NON-PARAMETRIC ESTIMATION OF LOCATION

BY

M. V. JOHNS, JR.

TECHNICAL REPORT NO. 41
SEPTEMBER 16, 1971

PREPARED UNDER THE AUSPICES
OF
NATIONAL SCIENCE FOUNDATION GRANT GP-27550

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA
NON-PARAMETRIC ESTIMATION OF LOCATION

by

M. V. Johns, Jr.

TECHNICAL REPORT NO. 41
September 16, 1971

PREPARED UNDER THE AUSPICES
OF
NATIONAL SCIENCE FOUNDATION GRANT GP-27550

Also partially supported by Office of Naval Research
Contract N00014-67-A-0112-0053

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA
NON-PARAMETRIC ESTIMATION OF LOCATION

by

M. V. Johns, Jr.

1. Introduction.

The problem of "efficiency-robust" estimation of location for symmetric continuous distributions has been treated by a number of authors (as noted in the text below) employing various optimality criteria. The present paper develops a sequence of estimators, indexed by an integer valued parameter \( k \), exhibiting the following rather strong notion of asymptotic "efficiency-robustness": For any \( k \) the corresponding estimator is consistent and asymptotically normally distributed (as the sample size \( n \) increases) for any \( F \) in a large subset \( \mathcal{F} \) of the class of symmetric continuous distributions. Furthermore, for any \( \epsilon > 0 \) the variance of the limiting normal distribution exceeds the Cramér-Rao bound for \( F \) by no more than \( \epsilon \) uniformly for all \( F \in \mathcal{F} \), for each sufficiently large \( k \). Thus, for large \( k \) the corresponding estimator is (nearly) Best Asymptotically Normal (BAN) for all \( F \in \mathcal{F} \).

This concept of optimality is distinguished from some others described below in that the class \( \mathcal{F} \) is "non-parametric" and indeed contains the simple parametric families used in other definitions.

Of greater significance is the fact that the simplest non-trivial estimator in the proposed sequence (corresponding to \( k = 2 \)) exhibits
quite high efficiencies for small to moderate sample sizes \( n = 10, 20, 40 \) for a collection of diverse distributions consisting of the normal, the Cauchy, the logistic, the double exponential, and the 10% contaminated normal. These efficiencies relative to the best linear unbiased estimates (BLUE) based on order statistics were obtained by Monte Carlo experiments. The estimator seems slightly less efficient for short-tailed distributions compared to some competitors designed particularly for use with such distributions, but it does considerably better for the long-tailed Cauchy case. The asymptotic efficiency of the estimator may be computed by hand without difficulty.

This estimator is no more difficult to compute than, say, a one sample \( t \)-statistic, and the calculation yields as a by-product a good estimate of the variance of the location estimator so that approximate confidence intervals and tests may be constructed. The author believes this computational simplicity to be one of the most important features of the proposed estimator since it enhances the possibility that it may be found useful by statistical practitioners dealing with actual data. Previously proposed procedures exhibiting comparable efficiency-robustness require complex arithmetic manipulations and some involve table lookups or matrix inversions. Considerations of simplicity have previously motivated interest in very easily computable estimators such as trimmed or Winsorized means, symmetric linear combinations of three order statistics, etc. (see, e.g. Gastwirth and Cohen [7]). Such estimators do not exhibit as broad a spectrum of efficiency-robustness as those considered in this paper, nor are they associated with natural
non-parametric estimates of their variances.

The various approaches to efficiency-robust estimation may for convenience be classified into two categories: (1) those achieving asymptotic optimality for each member of a specified finite or parametric family of distribution "shapes" (with unspecified scale parameters), and (2) those achieving asymptotic optimality for all members of a non-parametric family of distributions limited only by regularity conditions. Examples of the first category of estimation procedures are given by A. Birnbaum and his associates in [2], [3], [4], J. Gastwirth and H. Rubin [6], R. V. Hogg [8], and V. Miké [10]. In these papers the estimators depend explicitly on the parametric collection of distribution shapes chosen and "optimality" means the asymptotic achievement of the Cramér-Rao lower bound for the variance of the estimator, or the maximization of the minimum variance over the specified family of distribution shapes, or admissibility with respect to the family. In all of these cases it is reasonable to expect (and has to some extent been demonstrated) that the proposed estimators will be robust in the sense of performing well for distributions which do not differ too much from the families in terms of which they are defined. A related approach uses the data to select one estimator from a small specified collection of candidates. Such procedures have been proposed by P. Switzer [12] and L. Jaeckel [9].

The approaches falling into the second category mentioned above are inherently more ambitious in their goals since they aim at asymptotic optimality for all distributions in non-parametric families constrained only by regularity conditions. Estimators are regarded
as optimal if they achieve the Cramer-Rao bound asymptotically or if
they are DAN (or nearly DAN). The possibility of achieving this kind
of uniform efficiency was discussed as early as 1956 by C. Stein [11].
Recent proposals for such uniformly efficient estimators have been
made by P. K. Bhattacharya [1], C. Van Eeden [14] and K. Takeuchi [13].
The first two of these papers suggest estimators involving pointwise
estimates of density functions and are likely to require rather large
samples to be effective. Takeuchi's ingenious sequence of estimators
(indexed by an integer $k$) is not only asymptotically optimal for
large values of $k$, but shows good small sample efficiency-robustness
(for small values of $k$). A natural non-parametric estimate of the
variance of the estimator is also provided.

The estimators proposed in the present paper fall into the second
category described above and (like those of Hogg [6] and Takeuchi) they
are "quasi-linear" in that they are based on linear combinations of the
order statistics with the optimal coefficients being estimated from
the sample. The basic statistic from which the proposed estimators are
developed consists of a linear combination of the order statistics with
the coefficients constrained to be equal within each of several sets
corresponding to blocks of successive order statistics. Thus, the
number of distinct values assumed by the coefficients is some number $k$
which is less than the sample size $n$, and one can hope to use the
full sample to approximate the $k$ optimal values for the coefficients.
This type of statistic does not seem to have been considered before in
the literature although it is similar in spirit to linear combinations
of selected sample quantiles, and such statistics have been discussed
at length in many papers. The proposed estimation scheme and its asymptotic properties are detailed in Section 2. In Section 3 the special case \( k = 2 \) is discussed and some Monte Carlo results are given. The proof of the theorem stated in Section 2 is outlined in the Appendix.

2. The Proposed Estimator.

Let \( X_1, X_2, \ldots, X_n \) represent independent identically distributed observations with common distribution function \( F_\theta(x) = F(x-\theta) \) where \( F \) is a member of a family \( \mathcal{F} \) of symmetric continuous distributions satisfying certain regularity conditions described below. Thus, \( \theta \) represents the median of \( F_\theta \). Let \( Y_1 \leq Y_2 \leq \cdots \leq Y_n \) represent the ordered \( X_i \)'s (i.e., the "order statistics"). Suppose that \( n \) is even and let \( T_{0n}, T_{1n}, \ldots, T_{kn} \) be a partition of the integers \( 1, 2, \ldots, n/2 \), where \( T_{in} \) consists of \( t_{in} \) successive integers, so that \( \sum_{i=0}^{k} t_{in} = n/2 \).

We first consider estimators of \( \theta \) of the form

\[
\theta_{kn} = \frac{1}{n} \sum_{i=1}^{k} c_{in} S_{in}, \quad \text{where}
\]

\[
S_{in} = \sum_{j \in T_{in}} (Y_j + Y_{n-j+1}), \quad i = 1, \ldots, k,
\]

and the \( c_{in} \)'s are constants satisfying

\[
\sum_{i=1}^{k} c_{in} t_{in} = \frac{n}{2}.
\]

Condition (2) insures that \( \theta_{kn} \) is a location invariant estimator. Note that the \( t_{0n} \) smallest and largest order statistics have been "trimmed"
from the estimator \( \hat{\theta}_{kn} \). This permits the development of substantially simpler formulae than would otherwise be possible. Suppose further that
\[
c_i \xrightarrow{n} c_i, \quad i = 1, 2, \ldots, k,
\]
and
\[
t_i \xrightarrow{\text{n}} p_i, \quad i = 0, 1, 2, \ldots, k, \quad \text{as} \quad n \to \infty,
\]
where for each \( i \), \( p_i > 0 \) and \( \sum_{i=0}^{k} p_i = \frac{1}{2} \). Then we must also have
from (2)
\[
\sum_{i=1}^{k} c_i p_i = cp' = \frac{1}{2},
\]
where \( c = (c_1, c_2, \ldots, c_k) \) and \( p = (p_1, p_2, \ldots, p_k) \). Under these assumptions and the regularity conditions on \( F \) given below, for fixed \( k \) the sequence \( \sqrt{n} (\hat{\theta}_{kn} - \theta) \) is asymptotically normally distributed with mean zero and variance \( \sigma_k^2 \) given by
\[
\sigma_k^2 = 2 \sum_{i=0}^{k} \sum_{j=0}^{k} c_i c_j a_{ij} = 2cAc',
\]
where \( A = ((a_{ij}))_{k \times k} \) and the \( a_{ij} \)'s are defined as follows: Let \( \xi_{i}\) be the \( (p_0 + p_1 + \cdots + p_i)^{th} \) quantile of \( F \), i.e., \( F(\xi_{i}) = p_0 + p_1 + \cdots + p_{i-1} \) for \( i = 1, 2, \ldots, k-1 \), and \( \xi_k = 0 \). Then for \( i = 1, 2, \ldots, k \)
\[
a_{ii} = 2\xi_i \int_{\xi_{i-1}}^{\xi_i} F(x) dx - 2 \int_{\xi_{i-1}}^{\xi_i} xF(x) dx,
\]
and
\[
a_{ij} = a_{ji} = (\xi_j - \xi_{j-1}) \int_{\xi_{i-1}}^{\xi_i} F(x) dx, \quad i < j.
\]
The asymptotic normality of $\theta_{kn}$ and the expression for $\sigma_k^2$ follow directly from Theorem 3, page 63, of Chernoff et al. [5].

At this point it is clear that since the $a_{ij}$'s are invariant under shifts, one could estimate them by replacing the $\xi_i$'s by their sample analogs and $F(x)$ by the sample c.d.f. $F_n(x)$. One could then choose the $c_i$'s so as to minimize (4) using the estimated $a_{ij}$'s and use the resulting $c_i$'s in $\theta_{kn}$. Such estimators would doubtless be (nearly) BAN for all $F \in \mathcal{F}$ for sufficiently large $k$ and small $p_0$ (the trimming proportion). The computation of such estimates would, however, be quite complicated and the minimization process would require the inversion of a $k \times k$ matrix. We proceed, therefore, to introduce some simplifying approximations.

For $i = 1, 2, \ldots, k$ let

(6) \quad d_i = \xi_i - \xi_{i-1}, \quad \text{and} \quad b_i = \frac{1}{2} p_j + \frac{1}{2} p_1.

Let

(7) \quad B = \begin{pmatrix}
  b_1, b_1, \ldots, b_1 \\
  b_1, b_2, \ldots, b_2 \\
  \vdots & \ddots & \ddots \\
  b_1, b_2, \ldots, b_k
\end{pmatrix},

and

(8) \quad R = \left( (a_{ij}/d_i d_j)_{1 \times k} - B \right).
where the $a_{ij}$'s are given by (5).

Then if $\tilde{c} = (\tilde{c}_1, \tilde{c}_2, \cdots, \tilde{c}_k) = (c_1 d_1, c_2 d_2, \cdots, c_k d_k)$, and

$\tilde{p} = (p_1/d_1, p_2/d_2, \cdots, p_k/d_k)$, we have from (4), (7) and (8)

$$\sigma_k^2 = \tilde{c} (B^T R \tilde{c})'$$

and condition (3) becomes

$$\tilde{c} \tilde{p}' = \frac{1}{2}.$$

It is shown in the Appendix that for the cases of interest the contribution of $\tilde{c} B \tilde{c}'$ to $\sigma_k^2$ is negligible for large $k$. Hence we may minimize $\sigma_k^2$ approximately by minimizing $\tilde{c} B \tilde{c}'$ subject to condition (10).

A matrix of the form (7) is non-singular if the $b_i$'s are distinct, and a straightforward Lagrange multiplier argument establishes that the vector $\tilde{c}$ minimizing $\tilde{c} B \tilde{c}'$ subject to (10) is given by $\tilde{c} = \frac{1}{2}(\tilde{p} B^{-1} \tilde{p}')^{-1} \tilde{p}'$, and the corresponding approximate minimum of $\sigma_k^2$, denoted by $\tilde{\sigma}_k^2$, is given by $\tilde{\sigma}_k^2 = 2\tilde{c} B \tilde{c}' = \frac{1}{2}(\tilde{p} B^{-1} \tilde{p}')^{-1}$. Now letting $e = (e_1, e_2, \cdots, e_k)$ where $e_i = d_i^{-1} x$ ($i^{th}$ component of $B^{-1} \tilde{p}'$), we may write

$$\tilde{\sigma}_k^2 = \frac{1}{2} (e p')^{-1},$$

and recalling that $c_i = \tilde{c}_i / d_i$, the (approximately) minimizing vector $c$ is

$$c = \tilde{\sigma}_k^2 e.$$
The matrix $B$ given by (7) is a Green's matrix and hence readily invertible yielding the Jacobi matrix $B^{-1} = ((b^{ij}))_{k \times k}$ with

$$b^{i,i+1} = \frac{-2}{p_i + p_{i+1}}, \quad i = 1, 2, \ldots, k-1,$$

$$b^{11} = \frac{4p_0 + 4p_1 + 2p_2}{(2p_0 + p_1)(p_1 + p_2)},$$

$$b^{ii} = \frac{2p_{i-1} + 4p_i + 2p_{i+1}}{(p_{i-1} + p_1)(p_i + p_{i+1})}, \quad i = 2, 3, \ldots, k-1,$$

$$b^{kk} = \frac{2}{p_{k-1} + p_k},$$

and all other $b^{ij} = 0$. Thus the components of the vector $e$ are

$$e_1 = \left(\frac{2p_1(2p_0 + 2p_1 + p_2)}{(2p_0 + p_1)(p_1 + p_2)}\right) \frac{1}{d_1^2} - \left(\frac{2p_2}{p_1 + p_2}\right) \frac{1}{d_1 d_2},$$

$$(14) \quad e_i = \frac{1}{d_i} \left[\left(\frac{2p_i(p_{i-1} + 2p_i + p_{i+1})}{(p_{i-1} + p_i)(p_i + p_{i+1})}\right) \frac{1}{d_i^2} - \left(\frac{2p_{i-1}}{p_{i-1} + p_i}\right) \frac{1}{d_i d_{i-1}} - \left(\frac{2p_{i+1}}{p_i + p_{i+1}}\right) \frac{1}{d_i d_{i+1}}\right],$$

$$i = 2, 3, \ldots, k-1,$$

$$e_k = \left(\frac{2p_k}{p_{k-1} + p_k}\right) \frac{1}{d_k^2} - \left(\frac{2p_{k-1}}{p_{k-1} + p_k}\right) \frac{1}{d_k d_{k-1}}.$$

Note that for the special case $p_1 = p_2 = \cdots = p_k$, the formula for $e_i$ for $i = 2, 3, \ldots, k-1$ simplifies to

$$e_i = \frac{1}{d_i} \left(\frac{2}{d_i^2} - \frac{1}{d_{i-1}} - \frac{1}{d_{i+1}}\right).$$

(15)
The only quantities appearing in (14) which depend on the distribution $F$ are the $d_i$'s which are differences of quantiles which may be effectively estimated by their sample analogs. For each $i$ let $\hat{\xi}_i$ be the sample quantile corresponding to $\xi_i$ and let $\hat{d}_i = \hat{\xi}_i - \hat{\xi}_{i-1}$.

(More explicit formulas for the $\hat{d}_i$'s are offered in Section 3). Let $\hat{e}$ be the vector whose components $\hat{e}_i$, $i = 1, 2, \ldots, k$, are obtained from (15) by replacing the $d_i$'s by $\hat{d}_i$'s. Then following (11) and (12) let

\[
\hat{\sigma}_{kn}^2 = \frac{1}{\hat{e}(\hat{e}^p)^{-1}}, \text{ and}
\]

\[
\hat{c} = \hat{\sigma}_{kn}^2 \hat{e}.
\]

The proposed estimator of $\theta$ (by analogy with (1)) is

\[
\hat{\theta}_{kn} = \frac{1}{n} \sum_{i=1}^{k} \hat{c}_i S_{i, n-1},
\]

where $\hat{c}_i$ is the $i^{th}$ component of $\hat{c}$.

It has been assumed so far that $n$ is even. The case of $n$ odd may be treated in various ways, perhaps the simplest of which is to delete the sample median from the order statistics and compute $\hat{c}$ on the basis of the even sample size $n-1$. The coefficient $\hat{c}_k$ should then be modified by multiplying by $t_{k, n-1}/(t_{k, n-1} + 1)$ and the sample median should be added to $S_{k, n-1}$. Expression (18) may then be used with the $S_{i, n-1}$'s replacing the $S_{i, n}$'s.
To facilitate the discussion of the asymptotic behavior of \( \hat{\theta}_{kn} \) we shall assume that for each \( k \),

\[
q_1 = q_2 = \cdots = q_k = k^{-1} \left( \frac{1}{2} - p_0 \right) = q_k,
\]

i.e., each \( S_{in} \) for \( i \geq 1 \) is a sum of (approximately) \( \sum q_k \) successive order statistics.

For any continuous, twice differentiable symmetric distribution \( F(x) \) with density \( f(x) \), the Fisher information for a location parameter when the \( np_0 \) smallest and \( np_0 \) largest order statistics have been deleted ("trimmed") from a sample of size \( n \) is given by

\[
I(F, p_0) = 2 \int_{-\infty}^{0} (f'(x))^2 (f(x))^{-1} dx + 2 f^2(\xi_0) p_0^{-1},
\]

where, as before, \( F(\xi_0) = p_0 \) (see, e.g. [5]). The untrimmed (full sample) case is obtained by setting \( p_0 = 0 \) (\( \xi_0 = \infty \)), and eliminating the second term on the right of (20). The Cramér-Rao lower bound for the variance of any estimator of location based on a trimmed sample of size \( n \) is then \( n I(F, p_0)^{-1} \).

We now define more precisely the family of distributions for which the proposed estimator is shown to be asymptotically (nearly) efficient. **Definition.** Let \( \mathcal{F} \) be a family of symmetric continuous distribution functions \( F(x) \) with corresponding density functions \( f(x) = F'(x) \) such that
(a) \( f(x) \) and its first three derivatives exist and are continuous with \( f'(x) > 0 \) for all \( x \), and \( xf'(x) \to 0 \) as \( x \to \pm \infty \),

(b) \( f(x) \) and its first three derivatives are uniformly bounded for all \( F \in \mathcal{F} \),

(c) \( I(F,0) \) given by (20) is finite and uniformly bounded away from zero for all \( F \in \mathcal{F} \), and

(d) for each \( F \in \mathcal{F} \) the conditions of Theorem 3 of [5] are satisfied for estimators of the form (1).

The asymptotic behavior of the proposed estimator is summarized in the following

**Theorem:** If \( \hat{\theta}_{kn} \) and \( \hat{\sigma}_{kn}^2 \) are given by (13) and (16) respectively and (19) is satisfied with the trimming proportion \( p_0 \) held fixed, then for any \( \epsilon > 0 \) there exists a \( k_\epsilon \) such that for each \( k > k_\epsilon \), for all \( F \in \mathcal{F} \), the sequence \( \sqrt{n}(\hat{\theta}_{kn} - \theta_n) \) is asymptotically normally distributed with mean zero and variance \( r_k^2(F) \), as \( n \to \infty \), where

\[
(21) \quad r_k^2(F) \leq \frac{1}{I(F,p_0)} + \epsilon
\]

and furthermore, with probability approaching one as \( n \to \infty \),

\[
(22) \quad |\hat{\theta}_{kn}^2 - r_k^2(F)| < \epsilon.
\]

The proof of the theorem is deferred to the Appendix.
If the trimming proportion \( p_0 \) is allowed to decrease we have the following immediate

**Corollary:** Under the conditions of the theorem, for any \( \epsilon > 0 \), there exist a \( k_\epsilon \) and \( p_\epsilon \) such that if the trimming proportion \( p_0 = p_\epsilon \), then for each \( k > k_\epsilon \), for all \( F \in \mathcal{F} \) the conclusions of the theorem hold with \( I(F,p_0) \) replaced by \( I(F,0) \) in (21).

Thus, for a suitable choice of \( k \) and \( p_0 \) one can make \( \hat{\sigma}_{k,n} \) as near to being BAN as desired, uniformly for all \( F \in \mathcal{F} \).

The actual family \( \mathcal{F} \) for which the results of the theorem and its corollary hold is larger than that described in the above definition. It is shown in the Appendix that these results hold for the double exponential distribution (which is not in \( \mathcal{F} \) as defined), and by implication, for other distributions having a finite number of simple discontinuities in \( f \) and \( f' \).

One can conclude from the theorem that for any particular \( F \in \mathcal{F} \), there exist sequences \( k_n \uparrow \infty \) and \( p_{0n} \downarrow 0 \), as \( n \to \infty \), such that the corresponding sequence of estimators \( \hat{\sigma}_{k_n,n} \) (using trimming proportions \( p_{0n} \)) is BAN for \( F \). Also \( \hat{\sigma}_{k_n,n}^2 \) is a consistent estimator of \( (I(F,0))^{-1} \).

In order to assert the existence of a single pair of such sequences yielding a sequence of estimators which is BAN for all \( F \in \mathcal{F} \) it is necessary to further restrict \( \mathcal{F} \) so as to insure uniform convergence to normality in the theorem cited from [5]. Questions concerning the rate behavior of the \( k_n \) and \( p_{0n} \) sequences necessary for this uniform BAN property can also be raised. These questions seem somewhat uninteresting in view of the results of the next section wherein it is shown that very small values of \( k \) suffice to produce good asymptotic efficiencies as
well as good small sample performance.

3. The Cases \( k = 2 \) and \( k = 3 \). Asymptotic efficiency-robustness is only a theoretical curiosity unless it is accompanied by satisfactory performance for small and moderate sample sizes. The estimator \( \hat{\theta}_{kn} \) given by (18) reduces to the familiar trimmed mean for the case \( k = 1 \), so we concentrate on the next simplest cases where \( k = 2 \) or \( 3 \), with particular emphasis on \( k = 2 \).

It is convenient to introduce certain notational simplifications appropriate to the case of small \( k \) and \( n \). Referring to the quantities involved in (1), let \( r = t_{0n} \), \( s = t_{ln} = t_{2n} = \ldots = t_{k-1,n} \), and \( t = t_{kn} = \frac{n}{2} - r - (k-1)s \). Thus, for \( k = 2 \),

\[
S_{0,n} = \sum_{j=1}^{r} (Y_j + Y_{n+1-j}) ,
\]

\[
S_{1,n} = \sum_{j=r+1}^{r+s} (Y_j + Y_{n+1-j}) ,
\]

\[
S_{2,n} = \sum_{j=r+s+1}^{n/2} (Y_j + Y_{n+1-j}) .
\]

(23)

We now present explicit formulas for the estimates \( \hat{d}_1 \) of \( d_1 = \xi_i - \xi_{i-1} \). Any suitably defined sample analogs of \( \xi_i \) and \( \xi_{i-1} \) will suffice to define \( \hat{d}_1 \), but since \( \xi_i \) represents the quantile which is the upper endpoint of the quantile range corresponding to the sample quantiles contributing to \( S_{in} \) and the lower endpoint of the corresponding range for \( S_{i+1,n} \), it seems reasonable to choose for \( \hat{d}_1 \) the midpoint between the largest order statistic contributing to \( S_{in} \) and
the smallest contributing to $S_{i+1,n}$. This leads to estimates $\hat{d}_i$ given by

$$
\hat{d}_i = \frac{1}{4} \left( Y_{r+i}s + Y_{r+i}s+1 + Y_{n-r-(i-1)s} + Y_{n-r-(i-1)s+1} \right)
$$

(24) 

$$
\hat{d}_k = \frac{1}{4} \left( Y_{(n/2)+t} + Y_{(n/2)+1} + Y_{(n/2)-t} + Y_{(n/2)-1} \right).
$$

Thus, for $k = 2$ and even $n$, noting that $p_0 = r/n$, $p_1 = s/n$, and $p_2 = t/n$, we have from (14) and the definition of $\hat{e}$,

$$
\hat{e}_1 = \frac{2s(2r+s+t)}{(2r+s)(s+t)} \frac{1}{\hat{d}_1^2} - \frac{2t}{s+t} \frac{1}{\hat{d}_1^2} 
$$

(25)

$$
\hat{e}_2 = \frac{2t}{s+t} \frac{1}{\hat{d}_2^2} - \frac{2s}{s+t} \frac{1}{\hat{d}_2^2},
$$

and from (16), (17), and (18),

$$
\hat{e}_{2,n}^2 = \frac{n}{2(s\hat{e}_1^2 + t\hat{e}_2^2)}, \text{ and }
$$

(26)

$$
\hat{e}_{2,n} = \frac{\hat{e}_1 + \hat{e}_2}{2(s\hat{e}_1^2 + t\hat{e}_2^2)}.
$$

(27)

Similar simple formulas result for the case $k = 3$.

For given values of $p_0$, $p_1$, and $p_2$ and any particular distribution $F$, the asymptotic variance of $\hat{d}_{2,n}$ may be calculated by first calculating the four $a_{i,j}$'s given by (5) and $c = (c_1, c_2)$ given by (12).
and then applying (4). Asymptotic variances for five cases were computed and the efficiencies relative to \((I(F,0))^{-1}\) are shown below.

<table>
<thead>
<tr>
<th>Distribution</th>
<th>Normal</th>
<th>C.N.(.10)</th>
<th>Cauchy</th>
<th>Dble. Exp.</th>
<th>Logistic</th>
</tr>
</thead>
<tbody>
<tr>
<td>Efficiency (%)</td>
<td>94.4</td>
<td>93.6</td>
<td>97.2</td>
<td>84.5</td>
<td>97.5</td>
</tr>
</tbody>
</table>

Table 1

Asymptotic Efficiencies for \(\hat{\theta}_{2,n}\) when \(p_0 = .05, p_1 = p_2 = .225\)

A Monte Carlo experiment was performed for the cases \(k = 2\) and \(3\), using 4000 replications, for the normal, contaminated normal (C.N.(.10), Cauchy, double exponential, and logistic distributions. (The C.N.(.10) distribution is a standard normal with a 10% contamination by a normal with mean zero and variance = 9.) Sample sizes \(n = 10, 20, 40,\) and 80 were used, and the trimming proportion \(p_0\) (from each tail) was taken to be .05 for all cases except \(n = 10\) for which \(p_0 = .10\). For the case \(k = 2\), for \(n > 10\), the quantities \(s\) and \(t\) were chosen in three ways: (i) \(2s = t\), (ii) \(s = 2t\), and (iii) \(s = t\) (subject to the limitation that \(s\) and \(t\) must be positive integers with \(s + t = \frac{n}{2} - r\)). For \(n = 10\), \(s\) and \(t\) were set equal to two. The case \(k = 3\) was computed for \(n = 20, 40,\) and 80 with \(p_0 = .05\) and \(s = t\).

Table 2 shows the mean squared errors (MSE) and relative efficiencies of \(\hat{\theta}_{2,n}\) and the means of \(\hat{\theta}_{2,n}^2\) for the case \(k = 2, s \approx t\). Ideally the small sample efficiencies should be computed relative to the variances of the minimum variance location invariant (Pitman) estimators, but these were only available for the double exponential distribution (from
<table>
<thead>
<tr>
<th>Sample Size</th>
<th>Distribution</th>
<th>$\text{Var(BLUE)}$</th>
<th>$\text{MSE}(\hat{\theta}_{2,n})$</th>
<th>$\frac{\text{Var}(\hat{\theta}<em>{2,n})}{\hat{\theta}</em>{2,n}}$</th>
<th>Eff.($\hat{\theta}_{2,n}$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$n = 10$</td>
<td>Normal</td>
<td>$.1000$</td>
<td>$.1184$</td>
<td>$.0892$</td>
<td>84%</td>
</tr>
<tr>
<td></td>
<td>C.N.(.10)</td>
<td>$.1358$</td>
<td>$.1488$</td>
<td>$.1245$</td>
<td>91</td>
</tr>
<tr>
<td></td>
<td>Cauchy</td>
<td>$.3261$</td>
<td>$.4229$</td>
<td>$.4841$</td>
<td>77</td>
</tr>
<tr>
<td></td>
<td>Dble. Exp.</td>
<td>$.1399$</td>
<td>$.1645$</td>
<td>$.1399$</td>
<td>85</td>
</tr>
<tr>
<td></td>
<td>Logistic</td>
<td>$.3073$</td>
<td>$.3446$</td>
<td>$.2750$</td>
<td>89</td>
</tr>
<tr>
<td>$n = 20$</td>
<td>Normal</td>
<td>$.0500$</td>
<td>$.0579$</td>
<td>$.0525$</td>
<td>86</td>
</tr>
<tr>
<td></td>
<td>C.N.(.10)</td>
<td>$.0657$</td>
<td>$.0710$</td>
<td>$.0712$</td>
<td>93</td>
</tr>
<tr>
<td></td>
<td>Cauchy</td>
<td>$.1256$</td>
<td>$.1304$</td>
<td>$.1429$</td>
<td>96</td>
</tr>
<tr>
<td></td>
<td>Dble. Exp.</td>
<td>$.0637$</td>
<td>$.0703$</td>
<td>$.0707$</td>
<td>91</td>
</tr>
<tr>
<td></td>
<td>Logistic</td>
<td>$.1520$</td>
<td>$.1668$</td>
<td>$.1573$</td>
<td>91</td>
</tr>
<tr>
<td>$n = 40$</td>
<td>Normal</td>
<td>$.0250$</td>
<td>$.0284$</td>
<td>$.0281$</td>
<td>88</td>
</tr>
<tr>
<td></td>
<td>C.N.(.10)</td>
<td>$.0315^*$</td>
<td>$.0354$</td>
<td>$.0564$</td>
<td>89</td>
</tr>
<tr>
<td></td>
<td>Cauchy</td>
<td>$.0500^*$</td>
<td>$.0589$</td>
<td>$.0646$</td>
<td>85</td>
</tr>
<tr>
<td></td>
<td>Dble. Exp.</td>
<td>$.0297^{**}$</td>
<td>$.0333$</td>
<td>$.0353$</td>
<td>89</td>
</tr>
<tr>
<td></td>
<td>Logistic</td>
<td>$.0750^*$</td>
<td>$.0817$</td>
<td>$.0822$</td>
<td>92</td>
</tr>
<tr>
<td>$n = 80$</td>
<td>Normal</td>
<td>$.0125$</td>
<td>$.0137$</td>
<td>$.0147$</td>
<td>91</td>
</tr>
<tr>
<td></td>
<td>C.N.(.10)</td>
<td>$.0157^*$</td>
<td>$.0170$</td>
<td>$.0186$</td>
<td>92</td>
</tr>
<tr>
<td></td>
<td>Cauchy</td>
<td>$.0250^*$</td>
<td>$.0269$</td>
<td>$.0500$</td>
<td>93</td>
</tr>
<tr>
<td></td>
<td>Dble. Exp.</td>
<td>$.0142^{**}$</td>
<td>$.0155$</td>
<td>$.0174$</td>
<td>92</td>
</tr>
<tr>
<td></td>
<td>Logistic</td>
<td>$.0375$</td>
<td>$.0393$</td>
<td>$.0424$</td>
<td>95</td>
</tr>
</tbody>
</table>

* Cramer-Rao bound  ** Pitman estimator.
Table 3
Confidence Interval Error Probabilities

\[ P\left( |\hat{\theta}_n| > \gamma \frac{\sigma_{\theta}}{\sqrt{n}} \right), \ (k=2, \ s \approx t) \]

<table>
<thead>
<tr>
<th>Distribution</th>
<th>Sample Size(n)</th>
<th>(\gamma = 1.645)</th>
<th>(\gamma = 1.960)</th>
<th>(\gamma = 2.576)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Normal</td>
<td>10</td>
<td>.208</td>
<td>.143</td>
<td>.074</td>
</tr>
<tr>
<td></td>
<td>20</td>
<td>.145</td>
<td>.087</td>
<td>.039</td>
</tr>
<tr>
<td></td>
<td>40</td>
<td>.119</td>
<td>.066</td>
<td>.020</td>
</tr>
<tr>
<td></td>
<td>80</td>
<td>.095</td>
<td>.048</td>
<td>.011</td>
</tr>
<tr>
<td>C.N. (.10)</td>
<td>10</td>
<td>.196</td>
<td>.136</td>
<td>.068</td>
</tr>
<tr>
<td></td>
<td>20</td>
<td>.135</td>
<td>.083</td>
<td>.036</td>
</tr>
<tr>
<td></td>
<td>40</td>
<td>.117</td>
<td>.063</td>
<td>.021</td>
</tr>
<tr>
<td></td>
<td>80</td>
<td>.094</td>
<td>.050</td>
<td>.011</td>
</tr>
<tr>
<td>Cauchy</td>
<td>10</td>
<td>.148</td>
<td>.100</td>
<td>.047</td>
</tr>
<tr>
<td></td>
<td>20</td>
<td>.124</td>
<td>.076</td>
<td>.030</td>
</tr>
<tr>
<td></td>
<td>40</td>
<td>.112</td>
<td>.057</td>
<td>.017</td>
</tr>
<tr>
<td></td>
<td>80</td>
<td>.095</td>
<td>.048</td>
<td>.011</td>
</tr>
<tr>
<td>Dble. Exp.</td>
<td>10</td>
<td>.176</td>
<td>.120</td>
<td>.058</td>
</tr>
<tr>
<td></td>
<td>20</td>
<td>.128</td>
<td>.074</td>
<td>.028</td>
</tr>
<tr>
<td></td>
<td>40</td>
<td>.111</td>
<td>.061</td>
<td>.015</td>
</tr>
<tr>
<td></td>
<td>80</td>
<td>.092</td>
<td>.044</td>
<td>.090</td>
</tr>
<tr>
<td>Logistic</td>
<td>10</td>
<td>.200</td>
<td>.138</td>
<td>.070</td>
</tr>
<tr>
<td></td>
<td>20</td>
<td>.140</td>
<td>.088</td>
<td>.058</td>
</tr>
<tr>
<td></td>
<td>40</td>
<td>.117</td>
<td>.065</td>
<td>.019</td>
</tr>
<tr>
<td></td>
<td>80</td>
<td>.094</td>
<td>.048</td>
<td>.011</td>
</tr>
</tbody>
</table>

"t"-Dist.

| \(P(|t_{n-1}| > \gamma)\) | \(t\)       | \(s\)       | \(n\)       | \(s\)       |
|--------------------------|-------------|-------------|-------------|-------------|
| 10                       | .135        | .082        | .030        |
| 20                       | .117        | .065        | .019        |
| 40                       | .108        | .057        | .014        |
| 80                       | .104        | .054        | .012        |

Normal

| \(P(\left|\frac{\bar{X}_n}{\sigma}\right| > \gamma)\) | all \(n\) | \(s\)       | \(n\)       |
|--------------------------------------------------------|-------------|-------------|-------------|
| .100                                                   | .050        | .010        |
Table 4
"Best Case" Efficiencies (k = 2)

<table>
<thead>
<tr>
<th>Distribution</th>
<th>Sample Size</th>
<th>n=20</th>
<th>n=40</th>
<th>n=80</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Best(s,t)</td>
<td>Eff.($\hat{\theta}_{2,n}$)</td>
<td>Best(s,t)</td>
<td>Eff.($\hat{\theta}_{2,n}$)</td>
</tr>
<tr>
<td>Normal</td>
<td>(3,6)</td>
<td>90%</td>
<td>(6,12)</td>
<td>90%</td>
</tr>
<tr>
<td>C.N.(.10)</td>
<td>(3,6)</td>
<td>95</td>
<td>(6,12)</td>
<td>92</td>
</tr>
<tr>
<td>Cauchy</td>
<td>(6,3)</td>
<td>100</td>
<td>(12,6)</td>
<td>87</td>
</tr>
<tr>
<td>Dble. Exp.</td>
<td>(6,3)</td>
<td>92</td>
<td>(12,6)</td>
<td>93</td>
</tr>
<tr>
<td>Logistic</td>
<td>(4,5)</td>
<td>92</td>
<td>(9,9)</td>
<td>92</td>
</tr>
</tbody>
</table>

Table 5
Ratio of Variance for k = 3 to Variance for k = 2
for the Case s ≈ t

<table>
<thead>
<tr>
<th>Distribution</th>
<th>Sample Size</th>
<th>n=20</th>
<th>n=40</th>
<th>n=80</th>
</tr>
</thead>
<tbody>
<tr>
<td>Normal</td>
<td></td>
<td>1.03</td>
<td>1.07</td>
<td>1.05</td>
</tr>
<tr>
<td>C.N.(.10)</td>
<td></td>
<td>1.04</td>
<td>1.05</td>
<td>1.05</td>
</tr>
<tr>
<td>Cauchy</td>
<td></td>
<td>1.19</td>
<td>1.11</td>
<td>1.08</td>
</tr>
<tr>
<td>Dble. Exp.</td>
<td></td>
<td>1.09</td>
<td>1.08</td>
<td>1.06</td>
</tr>
<tr>
<td>Logistic</td>
<td></td>
<td>1.03</td>
<td>1.07</td>
<td>1.06</td>
</tr>
</tbody>
</table>
V. Miké [10]). For the other distributions the variances of the BLUE's were used when available, and for the larger sample sizes the Cramér-Rao bounds were used which accounts for the apparent slight drop in several of the efficiencies at \( n=40 \) compared to those at \( n=20 \). The fact that the efficiencies shown for the double exponential case are relatively high compared to the asymptotic value given in Table 1 is a consequence of the rather slow rate at which the variance of the Pitman estimator approaches its asymptotic value as the sample size increases.

Table 3 shows the estimated error (non-coverage) probabilities for confidence intervals of the form \( \hat{\theta}_{2,n} \pm \gamma n^{-1/2} \hat{\sigma}_{2,n} \). The values of \( \gamma \) were taken (arbitrarily) to be percentage points of the standard normal distribution. For comparison purposes the error probabilities are given for each \( n \) for the corresponding confidence intervals based on the t-distribution (assuming normally distributed observations). The distribution of the "studentized" version of \( \hat{\theta}_{2,n} \) shows a somewhat greater spread than the corresponding t-distribution for the smaller sample sizes. For each pair of values of \( n \) and \( \gamma \) the error probabilities are rather similar for all five of the distributions considered, and they do not differ much from those for the corresponding t-distribution for \( n=40 \) and 80.

Table 4 shows for each distribution the most favorable pair of values of \( s \) and \( t \) (among the three possibilities tested for each \( n \)) together with the corresponding efficiencies for \( n = 20, 40 \) and 80. For all
three sample sizes the large values of \( t \), leading to large groups of 
central order statistics receiving weight \( \hat{c}_2 \), tend to be favorable for 
the relatively short-tailed distributions (normal, contaminated normal, 
and logistic), while the reverse is true for the long-tailed cases 
(Cauchy and double exponential). This provides considerable justifica-
tion for the choice \( s = t \) as a suitable compromise.

Table 5 shows the ratio of the variance of \( \hat{\Theta}_{3,n} \) to the variance 
of \( \hat{\Theta}_{2,n} \) when \( s = t \), for \( n = 20, 40, \) and \( 80 \), for each of the five 
distributions. The uniform superiority of the estimator using \( k = 2 \) 
must be attributed to the additional sampling fluctuation introduced 
when an additional coefficient is estimated in the case \( k = 3 \), together 
with the fact that the efficiencies are already high for the case \( k = 2 \). 
In a preliminary Monte Carlo investigation, larger values of \( k \) were 
considered and in all cases the variances of the estimators for a given 
sample size tended to increase with increasing \( k \).

In interpreting the Monte Carlo results given above it should be 
borne in mind that the standard deviation associated with each of the 
efficiencies is on the order of two percentage points. It is perhaps 
worth mentioning that the ordinary congruence-type pseudo-random numbers 
proved to be completely inadequate for this investigation, and it was 
found necessary to introduce a suitable "re-randomization" procedure.

The possibility exists that the performance of the estimator \( \hat{\Theta}_{2,n} \) 
could be slightly improved by a minor modification as indicated in the 
following 

Remark: The quantities \( \hat{c}_1 \) and \( \hat{c}_2 \) appearing in the definition (27)
of $\hat{\theta}_{2,n}$ are of the form

$$\hat{e}_1 = \alpha_1 \hat{d}_1^{-2} - \alpha_2 \left( \hat{d}_1 \hat{d}_2 \right)^{-1},$$

$$\hat{e}_2 = \beta_1 \hat{d}_2^{-2} - \beta_2 \left( \hat{d}_1 \hat{d}_2 \right)^{-1},$$

where $\alpha_i, \beta_i, i = 1,2$, are determined by (25). Since the formulae given in (25) were obtained by an asymptotic argument valid for large $k$, it is remarkable that they work well for $k = 2$. It seems entirely possible that a somewhat different choice for the $\alpha_i$'s and $\beta_i$'s could result in uniformly improved performance. The author conjectures that a slight modification of $\alpha_1$ (only) could lead to improved efficiency for short-tailed distributions like the normal without materially impairing the performance of the estimator for long-tailed distributions.

**Acknowledgement.** The author wishes to express his gratitude for the significant contribution of Mrs. Elizabeth Hinkley who performed the Monte Carlo experiments reported in this paper.
Appendix

To prove the theorem of Section 2 we first show that \( \sqrt{n} (\theta_{kn} - \theta) \) is asymptotically normal with mean zero and variance

\[
\sigma_k^2 \leq \frac{1}{I(F, p_0)} + \varepsilon,
\]

for arbitrary fixed \( \varepsilon > 0 \), for each sufficiently large \( k \), for all \( F \in \mathcal{F} \), when \( \theta_{kn} \) is given by (1) with the \( c_i \)'s given by (12). The validity of (20) follows from conditions (a) and (c) of the definition of \( \mathcal{F} \), and asymptotic normality follows from condition (d). By virtue of expression (9) it therefore suffices to show (i) that (28)

holds with \( \sigma_k^2 \) replaced by \( \bar{\sigma}_k^2 \) given by (11), and (ii) that \( |\bar{c} R \bar{c}'| < \varepsilon \) for all sufficiently large \( k \), for all \( F \in \mathcal{F} \).

To verify assertion (i), let \( F^{-1}(u) = G(u) \), and referring to (19) let \( r_i = p_0^+ (i-1) q_k \). Then by definition (6) of \( d_i \) we may write

\[
d_i = G(r_i + q_k) - G(r_i), \quad i = 1, 2, \ldots, k,
\]

\[
d_{i+1} = G(r_i + 2q_k) - G(r_i + q_k), \quad i = 1, 2, \ldots, k-1,
\]

\[
d_{i-1} = G(r_i) - G(r_i - q_k), \quad i = 2, 3, \ldots, k.
\]

Now let \( G_1', G_1'' \) and \( G_1''' \) represent the first three derivatives of \( G \) evaluated at \( r_i \). Then under conditions (a) and (b) of the definition of \( \mathcal{F} \), a straightforward Taylor's series expansion in terms of \( q_k \) yields
\[ d_{i-1} = q_k^{-1}(G_i')^{-1} - \frac{1}{2}(G_i')^{-2}G_i'' + q_k \left( \frac{1}{4}(G_i'')^2(G_i')^{-3} - \frac{1}{6} G_i'''(G_i')^{-2} \right) \xi q_k^2, \]

\[ d_{i+1} = q_k^{-1}(G_i')^{-1} - \frac{3}{2}(G_i')^{-2}G_i'' + q_k \left( \frac{9}{4}(G_i'')^2(G_i')^{-3} - \frac{7}{6} G_i'''(G_i')^{-2} \right) \xi q_k^2, \]

\[ d_{i-1} = q_k G_i' - \left( \frac{1}{2}(G_i')^{-2}G_i'' + q_k \frac{1}{4}(G_i'')^2(G_i')^{-3} - \frac{1}{6} G_i'''(G_i')^{-2} \right) \xi q_k^2, \]

where \( \xi \) represents a generic uniformly bounded function of \( i, k, \) and \( G. \) Hence by (15), for \( i=2,3,...,k-1, \)

\[ e_i = - \left[ 2(G_i'')^2(G_i')^{-4} - G_i'''(G_i')^{-3} \right] + \xi q_k. \]

Letting \( f_i, f_i', \) and \( f_i'' \) be \( f \) and its corresponding derivatives evaluated at \( G(r_i) \) we have

\[ G_i' = (f_i')^{-1}, \quad G_i'' = -f_i'(f_i')^{-3}, \quad G_i''' = 3(f_i')(f_i')^{-5} - f_i''(f_i')^{-4}, \]

so that \( e_i = -[f_i''(f_i')^{-1} - (f_i')^2(f_i')^{-2}] \xi q_k. \)

Also, writing

\[ d_k = G(\frac{1}{2}) - q_k, \]

\[ d_{k-1} = G(\frac{1}{2} - q_k) - G(\frac{1}{2} - 2q_k) \]

and referring to (14), we obtain by a similar calculation

\[ e_k = G'''(\frac{1}{2})[G'(\frac{1}{2})]^{-3} + \xi q_k = - \frac{f''(0)}{f(0)} + \xi q_k. \]
Hence, as $k \to \infty$,

$$ q_k \sum_{i=2}^{k} e_i \to - \int_{p_0}^{1/2} \left\{ \frac{r''(G(u))}{f(G(u))} - \left[ \frac{r'(G(u))}{f(G(u))} \right]^2 \right\} du, $$

$$ = - \int_{G(p_0)}^{0} \left\{ \frac{d}{dx} \log f(x) \right\} f(x) dx = \int_{G(p_0)}^{0} (f'(x))^2 (f(x))^{-1} dx + f_1', $$

and the convergence is uniform for all $F \in \mathcal{F}$. Again referring to (14), we obtain

$$ e_i = q_k^{-1} \left( \frac{1}{p_0} f_1 - f_1' \right)^{i_1}. $$

Thus, from (11), (19), (20), (30) and (31) we have

$$ \frac{1}{q_k^2} = 2q_k \sum_{i=1}^{k} e_i \to I(F,p_0) $$

as $k \to \infty$, uniformly for all $F \in \mathcal{F}$. This together with condition (c) of the definition of $\mathcal{F}$ implies the truth of assertion (i).

To prove assertion (ii), let the elements of the matrix $R$ given by (8) be denoted by $r_{ij}$, $i,j=1,2,\ldots,k$. Then by (5) and (6) we have for $i < j$

$$ r_{ij} = r_{ji} = d^{-1}_i \int_{G(r_i)}^{G(r_i+q_k)} f(x) dx - b_i, $$

and

25
\[ r_{ii} = 2d_1^{-2} \left\{ \int \frac{G(r_i + q_k)}{G(r_i)} F(x)dx - \int \frac{G(r_i + q_k)}{G(r_i)} x F(x)dx \right\} - b_1, \]

where by (6) and (19), \( b_1 = r_1 + \frac{1}{2} q_k. \)

Expansion in Taylor's series with respect to \( q_k \) yields for \( i < j, \)

\[ r_{ij} = r_i + \frac{1}{2} q_k G'_{1,1} + \xi q_i^2 - b_i = \xi q_k^2, \]

and

\[ r_{ii} = r_i + q_k G'_{1,1} - b_i + \xi q_k^2 = \frac{1}{2} q_k + \xi q_k^2. \]

Since \( q_k = k^{-1} (\frac{1}{2} - p_0) \) we see that \( r_{ij} = O(k^{-2}) \) uniformly in \( i, j \) and \( \mathcal{F} \) for \( i \neq j, \) and \( r_{ii} = O(k^{-1}) \) uniformly in \( i \) and \( \mathcal{F} \in \mathcal{F} \). Furthermore, since \( \tilde{c}_i = c_i d_i \) where the \( c_i \)'s are given by (12) we see that \( \tilde{c}_i = O(k^{-1}) \) uniformly in \( i \) and \( \mathcal{F} \in \mathcal{F}, \) so that \( \tilde{c} R \tilde{c}' = O(k^{-2}), \) uniformly in \( \mathcal{F} \in \mathcal{F} \) which proves assertion (ii).

Conclusion (21) of the theorem follows from (28) provided that for any fixed \( k, \sqrt{n} \left( \hat{\theta}_{kn} - \theta_{kn} \right) \rightarrow 0, \) in probability, as \( n \rightarrow \infty, \) since then \( r_k^2(F) = \sigma_k^2. \) But from (18) we have

\[ \sqrt{n} \left( \hat{\theta}_{kn} - \theta_{kn} \right) = \sum_{i=1}^{k} \left( \hat{c}_i - c_i \right) \frac{1}{\sqrt{n}} S_{1n}. \]

Now the \( \hat{c}_i \)'s given by (17) are consistent estimates of the corresponding \( c_i \)'s given by (12) since the \( \hat{d}_i \)'s are consistent estimates of the corresponding \( d_i \)'s. The desired result follows from the
boundedness, in probability, of the quantities \( \frac{1}{\sqrt{n}} S_{in} \), as \( n \to \infty \).

Conclusion (22) now follows from the fact that \( \hat{\sigma}^2_k \) is a consistent estimate of \( \sigma^2_k \) which by assertion (ii) is uniformly close to \( \sigma^2_k \) for sufficiently large \( k \). This completes the proof of the theorem.

We consider now the double exponential distribution with density

\[ f(x) = \frac{1}{2} e^{-|x|}. \]

The expressions for \( e_i \) obtained above are valid for \( i=1,2,\ldots,k-1 \). From (31) we see that \( e_1 = \zeta \) for this case. Also

\[ \frac{d^2}{dx^2} \log f(x) = 0, \text{ for } x < 0, \text{ so that } q_k \sum_{i=1}^{k-1} e_i \to 0, \text{ as } k \to \infty. \]

Expression (29) for \( e_k \) is not valid for this case since \( f'(x) \) is not continuous at \( x = 0 \). Direct calculation shows that \( e_k = 4q_k^{-1} + o(1) \).

Therefore, in the limit, the only non-zero \( c_i \) is \( c_k \), so that, as \( k \) increases, \( \theta_{kn} \) is essentially an average of a decreasing proportion of central order statistics. The conclusions of the theorem are clearly correct for the corresponding sequence \( \hat{\theta}_{kn} \) even though the double exponential distribution is not in the family \( \mathcal{F} \) as defined.
REFERENCES


