2. EDGEWORTH EXPANSIONS

2.1. INTRODUCTION

Given a central limit theorem result for a statistic, one may hope to obtain an estimate for the error involved, that is the difference between the exact distribution $F_n(t)$ of the standardized statistic and the standard normal distribution $\Phi(t)$. Typically, this result for the distribution of the sum of n iid random variables is known as the Berry-Esseen theorem and takes the form

$$\sup_t |F_n(t) - \Phi(t)| \leq \frac{C_0}{\sqrt{n}}, \tag{2.1}$$

where the constant C_0 depends on the statistic and on the underlying distribution of the observations but not on the sample size n. We will discuss this result for the simplest case and mention some generalizations in section 2.2.

The inequality (2.1) suggests a way to improve the approximation of F_n by considering a complete *asymptotic expansion* of the form

$$\sum_{j=0}^{\infty} \frac{A_j(t)}{n^{j/2}}, \tag{2.2}$$

where the error incurred by using the partial sum is of the same order of magnitude as the first neglected term,
i.e.

$$|F_n(t) - \sum_{j=0}^{r} \frac{A_j(t)}{n^{j/2}}| \leq \frac{C_r(t)}{n^{(r+1)/2}}. \tag{2.3}$$

Of course, in our case $A_0(t) = \Phi(t)$ the cumulative of the standard normal distribution and $C_0(t) \equiv C_0$ is the constant given by the Berry-Esseen theorem. Note that for any fixed n and t (2.2) may or may not exist and that we are just using the property (2.3) of the partial sums to approximate F_n. These asymptotic expansions are common in numerical analysis where they are used to approximate a variety of special functions, including Bessel functions. For a good discussion of theoretical and numerical aspects, see Henrici (1977), Ch. 11. The following example presents some typical numerical aspects of these approximations.

Example 2.1

We consider the approximation of the Binet function

$$J(z) = \log\Gamma(z) - \frac{1}{2}\log(2\pi) - (z - \frac{1}{2})\log z + z$$

through the partial sums of the asymptotic expansion

$$\sum_{j=1}^{\infty} \frac{B_{2j}}{2j(2j-1)} z^{-2j+1},$$

where $\Gamma(z)$ is the Gamma function and B_{2j} are the Bernoulli numbers given by

2j	2	4	6	8	10	12	...
B_{2j}	1/6	-1/30	1/42	-1/30	5/66	-691/2730	...

Exhibit 2.1 shows the error curves as functions of z.

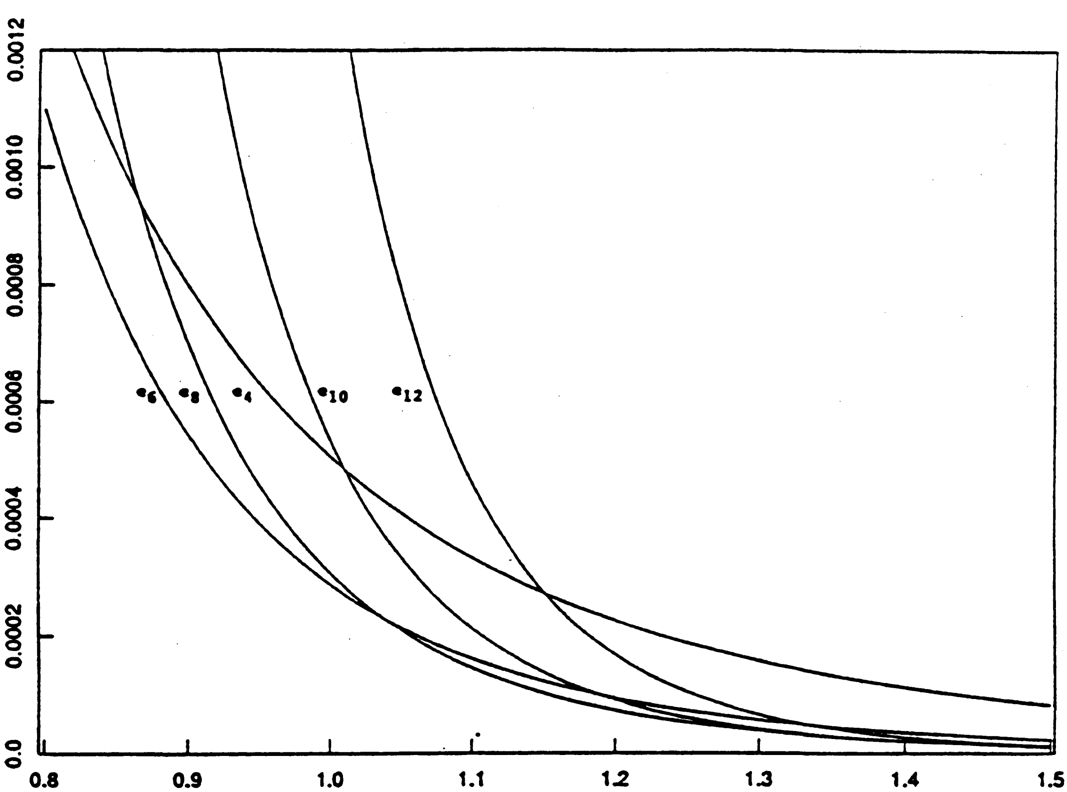

Exhibit 2.1

Error curves $e_r(z)$ for $r = 4, \cdots, 12$ in the approximation of the Binet function.

Exhibit 2.2 shows the error

$$e_r(z) = \left| J(z) - \sum_{j=1}^{r} \frac{B_{2j}}{2j(2j-1)} z^{-2j+1} \right|$$

as a function of the number of terms used (r) for various values of z.

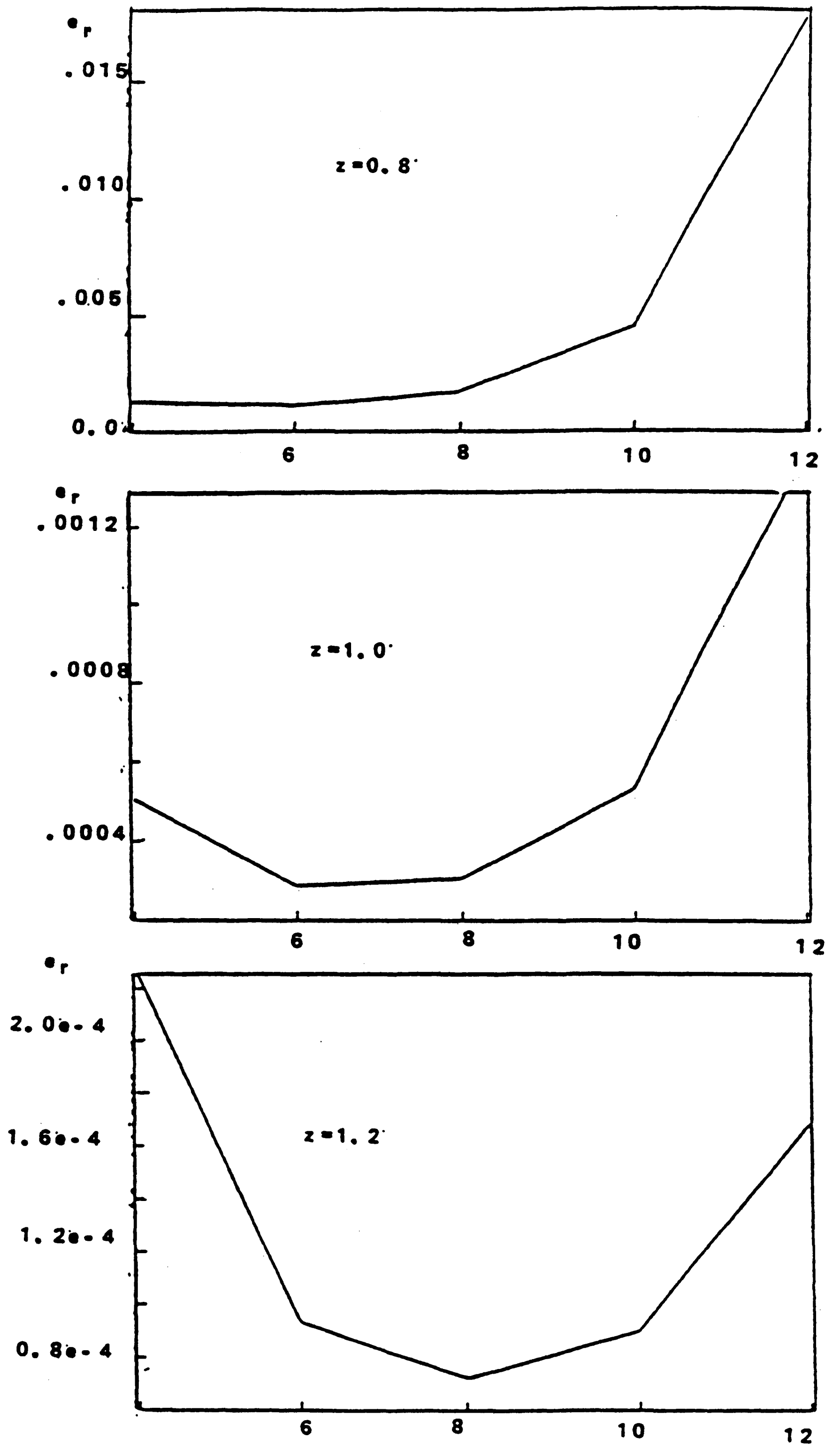

Exhibit 2.2

Error e_r as a function of the number of terms (r) in the asymptotic expansion for the Binet function $J(z)$.

Both exhibits show the same story. As we add more terms to the series the error first decreases. When the number of terms r reaches a certain point, adding more terms increases the error and an explosion takes place. The explosion point depends on z (which plays in this example the role of n in (2.2) and (2.3)) and increases as z increases. Thus, in a sense as $z \to \infty$ the series has a "convergent behavior" and represents $J(z)$ asymptotically.

Edgeworth expansions are expansions of the form (2.2) with the property (2.3) holding uniformly in t (i.e. C_r is independent of t) . They play a basic role in statistical theory and practice. Since they also form the basis of one approach to saddlepoint expansions (see sections 3.4, 5.2 and chapter 4), we present in section 2.3 and 2.4 a derivation for a simple case, namely the mean of n iid random variables. In sections 2.5 and 2.6 we then discuss briefly the huge amount of literature concerning Edgeworth expansions for general statistics both in univariate and multivariate situations. We conclude with some numerical examples in section 2.7.

2.2. CENTRAL LIMIT THEOREM AND BERRY-ESSEEN BOUND

The *central limit theorem* is the basic tool which drives asymptotic normality proofs. The resulting asymptotic distribution (the normal distribution) is used very often as an approximation to the exact distribution of some statistic T_n of interest. Nowadays, central limit theorems are available for very general spaces of the observations and for very general problems. It is not our goal here to review and discuss these results that can be found in Bhattacharya and Rao (1976) and Serfling (1980). Also Pollard (1985) describes some techniques (taken from the theory of empirical processes) for proving asymptotic normality under very general conditions. In particular, the usual heavy assumptions of higher-order pointwise differentiability of the underlying density function and/or the score function defining the statistic can be dispensed with for a very broad class of statistics.

Here we want to focus on the quality of the asymptotic normal distribution as an approximation of the exact distribution. Whereas in many situations this approximation is accurate even for moderate sample sizes n, there are cases where it is inaccurate in the tails even for large n. The following example taken from Ritcey (1985) illustrates this point. The statistic of interest is

$$T_n = \frac{1}{2}\sum_{i=1}^{n}(x_i^2 + y_i^2),$$

where x_i and y_i are independent standard normal random variables. This test statistic is used for instance in signal detection problems where the tail probabilities of interest $P[T_n > t_0]$ are in the range $10^{-9} - 10^{-10}$ (false alarm probability). Exhibit 2.3 shows the exact probabilities (calculated from a $\frac{1}{2}\chi^2_{2n}$ distribution) and the normal approximation for various values of n and t_o. It is clear that in this case the normal approximation is useless far out in the tails even for large n.

n	t_o	P	$\hat{P}$	Relative error (%)
10	15	$6.98 \cdot 10^{-2}$	$5.69 \cdot 10^{-2}$	18
	20	$4.99 \cdot 10^{-3}$	$7.83 \cdot 10^{-4}$	84
	25	$2.21 \cdot 10^{-4}$	$1.05 \cdot 10^{-6}$	99
100	125	$9.38 \cdot 10^{-3}$	$6.21 \cdot 10^{-3}$	34
	150	$5.92 \cdot 10^{-6}$	$2.87 \cdot 10^{-7}$	95
	175	$2.78 \cdot 10^{-10}$	$3.19 \cdot 10^{-14}$	99
500	550	$1.15 \cdot 10^{-2}$	$1.27 \cdot 10^{-2}$	10
	600	$1.23 \cdot 10^{-5}$	$3.87 \cdot 10^{-6}$	68
	625	$1.01 \cdot 10^{-7}$	$1.13 \cdot 10^{-8}$	89

Exhibit 2.3

Exact tail probabilities $P = P[T_n > t_0]$, the normal approximation $\hat{P}$, and the relative error $|P - \hat{P}|/P$ for the statistic $T_n = \frac{1}{2}\sum_{i=1}^{n}(x_i^2 + y_i^2)$, where x_i and y_i are independent standard normal random variables.

A first step in trying to improve the normal approximation is to assess the error involved. This is basically the Berry-Esseen bound. We first present this result for a simple case, namely the mean of n iid random variables.

Theorem 2.1 (Berry-Esseen)

Let $X_1, \cdots, X_n$ be n iid random variables with distribution F such that $EX_i = 0$, $EX_i^2 = \sigma^2 > 0$, $E|X_i|^3 = \rho < \infty$. Denote by F_n the distribution of the standardized statistic $n^{-1/2}\sum_{i=1}^{n} X_i/\sigma$. Then, for all n

$$\sup_t |F_n(t) - \Phi(t)| \leq \frac{3\rho}{\sigma^3\sqrt{n}} \tag{2.4}$$

Proof: see, for instance, Feller (1971), p. 543 ff.

This result was discovered (with two different proofs) by Berry (1941) and Esseen (1942). The surprising aspect is that the Berry-Esseen bound (the right hand side of inequality (2.4)) depends only on the first three moments of the underlying distribution.

Many generalizations of this result are available today. The best known constant (which replaces 3 in (2.4)) is 0.7975, see Bhattacharya and Rao (1976), p. 110. The result can be generalized to underlying distributions F without third moment and to non iid random variables. A Berry-Esseen theorem for U-statistics has been established under different sets of conditions and in increasing generality by Bickel (1974), Chan and Wierman (1977), Callaert and Janssen (1978), and Helmers and van Zwet (1982). Bjerve (1977) and Helmers (1977) proved the same result for linear combinations of order statistics. Finally, van Zwet (1984) proved a Berry-Esseen theorem for a broad class of statistics, namely symmetric functions of n iid random variables.

2.3. CHARLIER DIFFERENTIAL SERIES AND FORMAL EDGEWORTH EXPANSIONS

The inequality (2.4) suggests a way to improve the approximation of F_n by considering a complete asymptotic expansion in power of $n^{-1/2}$. This is in fact the idea of an Edgeworth expansion. Because of its conceptual simplicity, we choose in our exposition the approach via Charlier differential series as presented in Wallace (1958).

2.3.a Charlier Differential Series

Consider two distribution functions $H(x)$ and $G(x)$ with characteristic functions $\chi(u)$ and $\xi(u)$ and cumulants β_r and γ_r, $r = 1, 2, \cdots$. Recall that the r-th cumulant β_r is the r-th derivative at 0 of the cumulant generating function, i.e.

$$\beta_r = \frac{d^r}{du^r} \log \int e^{ux} dH(x)\Big|_{u=0} = (-i)^r \frac{d^r}{du^r} \log \chi(u)\Big|_{u=0}$$

Suppose that all derivatives of G vanish at the extremes of the range of x. Then, by formal Taylor expansion we have

$$\log \frac{\chi(u)}{\xi(u)} = \log \chi(u) - \log \xi(u) = \sum_{r=1}^{\infty} (\beta_r - \gamma_r) \frac{(iu)^r}{r!}$$

and

$$\chi(u) = exp\left\{ \sum_{r=1}^{\infty} (\beta_r - \gamma_r) \frac{(iu)^r}{r!} \right\} \cdot \xi(u). \tag{2.5}$$

By integration by parts we can easily see that $(iu)^r \xi(u)$ is the characteristic function of $(-1)^r G^{(r)}(x)$, and by Fourier inversion of (2.5) we obtain

$$H(x) = exp\left\{ \sum_{r=1}^{\infty} (\beta_r - \gamma_r) \frac{(-D)^r}{r!} \right\} \cdot G(x), \tag{2.6}$$

where D denotes the differential operator and $e^D := \sum_{j=0}^{\infty} D^j / j!$. We can now choose a distribution function G which we use to develop an expansion for H. In fact, given a developing function G and the cumulants β_r, one can formally obtain H by expanding the right hand side of (2.6). Such an expansion is called Charlier differential series from Charlier (1906).

A natural developing function (but by no means the only one) is the normal distribution. In this case $G(x) = \Phi(x)$ and $(-D)^r \varphi(x) = H_r(x) \cdot \varphi(x)$, where $\varphi(x)$ is the density of the standard normal distribution and $H_r(x)$ is the Hermite polynomial of degree r. Chebyshev (1890) and Charlier (1905) used the normal distribution and developed (2.6) by collecting terms according to the order of derivatives. A breakthrough came from Edgeworth (1905) who applied this expansion to the distribution of the sum of n iid random variables and expanded (2.6) using the normal distribution but by collecting terms according to the powers of n. In this way he obtained what we call today an Edgeworth expansion.

2.3.b Edgeworth Expansions

We now formally derive the Edgeworth expansion by applying (2.5) and (2.6) to the distribution of a standardized sum of n iid random variables. Let $X_1, \cdots, X_n$ be n iid random varibles with distribution $F(x)$ and $EX_i = 0$, $varX_i = \sigma^2 > 0$, and cumulants $\kappa_r(X_i) = \lambda_r\sigma^r, r \geq 3$. Denote by $F_n(t)$ the distribution function of $T_n = n^{-1/2}\sum_{i=1}^n X_i/\sigma$ and by $\psi_n(u)$ its characteristic function. We choose the standard normal distribution as developing function, that is $G(t) = \Phi(t)$, $\xi(u) = exp(-u^2/2)$, and $\gamma_1 = 0$, $\gamma_2 = 1$, $\gamma_r = 0$, $r \geq 3$. Finally, if β_r denotes the r-th cumulant of T_n we have

$$\beta_1 = 0, \quad \beta_2 = 1,$$
$$\beta_r = \kappa_r(T_n) = n^{-r/2}\sigma^{-r}n\kappa_r(X_1) = \lambda_r n^{-(r/2-1)},$$

for $r \geq 3$. By applying (2.5) we obtain

$$\psi_n(u) = \exp\left\{\sum_{r=3}^{\infty}\frac{\lambda_r}{n^{r/2-1}}\frac{(iu)^r}{r!}\right\}e^{-u^2/2}$$
$$= \exp\left\{\frac{\lambda_3}{\sqrt{n}}\frac{(iu)^3}{3!} + \frac{\lambda_4}{n}\frac{(iu)^4}{4!} + \frac{\lambda_5}{n^{3/2}}\frac{(iu)^5}{5!} + \cdots\right\}e^{-u^2/2}$$

and by expanding $\exp\{\cdots\}$ we get

$$\psi_n(u) = \left\{1 + \frac{1}{\sqrt{n}}\lambda_3\frac{(iu)^3}{3!} + \frac{1}{n}\left[\frac{1}{2}\lambda_3^2\frac{(iu)^6}{(3!)^2} + \lambda_4\frac{(iu)^4}{4!}\right] + \frac{1}{n^{3/2}}[\cdots] + \cdots\right\}e^{-u^2/2} \tag{2.7}$$

Finally, the Fourier inversion of (2.7) leads to the following expansion

$$F_n(t) = \Phi(t) + \sum_{r=3}^{\infty}\frac{P_r(-\Phi(t))}{n^{r/2-1}}, \tag{2.8}$$

where $P_r(\cdot)$ is a polynomial of degree $3(r-2)$ with coefficients depending only on $\lambda_3, \lambda_4, \cdots,$ λ_r. (The powers of these polynomials in (2.8) should be interpreted as derivatives.) For instance,

$$P_3(z) = \frac{\lambda_3}{6}z^3,$$
$$P_4(z) = \frac{\lambda_4}{24}z^4 + \frac{\lambda_3^2}{72}z^6.$$

For the density $f_n(t)$ one obtains the expansion

$$f_n(t) = \varphi(t) + \sum_{r=3}^{\infty}\frac{P_r(-\varphi(t))}{n^{r/2-1}}$$
$$= \varphi(t) - \frac{1}{\sqrt{n}}\frac{\lambda_3}{6}\varphi^{(3)}(t) + \frac{1}{n}\left[\frac{\lambda_4}{24}\varphi^{(4)}(t) + \frac{\lambda_3^2}{72}\varphi^{(6)}(t)\right] + \cdots$$
$$= \varphi(t) + \frac{1}{\sqrt{n}}\frac{\lambda_3}{6}H_3(t)\varphi(t) + \frac{1}{n}\left[\frac{\lambda_4}{24}H_4(t) + \frac{\lambda_3^2}{72}H_6(t)\right]\varphi(t) + \cdots, \tag{2.9}$$

where

$$H_3(t) = t^3 - 3t,$$
$$H_4(t) = t^4 - 6t^2 + 3,$$
$$H_6(t) = t^6 - 15t^4 + 45t^2 - 15$$

are the Hermite polynomials of order 3,4, and 6. (2.9) is called Edgeworth expansion for the density of T_n.

Remark 2.1
(2.9) is an expansion in powers of $n^{-1/2}$. Note however that for $t = 0$ (the expectation of the underlying distribution) all coefficients corresponding to odd powers disappear because $H_r(0) = 0$ when r is odd. In this case the Edgeworth expansion becomes a series in power of n^{-1}:

$$f_n(0) = \frac{1}{\sqrt{2\pi}}\left\{1 + \frac{1}{n}\left[\frac{\lambda_4}{8} - \frac{5}{24}\lambda_3^2\right] + 0(n^{-2})\right\}. \tag{2.10}$$

This fact will play a key role in the derivation of the saddlepoint approximation; cf. section 3.4.

Remark 2.2
A similar expansion can be obtained for the $(1-\alpha)$ quantile of the distribution. This is called Fisher-Cornish expansion (see Kendall and Stuart, 1977, p. 177-179 and Cox and Hinkley, 1974, p. 464–465) and takes the form

$$q^*_{1-\alpha} + \frac{1}{\sqrt{n}}\frac{\lambda_3}{6}(q^{*2}_{1-\alpha} - 1) + \frac{1}{n}\left[\frac{\lambda_4}{24}\left(q^{*3}_{1-\alpha} - 3q^*_{1-\alpha}\right)\right.$$
$$\left. - \frac{\lambda_3^2}{36}\left(2q^{*3}_{1-\alpha} - 5q^*_{1-\alpha}\right)\right] + \cdots, \tag{2.11}$$

where $q^*_{1-\alpha}$ is the $(1-\alpha)$ quantile of the standard normal distribution, $\Phi(q^*_{1-\alpha}) = 1 - \alpha$. The $1/\sqrt{n}$ and $1/n$ terms of this expansion can be interpreted as corrections to the normal quantiles taking into account skewness and kurtosis. Another method for inverting a general Edgeworth expansion is given by Hall (1983).

2.4. PROOF AND DISCUSSION

We present the basic idea of the proof for the density as given in Feller (1971), Ch. XVI; see also Cramer (1962).

Theorem 2.2

Let $X_1, \cdots, X_n$ be n iid random variables with common distribution F and characteristic function ψ. Let

$$EX_i = \mu_1 = 0, varX_i = \sigma^2 < \infty,$$

and $F_n(t) = P\left[n^{-1/2}\sum_{i=1}^{n} X_i/\sigma < t\right]$ with density $f_n(t)$. Suppose that the moments $\mu_3, \cdots, \mu_k$ exist and that $|\psi|^v$ is integrable for some $v \geq 1$. Then, f_n exists for $n \geq v$ and as $n \to \infty$

$$f_n(t) - \varphi(t) - \varphi(t)\sum_{r=3}^{k} P_r(t)/n^{r/2-1} = o\left(n^{-(k/2-1)}\right), \tag{2.12}$$

uniformly in t. Here P_r is a real polynomial of degree $3(r-2)$ depending only on $\mu_1, \cdots, \mu_r$ but not on n and k (or otherwise on F).

Proof: We give an outline of the proof with the basic idea for $k = 3$. In this case we have to show that

$$f_n(t) - \varphi(t) - \frac{1}{\sqrt{n}}\frac{\mu_3}{6\sigma^3}(t^3 - 3t)\varphi(t) = o\left(\frac{1}{\sqrt{n}}\right), \tag{2.13}$$

uniformly in t.

Since $\psi^n(\cdot/\sqrt{n}\sigma)$ is the characteristic function of $f_n(\cdot)$, we obtain by Fourier inversion of the left hand side of (2.13)

$$\begin{aligned} &|\text{left hand side of (2.13)}| \\ &\le \frac{1}{2\pi}\int_{-\infty}^{+\infty}\left|\psi^n\left(\frac{u}{\sqrt{n}\sigma}\right) - e^{-u^2/2} - \frac{1}{\sqrt{n}}\frac{\mu_3}{6\sigma^3}(iu)^3 e^{-u^2/2}\right|du \\ &=: N_n. \end{aligned} \tag{2.14}$$

We have to show that $N_n = o\left(1/\sqrt{n}\right)$ as $n \to \infty$. For a given $\delta > 0$, we split N_n in two parts $N_n = N_n^{(1)} + N_n^{(2)}$, where

$$N_n^{(1)} = \frac{1}{2\pi}\int |\cdots| \cdot 1_{\{|u|>\delta\sqrt{n}\sigma\}}du,$$

and

$$N_n^{(2)} = \frac{1}{2\pi}\int |\cdots| \cdot 1_{\{|u|<\delta\sqrt{n}\sigma\}}du.$$

Since $|\psi(u)| < 1$ for $|u| \neq 0$ and $\psi(u) \to 0$ as $|u| \to \infty$, it can be shown easily that $N_n^{(1)}$ tends to 0 faster than any power of n^{-1} as $n \to \infty$. The computation of $N_n^{(2)}$ requires a little more work. First rewrite $N_n^{(2)}$ as

$$\begin{aligned} N_n^{(2)} = \frac{1}{2\pi}\int e^{-u^2/2}\bigg|\exp\left\{n\eta\left(\frac{u}{\sqrt{n}\sigma}\right) - 1\right\} \\ - \frac{1}{\sqrt{n}}\frac{\mu_3}{6\sigma^3}(iu)^3\bigg| \cdot 1_{\{|u|<\delta\sqrt{n}\sigma\}}du, \end{aligned} \tag{2.15}$$

where $\eta(u) = \log\psi(u) + \frac{1}{2}\sigma^2u^2$. Then, expand $\eta(\cdot)$ in a Taylor series in a neighborhood of $u = 0$,

$$\eta(u) = \frac{\eta'''(u^*)}{6}\cdot u^3, \tag{2.16}$$

where u^* is a point in the interval $]0,\delta[$. Finally, (2.16) allows us to approximate the right hand side of (2.15) to get $N_n^{(2)} = o(1/\sqrt{n})$ and this completes the proof. □

From the expansion for the density f_n given in Theorem 2.2, one can obtain by integration an expansion for the cumulative distribution function F_n. However, this method is not available when the integrability condition on $|\psi|^\nu$ fails. In this case the following important "smoothing technique" can be used.

Lemma (Esseen's smoothing Lemma).

Let H be a distribution with zero expectation and characteristic function χ. Suppose $H - G$ vanishes at $\pm\infty$ and that G has a derivative g such that $|g| \leq m$. Finally, suppose that g has a continuously differentiable Fourier transform ξ such that $\xi(0) = 1$ and $\xi'(0) = 0$. Then,

$$|H(x) - G(x)| \leq \frac{1}{\pi}\int_{-T}^{T}\left|\frac{\chi(u)-\xi(u)}{u}\right|du + \frac{24m}{\pi T} \tag{2.17}$$

holds for all x and $T > 0$.

Proof: see Feller (1971), Ch. XVI, section 3.

The following theorem establishes the Edgeworth expansion for the cumulative distribution function.

Theorem 2.3

Let $X_1, \cdots, X_n$ be n iid random variables with common distribution F. Let

$$EX_i = 0,\quad varX_i = \sigma^2 < \infty,\quad \text{and}\quad F_n(t) = P\left[n^{-1/2}\sum_{i=1}^{n} X_i/\sigma < t\right].$$

If F is not a lattice distribution and if the third moment μ_3 of F exists, then

$$F_n(t) - \Phi(t) - \frac{1}{\sqrt{n}}\frac{\mu_3}{6\sigma^3}(1-t^2)\phi(t) = o(n^{-1/2})$$

uniformly in t.

Proof: (Feller, 1971, Ch. XVI) Define $G(t) = \Phi(t) + 1/\sqrt{n}\mu_3/6\sigma^3(1-t^2)\phi(t)$. Then G satisfies the conditions of Esseen's Lemma with $\xi(u) = [1 + \mu_3(iu)^3/(\sqrt{n}\cdot 6\sigma^3)]e^{-u^2/2}$.

We now apply inequality (2.17) with $T = a\sqrt{n}$, where the constant a is chosen such that $24|g(t)| < \epsilon \cdot a$ for all t and a given ϵ. Then

$$|F_n(t) - G(t)| \leq \int_{-a\sqrt{n}}^{a\sqrt{n}}\left|\frac{\psi^n(u/\sqrt{n}\sigma)-\chi(u)}{u}\right|du + \frac{\epsilon}{\sqrt{n}}. \tag{2.18}$$

Now we can apply the same arguments as in the proof of Theorem 2.2 to the integral on the right hand side of (2.18) and the result follows. □

We conclude this section with a discussion of some points on Edgeworth expansions.

1) If the underlying distribution F has moments of all orders, one is tempted to let $k \to \infty$ in (2.12). Unfortunately, the resulting infinite series need not converge for any n. In fact Cramér showed that this series converges for all n if and only if $e^{x^2/4}$ is integrable with respect to F; see Feller (1971), p. 542. This is in agreement with the discussion in section 2.1 In particular, adding higher order terms does not necessarily improve the approximation and can be disastrously bad; cf. Exhibits 2.1 and 2.2.
2) In section 2.3 we derived Edgeworth expansions via Charlier differential series by using the normal distribution as developing function. When the asymptotic distribution is not normal, similar expansions may be obtained by using the asymptotic distribution as developing function.
3) The approximation provided by the Edgeworth expansion is in general reliable in the center of the distribution for moderate sample sizes; see Remark 2.1, section 2.3.b and the examples in section 2.7. This makes it a suitable tool for local asymptotic theory. Unfortunately, the approximation deteriorates in the tails where it can even become negative; see section 2.7. Moreover, the absolute error is uniformly bounded over the whole range of the distribution, but the relative error is in general unbounded. This is in contrast with saddlepoint techniques which give approximations with uniformly small relative errors; cf. chapters 3, 4, and 6.

2.5. EDGEWORTH EXPANSIONS FOR GENERAL STATISTICS

In this section we want to discuss briefly some results on Edgeworth expansions for more complicated statistics than the arithmetic mean. Given the relative importance of U-statistics in the literature, we focus our presentation on this class of statistics.

Given a statistic T_n one can in principle still use the approach via Charlier differential series presented in section 2.3.a and go through the steps (2.5) to (2.8). However, there are two points which make life hard. The first one is the fact that in general the cumulants of the statistics $\kappa_r(T_n)$ cannot be expressed in terms of $\kappa_r(X_1)$, the cumulants of the underlying distribution of the observations. The second point is that the validity of (2.8) must be proven. These problems can be overcome by replacing the exact cumulants κ_r by approximations of order $n^{-(r/2-1)}$ and by using Esseen's smoothing Lemma to prove the validity of the resulting expansion. Most of the work lies in the estimation of the integral on the right hand side of (2.17) in order to obtain an upper bound of the appropriate order in n. Let us now see how this idea applies to U-statistics. Let $X_1, \cdots, X_n$ be n iid random variables with common distribution F. Then a one-sample *U-statistic of degree 2* is defined by

$$U_n = \binom{n}{2}^{-1} \sum_{1 \leq i < j < n} h(X_i, X_j), \tag{2.19}$$

where h is a symmetric function of two variables with $Eh(X_1, X_2) = 0$ and $Eh^2(X_1, X_2) < \infty$.

Define

$$g(x) = E[h(X_1, X_2)|X_1 = x],$$
$$\psi(x, y) = h(x, y) - g(x) - g(y),$$
$$\hat{U}_n = \frac{1}{n}\sum_{i=1}^{n} 2g(X_i), \text{ and}$$
$$\Delta_n = \frac{2}{n(n-1)} \sum_{i \le i < j \le n} \psi(X_i, X_j).$$

Then

$$U_n = \hat{U}_n + \Delta_n. \tag{2.20}$$

Note that $2g(x) = IF(x; U, F)$, *the influence function of* U, defined by

$$IF(x; U, F) = \lim_{\epsilon \to 0}[U((1-\epsilon)F + \epsilon\Delta_x) - U(F)]/\epsilon,$$

where $U(\cdot)$ is the functional defined by $U(F) = E_F(U_n)$ and Δ_x is the distribution which puts mass 1 at a point x; see Hampel (1968, 1974), Hampel et al. (1986), and section 7.3.a. Thus, (2.20) is the linear representation of U_n based on the influence function and corresponding von Mises expansion; see von Mises (1947).

U-statistics were introduced by Hoeffding (1948) who also proved the asymptotic normality. Note the important special case $h(x, y) = 1_{\{x<y\}}$ which defines the Wilcoxon statistic. Subsequently Berry-Esseen bounds for U-statistics were established by several authors; cf. section 2.2. Finally, an Edgeworth expansion of order $o(n^{-1})$ was derived by Callaert, Janssen, and Veraverbeke (1980) and more recently by Bickel, Götze, and van Zwet (1986). We will present the result of the latter paper where the assumptions appear to be very mild.

Define

$$\sigma_g^2 = Eg^2(X_1),$$
$$\lambda_3 = \sigma_g^{-3}\left\{Eg^3(X_1) + 3Eg(X_1)g(X_2)\psi(X_1, X_2)\right\},$$
$$\lambda_4 = \sigma_g^{-4}\left\{Eg^4(X_1) - 3\sigma_g^4 + 12Eg^2(X_1)g(X_2)\psi(X_1, X_2)\right.$$
$$\left. + 12Eg(X_1)g(X_2)\psi(X_1, X_3)\psi(X_2, X_3)\right\},$$
$$\sigma_n^2 = var(U_n) = var(\hat{U}_n) + var(\Delta_n)$$
$$= \frac{1}{n}4\sigma_g^2 + \frac{2}{n(n-1)}E\psi^2(X_1, X_2),$$

and

$$F_n(t) = P[U_n/\sigma_n < t].$$

Then we have the following theorem.

Theorem 2.4 (Bickel, Götze, van Zwet, 1986).

Suppose that there exist a number $r > 2$ and an integer k such that $(r-2)(k-4) > 8$ and that the following assumptions are satisfied

$$E\left|\psi(X_1, X_2)\right|^r < \infty,$$

$$E\left|g(X_1)\right|^4 < \infty,$$

$$\limsup_{|t|\to\infty} \left|E\ exp\{itg(X_1)\}\right| < 1.$$

Let $|\lambda_1| \geq |\lambda_2| \geq \cdots$ be the eigenvalues of the kernel ψ, that is,

$$\int \psi(x_1, x_2) w_j(x_1) dF(x_1) = \lambda_j \cdot w_j(x_2), j = 1, 2, \cdots$$

and suppose that there exist k nonzero eigenvalues.

Then

$$F_n(t) - \Phi(t) + \phi(t)\left\{\frac{1}{\sqrt{n}}\frac{\lambda_3}{6}(t^2-1) + \frac{1}{n}\frac{\lambda_4}{24}(t^3-3t) + \frac{1}{n}\frac{\lambda_3^2}{72}(t^5 - 10t^3 + 15t)\right\} = o(n^{-1})$$

uniformly in t.

Note that $\lambda_3/\sqrt{n}$ and λ_4/n are approximations with error $o(n^{-1})$ to the standard third and fourth cumulant of U_n/σ_n. Basically the conditions on $g(\cdot)$ in Theorem 2.4 establish an Edgeworth expansion for the *linearized statistic* $\hat{U}_n$ while the moment assumption on $\psi(\cdot,\cdot)$ allows to correct the expansion for the remainder term Δ_n in (2.20).

Results on Edgeworth expansions for other classes of statistics abound. Bickel (1974) gives a complete account of the literature up to 1974. A basic paper is Albers, Bickel, and van Zwet (1976). Results on L-statistics are discussed in Helmers (1979, 1980) and van Zwet (1979) and will be the starting point in section 4.4. More references can be found in Skovgaard (1986), p. 169-170 and Hall (1983).

2.6. MULTIVARIATE EDGEWORTH EXPANSIONS

The techniques used to derive Edgeworth expansions for the distribution of one - dimensional statistics can be generalized to the multivariate case. The basic idea, i.e. expansion of the characteristic function and Fourier inversion, is the same. However, in the multivariate case the notation becomes more complex. The Edgeworth expansion for the multivariate mean is discussed, among others, in Barndorff-Nielsen and Cox (1979), Skovgaard (1986), and McCullagh (1987). Explicit expressions for the multivariate Hermite polynomials appearing in the terms of the expansion are given in Grad (1949), Barndorff-Nielsen and Pedersen (1979), and Holly (1986).

A basic result on multivariate Edgeworth expansions is that of Bhattacharya and Ghosh (1978). Given a sequence of n iid m-dimensional random vectors $Y_1, \cdots, Y_n$ and a sequence of k real functions $f_1, \cdots, f_k$ on $\mathbf{R}^m$, Bhattacharya and Ghosh derive the rates of convergence to normality and asymptotic expansions of the distribution of statistics of the form $W_n = \sqrt{n}(H(\bar{Z}) - H(\mu))$, where $Z_i = (f_1(Y_i), \cdots, f_k(Y_i))$, $\bar{Z} = n^{-1}\sum_{i=1}^{n} Z_i$, $\mu = EZ_i$, and

$H : \mathbf{R}^k \to \mathbf{R}^p$. An important application of this result yields the asymptotic expansion of the distribution of maximum likelihood estimators. Since in sections 4.2 and 4.5 this result will be used critically to derive saddlepoint approximations to the distribution of M-estimators, we refer the reader to those sections for a detailed discussion of the conditions on the validity of the expansions.

2.7. EXAMPLES

In this section we discuss two examples which show the numerical aspects of the approximations based on Edgeworth expansions.

In the first one we consider the approximation of the distribution of the mean of 5 uniform $[-1, 1]$ observations. In this case the density of the mean can be computed exactly; it is given in section 3.5. In order to compare with the Edgeworth approximation, we will consider here the density of the standardized mean $\bar{X}_n/\sigma_n$, where $\sigma_n^2 = var\bar{X}_n = 1/3n$. For the density we obtain

$$f_{\bar{X}_n/\sigma_n}(t) = \frac{n^n}{\sqrt{3n}2^n(n-1)!}\sum_{i=0}^{n}(-1)^i\binom{n}{i}\left\langle 1 - \frac{t}{\sqrt{3n}} - \frac{2i}{n}\right\rangle^{n-1} \tag{2.21}$$

and for the cumulative distribution function

$$F_{\bar{X}_n/\sigma_n}(t) = \frac{n^{n-1}}{2^n(n-1)!}\sum_{i=0}^{n}(-1)^i\binom{n}{i}$$

$$\left\{\left(2 - \frac{2i}{n}\right)^n - \left(1 - m(i,t) - \frac{2i}{n}\right)^n\right\}, \tag{2.22}$$

where $|t| \leq \sqrt{3n}$, $\langle z\rangle = max(z, 0)$, and $m(i,t) = min(t/\sqrt{3n}, 1-2i/n)$. The corresponding approximations based on Edgeworth expansions to terms of order n^{-1} are given by

$$\tilde{f}_{\bar{X}_n/\sigma_n}(t) = \left\{1 - \frac{1}{20n}(t^4 - 6t^2 + 3)\right\}\phi(t) \tag{2.23}$$

and

$$\tilde{F}_{\bar{X}_n/\sigma_n}(t) = \Phi(t) - \frac{1}{20n}(3t - t^3)\phi(t) \tag{2.24}$$

where $\phi(t)$ and $\Phi(t)$ are the density and the cumulative of the standard normal distribution respectively.

Exhibit 2.4 shows the error (exact-Edgeworth) in the approximation of the density whereas Exhibits 2.5 and 2.6 show the *percent relative error* for the cumulative distribution. In these exhibits, the horizontal axis is in standardized units.

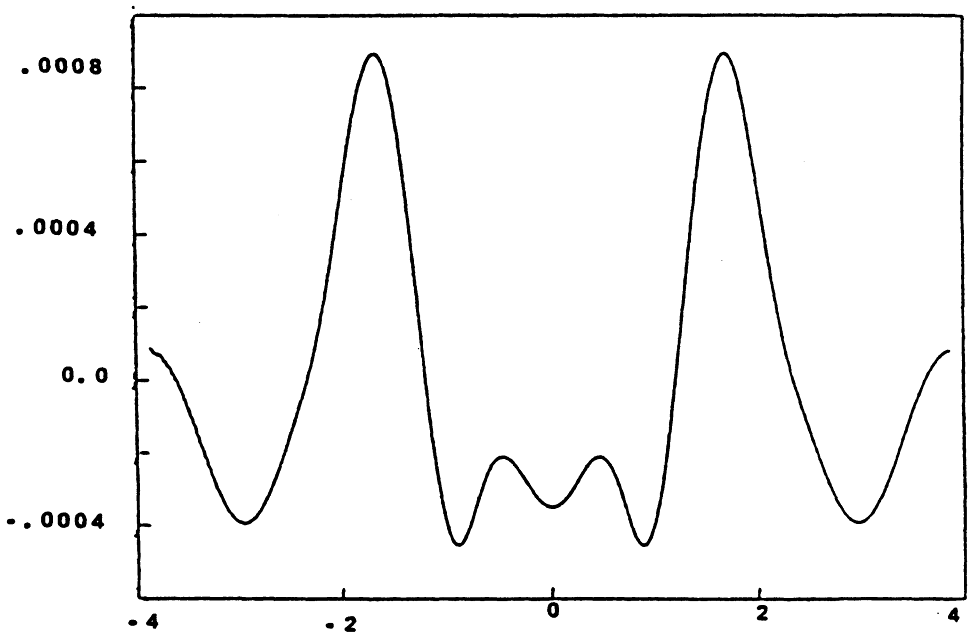

Exhibit 2.4
Error (exact-Edgeworth) in the approximation of the density of the mean of 5 uniform $[-1, 1]$ observations.

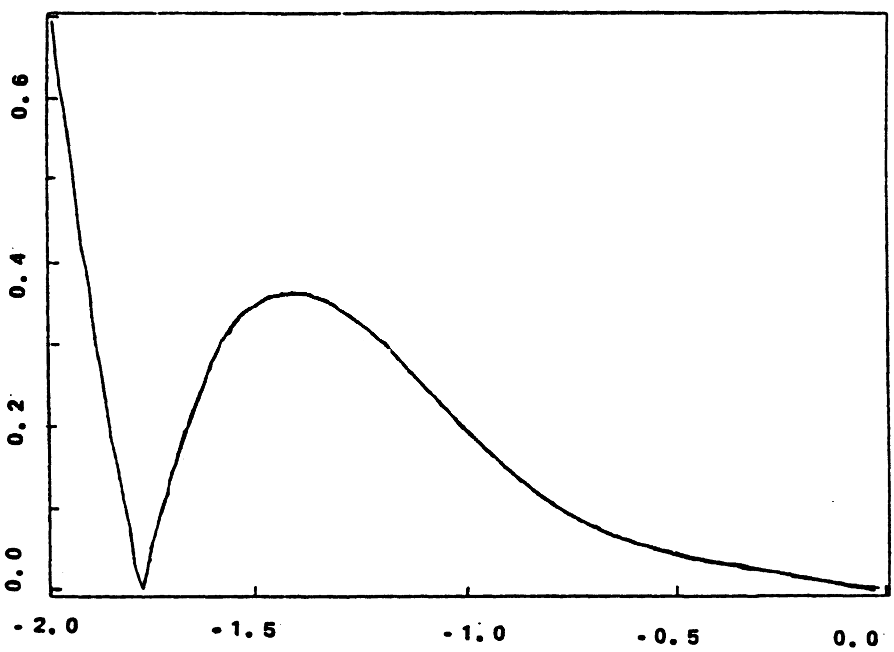

Exhibit 2.5
Relative error (%) in the Edgeworth approximation of the cumulative distribution function of the mean of 5 uniform $[-1, 1]$ observations.

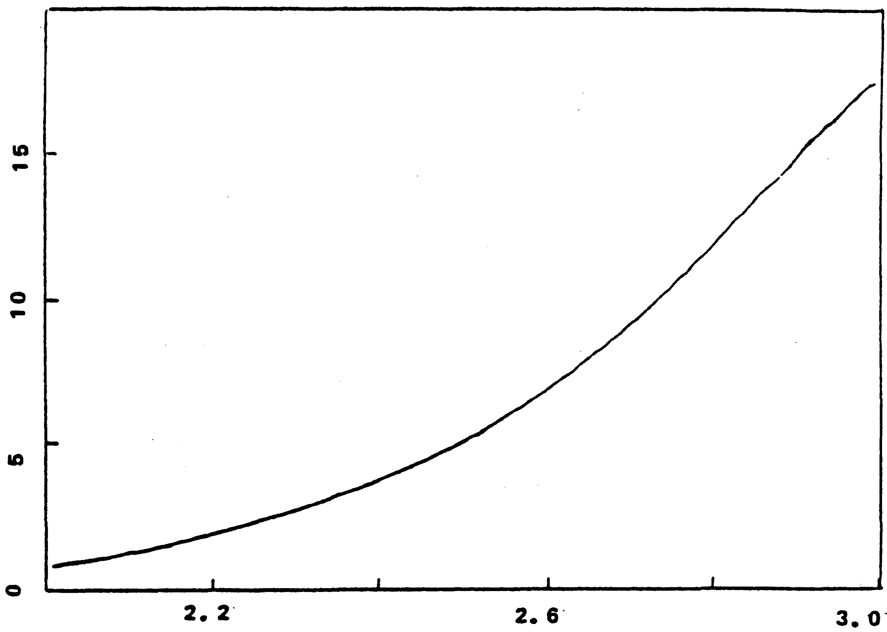

Exhibit 2.6

Relative error $\left(100 \times \frac{|\text{exact-Edgeworth}|}{1-\text{exact}}\right)$ in the Edgeworth approximation of the upper tail area of the distribution of the mean of 5 uniform $[-1,1]$ observations.

While in the range $(-2,0)$ the relative error is small, it is clear from Exhibit 2.6 that it can get up to 18% in the tail. Moreover, Exhibit 2.4 shows the typical polynomial behavior of the Edgeworth approximation. This contrasts with the uniformly small (over the whole range) relative error of the small sample asymptotic approximation, cf. Exhibits 3.7 and 3.8 and Hampel (1973). In particular, in this case the relative error is always smaller than 2.51% even in the extreme tails.

As a second example consider a Gamma distribution with shape parameter α (and scale parameter $\theta = 1$) as an underlying distribution:

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

The density of the mean of n iid observations with this underlying distribution is again a Gamma with shape parameter $n\alpha$ and scale parameter n. Moreover,

$$E\bar{X}_n = \alpha, \quad var\bar{X}_n = \alpha/n.$$

Exhibits 2.7 to 2.10 show the relative errors for $n = 4, 10$ and $\alpha = 2$. The same comments as in the first example apply here. Note that in this case the saddlepoint approximation is exact *(after renormalization)*; cf. Remark 3.3.

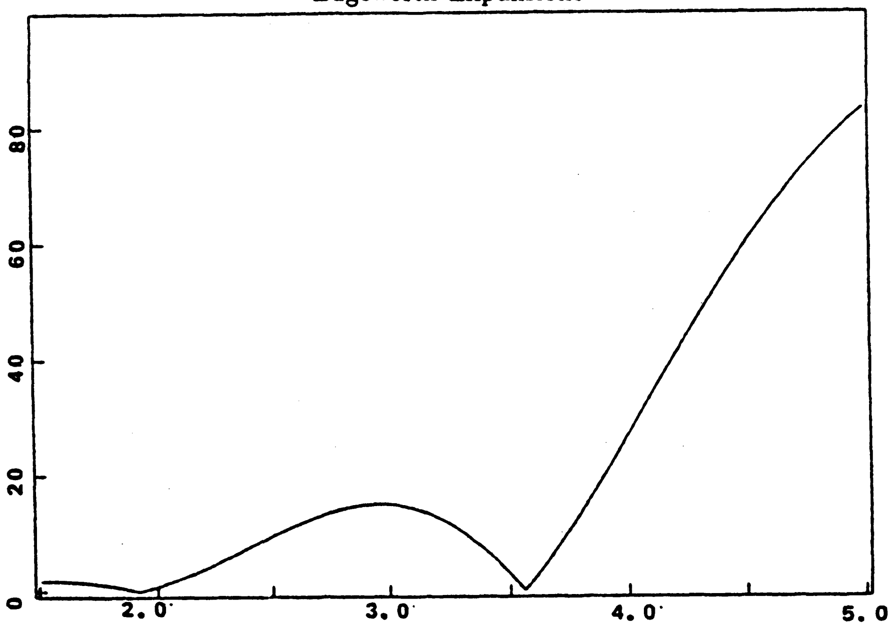

Exhibit 2.7

Relative error in % $\left(\frac{|\text{exact-Edgeworth}|}{1-\text{exact}} \times 100\right)$ in the Edgeworth approximation of the upper tail area of the distribution of the mean of 4 observations from a Gamma distribution with shape parameter $\alpha = 2$.

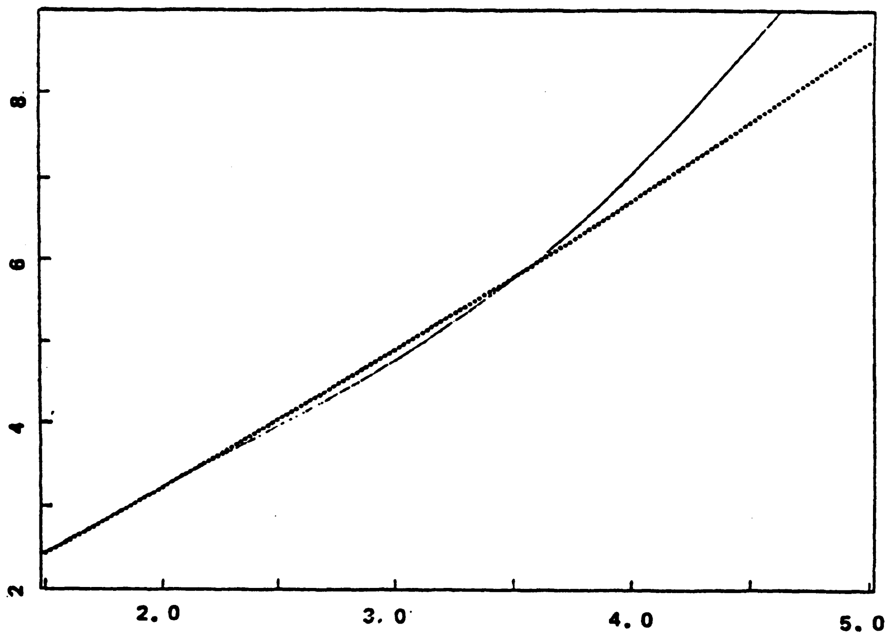

Exhibit 2.8

$\log(F/(1-F))$ for the exact $(\cdots)$ and the Edgeworth approximation (—) for the upper tail area for the mean of 4 observations from a Gamma distribution with shape parameter $\alpha = 2$.

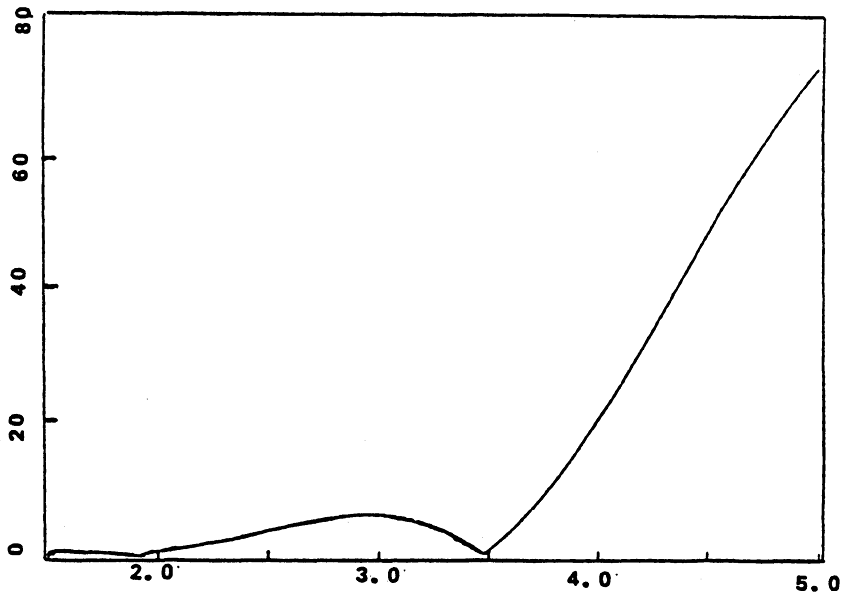

Exhibit 2.9
Relative error in % in the Edgeworth approximation of the upper tail area for the mean of 10 observations from a Gamma distribution with shape parameter $\alpha = 2$.

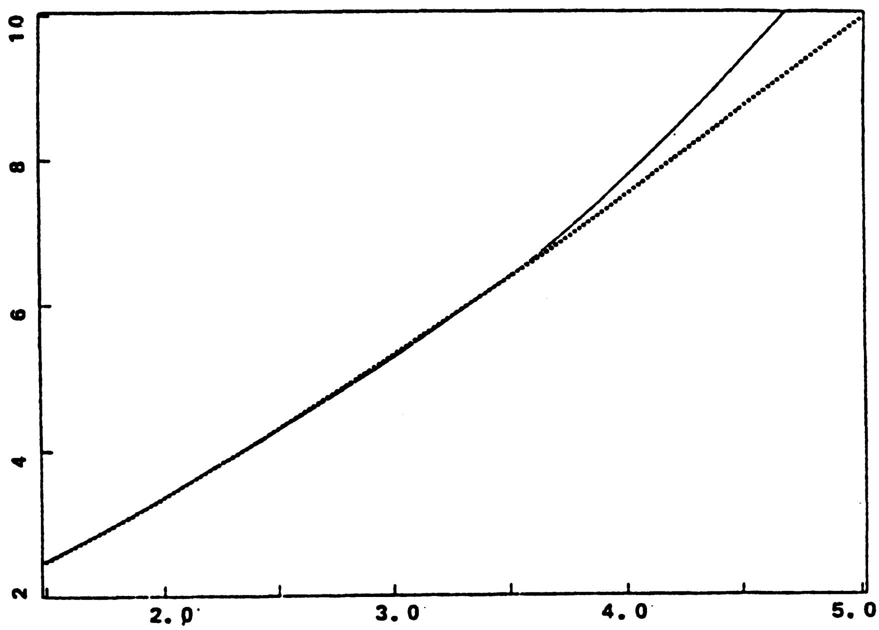

Exhibit 2.10
$\log(F/(1-F))$ for the exact $(\cdots)$ and the Edgeworth approximation (—) for the upper tail area of the distribution of the mean of 10 observations from a Gamma distribution with shape parameter $\alpha = 2$.

More examples can be found in sections 4.4 and 5.3.