The Annals of Statistics
1982, Vol. 10, No. 3, 672–689

SMALL SAMPLE ASYMPTOTIC EXPANSIONS FOR MULTIVARIATE *M*-ESTIMATES[1]

BY CHRISTOPHER FIELD

Dalhousie University

Asymptotic expansions are derived for the densities of multivariate M-estimates. The expansion is based on a saddlepoint technique and yields good accuracy in the tails for small sample sizes. Numerical results are given for robust estimation of location and scale and these are compared with known Monte Carlo results.

1. Introduction. This paper considers asymptotic expansions for the densities of multivariate M-estimates for small sample sizes. The technique is an adaptation of the saddlepoint technique developed by Daniels (1954) and the small sample method of Hampel (1973). Small sample approximations, very accurate even down to $n = 3$, were developed for M-estimates of location with a monotone ψ-function in Field and Hampel (1982). An approximating formula for a multivariate mean appears in Field (1978).

The approximating density at a fixed point is obtained by using a conjugate or associated density to re-center the underlying density about the fixed point and then to use the multivariate Edgeworth expansion locally to obtain an approximation to the density at the fixed point. In contrast to the usual technique of using a single high order Edgeworth expansion for all points at which the density is to be approximated, this approach uses a different low order Edgeworth expansion for each point, using the Edgeworth expansion only in the middle where it is very accurate.

Formally, consider observations $x_1, \cdots, x_n$ from an underlying density $f_\theta(x)$ where $\boldsymbol{\theta} = (\theta_1, \cdots, \theta_p)$, Note that the x_i's may be univariate or multivariate. To estimate $\boldsymbol{\theta}$, we use the M-estimate $\mathbf{T}_n$ which is a solution $\mathbf{t}$ of the system of equations:

$$(1) \qquad \textstyle\sum_{i=1}^n \psi_j(x_i, \mathbf{t}) = 0 \quad \text{for} \quad j = 1, \cdots, p.$$

The problem is to find an asymptotic expansion for the density $p_n(\mathbf{t})$ of $\mathbf{T}_n$.

This paper is related to a recent paper of Barndorff-Nielsen and Cox (1979). Their paper is similar in that they use saddlepoint or indirect Edgeworth expansions to approximate densities. They restrict attention to standardized sums and develop asymptotic expansions in the multivariate case for the standardized mean (eg. their formula (4.7)). From this point, they concentrate on approximating the density of part of the standardized mean conditional on the remainder of the standardized mean. This paper differs in that approximations for a much broader class of estimates are given. As has been often noted (cf. Huber, 1977), the class of M-estimates includes both the multivariate mean by setting $\psi_j(\mathbf{x}, \mathbf{t}) = (x_j - t_j)$ and maximum likelihood estimates by setting $\psi_j = (\partial f/\partial\theta_j)/f$, as well as the so-called robust M-estimates which have bounded score functions ψ_j. The univariate case of M-estimates for location with monotone ψ has been treated in Field and Hampel (1978). In that case, it is possible to use an approach similar that developed for the mean by Daniels (1953) and, in fact, Professor Daniels has carried out some computations for robust location estimates which are given in Field and Hampel (1982, Table 1).

The generalization of the saddlepoint approach to multivariate M-estimates requires a

Received July 1979; revised January 1982.

[1] Research supported in part by Natural Sciences and Engineering Research Council Grant A7675 and by Forschungsinsitut für Mathematik, ETH. Zurich.

AMS 1970 *subject classification.* Primary 62E15; secondary 62E20.

Key words and phrases. Direct and indirect Edgeworth expansions, local approximations, location, scale, robust estimation, saddlepoint approximation, studentizing.

local expansion of the estimate in terms of functions of means and an application of the asymptotic expansion for the means.

The steps in the argument are as follows:

(i) Show that the density of $\mathbf{T}_n$ at $\mathbf{t}_0$ under $f_\theta(x)$ can be computed as the density of $\mathbf{T}_n$ at $\mathbf{t}_0$ under a conjugate or associated density which has been centered at $\mathbf{t}_0$.
(ii) Find a Taylor series expansion for $\mathbf{T}_n - \mathbf{t}_0$ which is valid locally.
(iii) Use a multivariate Edgeworth expansion at $\mathbf{0}$ with the centered conjugate density to obtain an approximation to $p_n(\mathbf{t})$.

Step (i) enables us to approximate the density at $\mathbf{t}_0$ using an Edgeworth approximation only in the neighborhood of the expected value of $\mathbf{T}_n$. This ensures that at any point $\mathbf{t}_0$, we obtain the very good numerical accuracy of the Edgeworth approximation at its expected value. The result given in Lemma 1 is a generalization of equation (2.6) in Barndorff-Nielsen and Cox (1979) which in turn is essentially given by Daniels (1954). Steps (ii) and (iii) follow very closely arguments given in Bhattacharya and Ghosh (1978).

The next section of the paper develops the asymptotic formula for the density. Following this, the approximating formula is examined for the arithmetic mean and for robust estimates of location and scale. Numerical computations have been carried out for robust estimates of location and scale and the results are compared to existing Monte-Carlo results where appropriate. The percentiles of a "studentized" statistic are computed and the validity of using t-densities with reduced degrees of freedom to determine the percentiles is examined.

It should be noted that the asymptotic expansions obtained by Pfanzagl (1973) for univariate minimum contrast estimators are closely related to the expansions given by Bhattacharya and Ghosh (1978) for multivariate estimates and could be used as a starting point for saddle-point approximations for univariate M-estimates.

2. Asymptotic formula. In order to carry out the step (i) above, the following assumptions are made:

ASSUMPTION 1. The system of equations (1) has a unique solution.

ASSUMPTION 2. The joint density of $(\sum_{i=1}^n \psi_1(X_i, \mathbf{t}_0), \cdots, \sum_{i=1}^n \psi_p(X_i, \mathbf{t}_0), \mathbf{T}_n)$ exists and has Fourier transforms which are absolutely integrable both under f_θ and $h_{\mathbf{t}_0,\theta}$.

In general, Assumption 1 may be somewhat difficult to verify and is probably not absolutely essential. The question of how multiple roots might be handled is addressed in the final section. Results concerning the uniqueness of roots of a non-linear system of equations can be found in Ortego and Rheinboldt (1970, Section 5.4). Assumption 2 requires that $n > 2p$ and is used primarily in step (i) of the development; if it does not hold, the arguments can be modified to handle other possibilities. In fact, it is possible to give a direct proof of the lemma using only the density of $\mathbf{T}_n$ and certain regularity conditions.

For ease of notation, the underlying density f_θ to be used in evaluating the density of $\mathbf{T}_n$ at $\mathbf{t}_0$ will be denoted by f suppressing the dependence on θ. In situations in which there is some invariance, it may be necessary to evaluate the density only for a subset of f_θ's as would be the case in location or location/scale problems with invariant M-estimates.

To proceed, the density of $\mathbf{T}_n$ at some point $\mathbf{t}_0$, $p_n(\mathbf{t}_0)$, is to be approximated. The next lemma carries out step (i).

LEMMA 1. *If Assumptions* 1 *and* 2 *hold, and the underlying density is* f, *then*

$$p_n(\mathbf{t}_0) = C^{-n}(\mathbf{t}_0) q_n(\mathbf{t}_0),$$

where $q_n(\mathbf{t}_0)$ is the density of $\mathbf{T}_n$ with the underlying density

$$h_{t_0}(x) = f(x)\exp\{\textstyle\sum_{i=1}^{p} \alpha_i\psi_i(x, \mathbf{t}_0)\}C(t_0)$$

and

$$C^{-1}(\mathbf{t}_0) = \int f(x)\exp\{\textstyle\sum_{j=1}^{p} \alpha_j\psi_j(x, \mathbf{t}_0)\}\, dx.$$

PROOF. Let $\mathbf{Z} = (\sum_{k=1}^{n} \psi_1(X_k, \mathbf{t}_0), \cdots, \sum_{k=1}^{n} \psi_p(X_k, \mathbf{t}_0))$ and denote the density of $(\mathbf{Z}, \mathbf{T}_n)$ by $g(\mathbf{z}, \mathbf{t})$ under f and $g_1(\mathbf{z}, \mathbf{t})$ under $h_{\mathbf{t}_0}$. Writing $\mathbf{T}_n = (T_1(\mathbf{x}), \cdots, T_p(\mathbf{x}))$ with $\mathbf{x} = (x_1, \cdots, x_n)$, the moment generating function of $(\mathbf{Z}, \mathbf{T}_n)$ can be written as

$$M(\mathbf{u}, \mathbf{v}) = \int \cdots \int \exp\{\textstyle\sum_{k=1}^{n} \sum_{j=1}^{p} u_j\psi_j(x_k, \mathbf{t}_0) + \sum_{j=1}^{p} v_j t_j(\mathbf{x})\} \prod_{k=1}^{n} f(x_i)\, dx_1 \cdots dx_n.$$

Choose $\mathbf{u} = (\alpha_1 + iy_1, \alpha_2 + iy_2, \cdots, \alpha_p + iy_p) = \boldsymbol{\alpha} + i\mathbf{y}$, $\mathbf{v} = (iw_1, \cdots, iw_p) = i\mathbf{w}$.

Now

$$\begin{aligned} M(\boldsymbol{\alpha} + i\mathbf{y}, i\mathbf{w}) &= \int \cdots \int \exp\{\textstyle\sum_{k=1}^{n} \sum_{j=1}^{p} iy_j\psi_j(x_k, \mathbf{t}_0) + \sum_{j=1}^{p} iw_j t_j(x)\} \\ &\quad \times \exp\{\textstyle\sum_{k=1}^{n} \sum_{j=1}^{p} \alpha_j\psi_j(x_k, \mathbf{t}_0)\} \prod_{k=1}^{n} f(x_k)\, dx_1 \cdots dx_n \\ &= C^{-n}(\mathbf{t}_0) \int \cdots \int \exp\{\textstyle\sum_{k=1}^{n} \sum_{j=1}^{p} iy_j\psi_j(x_k, \mathbf{t}_0) \\ &\quad + \textstyle\sum_{j=1}^{p} iw_j T_j(x)\} \prod_{k=1}^{n} h_{t_0}(x_k)\, dx_1 \cdots dx_n = C^{-n}(\mathbf{t}_0)M_1(i\mathbf{y}, i\mathbf{w}), \end{aligned}$$

where M_1 is the moment generating function of $(\mathbf{Z}, \mathbf{T}_n)$ under h_{t_0}. Since both M and M_1 are absolutely integrable, we can apply the Fourier inversion formula to give

$$g(\mathbf{z}, \mathbf{t}) = \frac{1}{(2\pi i)^{2p}} \int \cdots \int \exp\{-\textstyle\sum_{j=1}^{p} u_j z_j - \sum_{j=1}^{p} v_j t_j\}M(\mathbf{u}, \mathbf{v})\, d\mathbf{u}\, d\mathbf{v},$$

where components of $\mathbf{u}$ and $\mathbf{v}$ are integrated along the path from $c - i\infty$ to $c + i\infty$ for some c. Choosing $\mathbf{u} = (\boldsymbol{\alpha} + i\mathbf{y})$ and $\mathbf{v} = i\mathbf{w}$, we have

$$\begin{aligned} g(\mathbf{z}, \mathbf{t}) &= \frac{1}{(2\pi)^{2p}} \int \cdots \int \exp\{-\textstyle\sum_{j=1}^{p} (\alpha_j + iy_j)z_j - \sum_{j=1}^{p} iw_j t_j\}M(\boldsymbol{\alpha} + i\mathbf{y}, i\mathbf{w})\, d\mathbf{y}\, d\mathbf{w} \\ &= \frac{C^{-n}(\mathbf{t}_0)}{(2\pi)^{2p}} \exp\{-\textstyle\sum_{j=1}^{p} \alpha_j z_j\} \int \cdots \int exp\{-\textstyle\sum_{j=1}^{p} iy_j z_j - \sum_{j=1}^{p} iw_j t_j\} \\ &\quad \cdot M_1(i\mathbf{y}, i\mathbf{w})\, d\mathbf{y}\, d\mathbf{w} \\ &= C^{-n}(\mathbf{t}_0)\exp\{-\textstyle\sum_{j=1}^{p} \alpha_j z_j\}\, g_1(\mathbf{z}, \mathbf{t}). \end{aligned}$$

Now

$$T_n(\mathbf{x}) = \mathbf{t}_0 \Leftrightarrow \textstyle\sum_{i=1}^{n} \psi_j(x_i, \mathbf{t}_0) = 0, \quad j = 1, \cdots, p \Leftrightarrow z = \mathbf{0}.$$

Hence, $g(\mathbf{z}, \mathbf{t}_0) = C^{-n}(\mathbf{t}_0)g_1(\mathbf{z}, \mathbf{t}_0)$ and from this the result follows. □

Before proceeding with step (ii), the Taylor series expansion of $\mathbf{T}_n - \mathbf{t}_0$, a particular value of $\boldsymbol{\alpha}$ will be chosen. For fixed $\mathbf{t}_0$, $\boldsymbol{\alpha}$ is chosen so that the random variables $\psi_j(\mathbf{X}, \mathbf{t}_0)$ will have mean zero under the conjugate density i.e. choose $\alpha_1, \cdots, \alpha_p$ so that

$$\int \psi_j(x, \mathbf{t}_0)\exp\{\textstyle\sum_{j=1}^{p} \alpha_j\psi_j(x, \mathbf{t}_0)\}f(x)\, dx = 0, \quad j = 1, \cdots, p. \tag{2}$$

Although $\boldsymbol{\alpha}$ is dependent on $\mathbf{t}_0$, the dependence will be suppressed throughout the proof but will be exhibited in the final formula.

In order to carry out steps (ii) and (iii), we need the following assumptions which correspond to Assumptions A_1–A_6 in Bhattacharya and Ghosh (1978, page 439). For nonnegative integral vectors $\nu = (\nu^{(1)}, \cdots \nu^{(p)})$ write $|\nu| = \nu^{(1)} + \cdots + \nu^{(p)}$, $\nu! = \nu^{(1)}!\nu^{(2)}! \cdots \nu^{(p)}!$ and let $D^\nu = (D_1)^{\nu^1} \cdots (D_p)^{\nu^{(p)}}$ denote the νth derivative with respect to $\boldsymbol{\theta}$.

ASSUMPTION 3. (i). $\{X_i\}_{i\geq 1}$ are i.i.d. m-dimensional random variables such that there is an open subset U of R^m such that (a) for each $\boldsymbol{\theta} \in \Theta \subset R^p$, $\int_U f_\theta(\mathbf{x})\, d\mathbf{x} = 1$ and (b) for each ν, $1 \leq |\nu| \leq s$, the functions ψ_j, $j = 1, \cdots, p$ have νth derivatives with respect to $\boldsymbol{\theta}$

(ii) For each compact $K \subset \Theta$ and each ν, $1 \leq |\nu| \leq s - 1$, $\sup_{\theta_0\in K} E_{\theta_0} |D^\nu \psi_j(X, \boldsymbol{\theta})|^{s+1} < \infty$; and for each compact K there exists $\varepsilon < 0$ such that

$$\sup_{\theta_0\in K} E_{\theta_0}\{\max_{|\theta-\theta_0|\leq\varepsilon} |D^\nu\psi_j(X, \boldsymbol{\theta})|\}^s < \infty$$

if $|\nu| = s$ for $j = 1, \cdots p$. Here E_{θ_0} denotes expected value with respect to the conjugate density h_{θ_0} with α chosen as in equation (2).

(iii) For each $\boldsymbol{\theta}_0 \in \Theta$, the matrices

$$I(\boldsymbol{\theta}_0) = \{-E_{\theta_0} D_i\psi_r(X, \boldsymbol{\theta}_0)\}, \quad D(\boldsymbol{\theta}_0) = \{E_{\theta_0}\psi_i(X, \boldsymbol{\theta}_0)\psi_r(X, \boldsymbol{\theta}_0)\}$$

are non-singular.

(iv) The functions $I(\boldsymbol{\theta})$, $E_\theta\{D^\nu\psi_i(X, \theta)D^{\nu'}\psi_r(X, \theta)\}$, $1 \leq |\nu|, |\nu'| \leq s - 1$, $1 \leq i, r \leq p$ are continuous on Θ.

(v) The map $\boldsymbol{\theta} \to F_\theta$, where F_θ is the cdf of f_θ, on Θ into the space of all probability measures on R^m is continuous when the latter space is given the (variation) norm topology.

(vi) For each $\boldsymbol{\theta}$ and each ν, $1 \leq |\nu| \leq s - 1$, $D^\nu\psi_j(\mathbf{x}, \boldsymbol{\theta})$ is continuously differentiable in y on U.

As can be seen by comparison, these conditions correspond to those in Bhattacharya and Ghosh (1978) with the convention that all expectations are with respect to the conjugate density and $D_i f$ is replaced by ψ_i.

For the cases of interest, robust M-estimates, these conditions create no difficulties since the score functions ψ_i are bounded and the underlying densities of interest are well-behaved. It should be noted that the conditions on differentiability can be weakened to allow piecewise differentiable score functions.

Although the arguments required to carry out steps (ii) and (iii) are essentially given in Bhattacharya and Ghosh (1978, cf. page 448 and page 435 noting misprint in (2.35)), it seems appropriate, in the interest of more complete exposition, to provide a short summary of the technique. We focus on the case when $s = 3$. Note that the expansion obtained will be assuming the underlying density is h_{t_0}. Lemma 1 can then be used to obtain the expansion under f.

To begin, consider the first two terms in the Taylor's series expansion of (1) about t_0

$$(3)\quad \begin{aligned} 0 = \textstyle\sum_{i=1}^n \psi_r(x_i, \mathbf{t})/n \approx \sum_{i=1}^n \psi_r(x_i, \mathbf{t}_0)/n + \sum_{j=1}^p (t_j - t_{0j}) \sum_{i=1}^n D_j\psi_r(x_i, \mathbf{t}_0)/n \\ + \textstyle\sum_{j,\ell} (t_j - t_{0j})(t_\ell - t_{0\ell}) \sum_{i=1}^n D_\ell D_j\psi_r(x_i, \mathbf{t}_0)/2n, \end{aligned}$$

where D_j represents differentiation with respect to t_j, and $D_\ell D_j$ is used to represent differentiation with respect to t_j and then t_ℓ instead of the more usual $D_{\ell j}$ to avoid notational confusion later on. The expansion we obtain will actually be for the solution of the right hand side of (3). As is shown by Bhattacharya and Ghosh (1978, page 449), the remainder term in (3) which has been ignored gives rise to terms with order of magnitude of the eventual error terms. With this in mind, we refer to the solution of the right hand of (3) as $\mathbf{T}_n$.

In terms of notation, let $Z_r = \psi_r(X, \mathbf{t}_0)$, $Z_{rj} = D_j\psi_r(X, \mathbf{t}_0)$, $Z_{rj\ell} = D_\ell D_j\psi_r(X, \mathbf{t}_0)$ and let E denote expectation with respect to h_{t_0} where α is chosen as in (2) above. Then $EZ_r = 0$ and we let $EZ_{rj} = \mu_{rj}$, $EZ_{rj\ell} = \mu_{rj\ell}$,

$$\mathbf{a} = (\overbrace{0, \cdots, 0}^{p}, \mu_{11}, \mu_{12}, \cdots, \mu_{pp}, \mu_{111}, \mu_{112}, \cdots, \mu_{ppp})$$

and $\mathbf{Z}^+ = (Z_1, \cdots, Z_p, Z_{11}, Z_{12}, \cdots, Z_{ppp})$. Consider the following system of equations, which correspond to the expansion in (3)

$$f_r(\mathbf{z}^+, \mathbf{t}) = z_r + \sum_{j=1}^{p} (t_j - t_{0j})z_{rj} + \frac{1}{2}\sum_{j,\ell} (t_j - t_{0j})(t_\ell - t_{0\ell})z_{rj\ell}$$

for $r = 1, \cdots, p$, i.e. $\mathbf{f} = (f_1, \cdots, f_p)$ maps R^{k+p} into R^p where $k = p + p^2 + p^3$. Since by Assumption 3, the matrix $A = \{\mu_{ij}\}_{i,j=1}^{p}$ is non-singular and $\mathbf{f}(\mathbf{a}, \mathbf{t}_0) = 0$, the implicit function theorem can be applied to prove there exists a unique differentiable function, $H(\mathbf{z}^+)$, $H: R^k \to R^p$, such that $\mathbf{f}(\mathbf{z}^+, H(\mathbf{z}^+)) = 0$ for $\mathbf{z}^+$ in a neighbourhood of $\mathbf{a}$ and $H(\mathbf{z}^+)$ in a neighbourhood of $\mathbf{t}_0$. From this it follows that

$$\text{(4)} \qquad \mathbf{T}_n = H(\bar{\mathbf{Z}}^+) \quad \text{and} \quad \mathbf{t}_0 = H(\mathbf{a}) \quad \text{where} \quad \bar{Z}_1 = \sum_{i=1}^{n} \psi(X_i, \mathbf{t}_0)/n, \quad \text{etc.}$$

The next step is to expand $H(\mathbf{Z}^+)$ in a Taylor series expansion about $\mathbf{a}$. Using the first two terms of the expansion, and using the relationship $f(\mathbf{Z}^+, H(\mathbf{Z}^+)) = 0$ to evaluate derivatives, it is straightforward to obtain the following expansion

$$\text{(5)} \quad (\mathbf{T}_n - \mathbf{t}_0)_i = \sum_j b_{ij}\bar{Z}_j - \sum_{j,\ell}\{\sum_r b_{ir}C_{j\ell}(r)\}\bar{Z}_j\bar{Z}_\ell + \sum_{j,\ell}\sum_r b_{ir}b_{\ell j}\bar{Z}_j\,\bar{Y}_{r\ell} + \text{error term},$$

where b_{ir} is the (i, r) element of $A^{-1} = \{\mu_{ir}\}^{-1}$, $\bar{Y}_{r\ell} = \bar{Z}_{r\ell} - \mu_{\ell r}$ and $C_{j\ell}(r) = \sum_{m,m_1} b_{m\ell}\mu_{rmm_1}b_{m_1 j}$.

The next step, working out the Edgeworth expansion, requires the cumulants of $n^{1/2}(\mathbf{T}_n - \mathbf{t}_0)$. The results of James and Mayne (1962) can be used here for the polynomial part. This will suffice since Bhattacharya and Ghosh (1978) show that the polynomial part determines the expansion. We denote the cumulant of order $(r_1, \cdots, r_p)$ for $n^{1/2}(\mathbf{T}_n - \mathbf{t}_0)$ by $\chi^{(r_1,\cdots r_p)}$ and the cumulants of $(\bar{Z}_1, \cdots, \bar{Z}_p, \bar{Y}_{11}, \bar{Y}_{12}, \cdots, \bar{Y}_{pp})$ by K^i, $K^{(r\ell)}$, K^{ij}, $K^{i(r\ell)}$ where

$$K^i = E\bar{Z}_i = 0, \quad K^{(r\ell)} = E\bar{Y}_{r\ell} = 0, \quad K^{ij} = E\bar{Z}_i\bar{Z}_j, \quad K^{i(r\ell)} = E\bar{Z}_i\bar{Y}_{r\ell} \quad \text{and so on}$$

where the expectation is with respect to h_{t_0}.

Since the random variables involved are all means, it is straightforward to verify that all the rth order cumulants $K^{i_1,\cdots,i_r}$ are of order n^{-r+1}. The cumulants of $n^{1/2}(\mathbf{T}_n - \mathbf{t}_0)$ can be expressed in terms of K's as follows (cf. James and Mayne, 1962, page 51):

$$\lambda^{s_1} \equiv \underset{\substack{\uparrow \\ s_1}}{\chi^{(0,\cdots,1,\cdots,0)}} = n^{1/2}\{\sum_{j,\ell}' (\sum_r b_{s_1 r}C_{j\ell}(r))K^{j\ell} + \sum_{j,r,\ell} b_{ir}b_{\ell j}K^{j(r\ell)}\} + O(n^{-3/2})$$
$$= d_{s_1}/n^{1/2} + O(n^{-3/2})$$

$$\lambda^{s_1s_2} \equiv \underset{\substack{\uparrow \quad \uparrow \\ s_1 \quad s_2}}{\chi^{(0,\cdots,1,\cdots,1,\cdots,0)}} = \sum_{j,\ell} b_{s_1}jb_{s_2\ell}K^{j\ell} + O(n^{-1})$$
$$= \{A^{-1}\textstyle\sum_{\mathbf{t}_0}(A^{-1})^T\}_{s_1,s_2} + O(n^{-1}) \equiv C_{s_1s_2} + O(n^{-1}).$$

where $\sum_{\mathbf{t}_0} = \{K^{j\ell}\}_{1\le j,\ell\le p}$, and

$$\lambda^{s_1s_2s_3} \equiv \underset{\substack{\uparrow \quad \uparrow \quad \uparrow \\ s_1 \quad s_2 \quad s_3}}{\chi^{(0,\cdots,1,\cdots,1,\cdots,1,\cdots,0)}} = d_{s_1s_2s_3}/n^{1/2} + O(n^{-3/2}).$$

All higher order cumulants can be shown to be of order $O(n^{-1})$ or higher.

The final step is to write out the characteristic function of $n^{1/2}(\mathbf{T}_n - \mathbf{t}_0)$, using the expressions for the cumulants above and then apply a Fourier inversion. (cf. Bhattacharya and Ghosh, 1978, page 436). Since Assumption 3 holds, the results of Theorem 3 with $s = 3$ in Bhattacharya and Ghosh (1978) can be applied to our problem and we have the result that for every compact set $K \subset \Theta$,

$$\sup_{\mathbf{t}_0 \in K} \Big| P_{\mathbf{t}_0}\{n^{1/2}(\mathbf{T}_n - \mathbf{t}_0) \in B\} - \int_B [1 + \textstyle\sum_{j=1}^p d_j D_j\{\phi_c(\mathbf{x})\}/n^{1/2}$$
$$+ \textstyle\sum_{j,k,l} d_{jkl} D_j D_k D_l \{\phi_c(\mathbf{x})\}/n^{1/2}]\phi_c(\mathbf{x})\, d\mathbf{x} \Big| = O(n^{-1/2}),$$

where the probability $P_{\mathbf{t}_0}$ is computed with the density $h_{\mathbf{t}_0}$ and ϕ_c is the multivariate normal density with mean $\mathbf{0}$ and covariance $C = A^{-1} \sum_{\mathbf{t}_0} (A^{-1})^T$, $A = \{ED_r\psi_l(\mathbf{x}, \mathbf{t}_0)\}_{1\le r,l \le p}$, $D_r\psi_l$ refers to differentiation with respect to t_r, $\sum_{\mathbf{t}_0} = \{E\psi_r(\mathbf{X}, \mathbf{t}_0)\psi_l(\mathbf{X}, \mathbf{t}_0)\}_{1\le r,l\le p}$ and all expectations are with respect to $h_{\mathbf{t}_0}$.

This result holds uniformly over every class $\mathscr{B}$ of Borel sets of R^p satisfying

$$\sup_{t_0 \in K} \sup_{B \in \mathscr{B}} \int_{(\partial B)^\varepsilon} \phi_c(x)\, dx = O(\varepsilon) \quad \text{as} \quad \varepsilon \downarrow 0$$

where $(\partial B)^\varepsilon$ is the ε-neighborhood of B.

By applying Theorem 3 with $s = 4$, it follows that the order of the error in the expression above is in fact $O(n^{-1})$. Since the above result holds uniformly over the class $\mathscr{B}$ above, it can be shown that the density of $n^{1/2}(\mathbf{T}_n - \mathbf{t}_0)$ under $h_{\mathbf{t}_0}$ at $\mathbf{x}$ is

$$g_n(\mathbf{x}) = \phi_c(\mathbf{x})[1 + \textstyle\sum_{j=1}^p d_j D_j\{\phi_c(\mathbf{x})\}/n^{1/2} + \sum_{j,k,l} d_{jkl} D_j D_k D_l\{\phi_c(x)\}/n^{1/2} + O(1/n)].$$

The density of $(\mathbf{T}_n - \mathbf{t}_0)$ under $h_{\mathbf{t}_0}$ at $\mathbf{0}$ is $q_n(\mathbf{t}_0) = n^{p/2} g_n(\mathbf{0})$. Using the results of Lemma 1, it follows that $p_n(\mathbf{t}_0) = \{C(\mathbf{t}_0)\}^n n^{p/2} g_n(\mathbf{0})$. Putting the results together, we have the following.

THEOREM 1. *If* $\mathbf{T}_n$ *represents the solution of* $\sum_{i=1}^n \psi_r(x_i, \mathbf{t}) = 0$, $r = 1, \cdots, p$, *and Assumptions* 1–3 *are satisfied, then an asymptotic expansion for the density of* $\mathbf{T}_n$, *say* p_n, *is*

$$(6)\ p_n(\mathbf{t}_0) = (n/2\pi)^{p/2} \left[\int \exp\{\textstyle\sum_{j=1}^p \alpha_j(\mathbf{t}_0)\psi_j(x, \mathbf{t}_0)\} f(x)\, dx \right]^n \left\{ |\det A|\, |\det \textstyle\sum|^{-1/2} + O\left(\dfrac{1}{n}\right)\right\},$$

where $\boldsymbol{\alpha}(\mathbf{t}_0)$ is the solution of

$$\int \psi_r(x, \mathbf{t}_0) \exp\{\textstyle\sum_{j=1}^p \alpha_j \psi_j(x, \mathbf{t}_0)\} f(x)\, dx = 0 \quad \text{for} \quad r = 1, \cdots, p,$$

$$A = \{E\partial\psi_l(x, \mathbf{t})/\partial t_r|_{\mathbf{t}=\mathbf{t}_0}\}_{1\le r,l\le p}, \qquad \textstyle\sum = \{E\psi_l(x, \mathbf{t}_0)\psi_r(x, \mathbf{t}_0)\}_{1\le r,l\le p},$$

and all expectations are with respect to the conjugate density

$$h_{\mathbf{t}_0}(x) = \exp\{\textstyle\sum_{j=1}^p \alpha_j(\mathbf{t}_0)\psi_j(x, \mathbf{t}_0)\} f(x) \left[\int \exp\{\sum_{j=1}^p \alpha_j(\mathbf{t}_0)\psi_j(y, \mathbf{t}_0)\} f(y)\, dy \right]^{-1}.$$

The error term holds uniformly for all $\mathbf{t}$ in a compact set.

NOTE. Following the practice advocated by Hampel (1974), the integrating constant $(n/2\pi)^{p/2}$ will not be used in numerical work but the constant will be determined numerically. Previous numerical experience suggests that this will substantially improve the accuracy of the approximation.

In applications of formula (6), it is usually probabilities obtained by the integration of $p_n(t)$ which are of interest. Since the approximation in (6) is uniform for all $\mathbf{t}$ in a compact set, we can state that for K compact

$$\int_K p_n(\mathbf{t})\,d\mathbf{t} = \left(\frac{n}{2\pi}\right)^{p/2} \int_K \left[\int \exp\{\textstyle\sum_{j=1}^p \alpha_j(\mathbf{t})\psi_j(\mathbf{t})\} f(x)\,dx\right]^n |\det A_{\mathbf{t}}|\,|\det \textstyle\sum_{\mathbf{t}}|^{-1/2}\,d\mathbf{t}$$

$$+\,O\left(\frac{1}{n}\right)\left(\frac{n}{2\pi}\right)^{p/2} \int_K C^{-n}(\mathbf{t})\,d\mathbf{t}.$$

To be more specific here, it is necessary to examine the behavior of $C^{-1}(\mathbf{t})$ over a compact set. In the cases which we have examined, $C^{-1}(\mathbf{t})$ is a continuous function so the last term above can be bounded on any compact set.

A more useful result here would be to have the above result hold for any arbitrary K. This would ensure that probability integrals of marginal densities would still be correct to order $O(1/n)$.

In the case of the univariate mean, Daniels (1954, see Section 7) shows that under certain conditions on the cumulants of the conjugate-density the saddlepoint approximation (formula (6)) holds uniformly over the whole parameter space in the sense that the relative error of (6) is uniformly bounded. Barndorff-Nielsen and Cox (1979, see Appendix) have a uniformity result for the multivariate mean provided that $\alpha(\mathbf{t})$ remains in a compact subset of the parameter space. Although at present no similar results are available for general $\psi(x, t)$, a careful generalization of the approach given by Daniels might yield interesting results. It should be noted that both the above approaches appear to use critically the special form of ψ for the case of the mean to parametrize the conjugate density by a parameter which remains in a compact set in many cases of interest. The analogue for the general ψ-function is not clear.

As has been pointed out by a referee, the key idea in the above development can be expressed as follows. We have a statistic $T(F_n)$ where F_n is the empirical distribution function whose density has an Edgeworth expansion and which has an expansion of the form

$$\mathbf{T}(F_n) = \mathbf{T}(F) + n^{-1} \textstyle\sum_{i=1}^n A^{-1}\psi(x_i, \boldsymbol{\theta}) + O_p(n^{-1/2}).$$

The conjugate density is determined by the linear part of the statistic and is followed by making an Edgeworth expansion. If the statistic is linear, we have the classical saddlepoint method. In this form, it seems likely that the method may have applicability to quite general non-linear statistics.

3. Special cases and numerical results. In what follows, the formula (6) will be applied to a number of specific settings and some numerical computations carried out where necessary. It should be noted that the motivation for developing (6) was to find a good approximation to the density of M-estimates. The comparisons involving standard results are presented here to give an idea of how the approximation works in classical situations.

(a) *One-dimensional location*: $p = 1$ and $\psi(x, t) = \psi(x - t)$. The score function, ψ, must be monotone to ensure uniqueness of the solution. This case has been studied extensively in Field and Hampel (1982) and numerical results are given showing the approximating density to be very accurate in the extreme tail down to $n = 3$. The derivation in Field and Hampel is based on an expansion of p_n'/p_n and uses the fact that $\Pr(T_n \le t) = \Pr\{\sum_{i=1}^n (X_i - t) \le 0\}$. Daniels, in an unpublished note, has derived the same formula using saddlepoint techniques.

(b) *Multivariate mean*: If $\mathbf{x} = (x_1, \cdots, x_p)$ and $\psi_r(\mathbf{x}, \mathbf{t}) = (x_r - t_r)$, then $\mathbf{T}_n = \bar{\mathbf{X}}_n$, the p-variate mean. We assume that $X_1, X_2, \cdots, X_n$ are independent and identically distributed. In Field (1978), this case is considered and an approximating formula is derived by means of an expansion of $p_n^{-1}(\partial p_n/\partial t_i)$. Except for the constant of integration, the formula is equivalent to (6) above for this case. The approximating formula becomes

$$(7)\quad p_n(\mathbf{t}_0) = \left(\frac{n}{2\pi}\right)^{p/2} |\det \textstyle\sum_{\mathbf{t}_0}|^{-1/2}\left[\displaystyle\int \cdots \int \exp\{\textstyle\sum_{i=1}^{p} \alpha_j(\mathbf{t}_0)(x_j - t_{0j})\} f(\mathbf{x})\, dx_1, \cdots, dx_p\right]^n.$$

If the components are independent, this formula is the product of the one-dimensional approximation for the mean given in Daniels (1954, cf. (2.6)).

For the case of observations from a multivariate normal,

$$f(\mathbf{x}) = (1/2\pi)^{p/2} (\det \textstyle\sum)^{-1/2} \exp\left(-\dfrac{1}{2}\mathbf{x}^T \textstyle\sum^{-1} \mathbf{x}\right),$$

the sample mean has density

$$p_n(\mathbf{t}_0) = (n/2\pi)^{p/2} (\det \textstyle\sum)^{-1/2} \exp\left(-\dfrac{n}{2}\mathbf{t}_0^T \textstyle\sum^{-1} \mathbf{t}_0\right).$$

The solution of the equations

$$\int \cdots \int (x_j - t_j) \exp\{\textstyle\sum_{j=1}^{p} \alpha_j (x_j - t_j)\} \exp\left(-\dfrac{1}{2}\mathbf{x}^T \textstyle\sum^{-1} x\right) d\mathbf{x} = 0$$

for $j = 1, \cdots, p$ is $\boldsymbol{\alpha}(\mathbf{t}_0)^T = \mathbf{t}_0^T \sum^{-1}$ and

$$(1/2\pi)^{p/2} (\det \textstyle\sum)^{-1/2} \displaystyle\int \cdots \int \exp\left\{\mathbf{t}_0^T \textstyle\sum^{-1} (\mathbf{x} - \mathbf{t}_0) - \dfrac{1}{2} x^T \textstyle\sum^{-1} \mathbf{x}\right\} dx_1 \cdots dx_p$$

$$= \exp\left(-\frac{1}{2}\mathbf{t}_0^T \textstyle\sum^{-1} \mathbf{t}_0\right).$$

This implies that the conjugate density

$$h_{\mathbf{t}_0}(\mathbf{x}) = (1/2\pi)^{p/2} (\det \textstyle\sum)^{-1/2} \exp\left\{\dfrac{1}{2}\mathbf{t}_0^T \textstyle\sum^{-1} \mathbf{t}_0\right\} \exp\left\{\mathbf{t}_0^T \textstyle\sum^{-1} (\mathbf{x} - \mathbf{t}_0) - \dfrac{1}{2}\mathbf{x}^T \textstyle\sum^{-1} \mathbf{x}\right\}$$

$$= (1/2\pi)^{p/2} (\det \textstyle\sum)^{-1/2} \exp\left\{-\dfrac{1}{2}(x - \mathbf{t}_0)^T \textstyle\sum^{-1} (\mathbf{x} - \mathbf{t}_0)\right\}.$$

From this it follows that $\sum_{\mathbf{t}_0} = \sum$ and the approximating formula (7) becomes

$$p_n(\mathbf{t}_0) = (n/2\pi)^{p/2} (\det \textstyle\sum)^{-1/2} \exp\left(-\dfrac{n}{2}\mathbf{t}_0^T \textstyle\sum^{-1} \mathbf{t}_0\right)$$

which is exact.

If the observations X_i are a $p \times p$ matrix with a Wishart density, $W(\sum, p, m)$ it is known that $\sum_{i=1}^{n} X_i$ is $W(\sum, p, mn)$ so that $\bar{X}_n$ has an exact density

$$p_n(t) = ncn^{(nm-p-1)/2} (\det t)^{(mn-p-1)/2} \exp\left\{-\frac{n}{2} \operatorname{tr} \textstyle\sum^{-1} t\right\} (\det \textstyle\sum)^{-mn/2}$$

where $c = [2^{mnp/2} \pi^{p(p-1)/4} \prod_{j=1}^{p} \Gamma((mn + 1 - j)/2)]^{-1}$ for $t > 0$. The solution of the equations

$$\int \cdots \int (x_{ij} - t_{ij}) \exp\{\textstyle\sum_{k,l} \alpha_{kl}(x_{kl} - t_{kl})\} (\det x)^{(m-p-1)/2} \exp\left(-\dfrac{1}{2} \operatorname{tr} \textstyle\sum^{-1} x\right) dx = 0$$

is given by $\alpha(t) = \sum^{-1}/2 - mt^{-1}/2$. The conjugate density can be shown to be Wishart, $W(t, p, m)$ and the integral in expression (7) is equal to

$$(\det \textstyle\sum)^{-m/2} \exp\left\{-\dfrac{1}{2} \operatorname{tr} \textstyle\sum^{-1} t\right\} (\det t)^{m/2}$$

up to a constant. It remains to compute the determinant of the covariance matrix of the conjugate density. Using results in Anderson (1958, page 162), it is possible to show that this deteminant is $(\det \sum_t)^{p+1}$. Upon substitution this yields the correct density except for the normalizing constant. For the case of the univariate mean, Daniels (1980) has shown that the normal, gamma and inverse normal are the only possible densities in which the renormalized saddlepoint approximation (i.e. (7) with the constant determined numerically) is exact. For the multivariate case, it would be interesting to determine those densities, if any, beyond the normal and Wishart for which the renormalized saddlepoint is exact. We now show some computations using formula (7) for a case in which the approximating formula is not exact. In particular, let $\mathbf{Z}_i = (Z_{1i}, Z_{2i})$, $i = 1, \cdots n$ where $\mathbf{Z} = B\mathbf{X}$ for (X_1, X_2) independent uniform random variables on $[-1, 1]$. Since the density of the mean of uniform observations is known, it is straight-forward to verify that the joint density of $(\bar{Z}_1, \bar{Z}_2)$ is

$$p_n(z_1, z_2) = \{n^n/2^n(n-1)!\}^2 \prod_{i=1}^{2} \{\textstyle\sum_{s=0}^{n} (-1)^s \binom{n}{s} < 1 - (B^{-1}Z)_i - 2s/n >^{n-1}\}(\det B)^{-1},$$

whenever $|B^{-1}\mathbf{z}| \leq (1, 1)^T$ and 0 otherwise. Note that $< x > = x$ if $x > 0$ and 0 otherwise. Formula (7) was used to determine the approximate density for sample sizes $n = 2$ to 8. In order to measure the accuracy of the approximation, cumulative distributions were evaluated by numerical integration for both the exact and the approximate, after adjustment for the constant of integration. For the particular case with

$$B = \begin{bmatrix} .25 & .75 \\ .67 & .33 \end{bmatrix},$$

the results of this computation are shown in Table 1 for selected values of (t_1, t_2). The value of the exact cumulative is given in the first column, the difference between the exact cumulative and the approximate cumulative is in the second column and the percent relative error is the third column. This last quantity is computed as

$$\frac{100(\text{exact cumulative} - \text{approximate cumulative})}{\min(\text{exact cumulative}, 1 - \text{exact cumulative})}.$$

As can be seen by an examination of the percent relative errors in the table, the approximation is very accurate in the extreme tail even down to $n = 2$. For instance, with $n = 2$, the percent relative errors are all less than 10% and at the point (.95, .95), the approximate cumulative differs from the exact cumulative of .99996 by 4 in the sixth decimal. For many practical purposes, this is already accurate enough. As n increases, the accuracy increases giving excellent agreement between the exact and the approximation at $n = 8$. This accuracy is consistent with the results found previously in the univariate case (cf. Field and Hampel, 1982).

(c) *Location and scale*: We assume $\boldsymbol{\theta} = (\mu, \sigma)$, $f_\theta(x) = f((x - \mu)/\sigma)$ and $\psi_i(x, \boldsymbol{\theta}) = \psi_i((x - \mu)/\sigma)$, $i = 1, 2$. In particular set $\psi_1(x) = \min\{k, \max(-k, x)\}$, $\psi_2(x) = \psi_1^2(x) - \beta$ with $\beta = E_\Phi(\psi_1^2(x))$. This corresponds to "Proposal 2" of Huber (1964) and gives translation and scale invariant estimates.

For $k = +\infty$, we have classical least squares with $\mathbf{T}_n = (\bar{X}, s)$ where $s^2 = \sum_{i=1}^{n} (x_i - \bar{x})^2/n$. As a first example assume $X_1, \cdots, X_n$ are independent observations from a $\mathcal{N}(0, 1)$ population. The equations

$$\int \psi_i(x, \mathbf{t})\exp[\alpha_1(x - t_1)/t_2 + \alpha_2\{(x - t_1)^2/t_2^2 - 1\} - x^2/2]\, dx = 0, \qquad i = 1, 2,$$

have solutions $\alpha_1(\mathbf{t}) = t_1 t_2$, $\alpha_2(t) = (t_2^2 - 1)/2$. In addition

$$\int \exp\{\textstyle\sum_{j=1}^{2} \alpha_j(t)\psi_j(x, t)\} f(x)\, dx = t_2 \exp\{-t_2^2/2 - t_1^2/2 + 1/2\}$$

TABLE 1
Exact cumulative and error in approximation for density of bivariate mean using formula (7) for $n = 2$ to 8

(t_1, t_2)	$n = 2$ Exact Cumulative	$n = 2$ Cumulative Error	$n = 2$ % Relative Error	$n = 4$ Exact Cumulative	$n = 4$ Cumulative Error	$n = 4$ % Relative Error
(0, 0)	.37178	−.37E-2	−1.00	.37261	−.22E-3	−.06
(.25, 0)	.47362	.23E-3	.05	.49189	.54E-4	.01
(.25, .50)	.46860	−.17E-4	−.00	.48941	−.30E-5	−.00
(.50, 0)	.67854	−.22E-3	−.07	.80041	−.36E-3	−.18
(.50, .50)	.90707	.21E-2	2.27	.98224	−.19E-3	−1.06
(.50, .75)	.93217	.15E-2	2.23	.98762	−.19E-3	−1.52
(.75, .50)	.94609	.12E-2	2.18	.99141	−.58E-4	−.68
(.75, .75)	.99046	.58E-3	6.06	.99988	.19E-5	1.55
(.75, .80)	.99157	.52E-3	6.22	.99990	.19E-5	1.65
(.80, .75)	.99266	.51E-3	6.99	.99994	.11E-5	1.97
(.80, .80)	.99407	.43E-3	7.31	.99998	.65E-6	2.57
(.80, .90)	.99530	.35E-3	7.38	.99998	.50E-6	2.73
(.90, .80)	.99626	.33E-3	8.68	.99999	.49E-6	3.19
(.90, .90)	.99947	.50E-4	9.51	1.00000	.62E-8	4.71
(.90, .95)	.99961	.37E-4	9.50	1.00000	.40E-8	5.18
(.95, .90)	.99975	.24E-4	9.60	1.00000	.30E-8	4.07
(.95, .95)	.99996	.40E-5	9.78	1.00000	−.63E-10	4.07

(t_1, t_2)	$n = 6$ Exact Cumulative	$n = 6$ Cumulative Error	$n = 6$ % Relative Error	$n = 8$ Exact Cumulative	$n = 8$ Cumulative Error	$n = 8$ % Relative Error
(0, 0)	.37289	−.57E-4	.02	.37268	−.19E-4	−.01
(.25, 0)	.49653	.38E-4	.01	.49793	.19E-4	.00
(.25, .50)	.49526	.29E-4	.01	.49729	.15E-4	.00
(.50, 0)	.86853	−.10E-3	.08	.91150	.70E-4	−.08
(.50, .50)	.99636	−.81E-4	−2.22	.99928	−.27E-4	−3.74
(.50, .75)	.99743	−.75E-4	−2.92	.99948	−.26E-4	−4.97
(.75, .50)	.99846	−.12E-4	−.76	.99972	−.18E-5	−.64
(.75, .75)	.99999	−.17E-7	.85	1.00000	.25E-9	.68
(.75, .80)	.99999	.13E-7	.79	1.00000	.16E-9	.49
(.80, .75)	1.00000	.54E-8	2.00	1.00000	.40E-10	−.44
(.80, .80)	1.00000	−.46E-9	.19	1.00000	−.53E-10	.40
(.80, .90)	1.00000	−.57E-9	.25	1.00000	−.48E-10	.41
(.90, .80)	1.00000	.99E-9	5.31	1.00000	−.15E-11	.11
(.90, .90)	1.00000	−.38E-11	1.67	1.00000	*	
(.90, .95)	1.00000	−.33E-11	1.58	1.00000	*	
(.95, .90)	1.00000	−.37E-12	2.60	1.00000	*	
(.95, .95)	1.00000	*		1.00000	*	

* This indicates that the error in the cumulative is less than 10^{-13} which is approaching the limits of accuracy available in this computation.

so that the conjugate density $h_t(x)$ is normal with mean t_1 and variance t_2^2. The matrix

$$\begin{bmatrix} E(-1/t_2) & -2E(X-t_1)/t_2^2 \\ -E(X-t_1)/t_2^2 & -2E(X-t_1)^2/t_2^3 \end{bmatrix} = \begin{bmatrix} -1/t_2 & 0 \\ 0 & -2/t_2 \end{bmatrix},$$

so that $\det A = 2/t_2^2$. Similarly $\det\Sigma = 2$. Hence the approximating formula (6) becomes

$$p_n(t_1, t_2) = (n/2\pi)t_2^{n-2}\exp(-nt_2^2/2 - nt_1^2/2 + n^{1/2})2^{1/2}.$$

TABLE 2

Approximation to joint density of robust estimates of location and scale using Huber's Proposal 2 with $k = 1.5$

t_1	t_2	$n = 5$				$n = 10$			
		normal	t_3	**slash**	**Cauchy**	**normal**	t_3	**slash**	**Cauchy**
0.00	.05	.012856	.009371	.007691	.005387	.000003	.000002	.000001	.000001
0.05	.05	.007407	.004776	.004672	.002197	.000001	0	0	0
1.00	.05	.001431	.000839	.001082	.000335	0	0	0	0
1.50	.05	.000096	.000087	.000107	.000048	0	0	0	0
2.00	.05	0	0	.000006	.000009	0	0	0	0
3.00	.05	0	0	0	.000001	0	0	0	0
4.00	.05	0	0	0	0	0	0	0	0
0.00	.50	.788738	.536291	.483449	.273190	.504604	.258916	.208629	.078980
0.50	.50	.436232	.284199	.286543	.134394	.156203	.074583	.075231	.019391
1.00	.50	.074330	.052233	.062209	.023784	.004716	.002719	.003850	.000663
1.50	.50	.003990	.005354	.005707	.003332	.000015	.000032	.000037	.000015
2.00	.50	.000070	.000472	.000294	.000564	0	0	0	0
3.00	.50	0	.000005	0	.000033	0	0	0	0
4.00	.50	0	0	0	.000004	0	0	0	0
5.00	.50	0	0	0	.000001	0	0	0	0
0.00	1.00	.941091	.651625	.693093	.693093	1.882181	1.199300	1.263642	.425929
0.50	1.00	.510072	.416783	.427590	.231780	.558563	.488809	.502669	.199933
1.00	1.00	.081463	.114013	.107231	.077477	.014700	.036937	.035675	.021036
1.50	1.00	.003873	.016072	.013034	.014750	.000035	.000778	.000614	.000799
2.00	1.00	.000056	.001607	.000977	.002492	0	.000009	.000004	.000026
3.00	1.00	0	.000017	—	.000127	0	0	0	0
4.00	1.00	0	0	.000005	.000014	0	0	0	0
5.00	1.00	0	0	0	.000002	0	0	0	0
0.00	2.00	.017740	.092733	.076215	.110021	001786	098294	.081975	.218988
0.50	2.00	.009514	.080232	.967794	.102566	.000516	.071258	.064659	.180062
1.00	2.00	.001468	.049090	.043112	.078223	.000012	.024586	.025895	.090319
1.50	2.00	.000065	.018907	.016886	.041991	0	.003333	.003854	.021953
2.00	2.00	.000001	.004405	.003957	.014694	0	.000172	.000198	.002430
3.00	2.00	0	.000080	.000074	.000900	0	0	.000004	.000010
4.00	2.00	0	.000002	.000001	.000069	0	0	0	0
5.00	2.00	0	0	0	.000010	0	0	0	0
0.00	3.00	.000003	.008352	.007853	.033976	0	.002132	.002132	.055246
0.50	3.00	.000002	.008156	.008281	.033681	0	.001962	.003645	.052255
1.00	3.00	0	.007339	.009137	.032197	0	.001446	.003787	.042583
1.50	3.00	0	.005824	.008654	.027821	0	.000701	.002738	.026525
2.00	3.00	0	.002925	.005405	.019309	0	.000172	.000869	.010439
3.00	3.00	0	.000241	.000489	.003687	0	.000001	.000005	.000303
4.00	3.00	0	.000007	.000016	.000359	0	0	0	.000003
5.00	3.00	0	0	.000001	.000037	0	0	0	0
0.00	4.00	0	.000985	.001927	.012603	0	.000063	.000448	.015383
0.50	4.00	0	.001001	.001927	.012679	0	.000063	.000475	.015168
1.00	4.00	0	.111034	.002205	.012830	0	.000055	.000647	.012404
2.00	4.00	0	.000913	.002960	.011952	0	.000032	.000604	.008956
3.00	4.00	0	.000274	.001183	.006007	0	.000002	.000058	.001541
4.00	4.00	0	.000021	.000010	.001190	0	0	0	.000052
5.00	4.00	0	.000001	.000005	.000160	0	0	0	.000001

TABLE 2—*continued*

t_1	t_2	n = 5 normal	n = 5 t_3	n = 5 slash	n = 5 Cauchy	n = 10 normal	n = 10 t_3	n = 10 slash	n = 10 Cauchy
0.00	5.00	0	.000162	.000645	.005495	0	.000003	.000101	.005064
1.00	5.00	0	.000179	.000743	.005692	0	.000003	.000122	.005040
3.00	5.00	0	.000168	.001259	.005229	0	.000002	.000125	.002213
5.00	5.00	0	.000005	.000028	.000460	0	0	0	.000012
10.00	5.00	0	0	0	0	0	0	0	0
0.00	10.00	0	0	.000031	.000346	0	0	.000001	.000109
1.00	10.00	0	0	.000032	.000355	0	0	.000001	.000113
3.00	10.00	0	0	.000048	.000431	0	0	.000001	.000137
5.00	10.00	0	0	.000114	.000597	0	0	.000003	.000151
10.00	10.00	0	0	.000001	.000018	0	0	0	0
0.00	15.00	0	0	.000006	.000065	0	0	0	.000010
1.00	15.00	0	0	.000006	.000066	0	0	0	.000012
5.00	15.00	0	0	.000010	.000094	0	0	0	.000016
10.00	15.00	0	0	.000017	.000116	0	0	0	.000007
0.00	20.00	0	0	0	.000019	0	0	0	.000002
1.00	20.00	0	0	0	.000020	0	0	0	.000002
3.00	20.00	0	0	0	.000022	0	0	0	.000002
5.00	20.00	0	0	0	.000025	0	0	0	.000003
10.00	20.00	0	0	0	.000056	0	0	0	.000006
15.00	20.00	0	0	0	.000016	0	0	0	0

This agrees with exact formula except for the constant terms which are in the ratio $n^{n/2-1}\pi^{1/2}2^{3/2-n/2}e^{-n/2}/\Gamma((n-1)/2)$. For $n = 9$, this ratio equals .897 so that the error from the constant term is relatively large, reemphasizing the need for a numerical determination of the constant.

For $k < \infty$, we have robust M-estimates of location and scale. Our choice of β above would be appropriate for models in which we expect the distribution to be in some neighborhood of the normal. The level of k represents the extent to which we wish to minimize the influence of outliers and usual values of k would be in the range from 1 to 2.

Numerical computations have been carried out using formula (6) to determine the approximate joint density of (T_1, T_2) where T_1 is the location estimate and T_2 the scale estimate. In what follows, samples of the joint density are reported, the marginal density of T_1 is computed and compared with Monte Carlo results from Andrews *et al* (1971), and finally the percentiles of a "studentized" ratio are computed.

In particular, computations have been carried out with $k = 1.5$ for the following underlying densities: normal, t_3, slash (ratio of normal and uniform on [0, 1]) and Cauchy. The first step in the computation is to solve the pair of non-linear equations:

$$\int \psi_i(x, \mathbf{t})\exp\{\alpha_1\psi_1(x, \mathbf{t}) + \alpha_2\psi_2(x, \mathbf{t})\}f(x)\,dx = 0 \quad \text{for} \quad i = 1, 2$$

for $\boldsymbol{\alpha}(\mathbf{t}) = (\alpha_1(\mathbf{t}), \alpha_2(\mathbf{t}))$ over a grid of points in the (t_1, t_2) plane. With these values, $p_n(t_1, t_2)$ can be evaluated in a straightforward and inexpensive way for any value of n. The marginal density of T_1 and the density of $R = T_1/T_2$ are determined by numerical integrations. Table 2 gives a sample of the values of $p_n(t_1, t_2)$ for $n = 5$ and 10 and for the four densities.

As can be seen from the table, the values of $p_n(t_1, t_2)$ vary considerably over the four densities. Figure 1 shows the regions where $p_n(t_1, t_2) > .5 \times 10^{-6}$ for each density.

By looking at Table 2, or simply at Figure 1, it can be seen that there is a lack of independence between T_1 and T_2 and that this dependence becomes more pronounced for longer-tailed densities. To illustrate this point, the following table gives the difference between the joint density and the product of the marginal densities of T_1 and T_2 relative to the value of the joint density. Let $q_n(t_1)$ and $r_n(t_2)$ be the marginal densities of T_1 and T_2 respectively.

Using the values in Table 3 as a rough measure of the degree of dependence between T_1 and T_2, we can see from the table that even for $n = 20$, there is little evidence of independence and in fact, not a substantial change from the situation with $n = 5$. Examinations of Table 2 and Figure 1 show that the main mass in $p_n(t_1, t_2)$ occurs along a strip moving to the right and upwards. The positive values in Table 3 correspond to the regions with relatively large values of $p_n(t_1, t_2)$ while the negative values occur in regions with relatively low $p_n(t_1, t_2)$. The results here show that although T_1 and T_2 are asymptotically independent, the rate at which this independence is approached may be very slow even with an underlying normal density.

The only results with which the approximation can be checked in this example are some simulation results for the marginal density of T_1 obtained during the Princeton Robustness Study and reported, in part, by Andrews *et al.* (1971). The complete results have been provided most kindly by F. Hampel. In Table 4, the values of the pseudovariances and n times the variance are reported both for the asymptotic approximation and the Monte Carlo results. The pseudovariance is defined as $n(t_{1,1-\alpha}/z_{1-\alpha})^2$ where $t_{1,1-\alpha}$ and $z_{1-\alpha}$ represent the $(1 - \alpha)$ quantile of T_1 and a standard normal variate respectively.

In order to make an assessment of whether the differences between the asymptotic and Monte Carlo results are to within the sampling errors in the Monte Carlo experiment, we can look at two bits of evidence. The first is provided by Exhibit 5-13 of Andrews *et al.* (1971) in which Monte Carlo and exact results for the percentage points are compared for the median with $n = 5$. Table 5 gives these differences in terms of the pseudo-variances and lists the differences from Table 4 above.

As can be seen from the table, for the normal case the differences between the asymptotic and Monte Carlo results from Table 4 are well within the errors inherent in

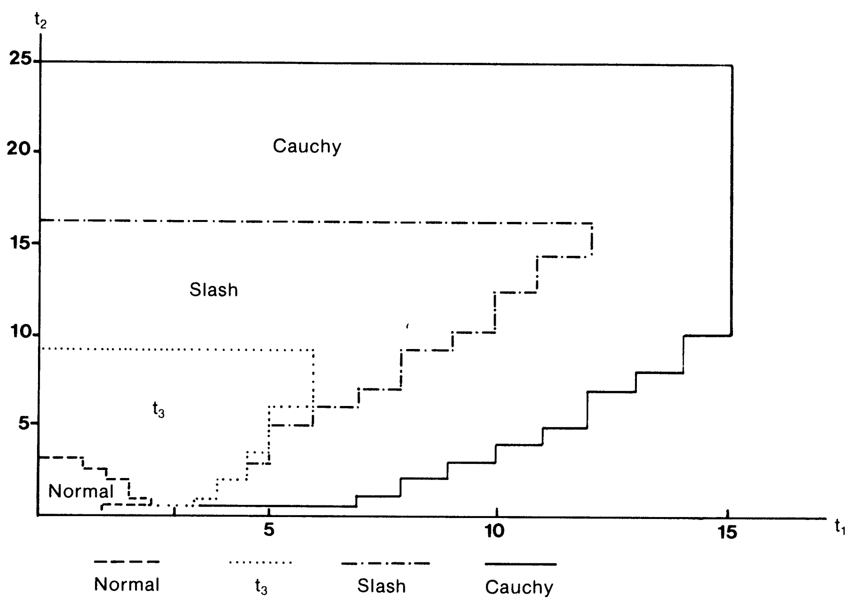

FIG. 1. *Region where joint density of robust location and scale estimates (Huber's Proposal 2, $k = 1.5$) exceeds $.5 \times 10^{-6}$ for $n = 5$.*

TABLE 3
Relative difference of joint and marginal densities $\{p_n(t_1, t_2) - q_n(t_1)r_n(t_2)\}/p_n(t_1, t_2)$ for Huber's Proposal 2, $k = 1.5$

t_1	t_2	$n = 5$				$n = 10$			
		normal	t_3	slash	Cauchy	normal	t_3	slash	Cauchy
0	.05	.538	.527	.627	.814	—	—	—	—
1	.05	.629	.729	.446	−.187	—	—	—	—
2	.05	—	—	−8.65	−12.21	—	—	—	—
3	1	.562	.425	.619	.716	.514	.555	.555	.708
1	1	.547	.832	.484	.521	.497	−1.8	−1.52	−15.00
2	1	.462	.936	−4.32	−3.35	—	−2.0 × E3	−5.6 × E5	−4.7 × E5
0	2	.565	.078	.339	.486	.520	.282	.174	.471
1	2	.531	.911	.756	.713	.359	.942	.955	.543
2	2	.328	.995	.754	.554	—	.912	−1.97	−44.25
0	5	—	−2.60	−2.421	−.549	—	−3.03	−4.97	−.705
3	5	—	1.00	.990	.926	—	1.00	1.00	.987
5	5	—	1.00	.898	.769	—	—	−4.69	−27.01
0	8	—	−8.86	−6.15	−1.95	—	—	−16.03	−2.59
3	8	—	1.00	.790	.897	—	—	1.00	.998
5	8	—	1.00	.997	.972	—	—	1.00	.999
0	15	—	—	−17.26	−6.17	—	—	−74.76	−10.45
4	15	—	—	.970	.875	—	—	1.00	.999
8	15	—	—	.992	.990	—	—	1.00	1.00

the Monte Carlo results. For the Cauchy, the differences observed in Table 4 for Huber's estimate with $k = 1.5$ are larger than the errors in the Monte Carlo for the median. To shed light on whether we can place any faith in the asymptotic results for the Cauchy, it is worth looking at the Monte Carlo results for $n = 10$. For this situation, there were two simulations carried out in the Princeton study and the replication gives an indication of the Monte Carlo errors. From Table 4, with the Cauchy and $n = 10$, the asymptotic results lie between the two Monte Carlo replications except for 0.1% and $n \times$ variance. This gives a strong indication of the reliability of the asymptotic results for $n = 10$. Until the exact marginal densities are computed in some fashion, or until additional Monte Carlo studies are done, further comparisons are difficult. To see that the large discrepancies for the Cauchy at $n = 5$ may be due to Monte Carlo variation, it is instructive to look at Figure 2.

From the graph, we note that the Monte Carlo results for $n = 5$ do not follow the pattern exhibited by the other values of n. In particular, it appears that the extreme percentiles for the Monte Carlo with $n = 5$ are not large enough. This would lead to the large differences observed in Table 5. While this is not a proof that the asymptotic results are accurate, it suggests that the precision of the asymptotic results may be very good even in the extreme case of the Cauchy with $n = 5$.

As a further step in this example, we consider the percentiles of a "studentized" version of T_1. Since the asymptotic variance of T_1 is

$$\sigma^2 E_f \psi^2(X) / \{E_f \psi'(X)\}^2,$$

an appropriate "studentized" version of T_1 would be $n^{1/2}T_1/(T_2\gamma)$ with $\gamma = E_\Phi \psi_1^2(X)/\{E_\Phi \psi_1'(X)\}^2$. This assumes that the estimate has been chosen as though the underlying density is normal. This was implicit in the definition of ψ_2 at the beginning of the example. In practice, it would be more desirable to replace γ by its estimated form where Φ is replaced by the empirical distribution. However the problem of working out the percentiles of this more complicated expression introduces some computational difficulties which are currently being worked on.

TABLE 4
Pseudovariances and asymptotic variances of T_1 as computed by approximation (A) *and Monte Carlo* (MC)

		$n = 5$		$n = 10$		$n = 20$		$n = 40$	
		A	MC	A	MC	A	MC	A	MC
Normal									
Pseudo-	25%	1.0345	1.0412	1.0357	1.0312	1.0368	1.036	1.0380	1.0392
variances	10%	1.0352	1.0411	1.0360	1.0303	1.0366	1.0356	1.0373	1.0380
	2.5%	1.0367	1.0432	1.0366	1.0306	1.0368	1.0357	1.0372	1.0380
	1%	1.0449	1.0462	1.0371	1.0310	1.0370	1.0359	1.0372	1.0382
	.5%	1.0385	1.0462	1.0374	1.0313	1.0372	1.0361	1.0372	1.0383
	.1%	1.0398	1.0496	1.0384	1.0321	1.0376	1.0366	1.0373	1.0386
	$n \times$ var	1.0360	1.0427	1.0364	1.0308	1.0369	1.0360	1.0375	1.0384
Slash									
Pseudo-	25%	1.7610	1.8597	1.6856	1.6599	1.6457	1.6559	1.6284	1.5953
variances	10%	1.9548	2.0715	1.7624	1.7265	1.6758	1.6902	1.6392	1.6076
	2.5%	2.8046	2.8036	1.9878	1.8941	1.7529	1.7679	1.6711	1.6341
	1%	4.5851	4.2350	2.2613	2.0682	1.8283	1.8234	1.6962	1.6535
	5%	7.3134	6.7839	2.5974	2.2829	1.8880	1.8876	1.7166	1.6686
	.1%	18.4117	19.7112	4.3116	4.5852	2.1149	2.0288	1.7755	1.7040
	$n \times$ var	3.549	3.8752	2.0776	3.5681	1.7419	1.7986	1.6629	1.6246
Cauchy									
Pseudo-	25%	4.607	3.75	4.590	4.5731	4.907	4.648	4.4852	4.0
variances					4.74*				
	10%	7.256	5.4060	5.834	5.8120	5.094	4.8625	4.7554	4.2781
					6.1554*				
	2.5%	17.405	11.590	9.392	9.2350	6.429	5.8338	5.3361	4.6673
					9.3463*				
	1%	30.747	19.2729	13.434	14.6001	7.634	6.8401	5.8013	4.9629
					12.1464*				
	.5%	44.252	26.6878	17.897	21.2734	8.752	7.7168	6.1902	5.2147
					14.9575*				
	1%	72.847	43.9043	35.365	47.2187	12.45	10.1046	7.2629	6.1221
					39.2633*				
	$n \times$ var	16.525	10.9373	9.592	10.2658	6.172	5.6630	5.161	4.5469
					23.7044*				
t_3									
Pseudo-	25%	1.6275		1.6569		1.6556	1.655	1.6522	
variances	10%	1.7670		1.7226		1.6870	1.6858	1.6671	
	2.5%	2.0690		1.8442		1.7488	1.7348	1.6968	
	1%	2.3447		1.9611		1.7948	1.7652	1.7158	
	.5%	2.5572		2.0478		1.8323	1.7873	1.7360	
	1%	3.2904		2.2680		1.9245	1.837	1.7789	
	$n \times$ var	1.9953		1.8097		1.7273	1.7132	1.6870	

* *replication*

The percentiles have been worked out by numerical integration of the joint density of T_1 and T_2 over the appropriate region of the plane. The results are tabulated in Table 6. It is important to remember that the estimate and γ have been chosen as though the underlying density is normal.

The first thing to check in Table 6 is the agreement of the percentiles under the normal with the percentiles of a t-density. There is a good, but not perfect, agreement with the t-density for degrees of freedom about $0.6n$. This seems to hold over the whole range of n values from 5 to 100. This result confirms some speculation that the "studentized" ratios

TABLE 5
Pseudovariances for median; n = 5 and differences from Table 4

	Normal			Cauchy		
$1-\alpha$	Exact	Monte Carlo	Difference from Table 4	Exact	Monte Carlo	Difference from Table 4
.25	1.426	1.446	.0067	2.459	2.361	.857
.10	1.427	1.463	.0059	3.173	3.024	1.850
.025	1.438	1.470	.0065	5.285	5.287	5.815
.01	1.432	1.474	.0071	7.781	8.802	11.474
.005	1.447	1.476	.0077	10.629	13.707	17.562
.001	1.496	1.482	.0098	23.117	33.793	28.943

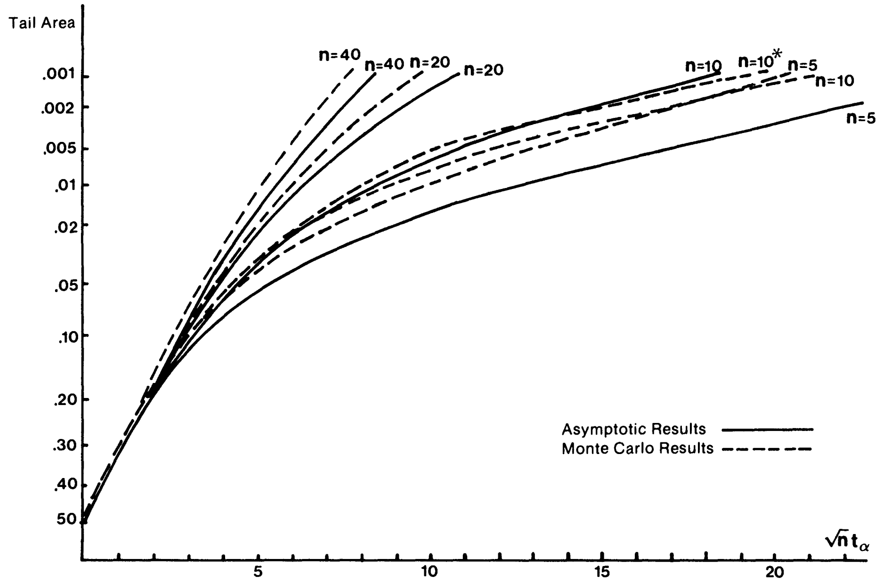

FIG. 2. *Plot of percentiles of $\sqrt{n}T_1$ on normal probability paper for Huber's Proposal 2, $k = 1.5$.*

behave like a t-density with reduced degrees of freedom, but the reduction may be larger than expected.

The important question of the stability of the percentiles as the underlying density varies can be examined using these results. As is to be expected, the largest variation occurs with small n and a Cauchy density. For $n = 5$, if we computed a 99% confidence interval, based on the normal figures, the interval would be 1.43 times longer than the correct interval for a Cauchy density while a 99.99% confidence interval would be 1.51 times longer than necessary. It should be noted that a procedure which estimates γ rather than leaving it fixed as though the underlying density is normal would give additional stability to these results. These results, as they are, are an order of magnitude improvement over results using a classical t-interval.

We now look at the marginal density of T_2^2 and determine whether its density is related to a Chi squared density with reduced degrees of freedom. Again we look at the results obtained under the normal density. Using the results from classical theory suggests that

TABLE 6
Percentiles of $n^{1/2}T_1/(\gamma T_2)$ using Huber's Proposal 2 with $k = 1.5$

	Tail area	normal	t_3	slash	Cauchy
$n = 5$	.25	.808	.831	.837	.871
	.10	1.729	1.657	1.647	1.547
	.05	2.491	2.288	2.297	1.999
	.025	3.382	3.020	3.075	2.514
	.01	4.860	4.249	4.389	3.387
	.005	6.297	5.456	5.673	4.271
	.001	11.269	9.667	10.139	7.450
	.0001	25.365	21.651	23.028	16.770
$n = 10$	.25	.732	.759	.759	.807
	.10	1.461	1.467	1.468	1.482
	.05	1.965	1.923	1.925	1.863
	.025	2.460	2.358	2.369	2.197
	.01	3.143	2.937	2.975	2.626
	.005	3.689	3.393	3.464	2.959
	.001	5.124	4.597	4.759	3.828
	.0001	7.749	6.823	7.147	5.472
$n = 20$	.25	.701	.728	.724	.775
	.10	1.361	1.393	1.388	1.448
	.05	1.783	1.802	1.797	1.835
	.025	2.173	2.168	2.166	2.162
	.01	2.665	2.613	2.620	2.536
	.005	3.027	2.937	2.953	2.795
	.001	3.866	3.664	3.714	3.395
	.0001	5.118	4.726	4.845	4.139
$n = 40$	.25	.686	.713	.707	.759
	.10	1.318	1.359	1.350	1.430
	.05	1.709	1.750	1.741	1.824
	.025	2.060	2.095	2.086	2.156
	.01	2.477	2.500	2.492	2.540
	.005	2.781	2.785	2.781	2.794
	.001	3.433	3.388	3.394	3.317
	.0001	4.307	4.174	4.205	3.959
$n = 100$	.25	.679	.705	.698	.750
	.10	1.296	1.341	1.329	1.419
	.05	1.668	1.722	1.708	1.817
	.025	1.993	2.055	2.039	2.158
	.01	2.383	2.445	2.428	2.550
	.005	2.649	2.710	2.693	2.917
	.001	3.213	3.266	3.250	3.355
	.0001	3.922	3.953	3.941	4.005

df $\times$ T_2^2 should have a density similar to that of a Chi squared density with degree of freedom df. To check this, Table 7 gives the percentiles of df $\times$ T_2^2 and that of a Chi squared density with df degrees of freedom.

The agreement appears to be quite reasonable even for $n = 5$. It is interesting that the reduction in the degrees of freedom ranges from 0.73 at $n = 40$ to 0.60 at $n = 5$. This constrasts with the constant reduction of degrees of freedom in the "studentized" ratio. Why this is so remains an open question at this moment. It may have to do with the lack of independence of T_1 and T_2. Some very preliminary computations suggest that the rate at which asymptotic independence is reached may be rather slow.

TABLE 7
Percentiles of df × T_2^2 and χ^2_{df} under normal density for Huber's Proposal 2, k = 1.5

Tail area	n = 40		n = 20	
	$29 \times T_2^2$	χ^2_{29}	$14 \times T_2^2$	χ^2_{14}
.10	38.57	39.09	20.41	21.06
.05	42.00	42.56	22.93	23.68
.025	45.13	45.72	25.26	26.12
.01	48.94	49.59	28.15	29.14
.005	51.65	52.33	30.22	31.32

Tail area	n = 10			n = 5		
	$7 \times T_2^2$	χ^2_7	χ^2_6	$3 \times T_2^2$	χ^2_3	χ^2_2
.10	11.35	12.02	10.64	5.40	6.25	4.61
.05	13.29	14.07	12.59	6.70	7.81	5.99
.025	15.13	16.01	14.45	7.97	9.35	7.39
.01	17.45	18.47	16.81	9.60	11.34	9.21
.005	19.15	20.28	18.55	10.81	12.84	10.60

4. Conclusion. As has been illustrated by the extensive example, the asymptotic approximation (6) can be successfully used to study the behavior of joint densities of robust statistics. Up to this point, the numerical techniques that have been used are relatively crude and it is anticipated that by refining the numerical methods, problems involving four or five dimensions will be feasible. The application of these techniques to maximum likelihood estimation and likelihood ratio tests is an area of considerable interest.

Acknowledgments. The author would like to thank Frank Hampel for very helpful comments at various stages of the work. In addition, I would like to acknowledge the careful computer work carried out by Donna Zwiers and Bruce Spencer. The comments of several referees were very helpful in revising the paper.

REFERENCES

ANDERSON, T. W. (1958). *Introduction to Multivariate Statistical Analysis.* Wiley, New York.

ANDREWS, D. F., BICKEL, P. J., HAMPEL, F. R., HUBER, P. J., ROGERS, W. H., and TUKEY, J. W. (1972). *Robust Estimates of Location.* Princeton University Press, Princeton.

BARNDORFF-NIELSEN, O. and COX, D. R. (1979). Edgeworth and Saddlepoint approximations with statistical applications. *J. Roy. Statist. Soc. B* **41** 279-312.

BHATTACHARYA, R. N. and GHOSH, J. K. (1978). On the validity of formal Edgeworth expansion. *Ann. Statist.* **6** 434-451.

DANIELS, H. E. (1954). Saddlepoint approximations in statistics. *Ann. Math. Statist.* **25** 631-650.

FIELD, C. A. (1978). Summary of small sample asymptotics. *Proceedings of the Second Prague Symposium on Asymptotic Statistics*, Prague, 173-179.

FIELD, C. A. and HAMPEL, F. R. (1982). Small sample asymptotic distributions of M-estimators of location. *Biometrika* **69** 29-46.

HAMPEL, F. R. (1973). Some small sample asymptotics. *Proceedings of the Prague Symposium on Asymptotic Statistics*, Prague, 109-126.

HUBER, P. J. (1977). Robust statistical procedures. *Society for Industrial and Applied Mathematics.*

JAMES, G. S. and MAYNE, A. J. (1962). Cumulants of functions of random variables. *Sankhya Ser. A* **24** 47-54.

ORTEGO, J. M. and RHEINBOLDT, W. C. (1970). *Iterative Solutions of Non-linear Equations in Several Variables.* Academic, New York.

PFANZAGL, J. (1973). Asymptotic expansions related to minimum contrast estimators. *Ann. Statist.* **1** 993-1026.

DEPT. OF MATHEMATICS,
STATISTICS AND COMPUTING SCIENCE
DALHOUSIE UNIVERSITY
HALIFAX, N. S.
CANADA B3H 4H8