The Annals of Probability
1988, Vol. 16, No. 4, 1608–1622

ASYMPTOTIC NORMALITY OF TRIMMED MEANS IN HIGHER DIMENSIONS

BY R. A. MALLER

CSIRO Division of Mathematics and Statistics

A representation for the distribution of the trimmed sum of vector-valued random variables is obtained, generalising a one-dimensional formula. The trimming is with respect to observations falling outside a fixed family of sets, e.g., spheres. Asymptotic normality of the heavily trimmed sum, when normed and centered in different ways, is proved, and rates of convergence are given for some cases.

1. Introduction and notation. Let $\mathbf{X}, (\mathbf{X}_i)_{i \geq 1}$ be nondegenerate iid random vectors in $\mathbb{R}^d$. Let $\{S(y)\}_{y \geq 0}$ be a fixed family of subsets of $\mathbb{R}^d$ for which $S(y)$ and $\partial S(y)$ [the boundary of $S(y)$] are bounded and measurable, satisfying

$$\{0\} = S(0) \subseteq S(y_1) \subseteq S(y_2) \subseteq S(+\infty) = \mathbb{R}^d \tag{1.1}$$

for $0 < y_1 \leq y_2 < +\infty$. Assume $S(y)$ increase continuously in the sense that for $\mathbf{x} \neq 0$ there is a unique y such that $\mathbf{x} \in \partial S(y)$.

Let F denote both the distribution function of $\mathbf{X}$, and the measure induced by this distribution on the Borel subsets of $\mathbb{R}^d$. Assume F is continuous with respect to $\{S(y)\}$ in the sense that

$$\lim_{\delta \to 0} \{F(S(y + \delta)) - F(S(y))\} = F(\partial S(y)) = 0, \qquad y \geq 0. \tag{1.2}$$

So

$$h(y) = P\{\mathbf{X} \in S(y)\}, \qquad y \geq 0, \tag{1.3}$$

is a probability distribution on $[0, \infty)$, continuous on $(0, \infty)$ and continuous from the right at 0.

The family $S(y)$ induces an ordering on $\mathbf{X}_1, \dots, \mathbf{X}_n$, which by the continuity, is almost surely unique, defined as follows: Let

$$y_{n-r} = \inf\{y > 0 \colon S(y) \text{ contains exactly } n - r \text{ of } \mathbf{X}_1, \dots, \mathbf{X}_n\}, \qquad r \geq 1.$$

Then define

$$\mathbf{X}_n^{(1)} = \mathbf{X}_{i_n(1)}, \quad \text{where } i_n(1) = i \text{ such that } \mathbf{X}_i \notin S(y_{n-1}),$$

$$\vdots$$

$$\mathbf{X}_n^{(r)} = \mathbf{X}_{i_n(r)}, \quad \text{where } i_n(r) = i \text{ such that } \mathbf{X}_i \notin S(y_{n-r}),$$

$$i \neq i_n(1), \dots, i_n(r-1), 2 \leq r \leq n.$$

Received January 1987; revised July 1987.
AMS 1980 *subject classifications*. 60F05, 62H10.
Key words and phrases. Trimmed means, trimmed sums, asymptotic normality, rates of convergence.

In dimension 1, $\mathbf{X}_n^{(r)}$ is the term of the rth largest modulus when $S(y) = [-y, y]$; for $d > 1$ the ordering is essentially one-dimensional, indexed by $y \geq 0$. The prototypical examples of $\{S(y)\}$ are the *spheres* $S(y) = \{\mathbf{x} | |\mathbf{x}| \leq y\}$, where $|\mathbf{x}|$ denotes the Euclidean norm of a vector, and the *cubes* $C(y) = \{\mathbf{x}: |x_i| \leq y, 1 \leq i \leq d\}$, where $\mathbf{x}$ has components $x_1, \dots, x_d$. The $S(y)$ may be generated, for example, by $S(y) = yS$, $y \geq 0$, where S is a convex compact set of nonzero measure containing 0. The resulting ordering of $\mathbf{X}_1, \dots, \mathbf{X}_n$ is then with respect to Minkowski distance. In general, however, we do not require $S(y)$ to be convex.

The properties we require are easily carried over from the one-dimensional theory. Most importantly for our purposes, $\mathbf{X}_n^{(j)}$ satisfy the following Markov property: If $y_1 \geq y_2 \geq \cdots \geq y_r$, then

$$\begin{aligned} &\left\{\mathbf{X}_n^{(j)}: r+1 \leq j \leq n | \mathbf{X}_n^{(k)} \in \partial S(y_k), 1 \leq k \leq r\right\} \\ (1.4) \qquad &=_D \left\{\mathbf{X}_n^{(j)}: r+1 \leq j \leq n | \mathbf{X}_n^{(r)} \in \partial S(y_r)\right\} \\ &=_D \left\{\mathbf{X}_{n-r}^{(j)}(y_r): 1 \leq j \leq n-r\right\}, \end{aligned}$$

where

$$(1.5) \qquad \{\mathbf{X}_i(y)\} =_D \{\mathbf{X}_i | \mathbf{X}_i \in S(y)\},$$

and $\mathbf{X}_n^{(1)}(y_r), \dots, \mathbf{X}_n^{(n)}(y_r)$ is the ordering induced on $\mathbf{X}_1(y_r), \dots, \mathbf{X}_n(y_r)$ by $S(y)$. ("D" means "has the same distribution as.")

The preceding conditional probabilities are defined by $P(A | \mathbf{X}_i \in \partial S(y)) = P(A | X_i = y)$, where A is any measurable set and X_i are the real-valued random variables given by $X_i = y$ when $\mathbf{X}_i \in \partial S(y)$. These are iidrv's with distribution $h(y)$. The required Markov properties described previously then follow from one-dimensional methods.

Now define trimmed sums by

$$(1.6) \qquad {}^{(r)}\mathbf{S}_n = \mathbf{X}_n^{(n)} + \cdots + \mathbf{X}_n^{(r+1)}, \qquad 1 \leq r < n,$$

where addition of vectors is defined componentwise. A corollary of (1.4) is that $y_1 \geq y_2 \geq \cdots \geq y_r$ implies

$$\begin{aligned} (1.7) \qquad \left\{{}^{(r)}\mathbf{S}_n | \mathbf{X}_n^{(k)} \in \partial S(y_k), 1 \leq k \leq r\right\} &=_D \left\{{}^{(r)}\mathbf{S}_n | \mathbf{X}_n^{(r)} \in \partial S(y_r)\right\} \\ &=_D \left\{\mathbf{S}_{n-r}(y_r)\right\}, \end{aligned}$$

where $\mathbf{S}_n(y) = \mathbf{X}_1(y) + \cdots + \mathbf{X}_n(y)$ for $y > 0$, $n \geq 1$.

The purpose of this paper is to show that under "heavy" trimming, i.e., when $0 < \alpha < 1$ and $r = [n\alpha]$ (the integer part of $n\alpha$), ${}^{[n\alpha]}\mathbf{S}_n$ is asymptotically normal when normed and centered appropriately, without the requirement of moment or other conditions on the tail of F, but with certain smoothness conditions on F in the neighbourhood of the $(1-\alpha)$ quantile of h. With other than "natural" centering, ${}^{[n\alpha]}\mathbf{S}_n$ may converge to a mixture of normal distributions under smoothness conditions on F. These results are related to a theorem by Stigler (1973) as explained later.

Under similar conditions on F, we show that the rate of convergence of ${}^{[n\alpha]}\mathbf{S}_n$ to its limit can easily be estimated using the methods developed for measuring rates of convergence of ordinary multidimensional sums to normality; see Bhattacharya (1977) and Sweeting (1977) for recent results and discussions on these.

Our results are stated and proved in the next two sections. Here we introduce more notation and give two lemmas, which are required in the next section.

Assume the existence of centering constants defined as follows: There is a measurable $\mathbf{m}: \mathbb{R}^d \to \mathbb{R}^d$ such that when $\mathbf{z} \in \partial S(y)$, $y > 0$,

$$\begin{aligned} \mathbf{m}(\mathbf{z}) &= E(\mathbf{X}|\mathbf{X} \in S(y)) \\ &= \mu(y), \end{aligned} \tag{1.8}$$

where $\mu: [0, \infty) \to \mathbb{R}^d$ is given by

$$\begin{aligned} \mu(y) &= E\mathbf{X}(y) \\ &= \int_{S(y)} \mathbf{x}\, dF(\mathbf{x}) \Big/ \int_{S(y)} dF(\mathbf{x}), \end{aligned} \tag{1.9}$$

when $h(y) = \int_{S(y)} dF(\mathbf{x}) > 0$.

The $S(y)$ are essentially homotopies of the sphere, and there may exist a function $f: \mathbb{R}^d \to \mathbb{R}$ such that $S(y)$ is given by $S(y) = f^{-1}[0, y]$, $y > 0$. Then $\mathbf{m}$ can be taken as

$$\mathbf{m}(\mathbf{z}) = \int_{S(f(\mathbf{z}))} \mathbf{x}\, dF(\mathbf{x}) / h(f(\mathbf{z})), \qquad \mathbf{z} \in \mathbb{R}^d.$$

For example, when $S(y) =$ a sphere of radius y, $f(\mathbf{x}) = |\mathbf{x}|$. More generally, ellipses are given by $f(\mathbf{x}) = \mathbf{x}^{\mathrm{T}} V^{-1} \mathbf{x}$, V a fixed nonsingular matrix, $\mathbf{x}^{\mathrm{T}}$ the transpose of $\mathbf{x}$ and rectangles by $f(\mathbf{x}) = \max(c_1|x_1|, \dots, c_d|x_d|)$, where c_j are positive constants and $\mathbf{x} = (x_1, \dots, x_d)$.

As a result of (1.7) and (1.8), we have for any measurable $A \subseteq \mathbb{R}^d$ the following formula, which is the basis of our analysis:

$$\begin{aligned} &P\left\{ {}^{(r)}\mathbf{S}_n - (n - r)\mathbf{m}\left(\mathbf{X}_n^{(r)}\right) \in A \right\} \\ &\quad = \int_0^\infty P\{\mathbf{S}_{n-r}(y) - (n - r)\mu(y) \in A\}\, dP\left\{\mathbf{X}_n^{(r)} \in S(y)\right\} \end{aligned} \tag{1.10}$$

and since $E\mathbf{S}_{n-r}(y) = (n - r)\mu(y)$ we have the correct centering for $\mathbf{S}_{n-r}(y)$. The variance of $\mathbf{X}(y)$ is

$$\Sigma(y) = E\{\mathbf{X}(y) - E\mathbf{X}(y)\}\{\mathbf{X}(y) - E\mathbf{X}(y)\}^{\mathrm{T}}, \qquad y \geq 0,$$

and we let $A(y)$ be a symmetric matrix such that

$$A(y)\Sigma(y)A^{\mathrm{T}}(y) = I, \tag{1.11}$$

where I is the identity matrix in d dimensions.

The behaviour of $\mathbf{X}_n^{[n\alpha]}$ is related to the $(1 - \alpha)$ quantile of h defined by

$$a = \sup\{y: h(y) < 1 - \alpha\}. \tag{1.12}$$

The value a is not assumed to be taken uniquely; in fact, let

$$[a, b] = \{y: h(y) = 1 - \alpha\}. \tag{1.13}$$

Also let

$$A_\varepsilon = \{y: |h(y) - (1-\alpha)| \le \varepsilon\}, \qquad 0 < \varepsilon < 1. \tag{1.14}$$

Since $h(a) > 0$, we have that $\Sigma(a)$ is positive definite. Then by continuity of F, $\Sigma(y)$ is positive for $y \in A_\varepsilon$ if ε is small enough. Thus $A(y)$ is also nonsingular for $y \in A_\varepsilon$ and

$$\Sigma^{-1}(y) = A^{\mathrm{T}}(y)A(y). \tag{1.15}$$

Now $\Sigma(y)$ may be singular for some values of $y < a$ but since we will only be interested in values of y in A_ε we can define $\Sigma^{-1}(y)$ and $A(y)$ arbitrarily for $y \notin A_\varepsilon$.

We also need the notation of Bhattacharya (1977) and Sweeting (1977). Let B_1^d be the bounded measurable real-valued functions on $\mathbb{R}^d$. A sequence of probability measures $Q_n \to_D Q$ (weak convergence on $\mathbb{R}^d$) if

$$\int f(\mathbf{x})Q_n(d\mathbf{x}) \to \int f(\mathbf{x})Q(d\mathbf{x})$$

(integrals without limits are assumed to be over $\mathbb{R}^d$) for all $f \in B_1^d$, which satisfy

$$\omega_f(Q,\delta) = \int \omega_f(\mathbf{x},\delta)Q(d\mathbf{x}) \to 0, \qquad \delta \to 0+,$$

where

$$\omega_f(\mathbf{x},\delta) = \sup\{|f(\mathbf{y}) - f(\mathbf{z})|, \mathbf{y},\mathbf{z} \in B(\mathbf{x},\delta)\}$$

and $B(x,\delta)$ is the sphere of radius δ centered at $\mathbf{x}$. Bounds on

$$\left|\int fQ_n - \int fQ\right|,$$

in terms of ω_f, give rates of convergence of Q_n to Q.

Finally, we require two lemmas on the asymptotic behaviour of $\mathbf{X}_n^{[n\alpha]}$. The first is proved by the methods of Weiss (1970). The second is implicit in work of Bjerve (1977), who gave rates of convergence to normality of linear combinations of order statistics in dimension 1. Let $N(0,1)$ be the standard normal random variable in dimension 1, and $\mathbf{N}(0,I)$ or just $\mathbf{N}$ the same in d dimensions. Let Φ be the distribution function of $\mathbf{N}(0,I)$.

LEMMA 1. *Suppose $a < b$ in* (1.13). *Then for $y > 0$,*

$$P\{\mathbf{X}_n^{[n\alpha]} \in S(b + yn^{-1/2})\} \to P\{N(0,1) \le yh'_+(b)(\alpha(1-\alpha))^{-1/2}\}, \tag{1.16}$$

where

$$h'_+(b) = \lim_{y\to 0+} \frac{h(b+y) - h(b-)}{y} = \lim_{y\to 0+} \frac{h(b+y) - (1-\alpha)}{y},$$

finite or infinite. [*If infinite, interpret the right-hand side of* (1.16) *as* 1.] *Also, for $y \ge 0$,*

$$P\{\mathbf{X}_n^{[n\alpha]} \in S(b - yn^{-1/2})\} \to \tfrac{1}{2},$$

and similarly

$$P\{\mathbf{X}_n^{[n\alpha]} \in S(a - yn^{-1/2})\} \to P\{N(0,1) \le yh'_{-}(a)(\alpha(1-\alpha))^{-1/2}\},$$

where $h'_{-}(a)$ *is finite or infinite, and*

$$P\{\mathbf{X}_n^{[n\alpha]} \in S(a + yn^{-1/2})\} \to \tfrac{1}{2}.$$

If $a < c < b$ *and* $-\infty < y < +\infty$,

$$P\{\mathbf{X}_n^{[n\alpha]} \in S(c + yn^{-1/2})\} \to \tfrac{1}{2}.$$

If $a = b = h^{-1}(1-\alpha)$, *then*

$$\begin{aligned} P\{\mathbf{X}_n^{[n\alpha]} \in S(a + yn^{-1/2})\} &\to P\{N(0,1) \le yh'_{+}(a)(\alpha(1-\alpha))^{-1/2}\}, && y > 0, \\ &\to \tfrac{1}{2}, && y = 0, \\ &\to P\{N(0,1) \le yh'_{-}(a)(\alpha(1-\alpha))^{-1/2}\}, && y < 0, \end{aligned}$$

with appropriate interpretations if $h'_{+}(a)$ *or* $h'_{-}(a)$ *is infinite. If* $a = b$ *and* $h'_{+}(a) = h'_{-}(a)$, *finite and nonzero, then for* $-\infty < y < +\infty$,

$$P\{\mathbf{X}_n^{[n\alpha]} \in S(a + yn^{-1/2})\} \to P\{N(0,1) \le yh'(a)(\alpha(1-\alpha))^{-1/2}\}.$$

PROOF. For y in a neighbourhood of $[a, b]$ such that $0 < h(y) < 1$,

$$\begin{aligned} P\{\mathbf{X}_n^{[n\alpha]} \in S(y)\} &= P\{\text{at least } n - [n\alpha] + 1 \text{ of } \mathbf{X}_1, \dots, \mathbf{X}_n \text{ are in } S(y)\} \\ &= P\left\{\sum_{i=1}^{n} I\{\mathbf{X}_i \in S(y)\} \ge n - [n\alpha] + 1\right\} \\ &= P\left\{T_n(y) \ge \frac{n - [n\alpha] + 1 - nh(y)}{(nh(y)H(y))^{-1/2}}\right\}, \end{aligned} \tag{1.17}$$

where $H(y) = 1 - h(y)$ and

$$T_n(y) = \frac{\sum_{i=1}^{n} I\{\mathbf{X}_i \in S(y)\} - nh(y)}{(nh(y)H(y))^{1/2}}$$

has mean 0 and variance 1. The central limit theorem as in Weiss (1970) can now be used to complete the proof [see also Reiss (1976)]. □

LEMMA 2. *There is a constant* $c > 0$ *such that*

$$\int_{y \notin A_\varepsilon} dP\{\mathbf{X}_n^{[n\alpha]} \in S(y)\} = O(e^{-cn}), \qquad n \to \infty.$$

PROOF. Let X_i be iid with distribution $h(y)$. Then for some $\delta_1 > 0$, $\delta_2 > 0$,

$$\begin{aligned} \int_{y \notin A_\varepsilon} dP\{\mathbf{X}_n^{[n\alpha]} \in S(y)\} &\le P\{X_n^{[n\alpha]} \notin [a - \delta_1, b + \delta_2]\} \\ &= O(e^{-cn}), \end{aligned}$$

by the argument in Bjerve (1977), page 360, paragraph 2, where $X_n^{(i)}$ denote the order statistics of $X_1, \dots, X_n$. □

2. Heavy trimming.

THEOREM 1. *Assume* (1.2) *and* (1.8) *and let* $f \in B_1^d$. *Then there are constants* $c_1 > 0$, $c_2 > 0$ *for which*

$$\left|\int f(\mathbf{x}) P\left\{A(a)\left({}^{[n\alpha]}\mathbf{S}_n - (n - [n\alpha])\mathbf{m}(\mathbf{X}_n^{[n\alpha]})\right)(n - [n\alpha])^{-1/2} \in d\mathbf{x}\right\} - \int f(\mathbf{x})\Phi(d\mathbf{x})\right| \tag{2.1}$$

$$\le c_1\left[\eta + \int \omega_f\left(c_2 n^{-1/2}, \mathbf{x}\right)\Phi(d\mathbf{x})\right], \tag{2.2}$$

if $n \ge n_0(\eta)$, *where* η *may be arbitrarily small; i.e., the weak convergence*

$$\frac{A(a)\left({}^{[n\alpha]}\mathbf{S}_n - (n - [n\alpha])\mathbf{m}(\mathbf{X}_n^{[n\alpha]})\right)}{(n - [n\alpha])^{1/2}} \to_D \mathbf{N}(0, I)$$

holds. Suppose, in addition,

$$0 < \liminf_{y \to 0+} \frac{h(b + y) - h(y)}{y} \le \limsup_{y \to 0+} \frac{h(b + y) - h(y)}{y} < +\infty, \tag{2.3a}$$

$$0 < \liminf_{y \to 0+} \frac{h(a) - h(a - y)}{y} \le \limsup_{y \to 0+} \frac{h(a) - h(a - y)}{y} < +\infty. \tag{2.3b}$$

Then for n sufficiently large, (2.1) *is bounded by*

$$c_1\left[n^{-1/2} + \int \omega_f\left(c_2 n^{-1/2}, \mathbf{x}\right)\Phi(d\mathbf{x})\right]. \tag{2.4}$$

PROOF. By (1.10) and change of variable, (2.1) equals

$$\begin{aligned}\Bigg|\int_0^\infty \int f(A(a)A^{-1}(y)\mathbf{x})\Big[P\big\{A(y)(\mathbf{S}_{n-[n\alpha]}(y) - E\mathbf{S}_{n-[n\alpha]}(y))\\ \times(n - [n\alpha])^{-1/2} \in d\mathbf{x}\big\} - P\{\mathbf{N} \in d\mathbf{x}\}\Big]\, dh_n(y)\\ + \int_0^\infty \int [f(A(a)A^{-1}(y)\mathbf{x}) - f(\mathbf{x})]\, P\{\mathbf{N} \in d\mathbf{x}\}\, dh_n(y)\Bigg|,\end{aligned} \tag{2.5}$$

where

$$h_n(y) = P\left\{\mathbf{X}_n^{[n\alpha]} \in S(y)\right\}, \qquad y \ge 0.$$

In the first term of (2.5), the integral over $y \notin A_\varepsilon$ is $O(n^{-1/2})$ by Lemma 2. To bound the integral over $y \in A_\varepsilon$, note that

$$\mathbf{Z} = A(y)\left(\mathbf{S}_{n-[n\alpha]}(y) - E\mathbf{S}_{n-[n\alpha]}(y)\right)(n - [n\alpha])^{-1/2}$$

is the sum of iid random vectors with mean 0 and variance I. Let $f^*(\mathbf{x}) = f(A(a)A^{-1}(y)\mathbf{x})$, which is in B_1^d. Sweeting's result [(1977), Corollary 3, page 39] gives

$$\left|\int_{a_\varepsilon}\int f^*(\mathbf{x})[P\{\mathbf{Z}\in d\mathbf{x}\}-P\{\mathbf{N}\in d\mathbf{x}\}]\,dh_n(y)\right|$$

$$\text{(2.6)}\qquad \le c_1\left[n^{-1/2}\int_{A_\varepsilon}\beta_3(y)\,dh_n(y)\right.$$

$$\left.+\int_{A_\varepsilon}\int\omega_{f^*}(\mathbf{x},c_2n^{-1/2}\beta_3(y))P\{\mathbf{N}\in d\mathbf{x}\}\,dh_n(y)\right],$$

where

$$\beta_3(y)=E|\mathbf{X}(y)|^3.$$

Notice that $c_3=\sup_{y\in A_\varepsilon}\beta_3(y)$ is finite. So (2.6) is bounded by

$$O(n^{-1/2})+\int_{A_\varepsilon}\int\omega_{f^*}(\mathbf{x},c_4n^{-1/2})P\{\mathbf{N}\in d\mathbf{x}\}\,dh_n(y).$$

We defer to the end of the proof the demonstration that this is of the form of the integral in (2.2) and (2.4).

Next, we deal with the second term of (2.5), which can be written as

$$\int_0^\infty\int f(A(a)A^{-1}(y)\mathbf{x})P\{\mathbf{N}\in dx\}\,dh_n(y)-\int_0^\infty\int f(\mathbf{x})P\{\mathbf{N}\in d\mathbf{x}\}\,dh_n(y)$$

$$=\int_0^\infty\int f(A(a)\mathbf{x})P\{A^{-1}(y)\mathbf{N}\in dx\}\,dh_n(y)$$

$$-\int_0^\infty\int f(A(a)\mathbf{x})P\{A^{-1}(a)\mathbf{N}\in d\mathbf{x}\}\,dh_n(y).$$

Again by Lemma 2 we need only consider values of $y\in A_\varepsilon$ in this. Let

$$A_\varepsilon=\{y:|h(y)-(1-\alpha)|<\varepsilon\}\subseteq[a-\delta_1,b+\delta_2],$$

for some $\delta_1(\varepsilon)>0$, $\delta_2(\varepsilon)>0$. So we need to estimate

$$\int_{-\delta_1}^0\int f(A(a)\mathbf{x})P\{A^{-1}(a+y)\mathbf{N}\in d\mathbf{x}\}\,dh_n(a+y)$$

$$\text{(2.7)}\qquad +\int_0^{\delta_2}\int f(A(a)\mathbf{x})P\{A^{-1}(b+y)\mathbf{N}\in dx\}\,dh_n(b+y)$$

$$-\left[\int_{-\delta_1}^0 dh_n(a+y)+\int_0^{\delta_2}dh_n(b+y)\right]\int f(A(a)\mathbf{x})P\{A^{-1}(a)\mathbf{N}\in d\mathbf{x}\}.$$

The elements of $\Sigma(y)$ are

$$\sigma_{jk}(y)=\frac{\int_{S(y)}x_jx_k\,dF(\mathbf{x})}{h(y)}-\frac{\int_{S(y)}x_j\,dF(\mathbf{x})\int_{S(y)}x_k\,dF(\mathbf{x})}{h^2(y)},$$

for $1\le j,k\le d$. Assuming only the (left) continuity of h at a and the boundedness of $S(y)$, we have if η is small

$$|\sigma_{jk}(a+y)-\sigma_{jk}(a)|\le\eta,\quad\text{for }-\delta_1\le y\le 0,$$

provided ε (and hence δ_1) are small enough. Assuming instead (2.3b) gives

$$|\sigma_{jk}(a+y)-\sigma_{jk}(a)| \le c|y|, \quad \text{for } -\delta_1 \le y \le 0,$$

for some $c > 0$ because, for example,

$$\left|\int_{S(a+y)-S(a)} x_j x_k \, dF(\mathbf{x})\right| \le \sup_{\mathbf{x}\in S(a+y)} |\mathbf{x}|^2 [h(a)-h(a+y)]$$
$$\le c|y|, \quad \text{for } y \to 0- .$$

Summarise these two possibilities as

$$|\sigma_{jk}(a+y)-\sigma_{jk}(a)| = O(\eta \text{ or } |y|), \quad \text{for } -\delta_1 \le y \le 0.$$

Similarly, we obtain

$$|\sigma_{jk}(b+y)-\sigma_{jk}(a)| = |\sigma_{jk}(b+y)-\sigma_{jk}(b)|$$
$$= O(\eta \text{ or } y), \quad \text{for } 0 \le y \le \delta_2,$$

assuming either that h is (right) continuous at b or (2.3a). Note that $\sigma_{jk}(a) = \sigma_{jk}(b)$ since h (and F) are constant on $[a, b]$ [and $S(b) - S(a)$].

Together these estimates imply

$$\|\Sigma(a+y)-\Sigma(a)\| = O(\eta \text{ or } |y|), \qquad -\delta_1 \le y \le 0, \tag{2.8a}$$

$$\|\Sigma(b+y)-\Sigma(b)\| = O(\eta \text{ or } y), \qquad 0 \le y \le \delta_2, \tag{2.8b}$$

where the matrix norm is defined by

$$\|\Sigma\| = \sup_{\mathbf{x}} (\mathbf{x}^{\mathrm{T}}\Sigma\mathbf{x})/(\mathbf{x}^{\mathrm{T}}\mathbf{x}).$$

Next, $A^{-1}(y)\mathbf{N}$ is multivariate normal $(0, \Sigma(y))$, so

$$\begin{aligned}
&\int f(A(a)\mathbf{x})\left[P\{A^{-1}(a+y)\mathbf{N}\in d\mathbf{x}\} - P\{A^{-1}(a)\mathbf{N}\in d\mathbf{x}\}\right] \\
&= (2\pi)^{-d/2}\int f(A(a)\mathbf{x})\left\{\frac{\exp\left(-\frac{1}{2}x^{\mathrm{T}}\Sigma^{-1}(a+y)\mathbf{x}\right)}{|\Sigma(a+y)|^{1/2}} - \frac{\exp\left(-\frac{1}{2}x^{\mathrm{T}}\Sigma^{-1}(a)\mathbf{x}\right)}{|\Sigma(a)|^{1/2}}\right\}d\mathbf{x} \\
&= (2\pi)^{-d/2}\int f(A(a)\mathbf{x})\frac{\exp\left(-\frac{1}{2}\mathbf{x}^{\mathrm{T}}\Sigma^{-1}(a)\mathbf{x}\right)}{|\Sigma(a+y)\Sigma(a)|^{1/2}} \\
&\quad\times\left\{\left(\exp\left(-\tfrac{1}{2}\mathbf{x}^{\mathrm{T}}(\Sigma^{-1}(a+y)-\Sigma^{-1}(a))\mathbf{x}\right)-1\right)|\Sigma(a)|^{1/2} + |\Sigma(a)|^{1/2} - |\Sigma(a+y)|^{1/2}\right\}d\mathbf{x}.
\end{aligned} \tag{2.9}$$

Also,

$$\left|\exp\left(-\tfrac{1}{2}\mathbf{x}^{\mathrm{T}}(\Sigma^{-1}(a+y)-\Sigma^{-1}(a))\mathbf{x}\right)-1\right|$$
$$\le \left|\mathbf{x}^{\mathrm{T}}(\Sigma^{-1}(a+y)-\Sigma^{-1}(a))\mathbf{x}\right|\exp\left(\left|\mathbf{x}^{\mathrm{T}}(\Sigma^{-1}(a+y)-\Sigma^{-1}(a))\mathbf{x}\right|\right),$$

whereas

$$\begin{aligned}\left|\mathbf{x}^{\mathrm{T}}(\Sigma^{-1}(a+y)-\Sigma^{-1}(a))\mathbf{x}\right| &= \left|\mathbf{x}^{\mathrm{T}}\Sigma^{-1}(a+y)(\Sigma(a)-\Sigma(a+y))\Sigma^{-1}(a)\mathbf{x}\right| \\ &\le \mathbf{x}^{\mathrm{T}}\mathbf{x}\|\Sigma^{-1}(a+y)\Sigma^{-1}(a)\|\,\|\Sigma(a+y)-\Sigma(a)\| \\ &= (\mathbf{x}^{\mathrm{T}}\mathbf{x})O(\eta \text{ or } c|y|), \qquad -\delta_1 \le y \le 0,\end{aligned}$$

by (2.8a). From the definition of the determinant, (2.8a) also implies

$$|\Sigma(a+y)| = |\Sigma(a)| + O(\eta \text{ or } |y|), \qquad -\delta_1 \le y \le 0.$$

Putting these together shows that (2.9) is

$$O(\eta \text{ or } |y|), \qquad -\delta_1 \le y \le 0;$$

here we again use the fact that $f \in B_1^d$.

A similar estimation holds for $0 \le y \le \delta_2$, so the difference of integrals in (2.7) is

$$O\left(\eta \text{ or } \int_{-\delta_1}^{0} |y|\, dh_n(a+y) + \int_0^{\delta_2} y\, dh_n(b+y)\right),$$

and this proves the part of the bound depending on η in (2.2). For (2.4), we use the notation of Lemma 1 of Section 1.

We have

$$\begin{aligned}\int_0^{\delta_2} y\, dh_n(b+y) &\le \int_0^{\delta_2}(1-h_n(b+y))\, dy \\ &= \int_0^{\delta_2} P\{\mathbf{X}_n^{[n\alpha]} \notin S(b+y)\}\, dy \\ &= \int_0^{\delta_2} P\left\{T_n(b+y) \le -n\frac{(h(b+y)-h(b))}{(nh(b+y)H(b+y))^{1/2}}\right. \\ &\qquad\qquad \left.+\frac{R_n}{(nh(b+y)H(b+y))^{1/2}}\right\} dy,\end{aligned}$$

by (1.17), where $H(y) = 1 - h(y)$ and

$$0 \le R_n = n - [n\alpha] + 1 - nh(b) = n - [n\alpha] + 1 - n(1-\alpha) \le 2.$$

By (2.3a), for some $c_1 > 0$,

$$h(b+y) - h(b) \ge c_1 y, \qquad 0 \le y \le \delta_2,$$

so by Chebyshev's inequality (T_n has mean 0 and variance 1),

$$\begin{aligned}
&\int_0^{\delta_2} y\,dh_n(b+y)\\
&\quad\le \int_0^{\delta_2} P\{T_n(b+y) \le -n^{1/2}c_2 y + c_3 n^{-1/2}\}\,dy\\
&\quad= n^{-1/2}\left[\int_0^1 + \int_1^{\delta_2 n^{1/2}}\right] P\{T_n(b+yn^{-1/2}) \le -c_2 y + c_3 n^{-1/2}\}\,dy \qquad (2.10)\\
&\quad\le O(n^{-1/2}) + n^{-1/2}\int_1^{\infty} \frac{dy}{(c_2 y - c_3 n^{-1/2})^2}\\
&\quad= O(n^{-1/2}).
\end{aligned}$$

A similar proof works for the integral over $[-\delta_1, 0]$ and proves the $O(n^{-1/2})$ in (2.4).

To complete the proof, we have to replace ω_{f^*} by ω_f in (2.6). Now if $\mathbf{z} \in B(\mathbf{x}, \delta)$ and $\mathbf{z}^* = A(a)A^{-1}(y)\mathbf{z}$, $\mathbf{x}^* = A(a)A^{-1}(y)\mathbf{x}$, then

$$\begin{aligned}
|\mathbf{z}^* - \mathbf{x}^*| &= |A(a)A^{-1}(y)(\mathbf{z}-\mathbf{x})|\\
&\le |\mathbf{z}-\mathbf{x}|\,\left\|(A^{-1}(y))^{\mathrm{T}} A^{\mathrm{T}}(a)A(a)A^{-1}(y)\right\|\\
&\le |\mathbf{z}-\mathbf{x}|\,\|\Sigma^{-1}(a)\|\,\|\Sigma^{-1}(y)\|\\
&\le c\delta,
\end{aligned}$$

for some c when $y \in A_\varepsilon$. Thus $\mathbf{z}^* \in B(\mathbf{x}^*, c\delta)$, so

$$\begin{aligned}
&\int_{A_\varepsilon}\int \sup\{|f(A(a)A^{-1}(y)\mathbf{z}_1) - f(A(a)A^{-1}(y)\mathbf{z}_2)|,\\
&\qquad\qquad \mathbf{z}_1, \mathbf{z}_2 \in B(\mathbf{x}, c_1 n^{-1/2})\}\Phi(d\mathbf{x})\,dh_n(y)\\
&\quad= \int_{A_\varepsilon}\int \sup\{|f(\mathbf{z}_1^*) - f(\mathbf{z}_2^*)|, \mathbf{z}_1^*, \mathbf{z}_2^* \in B(\mathbf{x}^*, cc_1 n^{-1/2})\}\Phi(d\mathbf{x})\,dh_n(y)\\
&\quad= \int_{A_\varepsilon}\int \omega_f(A(a)A^{-1}(y)\mathbf{x}, c_2 n^{-1/2})P\{\mathbf{N} \in d\mathbf{x}\}\,dh_n(y)\\
&\quad= \int_{A_\varepsilon}\int \omega_f(\mathbf{x}, c_2 n^{-1/2})P\{A(a)A^{-1}(y)\mathbf{N} \in d\mathbf{x}\}\,dh_n(y)\\
&\quad= \int_{A_\varepsilon}\int \omega_f(\mathbf{x}, c_2 n^{-1/2})P\{\mathbf{N} \in d\mathbf{x}\}\,dh_n(y)\\
&\qquad + \left[\int_{\delta_1}^{0} dh_n(a+y) + \int_0^{\delta_2} dh_n(b+y)\right] O(\eta \text{ or } c|y|),
\end{aligned}$$

where the last step follows by arguments like those following (2.9). Arguing as up to (2.10), the last bound is

$$\le \int \omega_f(\mathbf{x}, c_2 n^{-1/2})\Phi(d\mathbf{x}) + O(\eta \text{ or } n^{-1/2}),$$

which completes the proof. □

Remarks. In statistical practise, the more usual kind of trimming (in dimension 1) is "from above and below," i.e., the $[n\alpha]$ largest and $[n\beta]$ smallest terms are removed from the sum. In higher dimensions, though, trimming outside spheres or ellipses (in multivariate analysis) or cubes is commonly advocated, and occasionally in dimension 1 too, trimming is with regard to absolute values. The origin plays a special role in this procedure, which is not the case with trimming from above and below, and there results quite a striking difference in behaviour. The basic limit theorem for trimming above and below (with trimming proportions α and β) is due to Stigler (1973) [see also Bickel (1965)], and the limit law for the trimmed sum is not the normal but a mixture of normal distributions in the case where the upper (or lower) α (or β) quantile of the underlying distribution is not uniquely defined.

By contrast, Theorem 1 shows that "trimming absolutes" results in a normal limit, even with a $O(n^{-1/2})$ rate of convergence provided the random centering by $\mathbf{m}(\mathbf{X}_n^{[n\alpha]})$ is allowed. In fact, if F is symmetric with respect to $\{S(y)\}$ in the sense that $\mathbf{m}(\cdot) = 0$, this rate is always achieved if F satisfies the smoothness conditions (2.3), and, in addition,

$$\int \omega_f(\mathbf{x}, c_2 n^{-1/2})\Phi(d\mathbf{x}) = O(n^{-1/2}).$$

This holds, for example, when f is the indicator of a convex set in $\mathbb{R}^d$ [Sweeting (1977), page 40].

It is in replacing the random centering $\mathbf{m}$ by a constant centering, as would be required in practise, that the possibility of a nonnormal limit arises in our case, when the $(1-\alpha)$ quantile of h is not well defined. This is the subject of Theorem 2. After this, it is clear that rates of convergence of ${}^{[n\alpha]}S_n$ to its (possibly nonnormal) limit are entirely dependent on the smoothness of F at the α quantile.

The trimming proportion $[n\alpha]$ can be replaced throughout Theorems 1 and 2 with a sequence of integers r_n such that $r_n/n \to \alpha$. If restrictions are placed on the rate of decrease of the tail of F, r_n may go to $+\infty$ more slowly and still preserve a proper limit for ${}^{[n\alpha]}\mathbf{S}_n$ when normed and centered appropriately. In the extreme case, when a fixed number of terms r is trimmed ("light" trimming), no essential effect on the convergence of ${}^{(r)}S_n$ occurs [see Maller (1982) and Mori (1984)]. In an intermediate case, when F is stochastically compact, $r_n \to +\infty$ arbitrarily slowly is sufficient trimming to control the sum (but there is again a problem with centering). See Pruitt (1986) for the one-dimensional case (following a suggestion of Griffin) and Hahn, Kuelbs and Samur (1987) for higher dimensions (although the latter use a "hybrid" trimming of absolute values above a certain level not closely related to our form of trimming).

As mentioned, for the usual applications $\{S(y)\}$ will be spheres, ellipses or cubes but much more general shapes are covered by our treatment. For example, let $\mathbf{u} \in U$, the unit vectors in $\mathbb{R}^d$. Suppose $a(\mathbf{u})$ and $b(\mathbf{u})$, the $\alpha/2$ and $1-\alpha/2$ quantiles in direction $\mathbf{u}$, are uniquely defined, and define the convex sets

$$S_0(\alpha) = \{\mathbf{x}: a(\mathbf{u}) \le \mathbf{u}^{\mathrm{T}}\mathbf{x} \le b(\mathbf{u}), \text{ for all } \mathbf{u} \in U\}.$$

Then $S(y) = S_0(1/(y+1))$ is an admissible family for trimming as in Section 1.

Alternatively, we could take $S_0(\alpha) = \{\mathbf{u}a(\mathbf{u}): \mathbf{u} \in \mathbf{U}\}$ whose boundaries are the projectionwise quantiles. These sets, of course, depend on the distribution function F. See Maller (1988) for other "projection pursuit" methods.

THEOREM 2. *Assume* (1.2) *and let* $f \in B_1^d$. *Suppose* (2.3) *holds with* " = " *replacing* " ≤ " *and, in addition,* $\mu'_-(a)$ *and* $\mu'_+(b)$ *exist in the sense that*

$$\lim_{y \to 0+} \frac{\mu(a) - \mu(a-y)}{y} = \mu'_-(a), \tag{2.11a}$$

$$\lim_{y \to 0+} \frac{\mu(b+y) - \mu(b)}{y} = \mu'_+(b) \tag{2.11b}$$

are finite, where $\mu(y)$ *is defined in* (1.9). *Then*

$$\begin{aligned}\int f(\mathbf{x})P\Big\{A(a)\big({}^{[n\alpha]}\mathbf{S}_n - (n-[n\alpha])\mu(a)\big)&\\ \times(n-[n\alpha])^{-1/2} \in d\mathbf{x}\Big\}&\\ \to \int f(\mathbf{x})\int_0^\infty P\{\mathbf{N} + yA(a)\mu'_+(b) \in d\mathbf{x}\}&\\ \times dP\Big\{N(0,1) \le yh'_+(b)\alpha^{-1/2}(1-\alpha)^{-1}\Big\}&\\ + \int f(\mathbf{x})\int_{-\infty}^0 P\{\mathbf{N} + yA(a)\mu'_-(a) \in d\mathbf{x}\}&\\ \times dP\Big\{N(0,1) \le yh'_-(a)\alpha^{-1/2}(1-\alpha)^{-1}\Big\},&\end{aligned}$$

provided $\omega_f(\varepsilon, \Phi) \to 0$ *for every* $\varepsilon > 0$ *as* $n \to +\infty$.

PROOF. By (1.10), Lemma 2 and Corollary 3 of Sweeting (1977) again, we have

$$\begin{aligned}&\int f(\mathbf{x})P\Big\{A(a)\big({}^{[n\alpha]}\mathbf{S}_n - (n-[n\alpha])\mu(a)\big)(n-[n\alpha])^{-1/2} \in d\mathbf{x}\Big\}\\ &\quad= \int_0^\infty \int f(\mathbf{x})P\Big\{A(a)\big(\mathbf{S}_{n-[n\alpha]}(y) - (n-[n\alpha])\mu(a)\big)\\ &\qquad\qquad \times(n-[n\alpha])^{-1/2} \in d\mathbf{x}\Big\}\, dh_n(y)\\ &\quad= \int_{A_\varepsilon}\int f^*(\mathbf{x})P\Big\{A(y)\big(\mathbf{S}_{n-[n\alpha]}(y) - E\mathbf{S}_{n-[n\alpha]}(y)\big)\\ &\qquad\qquad \times(n-[n\alpha])^{-1/2} \in d\mathbf{x}\Big\}\, dh_n(y) + O(n^{-1/2})\\ &\quad= \int_{A_\varepsilon}\int f^*(\mathbf{x})P\{\mathbf{N} \in d\mathbf{x}\}\, dh_n(y) + o(1), \qquad n \to +\infty,\end{aligned}$$

provided $f^* \in B_1^d$ and

$$\int_{A_\varepsilon} \omega_{f^*}(\varepsilon, \Phi)\, dh_n(y) \to 0,$$

where

$$f^*(\mathbf{x}) = f\{A(a)A^{-1}(y)\mathbf{x} + (n - [n\alpha])^{1/2} A(y)(\mu(y) - \mu(a))\}.$$

Deferring ω_{f^*} until the end of the proof, we deal with

$$\int_{A_\varepsilon}\int f^*(\mathbf{x}) P\{N \in d\mathbf{x}\}\, dh_n(y)$$

$$= \int_{A_\varepsilon}\int f(\mathbf{x}) P\{A(a)A^{-1}(y)\mathbf{N} + (n - [n\alpha])$$

$$\times A(a)(\mu(y) - \mu(a)) \in dx\}\, dh_n(y).$$

By (2.11), given η arbitrarily small,

$$|\mu(a + y) - \mu(a) - y\mu'_-(a)| \le \eta|y|, \qquad -\delta_1 \le y \le 0,$$

$$|\mu(b + y) - \mu(b) - y\mu'_+(b)| \le \eta y, \qquad 0 \le y \le \delta_2,$$

where δ_1, δ_2 are as in the proof of Theorem 1. For $y < 0$,

$$\int f(\mathbf{x}) P\{A(a)A^{-1}(a + y)\mathbf{N} + (n - [n\alpha])^{1/2} A(a)(\mu(a + y) - \mu(a)) \in d\mathbf{x}\}$$

$$= \int f(\mathbf{x}) P\{\mathbf{N} + y(n - [n\alpha])^{1/2} A(a)\mu'_-(a) \in d\mathbf{x}\} + O\big(\eta|y|(n - [n\alpha])^{1/2}\big),$$

by arguments like those of Theorem 1. It was also shown there that

$$\int_{-\delta_1}^{0} |y|\, dh_n(a + y) = O(n^{-1/2}), \qquad n \to +\infty,$$

so

$$\int_{-\delta_1}^{0} O\big(\eta|y|(n - [n\alpha])^{1/2}\big)\, dh_n(y) = O(\eta),$$

i.e., is arbitrarily small. Similar arguments hold for $0 \le y \le \delta_2$.

So we need to look at

$$\int_0^{\delta_2}\int f(\mathbf{x}) P\{\mathbf{N} + y(n - [n\alpha])^{1/2} A(a)\mu'_+(b) \in dx\}\, dh_n(y)$$

and the same integral over $-\delta_1 \le y \le 0$ but with $\mu'_-(a)$ replacing $\mu'_+(b)$.

For $0 \le y \le \delta_1$,

$$h_n\big(b + y(n - [n\alpha])^{-1/2}\big) \to P\{N(0,1) \le y h'_+(b)\alpha^{-1/2}(1 - \alpha)^{-1}\},$$

by Lemma 1. So the weak convergence

$$\int_0^{\delta_2(n-[n\alpha])^{1/2}} \int f(\mathbf{x}) P\{\mathbf{N} + yA(a)\mu'_+(b) \in d\mathbf{x}\}\, dh_n\big(b + y(n - [n\alpha])^{-1/2}\big)$$

$$\to \int_0^{\infty} \int f(\mathbf{x}) P\{\mathbf{N} + yA(a)\mu'_+(b) \in dx\}$$

$$\times dP\big\{N(0,1) \leq yh'_+(b)\alpha^{-1/2}(1-\alpha)^{-1}\big\}$$

holds, since the inner integral is a continuous bounded function of y. Similar arguments hold for $-\delta_1 \leq y \leq 0$.

To complete the proof, we have to show

$$\int_{A_\varepsilon} \omega_{f^*}(\varepsilon, \Phi)\, dh_n(y) \to 0,$$

if $\omega_f(\varepsilon, \Phi) \to 0$. This is done as in Theorem 1; we omit the details. □

REMARKS. (i) Theorem 2 shows that $({}^{[n\alpha]}\mathbf{S}_n - (n - [n\alpha])\mu(a))/(n - [n\alpha])^{1/2}$ is asymptotically a mixture of normal random variables if left and right derivatives of h and μ exist at a and b, and (1.2) holds. If $a = b$ and the derivatives of h and μ exist at a, and $h'(a) > 0$, then the mixture of normals reduces to a normal, and so the distribution of $\{{}^{[n\alpha]}\mathbf{S}_n - (n - [n\alpha])\mu(a)\}/(n - [n\alpha])^{1/2}$ converges to a normal distribution with mean 0 and covariance matrix

$$\Sigma(a) + \alpha(1-\alpha)^2 \mathbf{m}'(a)[\mathbf{m}'(a)]^{\mathrm{T}} / (h'(a))^2.$$

(ii) The results of Maller (1982) and Mori (1984) on "light" trimming, i.e., when the number of extreme terms is fixed as $n \to \infty$, can be generalised to the present setup, and shown that there is "no effect" of light trimming in the sense that the conditions required on the distribution are the same with or without trimming.

(iii) For rates of convergence in dimension 1, see Hall (1984) ("light" trimming) and de Wet (1976) and Egorov and Nevzorov (1975) ("heavy" trimming).

Acknowledgment. I am grateful to a referee for advice and reference to Weiss (1970).

REFERENCES

BHATTACHARYA, R. N. (1977). Refinements of the multidimensional central limit theorem and applications. *Ann. Probab.* **5** 1–27.

BICKEL, P. J. (1965). On some robust estimates of location. *Ann. Math. Statist.* **36** 847–858.

BJERVE, S. (1977). Error bounds for linear combinations of order statistics. *Ann. Statist.* **5** 357–369.

DE WET, T. (1976). Berry–Esseen results for the trimmed mean. *South African Statist. J.* **10** 77–96.

EGOROV, V. A. and NEVZOROV, V. B. (1975). On the rate of convergence of linear combinations of absolute order statistics to the normal law. *Theory Probab. Appl.* **20** 203–211.

HAHN, M. G., KUELBS, J. and SAMUR, J. D. (1987). Asymptotic normality of trimmed sums of ϕ-mixing random variables. *Ann. Probab.* **15** 1395–1418.

HALL, P. (1984). On the influence of extremes on the rate of convergence in the central limit theorem. *Ann. Probab.* **12** 154–172.

MALLER, R. A. (1982). Asymptotic normality of lightly trimmed means—a converse. *Math. Proc. Cambridge Philos. Soc.* **92** 535–545.

MALLER, R. A. (1988). Some consistency results on projection pursuit estimates of location and scale. Preprint.

MORI, T. (1984). On the limit distributions of lightly trimmed sums. *Math. Proc. Cambridge Philos. Soc.* **96** 507–516.

PRUITT, W. E. (1986). Sums of independent random variables with the extreme terms excluded. Preprint.

REISS, R. D. (1976). Asymptotic expansions for sample quantiles. *Ann. Probab.* **4** 249–258.

STIGLER, S. M. (1973). The asymptotic distribution of the trimmed mean. *Ann. Statist.* **1** 472–477.

SWEETING, T. J. (1977). Speeds of convergence for the multidimensional central limit theorem. *Ann. Probab.* **5** 28–41.

WEISS, L. (1970). Asymptotic distributions of quantiles in some nonstandard cases. In *Nonparametric Techniques in Statistical Inference* (M. L. Puri, ed.) 343–348. Cambridge Univ. Press, London.

CSIRO DIVISION OF MATHEMATICS AND STATISTICS
PRIVATE BAG
P.O. WEMBLEY
WESTERN AUSTRALIA 6014
AUSTRALIA