asymptotic Expansions for Ordinary Differential Equations.* New York: Wiley.