4. GENERAL SADDLEPOINT APPROXIMATIONS

4.1. INTRODUCTION

In this chapter, we use the approximations obtained for the mean to get approximations for more complicated statistics. The first section considers one-dimensional M-estimates i.e. T_n is the solution of $\sum_1^n \psi(x_i;t) = 0$. The estimate T_n is written locally as a mean and the saddlepoint approximation for the mean is used. In section 4.3, we consider a slightly different approach in that the moment generating function is approximated and a saddlepoint approximation used. The technique is applied to approximating the density of L-estimates in the next section. At this point, we turn to the problem for multivariate M-estimates. Techniques are similar to those used for one-dimensional M-estimates. Finally we modify the results to handle the case of regression using M-estimates. Throughout the chapter, there are numerical results illustrating the accuracy of these approximations even for small sample sizes.

In the cases considered in this chapter, our interest is to be able to say something about the density of an estimate. Although asymptotic results are available in most cases, we usually do not know whether these asymptotic distributions are good approximations for small or moderate sample sizes. For instance there are several proposals for using "t-statistics" based on robust location/scale estimates as a means for computing confidence intervals. We then need to know whether this t-approximation works reasonably and if it does, what are appropriate degrees of freedom. Some results in this direction are given in section 4.5.b.

4.2. ONE-DIMENSIONAL M-ESTIMATORS

To begin, consider the problem of finding a saddlepoint approximation for the density of a one-dimensional M-estimate. As developed by Huber (1964, 1967) M-estimates are defined as the solution T_n of

$$\sum_{i=1}^{n} \psi(x_i,t) = 0 \tag{4.1}$$

for observations $x_1, x_2, \cdots, x_n$. If the x_i's are independent observations from a density $f(x,\theta)$, then by setting $\psi(x,\theta) = \frac{\partial}{\partial\theta} \log f(x,\theta)$, T_n becomes the maximum likelihood estimate of θ. In much of the work on robustness, M-estimates play a central role. However the derivation of the exact density of such an estimate is usually intractable mathematically and it becomes essential to have a good approximation in order to carry out inference.

Denote the density of T_n when the x_i's are independent observations from a density f as $f_n(t)$. To approximate $f_n(t)$, we proceed by writing T_n as a mean up to a certain order and then using the saddlepoint approximation to the mean as derived in section 3.2. The approach follows closely that developed in Field (1982) for multivariate M-estimates which in turn uses critically results on multivariate Edgeworth expansions in Bhattacharya and Ghosh (1978). Field and Hampel (1982) give an alternate derivation for univariate M-estimates based on the log-derivative density. In this section, the approximation is developed for a one-dimensional M-estimate. The development for multivariate M-estimates is presented in section 4.5 .

In the development of the approximation, the conjugate density (cf. (3.22)) will play

a central role. For the case of an M-estimate the conjugate density for a fixed value of t is defined as

$$h_t(x,\theta) = c(t)\exp\{\alpha(t)\psi(x,t)\}f(x,\theta) \tag{4.2}$$

By an appropriate choice of $\alpha(t)$, we will use the conjugate density to center T_n at t in the sense that $E_{h_t}(T_n) = t$ up to first order. This enables us to use a low order Edgeworth expansion to approximate f_n at t. $\alpha(t)$ will be chosen so that

$$\int \psi(x,t)\exp\{\alpha(t)\psi(x,t)\}f(x,\theta)dy = 0. \tag{4.3}$$

The following assumptions on ψ and $f(x,\theta)$ will be required in the development of the approximation. D^ν denotes the ν^{th} derivative of $\psi(x,\theta)$ with respect to θ.

A4.1 The equation (4.1) has a unique solution T_n and equation (4.2) has a unique solution $\alpha(t)$.

A4.2 There is an open subset U of R such that

(i) for each $\theta \in \Theta$, $F_\theta(U) = 1$

(ii) $D\psi(x,\theta)$, $D^2\psi(x,\theta)$, $D^3\psi(x,\theta)$ exist.

A4.3 For each compact $K \subset \Theta$

(i) $$\sup_{\theta_0 \in K} E_{\theta_0}|D^2\psi(X,\theta_0)|^4 < \infty$$

(ii) there is an $\epsilon > 0$ such that

$$\sup_{\theta_0 \in K} E_{\theta_0}(\max_{|\theta-\theta_0|\leq\epsilon}|D^3\psi(X,\theta)|^3) < \infty$$

A4.4 For each $\theta_0 \in \Theta$, $E_{\theta_0}\psi(X,\theta_0) = 0$ and

$$A(\theta_0) = E_{\theta_0}[D\psi(X,\theta_0)] \neq 0.$$

A4.5 The functions $A(\theta)$ and $E_\theta\{[D^2\psi(X,\theta)]^2\}$ are continuous on Θ.

The approach is to now fix a point t_o at which to approximate f_n and construct the conjugate density for t_o as in (4.2) and (4.3). The next step is to approximate the density of T_n under the conjugate density. As we will show, it will suffice to use a normal approximation at this point. To complete the process, we need a result linking the density of T_n evaluated under the conjugate h_{t_o} with that under the density f. The following centering result provides that link. In order to simplify the proof of the theorem, we add the following assumption

A4.6 $|D\psi(x,\theta)|$ is bounded by k.

A proof of the theorem which does not require the assumption is given in Field (1982, p. 673). In the following we suppress the dependence of the density f on θ.

Theorem 4.1

If assumptions A4.1 and A4.6 hold, then

$$f_n(t_o) = c^{-n}(t_o)h_{t_o,n}(t_o) \tag{4.4}$$

where $h_{t_o,n}(t)$ is the density of T_n under the conjugate density $h_{t_o}(x.\theta)$ and

$$c^{-1}(t_0) = \int \exp\{\alpha(t_0)f(x,\theta)\}dx.$$

Proof: To illustrate the ideas, we first give a proof for the discrete case

$$f_n(t_0) = P_f[T_n = t_0] = \sum_{\{\mathbf{x}|\sum\psi(x_i,t_0)=0\}} \prod_1^n f(x_i)$$

$$= \sum_{\{\mathbf{x}|\sum\psi(x_i,t_0)=0\}} c^{-n}(t_0) \prod_1^n f(x_i) \exp\{\alpha(t_0)\psi(x_i,t_0)\} c(t_0)$$

$$= c^{-n}(t_0) P_{h_{t_0}}[T_n = t_0] = c^{-n}(t_0) h_{t_0,n}(t_0).$$

In the continuous case, let

$$A(\Delta t) = \{\mathbf{x}|\sum_{i=1}^n \psi(x_i,u) = 0,\ t_0 \le u \le t_0 + \Delta t\}.$$

Now

$$f_n(t_0) = \lim_{\Delta t \to 0} \frac{P[t_0 \le T_n \le t_0 + \Delta t]}{\Delta t}$$

$$= \lim_{\Delta t \to 0} \frac{1}{\Delta t} \int_{A(\Delta t)} \int \prod_i^n f(x_i) d\mathbf{x}.$$

Now

$$\sum_{i=1}^n \psi(x_i,u) = \sum_{i=1}^n \psi(x_i,t_0) + (t_0 - u) \sum_{i=1}^n D\psi(x_i, u(t_0))$$

where

$$t_0 \le u(t_0) \le u \le t_0 + \Delta t.$$

Hence

$$f_n(t_0) = \lim_{\Delta t \to 0} \frac{c^{-n}(t_0)}{\Delta t} \int_{A(\Delta t)} \int \prod_1^n h_{t_0}(x_i) \exp\{\alpha(t_0)(t_0 - u) \sum_{i=1}^n D\psi(x_i, u(t_0))\} d\mathbf{x}$$

$$\le c^{-n}(t_0) \lim_{\Delta t \to 0} \exp\{\alpha(t_0)(t_0 - u)k\} \lim_{\Delta t \to 0} \frac{1}{\Delta t} \int_{A(\Delta t)} \int \prod_1^n h_{t_0}(x_i) d\mathbf{x}$$

$$= c^{-n}(t_0) h_{t_0,n}(t_0).$$

Similarly $f_n(t_0) \ge c^{-n}(t_0) h_{t_0,n}(t_0)$ giving the required result. □

Using this result, the approximation to $f_n(t_0)$ can be obtained directly from an approximation to $h_{t_0,n}(t_0)$. It is important to note that for each point t_0, we use a different conjugate density. In the proof of the theorem, the property of $\alpha(t)$ specified in (4.3) has not been used, so that in fact the centering result (4.4) holds for arbitrary α.

In order to approximate $h_{t_0,n}(t_0)$, it is necessary to express T_n as an approximate mean and evaluate its cumulants. The development follows that of Field (1982) and Tingley (1987, p. 49-52). The first part is very similar to expansions used to demonstrate properties of maximum likelihood estimates. The result will be stated in terms of θ_0, the true value of θ and then will be modified for the case of the conjugate density h_{t_0} which is centered at t_0.

Theorem 4.2 (Bhattacharya and Ghosh, 1978)

Assume that A4.2–A4.5 hold. Then there is a sequence of statistics $\{T_n\}$ such that for every compact $K \subset \Theta$

$$\inf_{\theta_0 \in K} P_{\theta_0}(|T_n - \theta_0| < d_0 n^{-1/2}(\log n)^{1/2}, T_n \text{ solves (4.1)})$$

$$= 1 - 0(1/\sqrt{n})$$

where d_0 may depend on k.

Corollary

With the above assumptions $|T_n - \theta_0|^m$ is $0_p(n^{-a})$ on every compact set $K \subset \Theta$ provided that $m/2 - a > 0$.

The following construction enables us to write T_n as an approximate mean. Consider the second order Taylor expansion of (4.1) about θ_0:

$$0 = \frac{1}{n}\sum_{i=1}^{n} \psi(x_i, T_n)$$

$$= \frac{1}{n}\sum_{i=1}^{n}[\psi(x_i, \theta_0) + D\psi(x_i, \theta_0)(T_n - \theta_0) + \frac{1}{2}D^2\psi(X_i, \theta_0)(T_n - \theta_0)^2]$$

$$+ R_n(T_n) \tag{4.5}$$

where $R_n(T_n) = 0_p|T_n - \theta_0|^3 = 0_p(1/n)$.

Looking at the first three terms, let $Z = \psi(X, \theta_0)$, $Z_1 = D\psi(X, \theta_0)$ and $Z_2 = D^2\psi(X, \theta_0)$ and let $E_{\theta_0} Z_1 = \mu_1$, $E_{\theta_0} Z_2 = \mu_2$. Note that $E_{\theta_0} Z = 0$.

Define

$$\mathbf{a} = (0, \mu_1, \mu_2)$$
$$\mathbf{Z} = (Z, Z_1, Z_2)$$

and let $q : R^4 \to R$ be defined as

$$q(\mathbf{z}, t) = z + (t - \theta_0)z_1 + (t - \theta_0)^2 z_2/2.$$

Note that since $q(\mathbf{a}, \theta_0) = 0$,we can apply the implicit function theorem to prove that there is a unique three times differentiable function $H : R^3 \to R$ such that $H(\mathbf{a}) = \theta_0$ and $q(\mathbf{z}, H(\mathbf{z})) = 0$ for $\mathbf{z}$ in a neighbourhood of $\mathbf{a}$. We now have, setting $\bar{\mathbf{z}} = (\bar{z}, \bar{z}_1, \bar{z}_2)$:

Lemma

$$H(\bar{\mathbf{z}}) - T_n = 0_p(1/n).$$

Proof: cf. expression 2.39 and 2.40 of Bhattacharya and Ghosh (1978).

The next step is to expand H in a Taylor series expansion about $\mathbf{a}$. The result is the following expression

$$H(\bar{\mathbf{z}}) - \theta_0 = -\bar{z}/A(\theta_0) + A(\theta_0)\mu_2 \bar{z}^2/2 + \bar{z}(\bar{z}_1 - \mu_1)/A^2(\theta_0) + 0_p(1/n).$$

Putting this equation together with the lemma, if A4.1–A4.5 hold, then for every θ_0 in a compact subset of Θ

$$T_n - \theta_0 = \bar{Z}/A(\theta_0) + A(\theta_0)\mu_2\bar{Z}^2/2 + \bar{Z}(\bar{Z} - \mu_1)/A^2(\theta_0) + 0_p(1/n). \tag{4.6}$$

In expression (4.6), all expectations are with respect to θ_0 and are computed with the original density. If we now replace θ_0 by t_0 and compute all expectations with respect to the conjugate density h_{t_0}, we have that $E_{t_0}Z = E_{t_0}\psi(X, t_0) = 0$ by the choice α in (4.3). Hence

$$T_n - t_0 = -\bar{Z}/A(t_0) + A(t_0)\mu_2\bar{Z}^2/2 + \bar{Z}(\bar{Z}_1 - \mu_1)/A^2(t_0) + 0_p(1/n). \tag{4.7}$$

In order to apply the Edgeworth expansion, we need to evaluate the cumulants of the left hand side of (4.7). For this purpose the results of James and Mayne (1962) are appropriate. Denote the cumulant of order r of $n^{1/2}(T_n - t_0)$ by λ^r and of $\bar{Z}$ and $\bar{Z}_1$ by K^r and K_1^r. Since we are working with means it follows that K^r and K_1^r are of order n^{-r+1}. The cumulants of $n^{1/2}(T_n - t_0)$ can be expressed in terms of K as follows: (cf. James and Mayne, 1962, p. 51).

$$\begin{aligned}
\lambda^1 &= n^{1/2}\{A(t_0)\mu_2K^2/2 + K_1^1/A^2(t_0)\} + 0(n^{-3/2}) \\
&= d_1/n^{1/2} + 0(n^{-3/2}) \qquad (say) \\
\lambda^2 &= \sigma^2(t_0)/A^2(t_0) + 0(n^{-1}) \qquad \text{where} \qquad \sigma^2(t_0) = E_{t_0}[\psi^2(X, t_0)] \\
\lambda^3 &= d_3/n^{1/2} + 0(n^{-3/2}).
\end{aligned}$$

All higher order cumulants are of order $0(n^{-1})$ or higher .

The characteristic function of $n^{1/2}(T_n - t_0)$ can be written as

$$\varphi(u) = \exp\left\{\sum_{r=0}^{\infty} \lambda^r(iu)^r/r!\right\}$$

Using the above results we have

$$\varphi(u) = \exp\left\{d_1(iu)/n^{1/2} - \sigma^2(t_0)u^2/A^2(t_0) + d_3(iu)^3/3!n^{1/2} + 0(1/n)\right\}.$$

We use the result that

$$\int \exp\left\{-u^2\sigma^2(t_0)/A^2(t_0)\right\}\pi_r(iu)e^{-iux}du = \pi_r(-D)\phi_{t_0}(x)$$

where π_r is a polynomial of order r, D represents differentiation and ϕ_{t_0} is the normal density with mean 0 and variance $\sigma^2(t_0)/A^2(t_0)$. From this it follows that the density of $n^{1/2}(T_n - t_0)$ under h_{t_0} at x is given by

$$h(x) = \phi_{t_0}(x)\left[1 + dD(\phi_{t_0}(x))/n^{1/2} + d_3D^{(3)}(\phi_{t_0}(x))/n^{1/2} + 0(1/n)\right].$$

For $x = 0$, the terms of order $n^{-1/2}$ drop out giving the density of T_n at t_0 under h_{t_0} as

$$n^{1/2}h(0) = n^{1/2}\phi_{t_0}(0)[1 + 0(1/n)].$$

Using the results of Theorem 4.1, it follows that

$$f_n(t_0) = (n/2\pi)^{1/2} c^{-n}(t_0)[A(t_0)/\sigma(t_0) + 0(1/n)].$$

To summarize:

Theorem 4.3

If T_n represents the solution of $\sum_{i=1}^{n} \psi(x_i, t) = 0$ and A4.1–A4.5 hold, then an asymptotic expansion for the density of T_n is

$$f_n(t_0) = (n/2\pi)^{1/2} c^{-n}(t_0) A(t_0)/\sigma(t_0)[1 + 0(1/n)] \tag{4.8}$$

where $\alpha(t_0)$ is the solution of $\int \psi(x, t_0) \exp\{\alpha\psi(x, t_0)\} f(x)dx = 0$, $c^{-1}(t_0) = \int \exp\{\alpha\psi(x, t_0)\} f(x)dx$, $\sigma^2(t_0) = E_{t_0}\psi^2(x, t_0)$, $A(t_0) = E_{t_0}[D\psi(x, t_0)]$ where E_{t_0} is expectation with respect to the conjugate density

$$h_{t_0}(x) = c(t_0) \exp\{\alpha(t_0)\psi(x, t_0)\} f(x).$$

From a practical point of view, the error bound on the density in the theorem above is not of direct use. What is usually needed is an error on the integral of the density over a region of interest. For example in testing, we need to compute tail areas for the calculation of P-values. In order to begin, a slightly stronger version of theorem 4.3 is needed. Following the argument on p. 677 of Field (1982) it can be seen that the error bound in (4.8) is uniform for all t in a compact set. We now use an argument very similar to that in Durbin (1980a, 310–316).

From the results on the cumulants of $n^{1/2}(T_n - \theta_0)$ we have that the fourth cumulant of $T_n - \theta_0$ is $0(n^{-3})$ and the variance is $0(n^{-1})$. This implies that the fourth moment is $0(n^{-2})$. We can now find a constant C_1, so that for n sufficiently large, $E(|T_n - \theta_0|^4) < C_1 n^{-2}$. Letting A be the region $|t - \theta_0| \leq d$, then

$$\begin{aligned} C_1 n^{-2} &> \int_A (t - \theta_0)^4 f_n(t)dt + \int_{A^c} (t - \theta_0)^4 f_n(t)dt \\ &> \delta^4 \int_A f_n(t)dt = \delta^4 P(|T_n - \theta_0| > \delta_2). \end{aligned}$$

This implies $P(|T_n - \theta_0| > \delta_2) = 0(n^{-2})$. Theorem 4.3 implies that

$$1 - C_2/n \leq f_n(t)/g_n(t) \leq 1 + C_2/n \text{ for } |t - \theta_0| \leq \delta_2$$

where $g_n(t) = (n/2\pi)^{1/2} c^{-n}(t) A(t)/\sigma(t)$ and C_2 does not depend, on n or t.

Let $\{B_n\}$ be a sequence of Borel sets such that $P(T_n \in B_n)$ converges to a positive limit.
Then

$$\left| \int_{B_n} (f_n(t) - g_n(t))dt \right| \leq \left| \int_{A \cap B_n} f_n(t)dt - \int_{A \cap B_n} g_n(t)dt \right| + \left| \int_{A^c \cap B_n} f_n(t)\left(1 - \frac{g_n(t)}{f_n(t)}\right)dt \right|$$

$$\leq 0(1/n^2) + \left| \int_A g_n(t)dt \right| + C_2/n.$$

We will verify in section 6.2,

$$\int_{\delta_2+\theta_0}^{\infty} g_n(t)dt = 0(1/n).$$

Conditional on this last result, we can then conclude

$$\int_{B_n} f_n(t)dt = \int_{B_n} g_n(t)dt + 0(n^{-1}) \tag{4.9}$$

This result shows that $g_n(t)$ can be integrated with an error which is at most $0(n^{-1})$. In many inference problems, it is an approximation to the tail area which is of interest. From this point of view (4.9) is a much more useful result than (4.8).

In actually using (4.8) for computational purposes, it is recommended that the constant of integration be determined numerically. Some further discussion of this point is given in section 5.2. This technique gives an important increase in the accuracy of the approximation. In fact, Durbin (1980a) p. 317 gives a heuristic argument that the renormalization reduces the magnitude of the error from n^{-1} to $n^{-3/2}$. Since we are using these approximations for very small n, often $n \leq 10$, any order terms on the error have to be keep in context. The usefulness of (4.8) and (4.9) is really determined by how well, they perform for small to moderate values of n. By the time, we have n of 20 in many situations, the normal approximation may suffice.

We turn now to questions of computation of $g_n(t)$. To evaluate $g_n(t)$ at a specific point, the main computational effort is in computing $\alpha(t)$ from (4.3). This is a non-linear equation which involves a numerical integration for each function evaluation. For any cases where we have done computations, secant methods have proven very satisfactory. As an example, we computed $\alpha(t)$, $c(t)$, $\sigma^2(t)$ over a grid of 90 points for the mean with an underlying uniform distribution. The computations took 5.6 seconds of CPU time on a VAX 785 running 4.3 BSC Unix. In order to compute probabilities a simple way is to evaluate g_n over a grid of points and then do numerical integration. If we are specifically interested in evaluating tail areas, very accurate integral approximations (uniform asymptotic expansions) are available which reduce the computational effort substantially. The basic reference is Lugannani and Rice (1980). In chapter 6, these ideas are developed more fully and the approximations to the tail area provided.

A useful feature of the computation is that once $\alpha(t)$, $c(t)$ and $\sigma(t)$ have been computed, we can evaluate $g_n(t)$ for any n. To give an indication of the accuracy of the approximation we consider the following situation.

The exact density for the Huber estimate with $\psi(x) = x$ if $|x| < k$, $= k$ sgn x if $|x| \geq k$ has been calculated by P. Huber for two contaminated normal distributions and by A. Marazzi for the Cauchy distribution (unpublished). For the contaminated normal, the results are obtained by direct convolution and have been checked in double precision by D. Zwiers. The exact results for the Cauchy were calculated with fast Fourier transforms via characteristic functions. The effort in obtaining the exact results is substantially more than that required for the approximation and must be recomputed for each new value of n.

To measure, the accuracy of our approximation, the relative percent errors of tail areas are computed for selected values of t for both the the contaminated normal and Cauchy. Note that the relative percent error for upper tails is computed as:

100 (approximate - exact cumulative)/(1− exact cumulative). The values are as given in Exhibit 4.1 and 4.2.

t \ n	1	2	3	4	5	6	7	8	9
0.1	-0.05	0.12	0.05	0.04	0.04	0.03	0.03	0.03	0.03
0.5	-0.70	0.90	0.40	0.40	0.30	0.30	0.30	0.20	0.20
1.0	-5.40	3.70	0.40	1.00	0.70	0.70	0.70	0.60	0.70
1.5	-23.20	13.50	-3.60	-0.05	1.50	1.50	1.10	1.10	0.90
2.0	-60.80	30.50	-24.00	19.80	-10.20	12.50	-4.00		
2.5	-99.60	41.60	-64.60	37.80	-46.70				
3.0	-116.70	45.40	-90.20						

Exhibit 4.1a
Percent relative errors of the Huber estimate ($k = 1.4$) for contaminated normal ($\epsilon = .05$) versus t.

$100 \cdot (1 - G_n)$ \ n	1	2	3	4	5	6	7	8	9
5	-60.8	13.5	0.2	1.0	0.6	0.5	0.4	0.4	0.3
1			-12.8	4.0	0.7	0.8	0.7	0.6	0.7
0.1					-10.2	3.2	1.1	1.0	0.9
0.01							-6.5	2.8	1.4

Exhibit 4.1b
Percent relative errors of the Huber estimate ($k = 1.4$) for contaminated normal ($\epsilon = .05$) versus approximate percentage points determined by G_n, the cumulative of g_n.

A glance at the relative errors for the contaminated normal shows that for t values of 1 or less, the relative errors are all 1% or less even down to $n = 3$ and all remain under 10% (most under 3%) for $t = 1.5$. In terms of percentage points, the estimate is very accurate at the 1% level down to $n = 4$ with $\epsilon = 0.5$ and at the .1% level down to $n = 6$ with $\epsilon = .05$. It is only with small n and large t that the relative errors become larger and even here the estimate is fairly good. For instance with $n = 3$ and $t = 3.0$ (a relative error of 90%), the actual difference is $.002(.99795 - .99610)$. The results when $\epsilon = .10$ are similar to those above.

t \ n	1	2	3	4	5	6	7	8	9
1	-12.3	8.0	-4.4	0.8	-1.5	0.6	-0.7	-.03	-0.5
3	-21.0	23.3	-12.6	14.1	-7.0	8.5	-4.0	4.7	-2.6
5	-33.6	33.6	-24.9	24.9	-16.2	18.6	-12.2	13.0	-7.3
7	-43.5	40.3	-37.2	33.1	-28.0	27.8	-16.7	22.5	-16.7
9	-51.2	44.8	-47.8	38.6	-37.5	35.7	-29.8	31.0	-16.7

Exhibit 4.2
Percent relative errors of the Huber estimate ($k = 1.4$) for the Cauchy.

For the Cauchy, the relative errors remain well under control even well out into the extreme tails. For instance with $n = 7$ and $t = 9$ (relative error of 30%), the actual

difference is .00001(.99995 − .99994) so that the estimate would be very usable even at the .005% level. A graphical display of the results is given in Exhibit 4.3 and 4.4. The value of $\log(F/1-F)$ is plotted against t for the 5% contaminated normal and the Cauchy. From the graphs it can be seen that in terms of critical values the results for the contaiminated normal with $\epsilon = .05$ are very accurate to $n = 3$ for the 5% level, to $n = 5$ at the 1% level, to $n = 7$ at the .1% level and to $n = 9$ (or even $n = 7$) at the .01% level. Similar results hold for the Cauchy. These results imply that the approximations accurately reflect the distributional behaviour of the estimates and provide a very useful tool for determining small sample properties of interest.

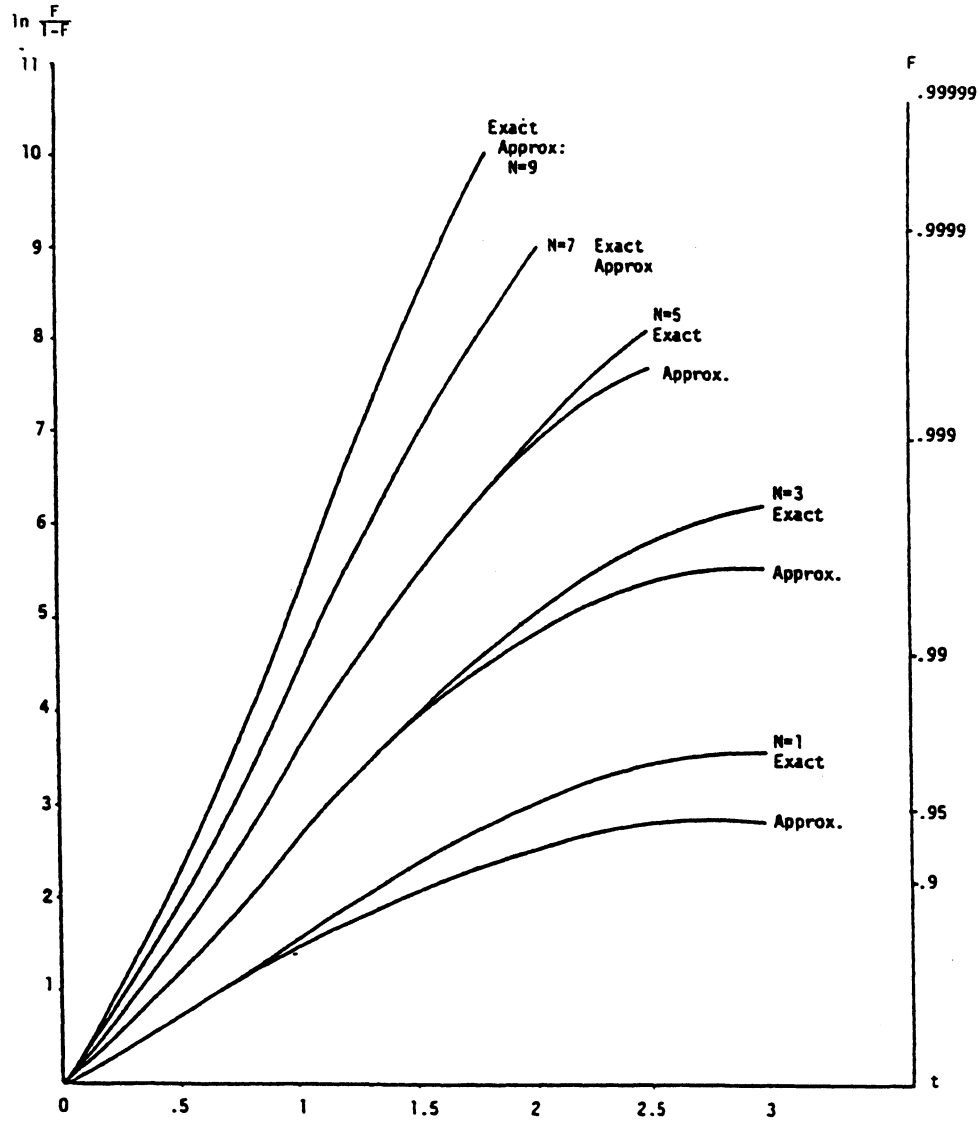

Exhibit 4.3
$\log(F_n/(1-F_n))$ versus t of Huber estimate ($k = 1.5$) for contaminated normal with ($\epsilon = .05$)

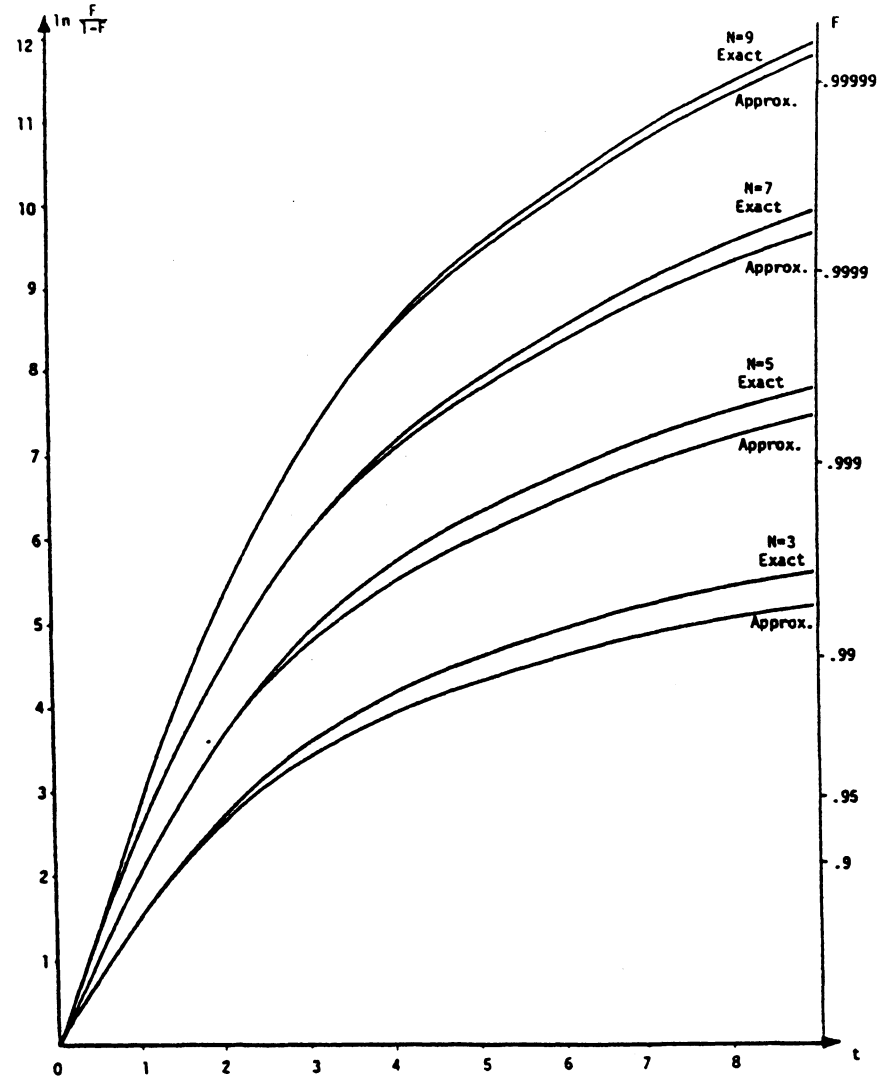

Exhibit 4.4
$\log(F_n/(1/F_n))$ versus t of Huber estimate ($k = 1.5$) for Cauchy

Some calculations with the mean suggest that any accuracy obtained by including the first neglected term is of the order of round off errors and so the first term in the expansion is all that is needed for good accuracy.

As another example, consider the M-estimate version of β-quantiles with $\psi(x) = \beta - 1$ for $x < 0$, $\psi(x) = 0$ for $x = 0$ and $\psi(x) = \beta$ for $x > 0$. For those n where the defining equation $\sum \psi(x_i - T_n) = 0$ has a unique solution, the exact density of the M-quantile is the density of the appropriate order statistic and is proportional to $(F(t))^{(n-1)\beta}(1 - F(t))^{(n-1)(\beta-1)}f(t)$. If we now compute the approximation given in (4.8), we find that

$$\alpha(t) = \log((1-\beta)F(t)/(\beta(1-F(t))),$$

$$c(t) = (1-\beta)\exp(-\beta\alpha(t))/(1-F(t)),$$

$$\sigma^2(t) = \beta(1-\beta) \quad \text{and} \quad A(t) = c(t)f(t)$$

and that in this case, the approximation is exact up to a normalizing constant. It should be noted that if we approximate f_n'/f_n instead of f_n as discussed in section 5.2, the approximation is exact.

As a final example, we could consider the example of logistic regression through the origin. This example will be developed in the next chapter where we compare several related approaches to obtaining an approximation.

4.3. GENERAL ONE-DIMENSIONAL SADDLEPOINT APPROXIMATIONS

In this section we consider a slightly different approach to derive saddlepoint approximations for general one-dimensional statistics. Basically, the moment generating function is approximated and a saddlepoint approximation is used by applying the techniques developed in section 3.3. We follow here Easton and Ronchetti (1986).

Suppose that $x_1, \cdots, x_n$ are n iid real valued random variables with density f and $T_n(x_1, \cdots, x_n)$ is a real valued statistic with density f_n. Let $M_n(\alpha) = \int e^{\alpha t} f_n(t)dt$ be the moment-generating function, $K_n(\alpha) = \log M_n(\alpha)$ be the cumulant-generating function, and $\rho_n(\alpha) = M_n(i\alpha)$ be the characteristic function of T_n. Further suppose that the moment-generating function $M_n(\alpha)$ exists for real α in some nonvanishing interval that contains the origin. By Fourier inversion (as in section 3.3),

$$\begin{aligned} f_n(t) &= \frac{1}{2\pi} \int_{-\infty}^{+\infty} M_n(ir) e^{-irt} dr \\ &= (n/2\pi i) \int_{\mathcal{I}} M_n(nz) e^{-nzt} dz \\ &= (n/2\pi i) \int_{\tau-i\infty}^{\tau+i\infty} \exp\{n[R_n(z) - zt]\} dz, \end{aligned} \tag{4.10}$$

where $\mathcal{I}$ is the imaginary axis in the complex plane and τ is any real number in the interval where the moment generating function exists, and

$$R_n(z) = K_n(nz)/n \ . \tag{4.11}$$

Note that if T_n is the arithmetic mean, then $R_n(z) = K(z)$, the cumulant generating function of the underlying density f, and in this case (4.10) equals formula (3.6). The problem in (4.10) is that in general R_n is *unknown*. The basic idea of the general saddlepoint approximation is to approximate $R_n(z)$ and then apply the saddlepoint technique to the integral in (4.10).

For a given problem, one can find sometimes special approximations to the cumulant generating function and to R_n. In very special situations these functions are even known exactly. However, since the goal of this approach is to derive a saddlepoint approximation for a *general* model, we do not want to assume any special structure in the problem. In this framework, classical Edgeworth expansions provide a very important source of general approximations to $\log \rho_n(\alpha)$ and therefore to $R_n(z)$. Moreover, they have the advantage

of being good local approximations around the center of the distribution. This property is exploited by the saddlepoint approximation which can be viewed as a low order Edgeworth approximation locally at each point (see section 3.4 and Remark 4.4 below). For these reasons we will use the Edgeworth approximation as a basic construction tool and will show that whenever an Edgeworth expansion for the density f_n of T_n is available, a saddlepoint approximation can be carrried out that will generally improve it.

Let us now work out the Edgeworth approximation for $R_n(z)$. Denote by $\tilde{f}_n$ the Edgeworth expansion for f_n up to and including the term of order n^{-1}. Let $\tilde{M}_n$ and $\tilde{K}_n$ be the moment generating function and the cumulant generating function of $\tilde{f}_n$, respectively. (Technically, $\tilde{M}_n$ and $\tilde{K}_n$ may not be moment generating and cumulant generating functions, since $\tilde{f}_n$ may not be a density. We will continue to use this terminology, however.) Let $\tilde{R}_n(z) = \tilde{K}_n(nz)/n$. Then, from the Edgeworth approximation (up to the term of order n^{-1}) one can obtain an approximation for $\log \rho_n(\alpha)$ and, therefore, $R_n(z)$ in terms of the first four cumulants. That is, $R_n(z)$ can be approximated by

$$\tilde{R}_n(z) = \mu_n z + \frac{n\sigma_n^2 z^2}{2} + \frac{\kappa_{3n}\sigma_n^3 n^2 z^3}{6} + \frac{\kappa_{4n}\sigma_n^4 n^3 z^4}{24}, \tag{4.12}$$

where μ_n is the mean, σ_n^2 the variance of T_n and κ_{jn} are the cumulants of T_n/σ_n. Note that $\mu_n = 0(1)$, $\sigma_n = 0(n^{-1/2})$, and $\kappa_{jn} = 0(n^{-j/2+1})$ for $j = 3,4$, since we have assumed that the Edgeworth expansion up to and including the term of order n^{-1} for f_n exists. In general, μ_n, σ_n and κ_{jn} are not known exactly, but expansions up to the appropriate order will suffice to keep the same order in the approximation. Applying the saddlepoint technique to the integral in (4.21) gives the saddlepoint approximation of f_n with uniform error of order n^{-1}:

$$g_n(t) = \left[\frac{n}{2\pi\tilde{R}_n''(\alpha_0)}\right]^{1/2} \exp\{n[\tilde{R}_n(\alpha_0) - \alpha_0 t]\}, \tag{4.13}$$

where α_0 is the saddlepoint determined as a root of the equation

$$\tilde{R}_n'(\alpha_0) = t, \tag{4.14}$$

$\tilde{R}_n(\alpha) = \tilde{K}_n(n\alpha)/n$ is given by (4.12) and $\tilde{R}_n'$ and $\tilde{R}_n''$ denote the first two derivatives of $\tilde{R}_n$.

As usual the approximation $g_n(t)$ given by (4.13) can be improved by renormalization, that is by computing numerically $C_n = \int_{-\infty}^{+\infty} g_n(t)dt$ to obtain $g_n(t)/C_n$ which integrates to 1. Tail areas can be obtained from (4.13) by numerical integration. However, sometimes it is convenient to have a direct saddlepoint approximation for the tail area $P[T_n > a]$. By the same argument as in Daniels (1983) (see section 6.2), the saddlepoint equation (4.14) can be used as a change of variables to obtain the approximation

$$P(T_n > a] \cong \int_{\alpha_\ell}^{\infty} \left[\frac{n\tilde{R}_n''(\alpha_0)}{2\pi}\right]^{1/2} \exp\{n[\tilde{R}_n(\alpha_0) - \alpha_0\tilde{R}_n'(\alpha_0)]\}d\alpha_0, \tag{4.15}$$

where $\tilde{R}_n'(\alpha_\ell) = a$.

Remark 4.1

In the case where the cumulant generating function K_n is known exactly, one can apply directly the saddlepoint method to (4.10). In this case, the saddlepoint approximation is given by (4.13) and (4.14), with $\tilde{R}_n$ replaced by the exact $R_n(\alpha) = K_n(n\alpha)/n$.

In order to show a connection with a result by Chaganty and Sethuraman (1985), we now compute the approximation (4.13) for the statistic T_n/n and *in the case where K_n is known exactly.* Let us denote by f_n, g_n, K_n, R_n the functions for T_n and by f_n^*, g_n^*, K_n^*, R_n^* the corresponding functions for T_n/n. Then we have

$$\begin{aligned} R_n^*(\alpha) &= K_n(\alpha)/n, \\ R_n^{*\prime}(\alpha) &= K_n'(\alpha)/n, \\ R_n^{*\prime\prime}(\alpha) &= K_n''(\alpha)/n, \end{aligned}$$

and the saddlepoint approximation for T_n/n evaluated at a point t_n

$$g_n^*(t_n) = \left[\frac{n}{2\pi R_n^{*\prime\prime}(\alpha_n)}\right]^{1/2} \exp\{n[R_n^*(\alpha_n) - \alpha_n t_n]\}, \tag{4.16}$$

where α_n is determined by the equation

$$R_n^{*\prime}(\alpha_n) = t_n.$$

Equation (4.16) is exactly equation (2.1) of Theorem 2.1 in Chaganty and Sethuraman (1985). The same authors generalize the approximation (4.16) to an arbitrary multidimensional statistics under the condition that the cumulant generating function is known; see Chaganty and Sethuraman (1986).

Remark 4.2
Since $\tilde{R}_n'(\alpha)$ is a third degree polynomial, the solution of equation (4.14) does not pose any computational problems. However, this equation can have multiple real solutions and only the solution α_0 with $\tilde{R}_n''(\alpha_0) > 0$ can be used for the saddlepoint approximation. A simple example in which this happens is the approximation to the density of $s^2 = (n-1)^{-1}\sum_{i=1}^n (x_i - \bar{x})^2$ where $x_1, \cdots, x_n$ are n iid observations from a $N(0,1)$.

Remark 4.3
Another possible approximation for $R_n(z)$ can be obtained using $\tilde{\rho}_n(\alpha)$ as given by the Edgeworth approximation instead of the expansion of $\log \tilde{\rho}_n(\alpha)$. This amounts to approximating $R_n(z)$ by

$$\begin{aligned} \tilde{R}_n(z) &= \mu_n z + n\sigma_n^2 z^2/2 \\ &\quad + \frac{1}{2}\log\left(1 + \frac{n^3\sigma_n^3 z^3 \kappa_{3n}}{6} + \frac{3\kappa_{4n} n^4 \sigma_n^4 z^4 + \kappa_3^2 n^6 \sigma_n^6 z^6}{72}\right). \end{aligned}$$

We have no numerical experience with this approach.

Remark 4.4
One can use the same kind of computations as in section 4.2 to express f_n by means of its conjugate density, namely,

$$f_n(t) = e^{n(R_n(\tau) - \tau t)} h_{\tau,n}(t)$$

where $h_{\tau,n}(t)$ is the density of T_n with the underlying conjugate density. The choice $\tau = \alpha_0$ and an Edgeworth expansion of $h_{\tau,n}(t)$ leads to the saddlepoint approximation of f_n. This approach also requires knowledge of the exact cumulant-generating function. The general

saddlepoint approximation described earlier corresponds to using the Edgeworth expansion to obtain an approximation to the conjugate density. Thus,

$$f_n(t) = e^{n(\tilde{R}_n(\tau)-\tau t)}\tilde{h}_{\tau,n}(t) + D_n(t)$$

where $\tilde{R}_n(z)$ is given by (4.12), $\tilde{h}_{\tau,n}(u)$ is the conjugate density of $\tilde{f}_n$, and $D_n(t) = f_n(t) - \tilde{f}_n(t)$. Note that the term of order $n^{-1/2}$ disappears because $\tilde{f}_n$ is recentered at t through $\tilde{h}_{\tau,n}$; that is,

$$\tilde{R}'_n(\tau) - t = 0$$

if $\tau = \alpha_0$, the saddlepoint.

From the conjugate density point of view, the development of this approach is similar in spirit to that of Durbin (1980a) and Barndorff-Nielsen (1983), (see section 5.3), but we do not restrict ourselves to sufficient statistics or to maximum likelihood estimators nor do we assume any underlying parametric model. It should be noted, however, that when a special structure exists, saddlepoint approximations that have been developed to exploit this structure should be used, as they will generally perform better than this general approach.

Remark 4.5

As was pointed out by J. W. Tukey and B. Efron, this technique could be applied iteratively as follows. Start with an approximation $\tilde{R}_n^{(1)}$ for R_n (given for instance by an Edgeworth approximation) and apply the saddlepoint technique to (4.10) to get the saddlepoint approximation g_n in (4.13). Now, by numerical integration, compute a new approximation for R_n,

$$\tilde{R}_n^{(2)}(z) = (1/n)\log \tilde{M}_n^{(2)}(nz),$$

where $\tilde{M}_n^{(2)}(\alpha) = \int e^{\alpha t} g_n(t)dt$, then compute a new saddlepoint approximation for the density. This can be repeated until convergence is reached. This iteration process may improve the original approximation but its performance is an open question.

Remark 4.6

An alternative way of approximating the density of a general statistic by means of saddlepoint techniques is the following (see Field 1982 and section 4.5). Suppose T_n can be written as a functional T of the empirical distribution function $F^{(n)}$; that is, $T_n = T(F^{(n)})$. First, linearize T_n using the first term of a von Mises expansion (see von Mises 1947),

$$T_n \cong T(F) + L_n(T, F), \tag{4.17}$$

where F is the underlying distribution of the observations,

$$L_n(T, F) = \frac{1}{n}\sum_{i=1}^{n} IF(x_i; T, F), \tag{4.18}$$

and $IF(x; T, F)$ is the influence function of T at F (cf. Hampel 1968, 1974 and section 2.5). Then apply the classical saddlepoint approximation to $L_n(T, F)$, which is just an average of iid random variables. For instance, Tingley and Field (1988) apply the Lugannani and Rice approximation for the tail areas to (4.18); see section 6.3.

4.4. L-ESTIMATORS

In this section we apply the general saddlepoint technique to derive approximations to the density of linear combinations of order statistics.

We consider statistics of the form

$$T_n = \frac{1}{n}\sum_{i=1}^{n} c_{in} x_{(i)}, \tag{4.19}$$

where $x_{(1)} \leq x_{(2)} \leq \cdots \leq x_{(n)}$ are the order statistics and $c_{1n}, \cdots, c_{nn}$ are weights generated by a function $J : (0,1) \to R$,

$$c_{in} = J[i/(n+1)], \qquad i = 1, \cdots, n.$$

Typically the conditions imposed on J are those that guarantee the existence of an Edgeworth expansion.

The distribution properties of L-statistics have been investigated by many authors. Exact distributions under special underlying distributions can be found in Weisberg (1971), Cicchitelli (1976), and David (1981). Asymptotic normality of these statistics has been shown under different sets of conditions (e.g., see Chernoff, Gastwirth and Johns, 1967; Shorack 1969, 1972; Stigler, 1969, 1974; David, 1981). Parr and Schucany (1982) investigated the small sample behavior of L-estimators via jackknifing. Finally, Helmers (1979, 1980) and van Zwet (1979) derived Edgeworth expansions for L-statistics with remainder $0(n^{-3/2})$. These will be the basic elements of our approximation, which we use in conjunction with the saddlepoint technique. In this section we summarize the numerical results obtained via saddlepoint approximation by Easton and Ronchetti (1986) for two L-estimators. These results show the great accuracy of this approximation down to very small sample sizes. These techniques can be used to approximate the distribution of more general L-statistics. One possible application is to so-called broadened letter values or "bletter values" (means of blocks of order statistics) suggested by Tukey (1977) as an improvement of the usual letter values in exploratory data analysis.

4.4.a The Asymptotically Most Efficient L-Estimator Under The Logistic Distribution

In this example we consider the asymptotically first-order efficient L-estimator for the center θ of the logistic distribution

$$F(x-\theta) = 1/[1+\exp(-(x-\theta))].$$

This L-estimator is of the form (4.19) with the weight function $J(s) = 6s(1-s)$ and weights $c_{in} = 6i(1 - i/(n+1))/(n+1)$. We apply the technique presented in 4.3 to compute an approximation to the distribution of the statistic $n^{1/2}(T_n - \mu)/\sigma$, where $\mu(=0)$ and σ are the asymptotic mean and variance of T_n under the logistic F. The Edgeworth expansion required in our formula is taken from Helmers (1980). It should be noted that the third moment equals 0 because of symmetry, so the term of order $n^{-1/2}$ disappears in the Edgeworth expansion. Thus the latter is of order n^{-1} and should be very competitive with the saddlepoint approximation.

Numerical results for the cumulative distribution for sample sizes $3, 4, 10$ are given in Exhibits 4.5, 4.6, 4.7 for the right half of the distribution, since the density is symmetric.

The exact values are taken from Helmers (1980) who computed them by numerical integration for sample sizes 3 and 4, and by Monte Carlo simulation for sample size 10. The saddlepoint approximation for the cumulative is obtained by numerical integration from the saddlepoint approximation for the density computed using (4.13) and (4.14) for about 500 t values. Exhibit 4.8 shows the residuals from the exact density for the rescaled saddlepoint, Edgeworth, and normal approximations. This plot clearly indicates that the rescaled saddlepoint approximation overall improves the Edgeworth approximation.

x	Exact	Rescaled Saddlepoint	Unscaled Saddlepoint	Edgeworth	Normal
2	.5640	.5617	.5735	.5536	.5793
.4	.6262	.6217	.6320	.6069	.6554
.6	.6850	.6787	.6874	.6592	.7257
.8	.7391	.7314	.7387	.7099	.7881
1.0	.7875	.7790	.7850	.7582	.8413
1.2	.8248	.8210	.8259	.8032	.8849
1.4	.8658	.8572	.8610	.8439	.9192
1.6	.8958	.8877	.8908	.8796	.9452
1.8	.9202	.9130	.9154	.9100	.9641
2.0	.9397	.9335	.9353	.9348	.9772
2.2	.9550	.9499	.9513	.9543	.9861
2.4	.9669	.9628	.9638	.9691	.9918
2.6	.9758	.9727	.9734	.9798	.9953
2.8	.9825	.9802	.9807	.9873	.9974
3.0	.9875	.9858	.9862	.9923	.9987

Exhibit 4.5
Exact cumulative distribution and approximations for sample size 3 for the asymptotically best L-estimator under logistic distribution.

x	Exact	Rescaled Saddlepoint	Unscaled Saddlepoint	Edgeworth	Normal
.2	.5663	.5650	.5750	.5601	.5793
.4	.6307	.6281	.6366	.6190	.6554
.6	.6919	.6877	.6949	.6758	.7257
.8	.7469	.7424	.7484	.7295	.7881
1.0	.7963	.7914	.7962	.7790	.8413
1.2	.8391	.8341	.8379	.8236	.8849
1.4	.8752	.8703	.8732	.8627	.9192
1.6	.9049	.9003	.9026	.8960	.9452
1.8	.9287	.9247	.9264	.9235	.9641
2.0	.9474	.9440	.9453	.9454	.9772
2.2	.9618	.9591	.9600	.9622	.9861
2.4	.9726	.9705	.9712	.9748	.9918
2.6	.9807	.9791	.9796	.9837	.9953
2.8	.9865	.9854	.9857	.9898	.9974
3.0	.9907	.9899	.9902	.9939	.9987

Exhibit 4.6
Exact cumulative distribution and approximations for sample size 4 for the asymptotically best L-estimator under the logistic distribution.

x	Exact	Rescaled Saddlepoint	Unscaled Saddlepoint	Edgeworth	Normal
.2	.5734	.5725	.5776	.5716	.5793
.4	.6445	.6426	.6468	.6409	.6554
.6	.7089	.7080	.7115	.7058	.7257
.8	.7680	.7670	.7698	.7647	.7881
1.0	.8196	.8186	.8208	.8164	.8413
1.2	.8629	.8622	.8638	.8604	.8849
1.4	.8985	.8978	.8990	.8966	.9192
1.6	.9275	.9260	.9269	.9255	.9452
1.8	.9486	.9477	.9483	.9478	.9641
2.0	.9646	.9639	.9644	.9645	.9772
2.2	.9764	.9757	.9760	.9766	.9861
2.4	.9845	.9840	.9842	.9850	.9918
2.6	.9905	.9897	.9898	.9907	.9953
2.8	.9937	.9935	.9936	.9944	.9974
3.0	.9959	.9960	.9961	.9968	.9987

Exhibit 4.7
Exact cumulative distribution and approximations for sample size 10 for the asymptotically best L-estimator under logistic distribution.

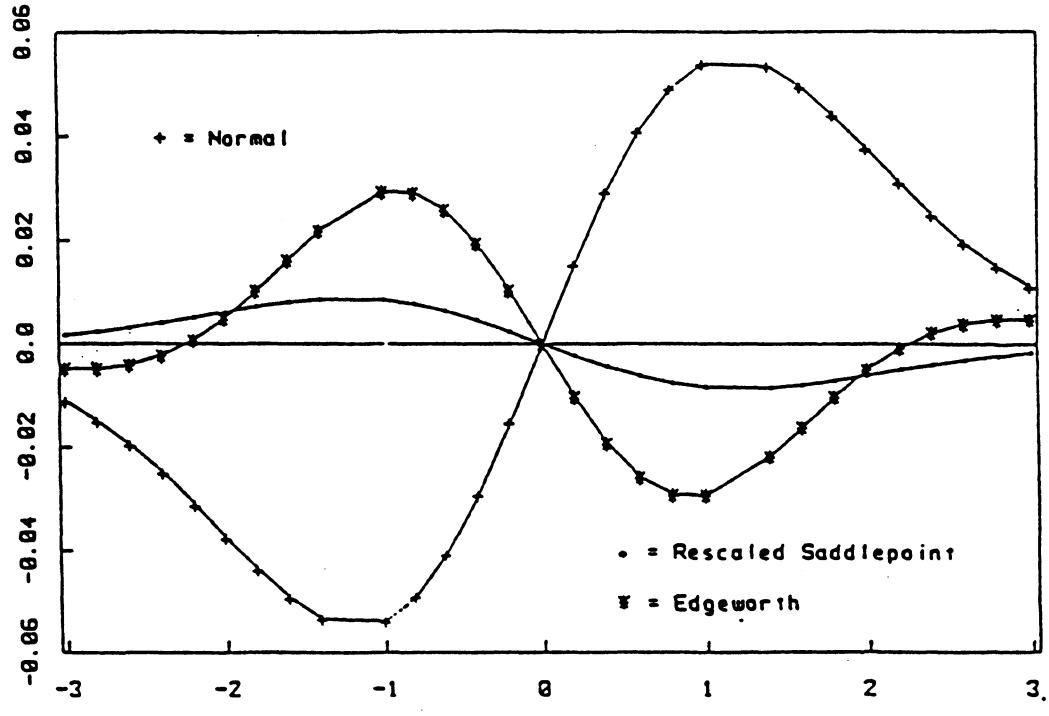

Exhibit 4.8
Error of the approximations of the density for sample size 3 for the asymptotically best L-estimator under logistic distribution.

In addition, unlike the Edgeworth and normal approximations, the Exhibits show that the rescaled saddlepoint approximation is wider tailed than the exact distribution, so its error is in the direction of giving conservative tests and confidence intervals. The same pattern can be seen for sample size 4 (not shown).

Although we are not approximating the distribution function directly, in practice these approximations may be used for calculating tail areas. Thus it is of interest to see how the saddlepoint approximation performs in the tails. Exhibit 4.9 shows the right-tail probability error for the right half of the distribution (for sample size 3) for the unscaled saddlepoint, rescaled saddlepoint, Edgeworth and normal approximation. The same pattern can be seen for sample size 10 (not shown).

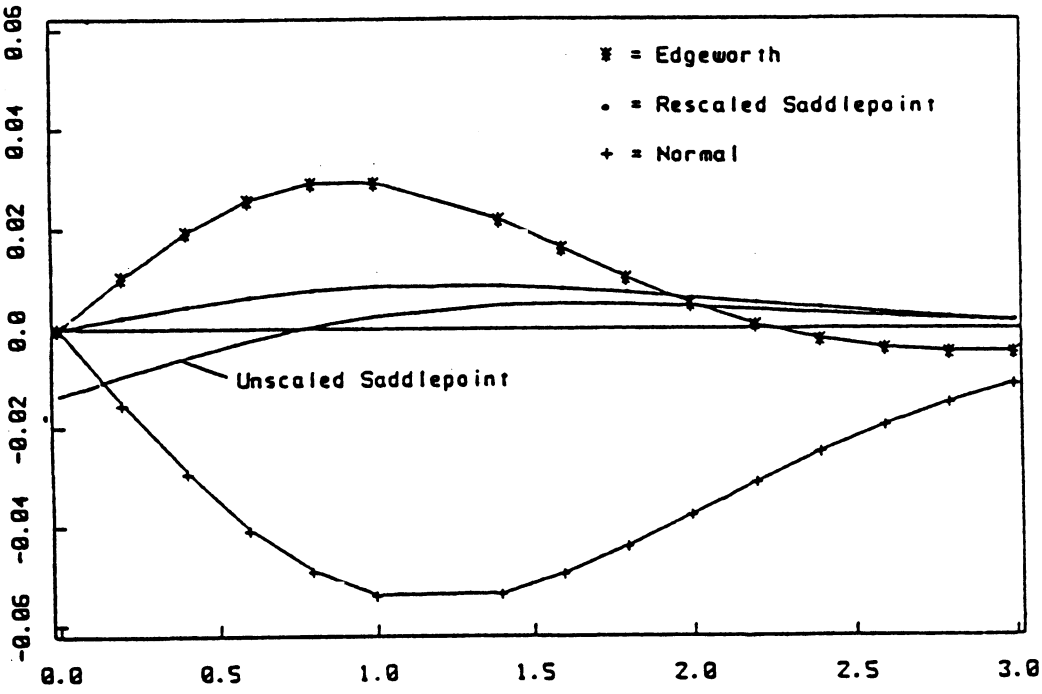

Exhibit 4.9
Error for the tail area for sample size 3 (cf. Exhibit 4.5).

Overall it appears that the rescaled saddlepoint technique generally improves on the Edgeworth approximation with respect to tail area and tends to err in the direction which produce conservative tests and confidence intervals.

4.4.b Trimmed Means of Exponential Observations

This example considers approximations to the distribution of trimmed means of exponential observations. Let α_ℓ and α_u be the fraction of the observations trimmed from the upper and lower tails, respectively. We consider statistics of the form (4.19), where

$$\begin{aligned} c_{in} &= 0 \qquad \text{for } i \le n\alpha_\ell \quad \text{or} \quad i \ge n(1-\alpha_u) \\ &= n/k \qquad \text{otherwise,} \end{aligned}$$

where k is the number of nonzero weights. Note that $(1/n)\sum_{i=1}^{n} c_{in} = 1$.

Helmers (1979) derived the Edgeworth expansion for the distribution of $(T_n - \mu_n)/\sigma_n$, for trimmed linear combinations of order statistics with general weights on the observations between the α_ℓ and $1 - \alpha_u$ sample quantiles and with zero weights on the remaining observations. This expansion forms the basis for our general saddlepoint approximation.

The density of certain linear combinations of exponential order statistics can be written explicitly (see David, 1981). In our case the exact density of T_n is

$$f_n(t) = \sum_{i=1}^{n} \frac{w_{in}}{a_{in}} \exp\left(-\frac{t}{a_{in}}\right), \tag{4.20}$$

where

$$w_{in} = a_{in}^{n-1} / \prod_{h \neq i} (a_{in} - a_{hn}) \quad \text{and} \quad a_{in} = \frac{1}{n-i+1} \frac{1}{n} \sum_{j=i}^{n} c_{jn}$$

$$\text{for } i = 1, \cdots, n, \text{ provided } a_{in} \neq a_{jn} \text{ for } i \neq j.$$

Note that, given the relative numerical instability of (4.20) for moderate sample sizes, a simple and accurate approximation like the saddlepoint approximation can be a good alternative even in this case where f_n is known exactly .

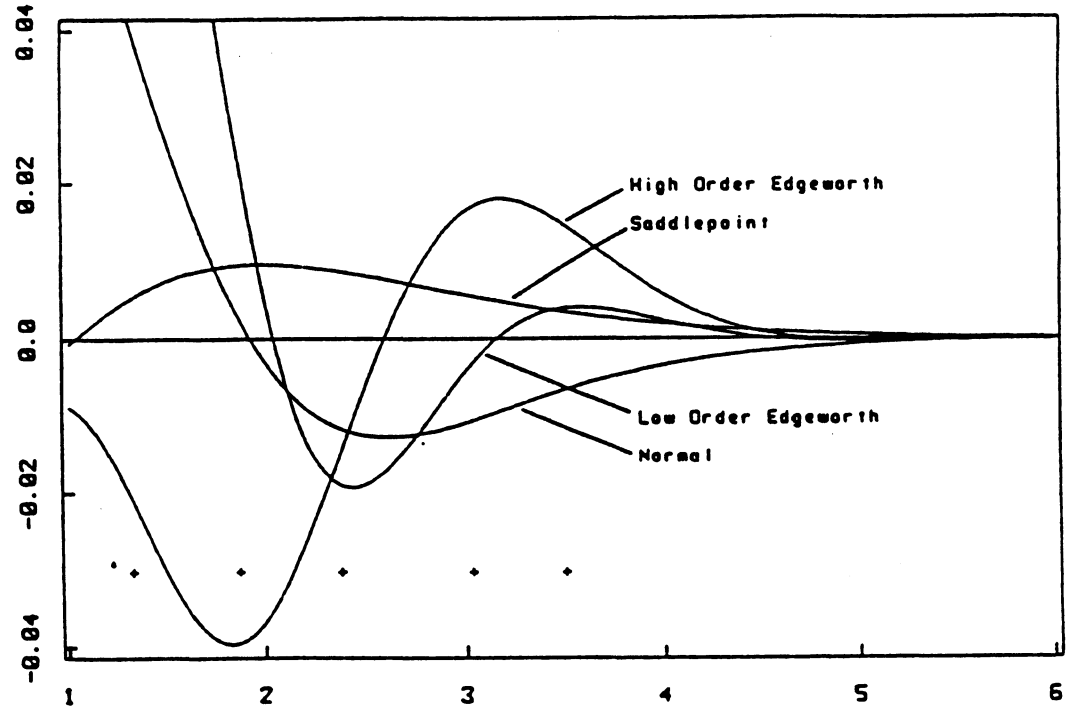

Exhibit 4.10
Error from the exact density in the right tail for 20% trimmed mean of 5 exponential observations. High (low) Edgeworth includes terms to order $n^{-1}(n^{-1/2})$. + marks denote .90, .95, .975, .99, .995 quantiles of exact distribution.

The figure shows that the saddlepoint approximation tends to be fairly stable and generally slightly wide throughout the tail. Both of the Edgeworth approximations show polynomial-like waves (and also become negative). The low order Edgeworth crosses the exact distribution a couple of times in the tail switching from being too wide to too narrow and back. The low-order Edgeworth approximation performs much better than the high-order Edgeworth approximation throughout this region and is competitive with the saddlepoint

approximation in the 5% tail. It is sometimes too narrow, however. In the 5% tail, the error in the normal approximation is only slightly larger in absolute value than the error in the saddlepoint approximation, but the normal approximation is uniformly narrow.

Exhibit 4.11 plots the error in the approximate tail areas for the right 10% tail. As in the density case, the error in the Edgeworth approximations shows wavy behavior whereas the saddlepont approximation is uniformly wide.

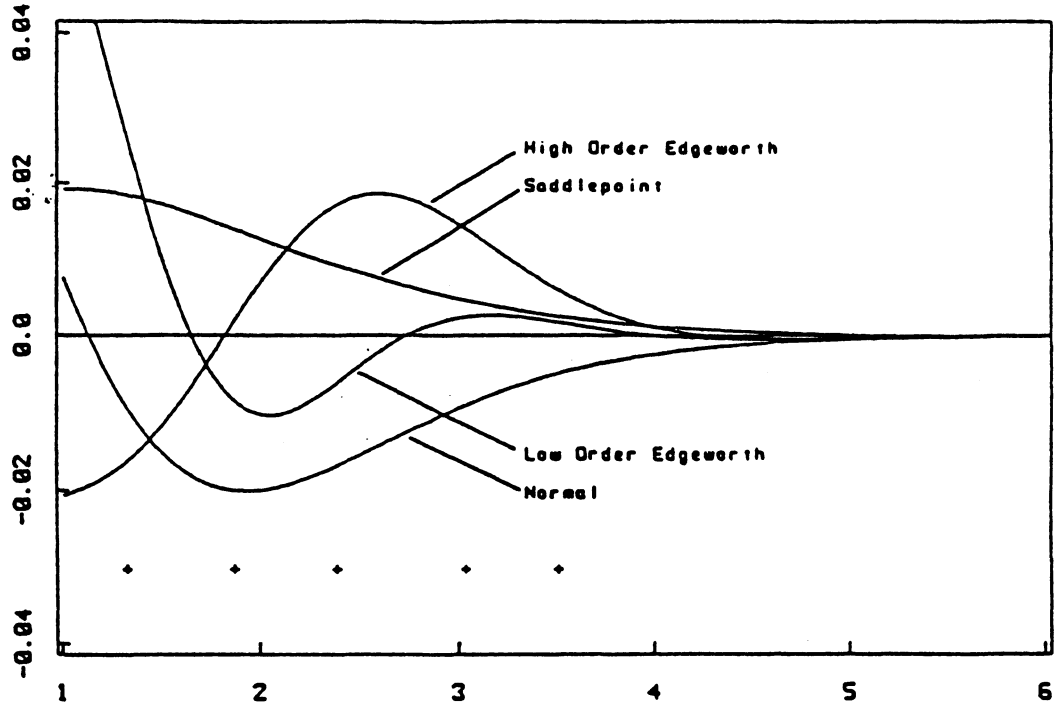

Exhibit 4.11
Error from the exact distribution in the right tail for the 20% trimmed mean of 5 exponential observations.

This example shows once more that the saddlepoint approximation exhibits some desirable properties that the Edgeworth approximations do not have. First, the saddlepoint approximation is unimodal and does not show the polynomial-like waves exhibited by the Edgeworth approximations. Thus the error in the saddlepoint approximation tends to be stable locally. Finally, the saddlepoint approximation tends to be wide in the tails so that error is in the direction of giving conservative tests and confidence intervals.

4.5. MULTIVARIATE M-ESTIMATORS

4.5.a Approximation

We return to the case of M-estimates as in section 4.2 but consider now the multiparameter problem. The M-estimate $\mathbf{T}_n$ for $\theta = (\theta_1, \theta_2, \cdots, \theta_p)$ is the solution of the system of equations:

$$\sum_{i=1}^{n} \psi_j(x_i, \mathbf{t}) = 0 \quad \text{for} \quad j = 1, \cdots, p. \tag{4.21}$$

The problem, as before, is to approximate the density $f_n(\mathbf{t})$ of $\mathbf{T}_n$. The development is similar to that presented in section 4.2 in that $\mathbf{T}_n$ is expanded as a multivariate mean, the density is centered and a multivariate lower order Edgeworth expansion is used. The centering parameter $\alpha(\mathbf{t})$ (cf. (4.3)) is replaced by a p-dimensional vector $\alpha(\mathbf{t})$ which is obtained as the solution of

$$\int \psi_j(x, \mathbf{t}) \exp\left\{\sum_{j=1}^{p} \alpha_j \psi_j(x, \mathbf{t})\right\} f(x) dx = 0 \quad j - 1, \cdots, p. \tag{4.22}$$

Although we write x as univariate, the results hold for $\mathbf{x}$ multivariate. The following assumptions are multivariate versions of A4.1 to A4.5 of section 4.2. Note that D_j denotes differentiation with respect to θ_j.

A4.1M The system of equations (4.21) has a unique solution.

A4.2M There is an open subset U of R^m such that

(i) for each $\theta \in \Theta$ one has $F_\theta(U) = 1$ and

(ii) the derivatives $D_j \psi_r(x, \theta)$, $D_k D_j \psi_r(x, \theta)$, $D_l D_k D_j \psi_r(x, \theta)$ exist for $1 \leq r, j, k, l \leq p$.

A4.3M For each compact $K \subset \Theta$,

(i) for $0 \leq j, k \leq p$, $1 \leq r \leq p$,

$$\sup_{\theta_0 \in K} E_{\theta_0} |D_k D_j \psi_r(X, \theta_0)|^4 < \infty.$$

(ii) there is an $\epsilon > 0$ such that for $1 \leq r, j, k, l \leq p$,

$$\sup_{\theta_0 \in K} E_{\theta_0} \Big(\max_{|\theta - \theta_0| < \epsilon} |D_l D_k D_j \psi_r(X, \theta)| \Big)^3 < \infty.$$

A4.4M For each $\theta_0 \in \Theta$

$$E_{\theta_0} \psi_r(X, \theta_0) = 0$$

and the matrices

$$A(\theta_0) = E_{\theta_0} \left[\frac{\partial \psi}{\partial \theta}(X, \theta_0) \right]$$

$$C(\theta_0) = E_{\theta_0} [\psi(X, \theta_0) \psi^T(X, \theta_0)]$$

are non singular.

A4.5M The functions $A(\theta)$ and $E_\theta[(D_{k_1} D_{j_1} \psi_{r_1})(D_{k_2} D_{j_2} \psi_{r_2})]$, $0 \leq j_1, j_2, k_1, k_2 \leq p$, $k_1 + j_1 \geq 1$, $k_2 + j_2 \geq 1$, $1 \leq r_1, r_2 \leq p$, are continuous on Θ.

At this point, a multivariate centering result is required. For a fixed $\mathbf{t}$, the conjugate density is

$$h_{\mathbf{t}}(x, \theta) = c(\mathbf{t}) \exp\left\{\sum_{j=1}^{p} \alpha_j \psi_j(x, \mathbf{t})\right\} f(x, \theta).$$

Theorem 4.4

Assume that the joint density of $\sum_{i=1}^{n}\psi_1(X_i,\mathbf{t})\cdots\sum_{i=1}^{n}\psi_p(X_i,\mathbf{t})$ exists and has Fourier transforms that are absolutely integrable both under f and $h_{\mathbf{t}}$. If we let $h_{\mathbf{t},n}$ be the density of $\mathbf{T}_n$ with underlying density $h_{\mathbf{t}}$, then

$$f_n(\mathbf{t}) = c^{-n}(\mathbf{t})h_{\mathbf{t},n}(\mathbf{t}). \tag{4.23}$$

Proof: Let $\mathbf{Z} = \left(\sum_{k=1}^{n}\psi_1(X_k,\mathbf{t}_0),\cdots,\sum_{k=1}^{n}\psi_p(X_k,\mathbf{t}_0)\right)$ 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

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

Choose $\mathbf{u} = (a_1+iy_1, a_2+iy_2,\cdots,a_p+iy_p) = \mathbf{a}+i\mathbf{y}$, $\mathbf{v} = (iw_1,\cdots,iw_p) = i\mathbf{w}$. Now

$$\begin{aligned}
M(\mathbf{a}+i\mathbf{y},i\mathbf{w}) &= \int_{\cdots}\int \exp\left\{\sum_{k=1}^{n}\sum_{j=1}^{p}iy_j\psi_j(x_k,\mathbf{t}_0) + \sum_{j=1}^{p}iw_jt_j(\mathbf{x})\right\} \\
&\quad\times \exp\left\{\sum_{k=1}^{n}\sum_{j=1}^{p}a_j\psi_j(x_k,\mathbf{t}_0)\right\}\prod_{k=1}^{n}f(x_k)dx_1\cdots dx_n \\
&= c^{-n}(\mathbf{t}_0)\int_{\cdots}\int \exp\left\{\sum_{k=1}^{n}\sum_{j=1}^{p}iy_j\psi_j(x_k,\mathbf{t}_0) + \sum_{j=1}^{p}iw_jt_j(\mathbf{x})\right\} \\
&\quad\prod_{k=1}^{n}h_{\mathbf{t}_0}(x_k)dx_1\cdots dx_n \\
&= c^{-n}(\mathbf{t}_0)M_1(i\mathbf{y},i\mathbf{w}),
\end{aligned}$$

where M_1 is the moment generating function of $(\mathbf{Z},\mathbf{T}_n)$ under $h_{\mathbf{t}_0}$. Since both M and M_1 are absolutely integrable, we can apply the Fourier inversion formula to give

$$g(\mathbf{z},\mathbf{t}) = \frac{1}{(2\pi i)^{2p}}\int_{\cdots}\int \exp\left\{-\sum_{j=1}^{p}u_jz_j - \sum_{j=1}^{p}v_jt_j\right\}M(\mathbf{u},\mathbf{v})d\mathbf{u}d\mathbf{v},$$

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 $\mathbf{u} = (\mathbf{a}+i\mathbf{y})$ and $\mathbf{v} = i\mathbf{w}$, we have

$$\begin{aligned}
g(\mathbf{z},\mathbf{t}) &= \frac{1}{(2\pi)^{2p}}\int_{\cdots}\int \exp\left\{-\sum_{j=1}^{p}(a_j+iy_j)z_j - \sum_{j=1}^{p}iw_jt_j\right\}M(\mathbf{a}+i\mathbf{y},i\mathbf{w})d\mathbf{y}d\mathbf{w} \\
&= \frac{c^{-n}(\mathbf{t}_0)}{(2\pi)^{2p}}\exp\left\{-\sum_{j=1}^{p}a_jz_j\right\}\int_{\cdots}\int \exp\left\{\sum_{j=1}^{p}iy_jz_j - \sum_{j=1}^{p}iw_jt_j\right\}M_1(i\mathbf{y},i\mathbf{w})d\mathbf{y}d\mathbf{w} \\
&= c^{-n}(\mathbf{t}_0)\exp\left\{-\sum_{j=1}^{p}a_jz_j\right\}g_1(\mathbf{z},\mathbf{t}).
\end{aligned}$$

Now

$$\mathbf{T}_n(\mathbf{x}) = \mathbf{t}_0 \Longleftrightarrow \sum_{i=1}^{n} \psi_j(x_j, \mathbf{t}_0) = 0, \qquad j = 1, \cdots, p \Longleftrightarrow \mathbf{z} = 0.$$

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. □

We now go through the same steps as in section 4.2 modified for the multivariate case (details given in Field, 1982, p. 675–6). As a result we can write

$$(\mathbf{T}_n - \mathbf{t}_0)_i = \sum\nolimits_j b_{ij} Z_j - \sum_{j,l} \left\{ \sum_r b_{ir} C_{jl}(r) \right\} \bar{Z}_j \bar{Z}_l + \sum_{j,l} \sum_r b_{ir} b_{lj} \bar{Z}_j (\bar{Z}_{r_l} - \mu_{r_l}) + 0_p(1/n) \tag{4.24}$$

where $B = -A(t_0)^{-1}$, $A(\mathbf{t}_0) = E_{h_{t_0}} \left| \frac{\partial \psi}{\partial t_0}(X, \mathbf{t}_0) \right|$

$$Z_r = \psi_r(X, \mathbf{t}_0), \quad Z_{rj} = D_j \psi_r(X, \mathbf{t}_0), \quad E(Z_{rj}) = \mu_{rj}$$

$$C_{jl}(r) = \sum_{i_1 i_2} b_{ji_1} b_{li_2} \mu_{ri_1 i_2}.$$

All expected values are with respect to h_{t_0}. Of course the result also holds for $\mathbf{t}_0 = \theta_0$ with density f.

Again using the results of James and Mayne (1962) on cumulants, the cumulants behave as in the unvariate case. It then can be shown that the density of $n^{1/2}(\mathbf{T}_n - \mathbf{t}_0)$ under h_{t_0} at $\mathbf{x}$ is

$$h(\mathbf{x}) = \phi(x)\left[1 + \sum\nolimits_j d_j D_j \{\phi(\mathbf{x})\}/n^{1/2} + \sum\nolimits_{j,k,l} d_{jkl} D_j D_k D_l \{\phi(\mathbf{x})\}/n^{1/2} + 0(1/n)\right]$$

where d_j and d_{jkl} are constants determined by the cumulants; $\phi(\mathbf{x})$ is the multivariate p-dimensional normal density with mean 0 and covariance matrix

$$A(\mathbf{t}_0)^{-1} \Sigma(\mathbf{t}_0) (A(\mathbf{t}_0)^{-1})^T \quad \text{and} \quad \Sigma(\mathbf{t}_0) = \left\{ E\psi_r(X, \mathbf{t}_0) \psi_i(X, \mathbf{t}_0) \right\}_{1 \le r, i \le p}$$

and all expectations are with respect to h_{t_0}. The density of $(\mathbf{T}_n - \mathbf{t}_0)$ under h_{t_0} at 0 is $h_{t_0,n}(\mathbf{t}_0) = n^{p/2} h(0)$. From the centering lemma it follows that

$$f_n(\mathbf{t}_0) = (c(\mathbf{t}_0))^{-n} n^{p/2} h(0).$$

Putting the results together gives the multivariate version of Theorem 4.3.

Theorem 4.5

If $\mathbf{T}_n$ represents the solution of $\sum_{i=1}^{n} \psi_r(x_i, t) = 0$, $r = 1, \cdots, p$, and Assumptions A4.1M-A4.5M are satisfied, then an asymptotic expansion for the density of $\mathbf{T}_n$, say f_n, is

$$f_n(\mathbf{t}_0) = (n/2\pi)^{p/2} c^{-n}(\mathbf{t}_0) |\det A| |\det \Sigma|^{-1/2} \{1 + 0(1/n)\} \tag{4.25}$$

where $\alpha(t_0)$ is the solution of

$$\int \psi_r(x, t_0) \exp\left\{\sum_{j=1}^{p} \alpha_j \psi_j(x, t_0)\right\} f(x)dx = 0 \quad \text{for} \quad r = 1, \cdots, p,$$

$$c^{-1}(t_0) = \int \exp\left\{\sum_{j=1}^{p} \alpha_j(t_0)\psi_j(x, t_0)\right\} f(x)dx,$$

$$A = \left\{E\partial\psi(x, t)/\partial t_r\big|_{t=t_0}\right\}_{1\le r,j\le p}, \quad \Sigma = \left\{E\psi_j(x, t_0)\psi_r(x, t_0)\right\}_{1\le r,j\le p}$$

and all expectations are with respect to the conjugate density

$$h_{t_0}(x) = c(t_0) \exp\left\{\sum_{j=1}^{p} \alpha_j(t_0)\psi_j(x, t_0)\right\} f(x).$$

The error term holds uniformly for all t_0 in a compact set .

Approximation (4.25) is usually an intermediate step. More often we are interested in tail areas for a marginal distribution of one of the estimates or for some function of the estimates (eg. in tests of hypothesis). In order to do this sort of computations, it is necessary to compute $f_n(t)$ over a grid in R^p and then do a numerical integration to compute the density of $\lambda(T)$ (say). If $p > 3$, it is not computationally feasible to proceed in this fashion and even with $p = 3$, the computational effort required may be large.

As in the univariate case, it would be useful to have (4.25) hold for an arbitrary set so that integrals would still be correct to $0(1/n)$. Although it is probably possible to obtain such results using sophisticated integral approximations, a different approach has been taken in recent work by Tingley and Field (1990) . The details are provided in chapter 6. The techniques there provide a computationally feasible technique for handling higher dimensional problems. The starting point of this approach is the approximation above.

4.5.b Calculations

We now turn to the case of location and scale where direct calculations have been carried out with a view of computing percentiles of a studentized version of the location estimate. Given the percentiles, we can then construct approximate confidence intervals for the location parameter. We let $\theta = (\mu, \sigma)$, $f_\theta(x) = f((x-\mu)/\sigma)/\sigma$ and $\psi_i(x, \theta) = \psi_i((x-\mu)/\sigma)$, $i = 1, 2$. In particular, we set $\psi_1(x) = \min\{k, \max(-k, x)\}$, $\phi_2(x) = \psi_1^2(x) - \beta$ with $\beta = E_\phi \psi_1^2(x)$.

This corresponds to "Proposal 2" of Huber(1964) and gives translation and scale equivariant estimates.

For $k < \infty$, we have robust M-estimates with a choice of β suitable for a model in some neighborhood of the normal. The joint density of (T_1, T_2) was computed giving values in the following table with underlying densities: normal, t_3, slash (ratio of normal and uniform on $[0, 1]$) and Cauchy.

		$n = 5$				$n = 10$			
t_1	t_2	normal	t_3	slash	Cauchy	normal	t_3	slash	Cauchy
0.00	.05	.012856	.009371	.007691	.005387	.00003	.000002	.000001	.000001
0.05	.05	.007407	.004776	.004672	.002197	.00001	0	0	0
1.00	.05	.001431	.000839	.001082	.003335	0	0	0	0
1.50	.05	.000096	.000187	.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	.438449	.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	.0078537	.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	.00086	.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

		$n = 5$				$n = 10$			
t_1	t_2	normal	t_3	slash	Cauchy	normal	t_3	slash	Cauchy
0.00	4.00	0	.000985	.001927	.012927	.012603	0	.000063	.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
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	.000010
3.00	15.00	0	0	.000007	.000074	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

Exhibit 4.12
Approximation to the joint density of robust estimates of location and scale using Huber's proposal 2 with $k = 1.5$.

To check the accuracy of the results is difficult since there is no obvious method for computing the exact joint density. The approximation can be checked for the marginal density of T_1 using Monte Carlo results from the Princeton Robustness Study, reported in part by Andrews et al (1971). The complete results have been provided most kindly by F. Hampel. The values of the pseudovariances and n times the variance are reported both for the approximation (obtained by integrating numerically the joint density) and the Monte Carlo results in Exhibit 4.13. 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 the distribution of the estimator T_1 and a standard normal variate respectively.

	$n=5$		$n=10$		$n=20$		$n=40$	
Normal	**A**	**MC**	**A**	**MC**	**A**	**MC**	**A**	**MC**
Pseudo-variances:								
25%	1.0345	1.0412	1.0357	1.0312	1.0368	1.036	1.0380	1.0392
10%	1.0353	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×var	1.0360	1.0427	1.0364	1.0308	1.0369	1.0360	1.0375	1.0384
Slash								
Pseudo-variances:								
25%	1.7610	1.8597	1.6856	1.6599	1.6457	1.6559	1.6284	1.5953
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×var	3.549	3.8752	2.0776	3.5681	1.7419	1.7986	1.6629	1.6246
Cauchy								
Pseudo-variances:								
25%	4.607	3.75	4.590	4.5731 4.7400*	4.907	4.648	4.4852	4.0
10%	7.256	5.4060	5.834	5.8120 6.1554*	5.094	4.8625	4.7554	4.2781
2.5%	17.405	11.590	9.392	9.2350 9.3463*	6.429	5.8338	5.3361	6.6673
1%	30.747	19.2729	13.434	14.6001 12.1464*	7.634	6.8401	5.8013	4.9629
.5%	44.252	26.6878	17.897	21.2734 14.9575*	8.752	7.7168	6.1902	5.2147
.1%	72.847	43.9043	35.365	47.2187 39.2633*	12.45	10.1046	7.2629	6.1221
n×var	16.525	10.9373	9.592	10.2658	6.172	5.6630	5.161	4.5469

t_3		*replication			
Pseudo-variances:					
25%	1.6275	1.6569	1.6556	1.655	1.6522
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	1.8370	1.7789
n×var	1.9953	1.8097	1.7273	1.7132	1.6870

Exhibit 4.13
Pseudovariances and asymptotic variances of T_1 as computed by approximation (A) and Monte Carlo (MC)

In order to determine whether the difference between the Monte Carlo result and the approximation are to within the sampling errors in the Monte Carlo experiment, we can

look at two bits of evidence. In Exhibit 5.13, Andrews et al. (1971) give differences between exact and Monte Carlo results for the pseudo variance of the median, with $n = 5$. With the normal, the differences are in the range of .02 to .03. The differences in Exhibit 4.13 between the approximate and Monte Carlo pseudovariances for the normal, $n = 5$, all are less than .01 indicating differences are well within the errors inherent in the Monte Carlo results. For the Cauchy, the differences observed in Exhibit 4.13 are larger by a factor of up to 10 than those reported for the median (cf. Table 5, Field (1982)). 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 Exhibit 4.13, 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 Exhibit 4.14.

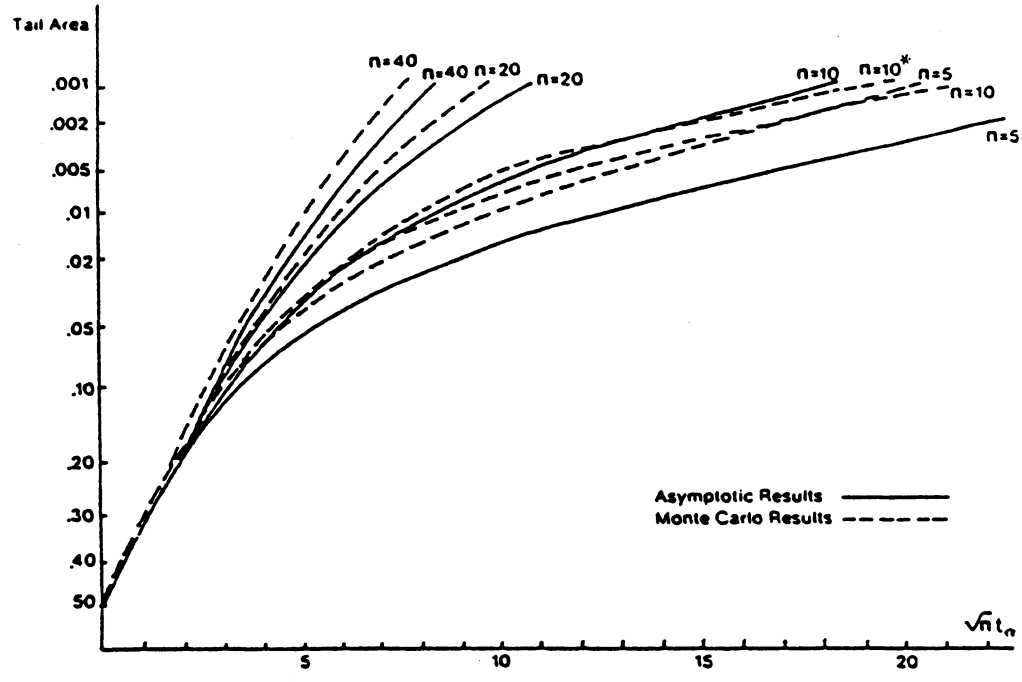

Exhibit 4.14
Plot of percentiles of $\sqrt{n}T_1$ on normal probability paper
for Huber's Proposal 2, $k = 1.5$.

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 Exhibit 4.13. 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 $n = 5$.

Given the approximation to the joint density, it is possible to examine in detail interesting characteristics of the robust estimates of location and scale. We illustrate the potential

with two different computations.

To continue with this example, we consider the percentiles of a "studentized" version of T_1. Since the asymptotic variance T_1 is

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

an appropriate "studentized" version of T_1 would be $n^{1/2} T_1 / (T_{2\gamma})$ with $\gamma = E_\phi \psi_1^2(X) / \left\{ E_\phi \psi_1' \right\}^2$. 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, we could 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.

The percentiles have been evaluated 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 Exhibit 4.15.

	Tail area	normal	t_3	slash	Cauchy
	.25	.808	.831	.837	.871
	.10	1.729	1.657	1.647	1.547
	.05	2.491	2.288	2.297	1.999
$n = 5$	.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
	.25	.732	.759	.759	.807
	.10	1.461	1.467	1.468	1.482
	.05	1.965	1.923	1.925	1.863
$n = 10$	.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
	.25	.701	.728	.724	.775
	.10	1.361	1.393	1.388	1.448
	.05	1.783	1.802	1.797	1.835
$n = 20$	.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

Exhibit 4.15a
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
	.25	.686	.713	.707	.759
	.10	1.318	1.359	1.250	1.430
	.05	1.709	1.750	1.741	1.824
$n = 40$	.025	2.060	2.095	2.086	2.156
	.01	1.477	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
	.25	.679	.705	.698	.750
	.10	1.296	1.341	1.329	1.419
	.05	1.668	1.722	1.708	1.817
$n = 100$	.025	1.993	2.055	2.039	2.158
	.01	2.383	2.445	2.428	2.550
	.005	2.649	2.710	1.693	2.917
	.001	3.213	3.266	3.250	3.355
	.0001	3.922	3.953	3.941	4.005

Exhibit 4.15b
Percentiles of $n^{1/2}T_1/\gamma T_2$ using Huber's Proposal 2 with $k = 1.5$

The first thing to check in Exhibit 4.15 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 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. These results, as they are, are an order of magnitude improvement over results using a classical t-interval.

As a second computation we consider the question of the degree of dependence between T_1 and T_2. For the normal with estimates $\bar{x}$ and s, we have independence and it is interesting to compare the behavior of T_1 and T_2 with this. There is no standard measure of dependence between two random variables. Renyi (1959) has proposed several measures which satisfy most of the properties he feels are natural. We compute two of these measures for the joint distribution of (T_1, T_2). The first of these γ_n is a normalized version of the mean square contingency,

$$C_n = \left(\int \int (k(x,y) - 1)^{-2} dP_{n,1}(x) dP_{n,2}(y) \right)^{1/2}$$

with $k(x,y) = p_n(x,y)/p_{n,1}(x)p_{n,2}(y)$, given as $\Gamma_n = C_n/(1 + C_n^2)^{1/2}$. The second measure L_n is based on information theoretical considerations and can be written as

$$L_n(T_1, T_2) = (1 - \exp(-2I(T_1, T_2))^{1/2}$$

where $I(T_1, T_2) = \int\int k(x,y) \log k(x,y) dP_{n,1}(x) dP_{n,2}(y)$ is the amount of information T_1 contains about T_2.

The calculations have been done for (T_1, T_2) giving the following results in Exhibit 4.16.

	normal		t_3		Cauchy	
n	Γ_n	L_n	Γ_n	L_n	Γ_n	L_n
5	.021	.017	.101	.090	.351	.326
10	.015	.016	.071	.067	.283	.271
15	.010	.010	.056	.054	.241	.233
20	.008	.008	.048	.047	.212	.207
25	.007	.007	.043	.042	.192	.187
30	.007	.007	.039	.038	.176	.171
35	.006	.006	.036	.035	.163	.158
40	.006	.006	.033	.032	.152	.147
45	.005	.005	.031	.031	.143	.137
50	.005	.005	.030	.029	.135	.129

Exhibit 4.16
Dependence measures for (T_1, T_2)

There are several interesting features of these results including the considerable variation in the dependence structure for different underlying densities. It is perhaps surprising to note such differences for rather similar underlying densities. Looking at the results as n increases, the dependence measures seem to be approaching 0 at a rather slow rate. For the case of the normal, it appears that Γ_n is decreasing at a rate of $1/n$ and in fact the relationship $\Gamma_n = 1/(10n)$ gives a good fit to the data.

The purpose in this example has not been to carry out an extensive study of all interesting properties of the robust location/scale but rather to illustrate the potential of the approximation for examining these types of questions.

It is worth noting that if we set $k = \infty$, we obtain the classical estimates of location and scale. If the underlying density is $N(0,1)$, then the equation of Theorem 4.5 can be solved explicitly giving $\alpha_1(\mathbf{t}) = t_1 t_2$ and $\alpha_2(\mathbf{t}) = (t_2^2 - 1)/2$. The conjugate density $h_{\mathbf{t}}(y)$ is normal with mean t_1 and variance t_2^2. It is easy to show that the approximating formula (4.25) becomes

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

This agrees with the 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, emphasizing the need for a numerical rescaling of the approximation.

4.5.c Regression

We now consider modifying the approximation developed in the previous section to the regression case. Let $y_i = \eta_i(\theta) + u_i$ be n iid random variables with scale parameter σ. The estimates $\mathbf{T} = (T_1, \cdots, T_{p+1})$ of (θ, σ) are the values which minimize (Huber, 1981, Ch.14)

$Q(\theta,\sigma) = \sum_{i=1}^{n} \rho((y_i - \eta_i(\theta))/\sigma)\sigma + a\sigma$; $\sigma \geq 0$ or equivalently solve

$$\sum_{i=1}^{n} \psi((y_i - \eta_i(\theta))/\sigma)\partial\eta_i/\partial\theta_j = 0, \quad j = 1,\cdots,p \tag{4.26}$$

$$\sum_{i=1}^{n} \chi((y_i - \eta_i(\theta))/\sigma) - a = 0$$

where $\psi = \rho'$ and $\chi(x) = x\psi(x) - \rho(x)$. Since the observations y_i are not identically distributed, the results of the previous section cannot be applied directly but must be modified. Note that $\rho(x) = x^2/2$ gives the standard least squares estimates. To begin the modifications required, let $\mathbf{t}$ be the point at which the density is to be evaluated, $\Delta_i = y_i - \eta_i(\mathbf{t})$, f_i the density of y_i, where f_i depends on the underlying value of (θ,σ), say $(\theta_{10},\cdots,\theta_{p0},\sigma_0)$, z_j, $j = 1,\cdots,p+1$, the left hand side of the equations in (4.26). To proceed with the centering result (cf. 4.23), let the conjugate density for the i^{th} observation be

$$h_{\mathbf{t}}^i(y) = c^i(\mathbf{t})f_i(y)\exp\left\{\sum_{j=1}^{p} \alpha_j\psi(\Delta_i/t_{p+1})\partial\eta_i/\partial t_j + \alpha_{p+1}\chi(\Delta_i/t_{p+1})\right\}$$

where $c^i(\mathbf{t})$ is the appropriate normalizing constant so that $\int h_{\mathbf{t}}^i(y)dy = 1$. Then it follows that $\phi(\alpha + i\mathbf{y}, i\mathbf{v}) = \prod_{i=1}^{n} c^i(\mathbf{t})M_1(i\mathbf{y}, i\mathbf{v})$ where $M(M_1)$ is the moment generating function of $(z_1,\cdots,z_{p+1},T_1,\cdots,T_{p+1})$ under density $f_i(h_{\mathbf{t}}^i)$, $i = 1,\cdots,n$. From this it follows that $f_n(\mathbf{t}) = \left(\prod_{i=1}^{n} c^i(\mathbf{t})\right)^{-1} h_{\mathbf{t},n}(\mathbf{t})$ where $h_{\mathbf{t},n}$ is the density under the conjugate density $(g_{\mathbf{t}}^1,\cdots,g_{\mathbf{t}}^n)$. The vector α solves the following set of $p+1$ equations in $p+1$ unknowns.

$$\sum_{i=1}^{n} \int \psi((y_i - \eta_i(\mathbf{t}))/t_{p+1})\partial\eta_i(\mathbf{t})/\partial t_j h_{\mathbf{t}}^i(y_i)dy_i = 0, \quad j = 1,\cdots,p$$

$$\sum_{i=1}^{n} \int \left[\chi((y_i - \eta_i(\mathbf{t}))/t_{p+1}) - a\right] h_{\mathbf{t}}^i(y_i)dy_i = 0 \tag{4.27}$$

The arguments expressing $\mathbf{T}_n$ as a mean and obtaining the multivariate expansion given in the previous section go through with minor notational changes. This leads to the following approximating density for $(T_1,\cdots,T_p,T_{p+1})$ where $T_1,\cdots,T_p$ are the estimates of θ and T_{p+1} estimates σ:

$$f_n(\mathbf{t}_0) = (n/2\pi)^{p/2}\left(\prod_{i=1}^{n} c_{\mathbf{t}_0}^i\right)^{-1} |\det A||\det \Sigma|^{-1/2}$$

where

$$A = \left\{E\frac{\partial}{\partial t_\ell} Z_r\Big|_{\mathbf{t}=\mathbf{t}_0}\right\}_{1\leq \ell,r\leq p+1} \tag{4.28}$$

and $\sum = \{EZ_\ell Z_r\}_{1\leq \ell,r\leq p+1}$ with the expectations $E\frac{\partial}{\partial t_\ell} z_r$ to be interpreted as $E\frac{\partial}{\partial t_\ell}\sum_{i=1}^{n} E_{h_{\mathbf{t}}^i}\left[\psi((y_i - f_i(\mathbf{t}))/t_{p+1})\partial f_i/\partial t_r\right]$ if $r \leq p$ and with χ in the square brackets for $r = p+1$. Similar interpretations hold for $EZ_\ell Z_r$.

As a special case, assume $\rho(x) = x^2/2$, $a = n-2$, $\eta_i(\theta) = \theta_1 + \theta_2(x_i - \bar{x})$, u_i's are independent $N(0,\sigma^2)$. The solution of the equations (4.26) yields the least squares estimates for straight line regression. Equations (4.27) become

$$\sum_{i=1}^{n} \int (y_i - t_1 - t_2(x_i - \bar{x}))h_t^i(y_i)dy_i/t_3 = 0$$

$$\sum_{i=1}^{n} \int (y_i - t_1 - t_2(x_i - \bar{x}))(x_i - \bar{x})h_t^i(y_i)dy_i/t_3 = 0$$

$$\sum_{i=1}^{n} \int \left[(y_i - t_1 - t_2(x_i - \bar{x}))^2/t_3^2 - (n-2)\right] h_t^i(y_i)dy_i = 0.$$

It can be seen that these equations will be satisfied if $h_t^i(y)$ is $N(t_1 + t_2(x_i - x), (n-2)t_3^2/n)$. By choosing $\alpha_1(\mathbf{t}) = (t_1 - \theta_1)t_3/\sigma^2$, $\alpha_2(\mathbf{t}) = (t_2 - \theta_2)t_3/\sigma^2$, $\alpha_3(\mathbf{t}) = (t_3^2/2\sigma^2 - n/2(n-2)$, $h_t^i(y)$ is $N(t_1 + t_2(x_i - \bar{x}),\ (n-2)t_3^2/n)$ when $f_i(y)$ is $N(\theta_1 + \theta_2(x_i - x), \sigma^2)$. Evaluating $c^i(\mathbf{t})$, we obtain

$$\prod_{i=1}^{n} c^i(\mathbf{t}) = ((n - 2/n)^{n-2}(t_3/\sigma)^n$$

$$\exp\left\{-n(t_1 - \theta_1)^2/2\sigma^2 - (t_2 - \theta_2)^2 \sum_{i=1}^{n}(x_i - \bar{x})^2/2\sigma^2 - (n-2)t_3^2/2\sigma_0^2 + n^2\right\}.$$

It can be shown that $\det A = (n-2)\sum_{i=1}^{n}(x_i - \bar{x})^2/2t_3^2$, $(\det \Sigma)^{1/2} \propto \left(\sum_{i=1}^{n}(x_i - \bar{x})^2\right)^{1/2}$. The approximating density (4.28) evaluated at $\mathbf{t} = (t_1, t_2, t_3)$ with underlying observations $Y_1, \cdots, Y_n$ where Y_i is $N(\theta_1 + \theta_2(x_i - \bar{x}),\ \sigma^2)$ is:

$$f_n(\mathbf{t}) = (t_3/\sigma)^n$$

$$\exp\{-n(t_1 - \theta_1)^2/2\sigma^2 - (t_2 - \theta_2)^2 \sum_{i=1}^{n}(x_i - \bar{x})^2/2\sigma^2 - (n-2)t_3^2/2\sigma^2\}$$

$$\left(\sum_{i=1}^{n}(x_i - \bar{x})^2\right)^{1/2}/t_3^2$$

up to a constant of integration. This agrees with the exact density up to the constant of integration.

For other choices of ρ, it is not possible to find explicit solutions. Even for the case of linear regression, we have to evaluate f_n over a three-dimensional grid. In many situations we are interested in the marginal density of one of the parameters. It's not usually feasible to evaluate $f_n(\mathbf{t})$ over a grid and then numerically integrate to obtain the marginal density. In section 6.3 we discuss a procedure which provides a one-dimensional technique to construct confidence intervals for a function of the parameters in multiparameter problem.

Spady (1987) computes the saddlepoint approximation to the density of a symmetrically trimmed least squares estimator for the censored regression model with an intercept and one regressor.