*Biometrika* (1982), **69**, 1, *pp.* 29–46
*Printed in Great Britain*



# Small-sample asymptotic distributions of *M*-estimators of location

BY CHRISTOPHER A. FIELD
*Department of Mathematics, Dalhousie University, Halifax, N.S., Canada*

AND FRANK R. HAMPEL
*Fachgruppe für Statistik, ETH Zürich, Switzerland*

SUMMARY

Asymptotic formulae for the distribution of $M$-estimators, i.e. maximum likelihood type estimators, of location, including the arithmetic mean, are derived which numerical studies show to give relative errors for densities and tail areas of the order of magnitude of 1% down to sample sizes 3 and 4 even in the extreme tails. The paper is the continuation of earlier work by the second author and is also closely related to Daniels's work on the saddlepoint approximation. The method consists in expanding the derivative of the logarithm of the unstandardized density of the estimator in powers of $1/n$ at each point, using recentring by means of conjugate distributions. This method yields a unified point of view for the comparison of other asymptotic methods, namely saddlepoint method, Edgeworth expansion and large deviations approach, which are also compared numerically.

*Some key words*: Arithmetic mean; Central limit theorem; Conjugate distributions; Edgeworth expansion; Huber-estimator; Large deviation; Pearson curve; Saddlepoint method; Small-sample asymptotics.

## 1. INTRODUCTION

The paper discusses asymptotic approximations to the distributions of certain estimators for very small sample sizes. It extends the applicability of a new method of asymptotic expansion (Hampel, 1974a) from the arithmetic mean to so-called $M$-estimators of location with a monotone $\psi$-function. If $\psi$ is also bounded, meaning essentially that the $M$-estimator is robust, the method is applicable to sufficiently smooth but arbitrarily long-tailed distributions, such as the Cauchy distribution, and yields a very accurate approximation even in the extreme tails and even for sample sizes around 3. The method has very close ties to the saddlepoint approximation used by Daniels (1954), but is more elementary and provides in some sense a complementary aspect of the same phenomenon. Besides supplying highly accurate approximations for the distributions of some estimators, which are considerably cheaper to obtain on the computer than the exact distributions, the method allows a unified comparison and better intuitive understanding of saddlepoint approximation, Edgeworth expansion and large deviation theory, and it provides a deeper intuitive understanding of the central limit theorem and various related topics.

Let $X_1, X_2, \ldots$ be independent and identically distributed with zero expectation, density $f$, and no longer than exponential tails, and let $p_n$ denote the density of the arithmetic mean, $T_n = \bar{X}_n$, of the first $n$ observations. Hampel (1973) showed that it is reasonable to consider the expansion of the logarithmic derivative of the density

$$-K_n(t) = p_n'(t)/p_n(t) = -n\alpha(t) - \beta(t) - \gamma(t)/n - \dots.$$

It was found empirically there that the first two terms, which for each $t$ are linear in $n$, provide already an excellent approximation. Integration, which can be done in closed form, yields $\log p_n$ up to the normalizing constant which is obtained by exponentiation and numerical integration.

In order to obtain the expansion, one has to recentre the density $f$ around each point $t$ by multiplying it with a suitable exponential function and restandardizing it so that the expectation of $X-t$ becomes 0 and the total mass remains 1. This corresponds to a mere shift of $f'/f$ and is the well-known trick of conjugate or associated distributions (Khinchin, 1949; Feller, 1966, p. 518). Now by somewhat tedious but elementary calculations $p_n'(t)/p_n(t)$ can be expressed as $n$ times a weighted average of $f'/f$, the weight function being a convolution integral of the conjugate distribution centred at $t$. If this convolution is approximated by the Edgeworth expansion at the expectation, one obtains the desired asymptotic expansion in powers of $1/n$. The first two terms require up to the third moments of the conjugate distributions, which are obtainable from the moment generating function; their integrals require only up to the second moments.

It turns out (Hampel, 1974a) that this method is formally nearly equivalent to the saddlepoint method as used by Daniels (1954) and can be viewed as another, more elementary, way of deriving its results, except for two slight differences: the infinite expansion of $f'/f$ leads to a series in the exponent which Daniels (1954, equations (2·5), (2·6)) expanded beyond the first terms into a sum; moreover, the saddlepoint method automatically yields the trivial constant of integration $n^{\frac{1}{2}}/(2\pi)^{\frac{1}{2}}$ while the new method automatically has to determine the best-fitting constant in each case by integration. When the saddlepoint approximation, i.e. the first two terms, which are not yet expanded and which appear as the first term of Daniels (1954, equation (2·6)), is renormalized to make the total probability unity the two methods give identical results. There is also an expansion for $p_n(t)$ directly which is technically simpler than that for $p_n'/p_n$ and strictly equivalent to the full saddlepoint expansion. By contrast, the classical Edgeworth expansion is an expansion of $p_n$ only around $t = 0$ and thus disastrously bad for large $|t|$; for small $|t|$, it is quite good but can apparently still be improved by putting the expansion back into the exponent where it arrives naturally by integration of the expansion for $K_n(t)$ around $t = 0$.

Finally, the large deviation expansion is only the expansion of the first term $\alpha(t)$ in powers of $t$ around $t = 0$ in the exponent for $p_n$, or the cumulative $P_n$. Since the first term alone yields a very bad fit (Hampel, 1974a, p. 118), the large deviation fit is very poor, often even worse than the normal approximation, even for small $|t|$, except for $t = 0$ itself when it coincides with the normal and the saddlepoint approximation.

Another class of local approximations is given by the system of Pearson curves which start out with a different form of approximation for $K_n(t)$; see, for example, Jeffreys (1961, Chapter 2). Formally, one can either match the local behaviour of $K_n(t)$ at $t = 0$, for example with that of the new second-order approximation, or match the first four moments, or, in some cases, use partly information about the limits of range. In general, of course, Pearson curves may be much less accurate than the present approaches, as they utilize basically only the first four moments of $f$ and none of the conjugate distributions.

So far, only the distribution of the arithmetic mean $\bar{X}_n$ has been considered, partly for ease of description, partly for sake of its importance in connexion with the central limit problem and partly for ease of comparison with other asymptotic methods. The derivation of the new method for $\bar{X}_n$, as well as the most essential relations to other methods, are already described by Hampel (1974a) and are only partly reviewed and partly extended here for ease of reference and for completeness. The main new feature of this paper, however, is the full generalization of the new method to $M$-estimators of location in the sense of Huber (1964) with monotone $\psi$, as already referred to briefly by Hampel (1974a).

An $M$-estimate of location is defined via some function $\psi$ as the solution $T$ of the implicit equation $\Sigma\,\psi(X_i-T) = 0$, which is a slightly generalized form of the likelihood equation, for which $\psi = -f'/f$. Arithmetic mean and median are special cases with $\psi(x) = x$ and $\psi(x) = \operatorname{sgn}(x)$ respectively. If $\psi$ is monotone nondecreasing and takes on positive and negative values, the solution of the defining equation is essentially unique; it may be a unique interval, or there may be a unique point of transition from positive to negative values of the left-hand side. These $M$-estimators play a considerable role in the theory of robust estimation, and the most important condition for them to be robust, i.e. insensitive against gross errors and other deviations from an ideal parametric model, is that the $\psi$-function is bounded (Hampel, 1971, 1974b). Now, if $\psi$ is bounded with $\psi$ or $f$ sufficiently smooth, then the moment generating function of $\psi(X)$ as well as all conjugate distributions of the form $c_t \exp\{\alpha_t\, \psi(x-t)\} f(x)$ always exist, and we can abandon the artificial restriction to rather short-tailed, i.e. at most exponentially-tailed, $f$, which causes certain limitations in the central limit problem and which is intimately connected with the nonrobustness of the arithmetic mean.

It was noted empirically by Hampel (1974a) that a linear function of $n$ gave an excellent fit to $K_n(t)$ for the Huber-estimator $H_1(k)$, with $\psi(x) = x$ for $|x| \leqslant k$, $\psi(x) = k\operatorname{sgn}(x)$ otherwise (Huber, 1964), under several distributions, including the Cauchy distribution. Also, the first-order terms for $K_n(t)$ were given there. An exact second-order formula was found for quantiles regarded as $M$-estimators but the general second-order term contained an error and was not included. Now, the correct second-order term is available and numerical comparisons between exact and approximate distributions can be made.

There is an indirect way of obtaining the distribution of $M$-estimators with monotone $\psi$ by reducing the problem to that of the arithmetic mean. This fact was noted by Daniels and was used by him to compute the values shown under $G$ in Tables 1 and 2 by his kind permission; it was also proposed independently by Huber (1977, pp. 21, 22) for the generalization of the method of Hampel (1974a). The resulting approximations are different from the direct ones with or without renormalization; but they are of similar quality, as the tables show. Meanwhile, Daniels, in oral remarks and an internal research note of March 1978, was also able to find the analogue of our direct approach by applying the saddlepoint method and showed that his new result is again equivalent to our result for $p_n(t)$.

The present paper is organized as follows. After a section on the heuristic motivation for the approach used, the main body of the paper contains the derivation of the second-order formula for $M$-estimators of location with strictly monotone $\psi$ both for $p_n'/p_n$ and for $p_n$. Strict monotonicity is then relaxed to weak monotonicity to include cases like the median and the Huber-estimator. Special cases which reduce to known results are quantiles including the median, and the arithmetic mean. Following this theoretical part, the formula is applied to two situations: Huber-estimators under a 5% contaminated normal distribution and the Cauchy distribution, and compared both with the exact distributions and Daniels's indirect version of the saddlepoint method. The final sections discuss the relation of our method to the saddlepoint approximation, to the indirect approach, and last but not least to Edgeworth expansions and large deviations, with some new variants and two comparative examples.

## 2. HEURISTIC MOTIVATION

The method of approximation used in this paper differs from the more customary methods like Edgeworth expansions and large deviations in three respects. First, the distribution of $T_n$ is not blown up by the usual factor $n^{\frac{1}{2}}$, but rather is allowed to concentrate towards a point mass. While this is of course formally equivalent, it allows a more lucid description of what is happening with increasing sample size.

Secondly, instead of a high-order expansion around a single point, the expectation, a low-order expansion around each point is used. High-order expansions can be at most locally accurate, and the higher-order terms are superfluous for large $n$, while the other approach yields a very accurate fit globally even for small $n$.

Thirdly, and this is also a difference from the saddlepoint method, neither the density nor the cumulative are expanded, but rather the derivative of the log density $-K_n(t) = p_n'(t)/p_n(t)$ is approximated. This quantity permeates much of mathematical statistics as an auxiliary function, perhaps most noticeably as the score function of maximum likelihood estimators, but it is rarely considered in its own right. This is surprising, since the first great system of frequency curves, the Pearson curves, with all its important special cases, is based on a simple class of functions for $K(t) = -f'(t)/f(t)$. This $K(t)$ can also be regarded as a transform of a density function, like a characteristic function or a Laplace transform, with its own special properties; it even has a physical interpretation, namely as the local force in a field which under suitable circumstances causes a mass of particles to have density proportional to $\exp\{-\int K(t)\,dt\}$.

Another aspect is that asymptotic theory may be regarded as studying purely local properties of a distribution which are not affected by adding, deleting or shifting probability masses elsewhere. But neither the cumulative $F$ nor the density $f$ describe purely local properties.

A further argument is simplicity. It has been said that the role of the normal distribution in probability is similar to that of the straight line in geometry; however, there is not much in the form of the normal cumulative or density to support such a statement, while $K(t)$ is in fact a straight line. As just about every 'smooth' function is locally linear, about every 'smooth' distribution is locally normal and if the distribution is highly concentrated, only the local behaviour matters; this is the essence of the central limit theorem. And while the normal distribution is distinguished by its linearity of $K$ in $t$, our second-order approximation for $K_n(t)$ is linear in $n$ for each $t$, a form which can hardly be matched in simplicity and generality simultaneously.

Numerical computations confirm that we have found an asymptotic theory which can often be used down to $n = 1$, as is sometimes demanded of a good asymptotic theory and as is beautifully exemplified by Stirling's formula. For some more heuristic aspects, see an ETH Zurich Research Report. Field (1978), Barndorff-Nielsen & Cox (1979), Daniels (1980) and Durbin (1980a, b) also give related work.

## 3. Asymptotic formula for $p_n'(t)/p_n(t)$

It is assumed that $X_1, \dots, X_n$ are $n$ independent observations from a location family with density $f(x-\theta)$ and that the estimate $T$ of $\theta$ is defined as the implicit solution of the equation $\Sigma_i \psi(x_i - T) = 0$. In order to develop the formula, $f$ and $\psi$ must satisfy certain regularity conditions as follows.

I. The function $\psi$ is a strictly monotone increasing continuous function.

II. The functions $f$ and $\psi$ are piecewise differentiable.

III. The density $m_t(x) = c_t \exp\{\alpha\psi(x-t)\} f(x)$, with

$$c_t^{-1} = \int \exp\{\alpha\psi(x-t)\} f(x)\,dx,$$

where the integral is over $(-\infty, \infty)$, exists as do all moments of $\psi(x-t)$ computed with $m_t(x)$ for arbitrary $\alpha$.

IV. The following random variables have finite expected values with respect to the density $m_t(x)$:

$$\psi'(x-t),\ \psi'(x-t)\psi(x-t),\ \psi'(x-t)\,\psi^2(x-t),\ \psi'(x-t)\,f'(x)/f(x),$$

where $\psi'$ is to be interpreted as the piecewise derivative.

V. The function $f$ must be sufficiently regular so that

$$d/dt \int f(x+t)\,dx = \int f'(x+t)\,dx.$$

As will be shown, assumption I can be weakened by requiring only monotone, continuous $\psi$. The results given will also hold for discontinuous $\psi$ with some minor modifications in proofs to allow for point masses in the conjugate distribution. If $\psi$ is bounded, as it is for robust estimators, conditions III and IV may be satisfied even if $f$ has no moments as in the case of the Cauchy distribution. If $\psi$ is not bounded, as for the arithmetic mean, conditions III and IV restrict the length of the tails of the underlying distribution (Hampel, 1974a).

To begin, assume $\psi$ is strictly monotone with range $R$ and let $p_n(t)$ denote the density of $T$. Denote the density of $\psi(x-t)$ by $g_t(y)$, where

$$g_t(y) = \begin{cases} f\{\psi^{-1}(y)+t\}/\psi'\{\psi^{-1}(y)\} & \text{if } y \in R, \\ 0 & \text{otherwise.} \end{cases}$$

Then

$$\mathrm{pr}(T \leqslant t) = \mathrm{pr}\left\{\sum_{i=1}^{n} \psi(x_i - t) \leqslant 0\right\} = \int \dots \int \prod_{i=1}^{n} g_t(y_i)\,dy_1 \dots dy_n,$$

$$p_n(t) = n \int \dots \int \prod_{i=1}^{n-1} g_t(y_i)\,\partial/\partial t\{g_t(y_n)\}\,dy_1 \dots dy_n,$$

where the integrals are over the range

$$\left\{\sum_{i=1}^{n} y_i \leqslant 0\right\}.$$

But $\int \partial/\partial t\, g_t(y_i)\, dy = f\{\psi^{-1}(y)+t\}$ if $y \in R$. Hence

$$p_n(t) = n\int \dots \int f\left\{\psi^{-1}\left(-\sum_{i=1}^{n-1} y_i\right)+t\right\}\prod_{i=1}^{n-1} g_t(y_i)\, dy_1 \dots dy_{n-1}.$$

In this and following integrals, it is assumed that the integration is over the range for which the argument of $\psi^{-1}$ belongs to $R$. Thus

$$p_n'(t) = n(n-1)\int \dots \int f\left\{\psi^{-1}\left(-\sum_{i=1}^{n-1} y_i\right)+t\right\} f'/f\{\psi^{-1}(y_{n-1})+t\}$$
$$\times \prod_{i=1}^{n-1} g_t(y_i)\, dy_1 \dots dy_{n-1} + n\int \dots \int f'\left\{\psi^{-1}\left(-\sum_{i=1}^{n-1} y_i\right)+t\right\}\prod_{i=1}^{n-1} g_t(y_i)\, dy_1 \dots dy_{n-1}.$$

The next step is to recentre $g_t$ about $t$ by replacing $g_t$ with a conjugate or associated density $h_t$.

Let $h_t(y) = c_t \exp(\alpha_t y) g_t(y)$, where the constants $c_t$ and $\alpha_t$ are determined by $\int h_t(y)\, dy = 1$ and $\int y h_t(y)\, dy = 0$. Later we shall also need $\sigma_t^2 = \int y^2 h_t(y)\, dy$, $\lambda_{3,t} = \int y^3 h_t(y)\, dy/\sigma_t^3$. Note that $h_t(y)\, dy = c_t \exp\{\alpha_t \psi(x-t)\} f(x)\, dx = m_t(x)\, dx$, so that all the moments of $h_t$ exist by assumption III.

Now

$$p_n(t) = nc_t^{-n}\int \dots \int \psi'\left\{\psi^{-1}\left(-\sum_{i=1}^{n-1} y_i\right)\right\} h_t\left(-\sum_{i=1}^{n-1} y_i\right)\prod_{i=1}^{n-1} h_t(y_i)\, dy_1 \dots dy_{n-1},$$

$$p_n'(t) = n(n-1)\, c_t^{-n}\int \dots \int \psi'\left\{\psi^{-1}\left(-\sum_{i=1}^{n-1} y_i\right)\right\} f'/f\{\psi^{-1}(y_{n-1})+t\}\, h_t\left(-\sum_{i=1}^{n-1} y_i\right)$$
$$\times \prod_{i=1}^{n-1} h_t(y_i)\, dy_1 \dots dy_{n-1} + nc_t^{-n}\int \dots \int \psi'\left\{\psi^{-1}\left(-\sum_{i=1}^{n-1} y_i\right)\right\} f'/f\left\{\psi^{-1}\left(-\sum_{i=1}^{n-1} y_i\right)+t\right\}$$
$$\times h_t\left(-\sum_{i=1}^{n-1} y_i\right)\prod_{i=1}^{n-1} h_t(y_i)\, dy_1 \dots dy_{n-1}.$$

Denote the density of the sum of $v$ independent random variables each with density $h_t(y_i)$ by

$$j_{v,t}(r) = \int \dots \int h_t\left(r-\sum_{i=1}^{v-1} y_i\right)\prod_{i=1}^{v-1} h_t(y_i)\, dy_1 \dots dy_{v-1}.$$

In the first term of $p_n'(t)$, let $z_i = y_i$ $(i = 1, \dots, n-3)$, $r = \Sigma_i y_i$, where the sum is over $i = 1, \dots, n-1$, $s = y_{n-1}$ and in the second term of $p_n'(t)$ and in $p_n(t)$, let $z_i = y_i$ $(i = 1, \dots, n-2)$, $s = \Sigma_i y_i$ summed over $i = 1, \dots, n-1$.

Then

$$p_n(t) = nc_t^{-n}\int \psi'\{\psi^{-1}(-s)\}\, h_t(-s)\left\{\int \dots \int \prod_{i=1}^{n-2} h_t(z_i)\, h_t\left(s-\sum_{i=1}^{n-2} z_i\right) dz_1 \dots dz_{n-2}\right\} ds$$

$$= nc_t^{-n}\int j_{n-1,t}(s)\, \psi'\{\psi^{-1}(-s)\}\, h_t(-s)\, ds, \qquad (3\cdot 1)$$

$$p_n'(t) = n(n-1)\,c_t^{-n}\int\!\!\int \psi'\{\psi^{-1}(-r)\}\,f'/f\{\psi^{-1}(s)+t\}\,h_t(-r)\,h_t(s)$$

$$\times\left\{\int\!\cdots\!\int \prod_{i=1}^{n-3} h_t(z_i)\,h_t\!\left(r-s-\sum_{i=1}^{n-3} z_i\right)dz_1\ldots dz_{n-3}\right\}dr\,ds$$

$$+nc_t^{-n}\int \psi'\{\psi^{-1}(-s)\}\,f'/f\{\psi^{-1}(s)+t\}\,h_t(-s)$$

$$\times\left\{\int\!\cdots\!\int \prod_{i=1}^{n-2} h_t(z_i)\,h_t\!\left(s-\sum_{i=1}^{n-2} z_i\right)dz_1\ldots dz_{n-2}\right\}ds$$

$$= n(n-1)\,c_t^{-n}\int\!\!\int j_{n-2,t}(r-s)\,\psi'\{\psi^{-1}(-r)\}\,f'/f\{\psi^{-1}(s)+t\}\,h_t(-r)\,h_t(s)\,dr\,ds$$

$$+nc_t^{-n}\int j_{n-1,t}(s)\,\psi'\{\psi^{-1}(-s)\}\,f'/f\{\psi^{-1}(-s)+t\}\,h_t(-s)\,ds, \qquad (3\cdot 2)$$

whence $p_n'(t)/p_n(t)$ is obtained as a fraction not involving $c_t$.

When $n$ increases, $j_{n,t}(s)$ flattens out while the other terms in the integrands stay the same. Thus, only the local behaviour of a standardized sum of random variables at zero matters. To proceed further, $j_{n,t}(s)/j_{n,t}(0)$ is approximated by a series with the coefficients determined by the Edgeworth expansion of $j_{n,t}(s)$ at the origin. This enables us to use locally the good properties of the Edgeworth expansion at the origin.

Let $S$ be the sum of $n$ independent random variables with density $h_t(s)$ and let $f_t(s)$ be the density of $S/(n^{\frac{1}{2}}\sigma_t)$. The Edgeworth expansion (Cramér, 1946, p. 229) gives

$$f_t(s) = \phi(s)\,\{1+\lambda_{3,t}(s^2-3s)/(3!\,n^{\frac{1}{2}})+\lambda_{4,t}(s^4-6s^2+3)/(4!\,n)$$
$$+10\lambda_{3,t}^2(s^6-15s^4+45s^2-15)/(6!\,n)+O(1/n^{3/2})\},$$

where $\phi(s) = (2\pi)^{-\frac{1}{2}}\exp(-\frac{1}{2}s^2)$. Since $j_{n,t}(s)$ is the density of $S$, we have

$$j_{n,t}(s) = f_t\{s/(n^{\frac{1}{2}}\sigma_t)\}/(n^{\frac{1}{2}}\sigma_t).$$

Now

$$j_{n,t}(0) = \phi(0)\,\{1+(3\lambda_{4,t}/4!-15\lambda_{3,t}^2/72)\,n^{-1}+O(1/n^2)\}/(n^{\frac{1}{2}}\sigma_t),$$
$$j_{n,t}'(0) = \phi(0)\,\{-\lambda_{3,t}/(2\sigma_t n)+O(1/n^2)\}/(n^{\frac{1}{2}}\sigma_t),$$
$$j_{n,t}''(0) = \phi(0)\,\{-1/(n\sigma_t^2)+O(1/n^2)\}/(n^{\frac{1}{2}}\sigma_t),\quad j_{n,t}^{(3)}(0) = \phi(0)\,\{O(1/n^2)\}/(n^{\frac{1}{2}}\sigma_t).$$

In addition, we note that $j_{n,t}^{(4)}(s) = O(1/n^2)/(n^{\frac{1}{2}}\sigma_t)$ for any $s$. Now, with $0\leqslant\xi\leqslant s$,

$$j_{n,t}(s) = j_{n,t}(0)\,[1+j_{n,t}'(0)\,s/j_{n,t}(0)+j_{n,t}''(0)\,s^2/\{j_{n,t}(0)\,2!\}$$
$$+j_{n,t}^{(3)}(0)\,s^3/\{j_{n,t}(0)\,3!\}+j_{n,t}^{(4)}(\xi)s^4/\{j_{n,t}(0)\,4!\}]$$
$$= j_{n,t}(0)\,\{1-\lambda_{3,t}\,s/(2\sigma_t n)-s^2/(2\sigma_t^2 n)+O(1/n^2)\}.$$

Continuing this way, we obtain an infinite expansion of $j_{n,t}(s)/j_{n,t}(0)$ in powers of $1/n$, which is closely related to the Edgeworth expansion, but which yields only purely local properties and thus does not contain the absolute height $j_{n,t}(0)$, or normalizing constant, of the density. It is true that for somewhat larger $n$ only the local properties around $t = 0$

matter and thus the normalizing constant can also be approximated satisfactorily by the Edgeworth expansion, but for very small $n$ it is good to have an expansion that does not claim information about the tail areas which it cannot possess.

We insert this expansion into numerator and denominator of $p_n'(t)/p_n(t)$, where $j_{n,t}(0)$ essentially disappears. More precisely, we need in addition the expansion of $j_{n-2,t}(0)/j_{n-1,t}(0)$ which can also be obtained from the Edgeworth expansion. Dividing and multiplying these infinite series and recollecting terms we obtain the desired purely local expansion of the form

$$-K_n(t) = p_n'(t)/p_n(t) = -\alpha(t)\,n - \beta(t) - \gamma(t)/n - \dots .$$

We now determine the first- and second-order terms of this expansion. Define the following quantities in terms of the original variables. To obtain the expressions in terms of the transformed variables, let $s = \psi(x-t)$ and note that $h_t(y)\,dy = c_t \exp\{\alpha_t \psi(x-t)\} f(x)\,dx$. Then

$$A_{1,t} = \int \psi'(x-t)\,c_t \exp\{\alpha_t \psi(x-t)\} f(x)\,dx, \quad A_{2,t} = \int c_t \exp\{\alpha_t \psi(x-t)\} f'(x)\,dx,$$

$$A_{3,t} = \int \psi(x-t)\psi'(x-t)\,c_t \exp\{\alpha_t \psi(x-t)\} f(x)\,dx,$$

$$A_{4,t} = \int \psi(x-t)\,c_t \exp\{\alpha_t \psi(x-t)\} f'(x)\,dx,$$

$$A_{5,t} = \int \psi'(x-t)\,c_t \exp\{\alpha_t \psi(x-t)\} f'(x)\,dx, \quad A_{6,t} = \int \psi^2(x-t)\,c_t \exp\{\alpha_t \psi(x-t)\} f'(x)\,dx,$$

$$A_{7,t} = \int \psi^2(x-t)\psi'(x-t)\,c_t \exp\{\alpha_t \psi(x-t)\} f(x)\,dx,$$

where all the integrals are over the range $(-\infty, \infty)$.

Condition IV guarantees that these integrals exist. There are a number of relations between the integrals, such as $A_2 = -\alpha_t A_1$, $A_1 = A_4 - \alpha_t A_3$ and $2A_3 + \alpha_t A_7 + A_6 = 0$, if we drop temporarily the $t$ in $A_{i,t}$. With these terms the approximations to $p_n(t)$ and $p_n'(t)$, as given in (3·1) and (3·2) are as follows:

$$p_n(t) = nc_t^{-n} j_{n-1}(0)\,[A_1 + \{\lambda_{3,t} A_3/(2\sigma_t) - A_7/(2\sigma_t^2)\}/n + O(1/n^2)],$$

$$p_n'(t) = n(n-1)\,c_t^{-n} j_{n-2}(0)\,[A_1 A_2 + \{\lambda_{3,t} A_3 A_2/(2\sigma_t) + \lambda_{3,t} A_4 A_1/(2\sigma_t) - A_7 A_2/(2\sigma_t^2)$$
$$- A_4 A_3/(2\sigma_t^2) - A_1 A_6/(2\sigma_t^2)\}/n + O(1/n^2)] + nc_t^{-n} j_{n-1}(0)\,\{A_5 + O(1/n^2)\}.$$

Divide $p_n'(t)$ by $p_n(t)$ to obtain

$$p_n'(t)/p_n(t) = (n-1)\,j_{n-2}(0)/j_{n-1}(0)$$
$$\times [A_2 + \{\lambda_{3,t} A_4/(2\sigma_t) - A_4 A_3/(2\sigma_t^2 A_1) - A_6/(2\sigma_t^2)\}/n + O(1/n^2)]$$
$$+ A_5/A_1 + O(1/n).$$

But

$$j_{n-2}(0)/j_{n-1}(0) = \{(n-1)^{\frac{1}{2}}/(n-2)^{\frac{1}{2}}\}\,\{1 + c_1/n + O(1/n^2)\}/\{1 + c_1/n + O(1/n^2)\}$$
$$= 1 + 1/(2n) + O(1/n^2),$$

where $c_1$ is determined by the Edgeworth expansion. Hence finally we have that if $f$ and $\psi$ satisfy conditions I–V, and $p_n(t)$ is the density of $T$ where $T$ fulfills $\Sigma_i\,\psi(x_i - T) = 0$, then

$$p_n'(t)/p_n(t) = (n-\tfrac{1}{2})\,A_{2,t} + \lambda_{3,t}\,A_{4,t}/(2\sigma_t) - A_{4,t}\,A_{3,t}/(\sigma_t^2\,A_{1,t})$$
$$-A_{6,t}/(2\sigma_t^2) + A_{5,t}/A_{1,t} + O(1/n), \tag{3·3}$$

where $A_{1,t}$ to $A_{6,t}$ are given above, $c_t$ and $\alpha_t$ satisfy the equations

$$\int c_t \exp\{\alpha_t\,\psi(x-t)\} f(x)\,dx = 1, \quad \int \psi(x-t)\,c_t \exp\{\alpha_t\,\psi(x-t)\}\,f(x)\,dx = 0,$$

$$\sigma_t^2 = \int \psi^2(x-t)\,c_t \exp\{\alpha_t\,\psi(x-t)\}\,f(x)\,dx, \quad \lambda_{3,t} = \int \psi^3(x-t)\,c_t \exp\{\alpha_t\,\psi(x-t)\}\,f(x)\,dx/\sigma_t^3,$$

where all the integrals are over the range $(-\infty, \infty)$.

## 4. Asymptotic formula for $p_n(t)$

It is technically much simpler, though theoretically slightly less satisfactory, to obtain a slightly different expansion directly for $p_n(t)$. At first, we proceed exactly as in the previous section, except that we can ignore all formulae for $p_n'(t)$. Instead of dividing the expansions in $p_n'$ and $p_n$, we just multiply the expansions of $j_{n,t}(0)$ and $j_{n,t}(s)/j_{n,t}(0)$ and thus obtain an expansion of which the first few terms are

$$p_n(t) = n^{\frac{1}{2}}\,c_t^{-n}\,\phi(0)\,\sigma_t^{-1}\{n/(n-1)\}^{\frac{1}{2}}[A_{1,t} + \{(\lambda_{4,t}/8 - 5\lambda_{3,t}^2/24)\,A_{1,t} + \lambda_{3,t}\,A_{3,t}/(2\sigma_t)$$
$$-A_{7,t}/(2\sigma_t^2)\}/n + O(1/n^2)].$$

After replacing $\{n/(n-1)\}^{\frac{1}{2}}$ by its expansion, we have that if $f$ and $\psi$ satisfy conditions I–IV and $p_n(t)$ is the density of $T$ where $\Sigma\,\psi(x_i - T) = 0$, then an approximation for $p_n(t)$ is

$$p_n(t) = n^{\frac{1}{2}}\,\phi(0)\,c_t^{-n}\,\sigma_t^{-1}[A_{1,t} + \{(\tfrac{1}{2} + \lambda_{4,t}/8 - 5\lambda_{3,t}^2/24)\,A_{1,t}$$
$$+\lambda_{3,t}\,A_{3,t}/(2\sigma_t) - A_{7,t}/(2\sigma_t^2)\}/n + O(1/n^2)], \tag{4·1}$$

where $A_{1,t}$, $A_{3,t}$, $A_{7,t}$, $c_t$, $\alpha_t$, $\sigma_t^2$, $\lambda_{3,t}$ are given in the previous section and $\lambda_{4,t} = E\{\psi^4(x-t)\}/\sigma_t^4 - 3$.

Formula (4·1) corresponds to the third-order formula for $p_n'/p_n$, since the 'constant' contains terms of order $n$ and one. The precise relations are explained in § 8. To second order, we have

$$p_n(t) = n^{\frac{1}{2}}\,\phi(0)\,c_t^{-n}\,\sigma_t^{-1}\,A_{1,t}\{1 + O(1/n)\}. \tag{4·2}$$

For higher accuracy in numerical computations one will determine the normalizing constant empirically by numerical integration, and thus one needs only the approximation

$$p_n(t) \propto c_t^{-n}\,\sigma_t^{-1}\,A_{1,t}. \tag{4·3}$$

## 5. Relaxing the strict monotonicity condition

The formulae (3·3) and (4·1) can be shown to be valid if condition I, the strict monotonicity of $\psi$, is replaced by I′ that $\psi$ is a bounded monotone increasing continuous function.

This becomes important for some of the standard robust estimators such as those where $\psi$ is linear in an interval and constant outside the interval. The development of the formula just given cannot be directly extended because the distribution of $\psi(x-t)$ now has point masses as does the appropriate conjugate distribution.

The idea is to approximate $\psi$ by an increasing sequence of strictly monotone functions $\{\psi_m\}$ and to verify that the density of $T$ under $\psi_m$ converges to $p_n$; similarly for $p'_n$. Then one can show that all terms, in particular $\alpha_t$, in the approximating formulae for $\psi_m$ converge to the corresponding terms for $\psi$, and one can also verify that the remainder is still of the same order as in the formulae. For some more details, see the aforementioned report.

## 6. Special cases of $\psi$

Consider first the $M$-estimate version of $\alpha$-quantiles with $\psi(x) = \alpha - 1$ for $x < 0$, $\psi(x) = 0$ for $x = 0$ and $\psi(x) = \alpha$ for $x > 0$. For those $n$ where the defining equation $\Sigma\psi(x_i - T) = 0$ has a unique solution, the exact density of $M-\alpha$-quantiles is the density of the appropriate order statistic and hence well known. From this, the exact result (Hampel, 1974a) is

$$p'_n(t)/p_n(t) = (n-1)\,\alpha f(t)/F(t) - (n-1)\,(1-\alpha) f(t)/\{1-F(t)\} + f'/f(t).$$

That is, remarkably, we obtain precise linearity in $n$ for each $t$.

If the computations are carried out with (3·3), then after calculation we get the exact result for $p'_n/p_n$ given above. Hence for $M-\alpha$-quantiles, including the median for odd $n$, the second-order formula (3·3) is exact.

A second case is that of the arithmetic mean which can also be considered as an $M$-estimate with $\psi(x) = x$. As has been noted earlier, the unboundedness of $\psi$ necessitates conditions on the tail behaviour.

For $p'_n/p_n$ we obtain the simple exact form (Hampel, 1974a, p. 116)

$$p'_n(t)/p_n(t) = n\left\{\int j_{n-1}(s)\, g_t(-s)\, f'/f(t-s)\, ds\right\}\left\{\int j_{n-1}(s)\, g_t(-s)\, ds\right\}^{-1}. \tag{6·1}$$

This means that $K_n(t)$ is just $n$ times a weighted average of $f'/f$, with the weight function consisting of one fixed localizing part and one part that flattens out with increasing $n$. This remarkable fact may open the new possibility of deriving results about the central limit theorem by smoothing techniques and arguments.

It can be checked that $A_{1,t} = 1$, $A_{2,t} = -\alpha_t$, $A_{3,t} = 0$, $A_{4,t} = -1$, $A_{5,t} = -\alpha_t$, $A_{6,t} = -\alpha_t \sigma_t^2$ and $A_{7,t} = \sigma_t^2$.

The formula (3·3) gives the approximation (Hampel, 1974a)

$$p'_n(t)/p_n(t) = -n\alpha_t - \lambda_{3,t}/(2\sigma_t) + O(n^{-1}). \tag{6·2}$$

The third-order formula (4·1) for $p_n$ becomes

$$p_n(t) = n^{\frac{1}{2}}\, \phi(0)\, c_t^{-n}\, \sigma_t^{-1}\{1 + (\lambda_{4,t}/8 - 5\lambda_{3,t}^2/24)\, n^{-1} + O(n^{-2})\}. \tag{6·3}$$

## 7. Computations and comparison with exact results

The results given by the formula for the arithmetic mean have already been compared with some known exact results (Daniels, 1954; Hampel, 1974a); see also Table 4 and Fig. 1. The exact density for the Huber-estimator with known scale with $\psi(x) = x$ if

Table 1. *Cumulative distribution of Huber-estimator $H_1$ ($k = 1{\cdot}4$) under 5% contaminated normal with contamination at $\pm\infty$*; E, *exact*; N, *new approximation, from formula* (3·3) *or* (4·3); G, *indirect method via* $\bar{X}$ (§ 9; *Daniels*, 1954)

| | $n = 1$ | | | $n = 3$ | | | $n = 5$ | | | $n = 7$ | | $n = 9$ | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $t$ | E | N | G | E | N | G | E | N | G | E | N | E | N |
| 0·1 | 0·53779 | 0·53771 | 0·53414 | 0·56234 | 0·56260 | 0·56266 | 0·57974 | 0·57991 | 0·57996 | 0·59384 | 0·59397 | 0·60597 | 0·60607 |
| 0·5 | 0·68168 | 0·68020 | 0·66860 | 0·78124 | 0·78213 | 0·78159 | 0·84016 | 0·84065 | 0·84039 | 0·87979 | 0·88010 | 0·90808 | 0·90829 |
| 1·0 | 0·82399 | 0·81572 | 0·80347 | 0·93241 | 0·93284 | 0·93210 | 0·97201 | 0·97221 | 0·97199 | 0·98788 | 0·98796 | 0·99462 | 0·99465 |
| 1·5 | 0·91130 | 0·89227 | 0·88358 | 0·98112 | 0·98061 | 0·98027 | 0·99586 | 0·99587 | 0·99580 | 0·99906 | 0·99907 | 0·999783 | 0·999784 |
| 2·0 | 0·95326 | 0·92655 | 0·92125 | 0·99371 | 0·99237 | 0·99224 | 0·99912 | 0·99904 | 0·99902 | 0·999875 | 0·999870 | 0·999982 | 0·999982 |
| 2·5 | 0·96905 | 0·94000 | | 0·99710 | 0·99540 | | 0·99970 | 0·99957 | | 0·999967 | 0·999956 | 0·999996 | 0·999995 |
| 3·0 | 0·97370 | 0·94480 | | 0·99795 | 0·99627 | | 0·99982 | 0·99969 | | 0·999984 | 0·999972 | 0·999998 | 0·999997 |

Table 2. *Cumulative distribution of Huber-estimator $H_1$ ($k = 1{\cdot}5$) under Cauchy*; E, *exact*; N, *new, from* (3·3) *or* (4·3); G, *indirect* (*Daniels*; § 9)

| | $n = 1$ | | $n = 3$ | | | $n = 5$ | | | $n = 7$ | | | $n = 9$ | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $t$ | E | N | E | N | G | E | N | G | E | N | G | E | N | G |
| 1 | 0·75000 | 0·71918 | 0·83387 | 0·82653 | 0·82880 | 0·88715 | 0·88542 | 0·88716 | 0·92230 | 0·92175 | 0·92280 | 0·94578 | 0·94553 | 0·94610 |
| 3 | 0·89758 | 0·87603 | 0·97187 | 0·96832 | 0·96837 | 0·99175 | 0·99117 | 0·99116 | 0·99751 | 0·99741 | 0·99739 | 0·99924 | 0·99922 | 0·99921 |
| 5 | 0·93717 | 0·91608 | 0·98894 | 0·98619 | 0·98621 | 0·99790 | 0·99756 | 0·99754 | 0·99959 | 0·99954 | 0·99954 | 0·999918 | 0·999912 | 0·99991 |
| 7 | 0·95483 | 0·93156 | 0·99416 | 0·99199 | 0·99199 | 0·99918 | 0·99895 | 0·99894 | 0·99988 | 0·99986 | 0·99985 | 0·999982 | 0·999979 | 0·99998 |
| 9 | 0·96478 | 0·94673 | 0·99640 | 0·99468 | 0·99467 | 0·99960 | 0·99945 | 0·99944 | 0·999953 | 0·999939 | 0·99994 | 0·999994 | 0·999993 | 0·99999 |

$|x| < k, \psi(x) = k \operatorname{sgn} x$ if $|x| \geqslant k$ has been calculated by P. J. Huber by direct convolution and checked in double precision by D. Zwiers for two contaminated normal distributions and by A. Marazzi with fast Fourier transformation via characteristic functions for the Cauchy distribution. The results from our formulae, which were computed by A. Marazzi, using (3·3), and later recomputed by D. Zwiers, using (4·3), will be compared with these exact distributions.

The $\varepsilon$-contaminated normal used by Huber is a standard normal density with probability $1-\varepsilon$ plus point masses at $\pm\infty$, each with probability $\frac{1}{2}\varepsilon$. In our numerical approximations the contaminating point masses were replaced by standard normal distributions at $\pm 12$. For a fine grid of $t$ values, $\alpha_t$ and the other constants $(c_t, \sigma_t, \lambda_{3,t}, A_{i,t})$ were determined by an iterative search and by integrations using some special properties of the Huber $\psi$-function, in general by numerical integration. Note that these constants remain unchanged as $n$ varies, so they need be computed just once. The approximation (3·3) for $p_n'(t)/p_n(t)$ was numerically integrated, exponentiated and again integrated to obtain the normalizing constant and the cumulative distribution. The exact and the approximate values for the 5%-contaminated normal and the Cauchy distribution are given in Tables 1 and 2; more extensive, as well as additional tables can be found in the aforementioned report. Tables 1 and 2 contain also an indirect approximation by means of the saddlepoint method via the arithmetic mean; see § 9.

The approximations require significantly less time and storage on the computer than the exact computations. For example, even though the first numerical integration of (3·3), along with several constants, later turned out to be superfluous, since it can be replaced by (4·3), the exact computations for the Cauchy for $n = 1-9$ at only 5 points $t = 1, 3, 5, 7, 9$ required more than ten times as much costs and CPU time, 22 min versus 2 min on a CDC 6500, than the approximate computations using (3·3) for a grid of 108 points, $t = 0(0{\cdot}2)7(2)153$.

The relative percent error of the tail area $100(\mathrm{N}-\mathrm{E})/(1-\mathrm{E})$, where N is the approximate and E the exact cumulative, has been calculated. For the contaminated normal with $\varepsilon = 5\%$ the relative error is about or below 1% for $t = 0{\cdot}5$ down to $n = 1$, for $t = 1$ down to $n = 3$, for $t = 1{\cdot}5$ down to $n = 5$. In terms of percentage points, for the same distribution, the relative error is about or below 1% down to $n = 3$ at the one-sided 5%-point, $n = s$ at the 1% point, $n = 7$ at the 0·1%-point and $n = 9$ at the 0·01%-point. For the same critical values and sample sizes, the relative errors in the Cauchy case are about or below 7%. It is only with small $n$ and large $t$ that the relative errors become larger and even here the estimate is fairly good; see, for instance, the 5% contaminated normal with $n = 3$ and $t = 3{\cdot}0$, a relative error of 82%, the actual difference being 0·002 (0·99795 − 0·99627).

Relative tail area errors obtained by use of the first-order approximation $p_n'(t)/p_n(t) \doteq nA_{2,t}$, which is the basis for large deviations, § 10, and by use of the third-order formula (4·1) with renormalization have been computed. They show that a big

Table 3. *First and second order terms in* (3·3), *written as* $p_n'/p_n = nA_2 + B$, *along with* $\alpha_t$, *for* $H_1$ $(k = 1{\cdot}4)$ *under normal* 5%*-contaminated at* $\pm\infty$.

| $t$ | $A_2$ | $B$ | $\alpha_t$ |
|---|---|---|---|
| 0·2 | −0·15803 | −0·15129 | 0·19921 |
| 0·4 | −0·30867 | −0·30939 | 0·39425 |
| 0·8 | −0·55249 | −0·66703 | 0·75214 |
| 1·2 | −0·65315 | −1·06915 | 1·02918 |
| 1·6 | −0·57664 | −1·42561 | 1·19706 |
| 2·0 | −0·39724 | −1·69161 | 1·27391 |

improvement is achieved by inclusion of the second-order term, while the third-order term yields only a small and usually unimportant further improvement. As an example, for the contaminated normal with $\varepsilon = 5\%$, $n = 7$ and $t = 1$, the relative tail area errors are $-57\%$, $0{\cdot}66\%$ and $0{\cdot}25\%$ for the three different orders of approximation. Thus, the second-order formula (3·3), with its nice linearity in $n$, appears to be the most reasonable one.

The need for including the second-order terms becomes also clear from Table 3, which shows what happens in the $K_n(t)$ domain and which also demonstrates nicely the approximate proportionality of the constants with $t$ for small $t$, corresponding to normality, and the deviations from linearity, even up to downbending of $A_{2,t}$, for larger $t$. This complicated behaviour of $K_n(t)$ makes it hard for any expansion about a single point to achieve good accuracy in the tails.

## 8. Relationship to the saddlepoint method

The method closest to those used in this paper is the saddlepoint expansion (Daniels, 1954). Hampel (1974a) noted that integration of formula (6·2) for the arithmetic mean gave precisely Daniels's (1954, p. 633) saddlepoint approximation, apart from the normalizing constant. More recently, H. E. Daniels, in a private communication, has shown that (4·2) can also be obtained by means of a saddlepoint approximation. That this is possible is indicated by the close relationship between the saddlepoint approximation and conjugate distributions (Daniels, 1954, p. 639). In fact, the saddlepoint expansion and the expansion starting with (4·1) for $p_n(t)$ obviously yield identical results.

On the other hand, the expansion of $p_n'/p_n$ and the saddlepoint expansion, infinite, and without renormalization, differ in two minor aspects: first, the free constant of integration in $\log p_n$ to be determined by numerical integration of $p_n$ is fixed in the saddlepoint expansion to be $\log\{n^{\frac{1}{2}}\,\phi(0)\}$; secondly, integration and exponentiation of the expansion of $p_n'/p$ yields an expansion of $p_n$ entirely in the exponent, namely of the form, § 10,

$$p_n(t) \propto \exp\{-n\tilde{\alpha}(t)-\tilde{\beta}(t)-\tilde{\gamma}(t)/n-\ldots\}$$

with $\tilde{\alpha}'(t) = \alpha(t)$, etc., while in the saddlepoint expansion (2·6) of Daniels (1954) the third and further terms $\exp\{-\tilde{\gamma}(t)/n-\ldots\}$ are expanded into $\{1-\tilde{\gamma}(t)/n\pm\ldots\}$. This expansion of the exponent can cause finite sections of the saddlepoint expansion to yield negative densities. For the Edgeworth expansion, formulae (10·3), (10·4) and Table 4, an example is given later where expansion of an exponent roughly doubles the approximation error in the best range. However, in our case the first two terms give an excellent approximation, and here the methods differ only by their normalizing constant. The constant of the saddlepoint approximation can be improved (Hampel, 1974a) by using the third, that is $1/n$, term for $t = 0$; and as noted by Daniels (1954), it can be even more improved by exact renormalization using numerical or analytical integration. Thus, the second-order formula (3·3) and the saddlepoint approximation with renormalization give identical results.

## 9. Indirect approach via arithmetic mean

It is possible to utilize the older results for the arithmetic mean in order to derive asymptotic approximations for $M$-estimators with monotone $\psi$, by noting that the event $T \leqslant t$ is essentially the same as the event $\Sigma\,\psi\,(X_i-t) \leqslant 0$. Thus one can, for each $t$ separately, determine the approximate cumulative distribution of $\bar{Y}_t = \Sigma_i\,\psi(X_i-t)/n$ by using either the results of Daniels (1954) or Hampel (1974a) for the arithmetic mean of the $\psi$'s, and then read off $\mathrm{pr}(\bar{Y}_t \leqslant 0) = \mathrm{pr}(T \leqslant t)$. Obviously, this approach needs more computation than the direct approach; and in the case of $M-\alpha$-quantiles, the result is not exact even using renormalization, and $\bar{Y}_t$ does not even have a density. But, as H. E. Daniels has pointed out, if only a single tail area instead of the full density function is required, this approach is simpler. The accuracy achieved is comparable with that of the direct approach. This is also shown by the numerical results, G in Tables 1 and 2, by Daniels, which are included with his kind permission and which were computed by D. Guest using the saddlepoint approximation for $\bar{Y}_t$ with the renormalization. Both approximate cumulatives N, direct method, and G in Tables 1 and 2 are about equally close to E, exact cumulative, but they show different behaviour.

## 10. Relationship to Edgeworth expansion and large deviations

In this section, our formula for $p_n'/p_n$ and the saddlepoint method are compared to the classical methods of Edgeworth expansion (Cramér, 1946, pp. 133, 223, 229, etc.; Daniels, 1954, p. 635) and large deviations (Richter, 1957, pp. 212, 214; Feller, 1966, p. 520, etc.; Cramér, 1938). In addition, from the formula for $p_n'/p_n$ some other variants, including a new variant of the Edgeworth expansion, are derived.

First we recall that our expansions for $p_n'(t)/p_n(t)$ and $p_n(t)$ as well as the saddlepoint method differ from other methods by expanding not only at one point $t=0$, but rather at every $t$. For the arithmetic mean, this is merely a shift in $K_n$-space: $h_t'(x)/h_t(x) = f'(x)/f(x)+\alpha_t$. All three methods utilize the Edgeworth expansion only at the expectation, though in slightly different ways; see also Daniels (1954, p. 634) for relations between saddlepoint approximation and Edgeworth expansion.

To facilitate comparisons we consider the arithmetic mean $T_n = \bar{X}_n$ with density $p_n$ of independent identically distributed observations $X_i$ with underlying density $f$ and

$$E(X_i) = 0, \quad \mathrm{var}(X_i) = \sigma^2, \quad E(X_i^3)/\sigma^3 = \lambda_3, \quad E(X_i^4)/\sigma^4 - 3 = \lambda_4.$$

Assume $f$ fulfills all necessary regularity conditions such as those of Richter (1957, p. 208).

Assume that the following general form of expansion basic to this paper holds:

$$p_n'(t)/p_n(t) = -n\alpha(t)-\beta(t)-\gamma(t)/n-\ldots, \tag{10·1}$$

where $\alpha(t) = \alpha_t$ and $\beta(t) = \lambda_{3,t}/(2\sigma_t)$ in the previous notation. Now expand the terms into Taylor series around $t=0$: $\alpha(t) = \alpha(0)+\alpha'(0)t+\frac{1}{2}\alpha''(0)t^2+\ldots$, etc. Integration yields

$$\begin{aligned}\log p_n(t) = \Omega_n - n\alpha(0)t-\tfrac{1}{2}n\alpha'(0)t^2-n\alpha''(0)t^3/6-\ldots-\beta(0)t-\tfrac{1}{2}\beta'(0)t^2-\beta''(0)t^3/6-\ldots\\ -\gamma(0)t/n-\tfrac{1}{2}\gamma'(0)t^2/n-\ldots.\end{aligned} \tag{10·2}$$

We assume that the constant of integration can be expanded as

$$\Omega_n = \log\{n/(2\pi\sigma^2)\}^{\frac{1}{2}}(1+\omega_1/n+\ldots).$$

Exponentiation of (10·2) yields a doubly infinite expansion for $p_n(t)$, from which most other known expansions can be derived, as well as new ones.

The main question is at which rate $t\to 0$ as $n\to\infty$. If we aim for good accuracy at a fixed multiple of the standard deviation of the distribution of $T_n$, as does the Edgeworth expansion, we are led to the choice $nt^2 = \text{const} > 0$. This choice induces an ordering of the terms of (10·2) and thus an infinite expansion for $\log p_n(t)$ which appears to be new.

The leading terms up to third order using $\lambda_4$, which empirically seem to give a good fit near 0 even for small $n$, are, using $\alpha(0) = 0$ because of centring,

$$p_n(t) \doteq \{n/(2\pi\sigma^2)\}^{\frac{1}{2}} \exp\{-\tfrac{1}{2}n\alpha'(0)\,t^2\} \times \exp\{-\alpha''(0)\,nt^3/6 - \beta(0)\,t - \alpha^{(3)}(0)\,nt^4/24 - \tfrac{1}{2}\beta'(0)\,t^2 + \omega_1/n\}. \qquad (10{\cdot}3)$$

Expansion of the second exponential yields precisely the usual third-order Edgeworth approximation and corresponding expansion of the infinite series obviously yields the Edgeworth series:

$$p_n(t) \doteq \{n/(2\pi\sigma^2)\}^{\frac{1}{2}} \exp\{-\tfrac{1}{2}n\alpha'(0)\,t^2\} \times\{1 - \alpha''(0)\,nt^3/6 - \beta(0)\,t - \alpha^{(3)}(0)\,nt^4/24 - \tfrac{1}{2}\beta'(0)\,t^2 + \omega_1/n + \alpha''(0)^2\,n^2\,t^6/72 + \alpha''(0)\,\beta(0)\;nt^4/6 + \tfrac{1}{2}\beta(0)^2\,t^2\}. \qquad (10{\cdot}4)$$

Comparison of terms, e.g. with Cramér (1946, p. 229), provides the bridge to the moments of the $X_i$:

$$\alpha(0) = 0,\quad \alpha'(0) = 1/\sigma^2,\quad \alpha''(0) = -\lambda_3/\sigma^3,\quad \alpha^{(3)}(0) = 3\lambda_3^2/\sigma^4 - \lambda_4/\sigma^4,$$
$$\beta(0) = \lambda_3/(2\sigma),\quad \beta'(0) = \lambda_4/(2\sigma^2) - \lambda_3^2/\sigma^2,\quad \omega_1 = \lambda_4/8 - 5\lambda_3^2/24. \qquad (10{\cdot}5)$$

Formula (10·3), which we may call 'Edgeworth in exponent', proves in an example, Table 4, to be about twice as good as Edgeworth (10·4) in the central region where both are good. Outside about two standard deviations of $T_n$ from its mean both are very bad; so it matters little that (10·3) loses the norming of (10·4) and may sometimes explode, namely if $\lambda_4 - 3\lambda_3^2 \geqslant 0$ and not $\lambda_3 = \lambda_4 = 0$, while (10·4) may lead to negative densities.

There are, of course, many possibilities for the speed with which $t \to 0$. For instance, the expansion with $nt$ = const has as its leading term

$$p_n(t) \doteq \{n/(2\pi\sigma^2)\}^{\frac{1}{2}} \exp\{-\tfrac{1}{2}n\alpha'(0)\,t^2 - \beta(0)\,t + \omega_1/n\}.$$

This is the nonnormalized 'best fitting normal density' which uses $\lambda_3$ for a shift of the mean and in addition $\lambda_4$ for approximate adjustment of the height at $t = 0$. The normalized counterpart can be obtained from the expansion of (10·1) around $t = 0$ by integrating $p_n'(t)/p_n(t) \doteq -n\alpha'(0)\,t - \beta(0)$. These and other variants were studied for the

Table 4. *Comparison of approximations for density of mean of four independent exponentially distributed observations*

| $t$ | Exact density | EDG exp (10·3) | %err. | EDG (10·4) | % err. | Norm. approx. | % err. | Large dev. (10·6) | % err. |
|---|---|---|---|---|---|---|---|---|---|
| −0·25 | 0 | 0·00168 | | −0·03920 | | 0·0350 | | 0·0002 | |
| 0 | 0 | 0·04596 | | 0·02175 | | 0·1080 | | 0·0105 | |
| 0·25 | 0·24525 | 0·29544 | 20·5 | 0·29765 | 21·4 | 0·2590 | 5·6 | 0·1076 | −56·1 |
| 0·5 | 0·72179 | 0·70413 | −2·4 | 0·69231 | −4·1 | 0·4840 | −11·9 | 0·3848 | −46·7 |
| 0·75 | 0·89617 | 0·89126 | −0·5 | 0·88857 | −0·8 | 0·7042 | −21·4 | 0·6869 | −23·4 |
| 1 | 0·78147 | 0·78143 | −0·005 | 0·78126 | −0·02 | 0·7979 | 2·1 | 0·7979 | 2·1 |
| 1·25 | 0·56150 | 0·56358 | 0·4 | 0·56584 | 0·8 | 0·7042 | 25·4 | 0·7162 | 27·6 |
| 1·5 | 0·35694 | 0·36151 | 1·3 | 0·36968 | 3·6 | 0·4840 | 35·6 | 0·5371 | 50·5 |
| 1·75 | 0·20852 | 0·20305 | −2·6 | 0·20052 | −3·8 | 0·2590 | 24·2 | 0·3313 | 58·9 |
| 2 | 0·11451 | 0·08952 | −21·8 | 0·09373 | −18·1 | 0·1080 | −5·7 | 0·1507 | 31·6 |

Approximation (3·3) is exact in this case (Daniels, 1954, p. 636), and the saddlepoint approximation without renormalization has constant relative error of +2·1%.

EDG exp, Edgeworth in exponent; norm. approx., normal approximation; large dev., large deviations; % err., % error.

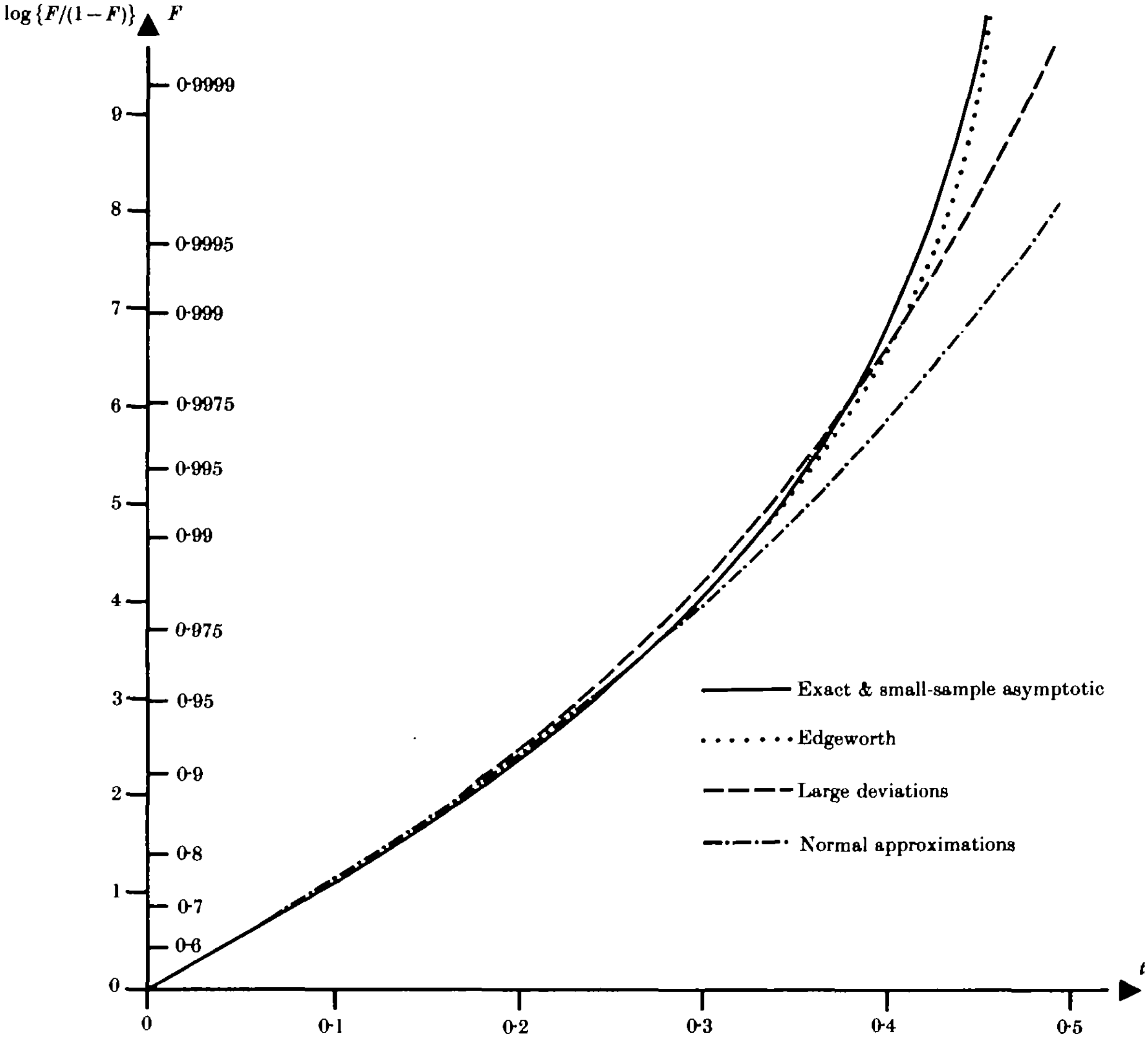


Fig. 1. Comparison of approximations for the cumulative of the mean of four independent uniformly distributed data, in logistic scale. Compared with the exact cumulative of $\bar{X}_4$ under $U(-\frac{1}{2},\frac{1}{2})$ are Edgeworth (10·4), large deviations (10·6) and normal approximation; the new approximation (3·3) or (4·3) coincides with the exact cumulative within drawing accuracy.

example of Table 4, but are not pursued here because of lack of space. Some conclusions are that the precise aims of mediocre simple fits have to be selected first and that renormalization may be much less suitable here than for the globally good approximations related to the saddlepoint technique.

An extreme case of asymptotic 'directions' in (10·2) is first to let $n \to \infty$ and then $t \to 0$ or the limiting case of $n^c t = \text{const}$ for $c \to 0$. This amounts to keeping only the leading constant term and the expansion of $\alpha(t)$ and leads precisely to the large deviations expansion for $p_n(t)$. The usual large deviations approximation using $\lambda_4$ becomes (Richter, 1957)

$$p_n(t) \doteq \{n/(2\pi\sigma^2)\}^{\frac{1}{2}} \exp\{-\tfrac{1}{2}n\alpha'(0)\,t^2\} \exp\{-\alpha''(0)\,nt^3/6 - \alpha^{(3)}(0)\,nt^4/24\}. \quad (10·6)$$

See (10·5) and compare with (10·3).

Now it is known from a number of examples that for small $n$ the $\beta$ terms are definitely needed to give a good approximation. Hence it can be expected that even the full infinite large deviations expansion will give a bad fit except for very large $n$, but it is still remarkable that in the two examples computed in this paper, Table 4 and Fig. 1, formula (10·6) is even worse than the normal approximation. The large deviation formulae for the cumulative distribution of $T_n$, as given, for example, by Feller (1966, p. 520), contain an additional approximation, namely of the normal tail area, and hence are not likely to fit any better.

If we go to the other extreme in the choice of asymptotic 'directions' and let first $t \to 0$ and then $n \to \infty$, we obtain the expansion of $\Omega_n$, albeit only for $\log p_n(0)$. This is almost the same as the saddlepoint series expansion (2·6) of Daniels (1954) for $t = 0$, there $\bar{x} = 0$, the latter differing again in having the exponential expanded into a sum. However, the expansion of $\Omega_n$ is still unknown beyond the first two terms.

For completeness, we note the obvious facts that the first terms of (10·3), (10·4) and (10·6), up to using $\alpha'(0)$, are merely the normal approximation; that the latter, (10·6) and the saddlepoint approximation, the leading term of the saddlepoint expansion, coincide for $t = 0$; and that from (10·1) $\alpha'(0) = 1/\sigma 2$ is the limit of the derivative of $p_n'(t)/p_n(t)$ divided by $n$ at $t = 0$, namely of the inverse standardized local variance at $t = 0$.

The authors are very grateful to H. E. Daniels, the first worker in the field, for very pleasant cooperation, many stimulating discussions and for contributing the computations with the indirect method; to P. J. Huber for his permanent interest, support of the second author and for supplying his exact computations for $H_1(1{\cdot}4)$; to the Forschungsinstitut für Mathematik, ETH Zürich for their support of the first author; to A. Marazzi, D. Guest and D. Zwiers for careful computer computations; and to a referee for very thorough and detailed comments which led to considerable changes and improvements of the earlier research report.

## References

Barndorff-Nielsen, O. & Cox, D. R. (1979). Edgeworth and saddle-point approximations with statistical applications (with discussion). *J. R. Statist. Soc. B* **41**, 279–312.

Cramér, H. (1938). Sur un nouveau théorème-limite de la théorie des probabilités. *Actualités Scientifiques et Industrielles* 736. Paris: Hermann & Cie.

Cramér, H. (1946). *Mathematical Methods of Statistics*. Princeton University Press.

Daniels, H. E. (1954). Saddlepoint approximations in statistics. *Ann. Math. Statist.* **25**, 631–50.

Daniels, H. E. (1980). Exact saddlepoint approximations. *Biometrika* **67**, 59–63.

Durbin, J. (1980a). Approximations for densities of sufficient estimators. *Biometrika* **67**, 311–33.

Durbin, J. (1980b). The approximate distribution of partial serial correlation coefficients calculated from residuals from regression on Fourier series. *Biometrika* **67**, 335–49.

Feller, W. (1966). *An Introduction to Probability Theory and its Applications*, Vol. II. New York: Wiley.

Field, C. A. (1978). Summary of small sample asymptotics. In *Proceedings of the Second Prague Symposium on Asymptotic Statistics*, Ed. P. Mandl and M. Hušková, pp. 173–80. Prague: Society of Czechoslovak Mathematicians and Physicists, Charles University.

Hampel, F. R. (1971). A general qualitative definition of robustness. *Ann. Math. Statist.* **42**, 1887–96.

Hampel, F. R. (1974a). Some small sample asymptotics. In *Proceedings of Prague Symposium on Asymptotic Statistics, 1973*, Vol. II, Ed. J. Hájek, pp. 109–26. Prague: Charles University.

Hampel, F. R. (1974b). The influence curve and its role in robust estimation. *J. Am. Statist. Assoc.* **69**, 383–93.

Huber, P. J. (1964). Robust estimation of a location parameter. *Ann. Math. Statist.* **35**, 73–101.

Huber, P. J. (1977). *Robust Statistical Procedures*. CBMS-NSF Regional Conference Series in Applied Mathematics No. 27. Philadelphia: Society for Industrial and Applied Mathematics.

Jeffreys, H. (1961). *Theory of Probability*, 3rd edition. Oxford: Clarendon Press.

KHINCHIN, A. I. (1949). *Mathematical Foundations of Statistical Mechanics*. New York: Dover.
RICHTER, W. (1957). Local limit theorems for large deviations. *Theory Prob. Applic.* **2**, 206–20.

[*Received September* 1978. *Revised June* 1981]