JOURNAL OF MULTIVARIATE ANALYSIS **63**, 136–179 (1997)
ARTICLE NO. MV971698

Random Deletion Does Not Affect Asymptotic Normality or Quadratic Negligibility

Harry Kesten and R. A. Maller

Cornell University and the University of Western Australia, Nedlands, Australia

Suppose a number of points are deleted from a sample of random vectors in $\mathbb{R}^d$. The number of deleted points may depend on the sample size n, and on any other sample information, provided only that it is bounded in probability as $n \to \infty$. In particular, "extremes" of the sample, however defined, may be deleted. We show that this operation has no effect on the asymptotic normality of the sample sum, in the sense that the sum of the deleted sample is asymptotically normal, after norming and centering, if and only if the sample sum itself is asymptotically normal with the same norming and centering as the deleted sum. That is, the sample must be drawn from a distribution in the domain of attraction of the multivariate normal distribution. The domain of attraction concept we employ uses general operator norming and centering, as developed by Hahn and Klass. We also show that random deletion has no effect on the "quadratic negligibility" of the sample. These are conditions that are important in the robust analysis of multivariate data and in regression problems, for example.

1. INTRODUCTION

In this paper we bring together two strands of research concerned with the asymptotic behaviour of estimates from a multivariate sample. The first of these relates to the random deletion of points from the sample and the effect this operation has on the large sample properties of the estimators. This question arises, for example, when a sample is "trimmed" by the deletion of the sample extremes, however they may be defined for multivariate observations—usually with the hope that the behaviour of estimates such as the sample mean and covariance matrix may be "improved" in this way. Thus the trimmed sample mean may be asymptotically normal, after centering and norming, even though the original observations were not

Received February 14, 1996; revised June 2, 1997.

AMS 1991 subject classifications: primary 60F05; secondary 60E20, 62H12.

Key words and phrases: asymptotic normality, random deletion of observations, trimming, extreme values, quadratic negligibility, operator norming, sums and maxima of random variables, sum of squares and products matrix, covariance matrix.

0047-259X/97 $25.00

drawn from a distribution in the domain of attraction of the normal distribution.

Indeed, Csörgő, Horváth, and Mason [3] (see also Griffin and Pruitt [9] and the references in both these articles) showed that in the one-dimensional case the trimmed sum of i.i.d. random variables in the domain of attraction of a stable law is asymptotically normal provided that the number of trimmed terms goes to infinity with the sample size. On the other hand, Maller [17] and Mori [21] proved that this cannot occur (still in the one-dimensional case) when the number of extremes removed from the sample stays bounded. More precisely, they proved that removal of a fixed number, r, say, of extremes, does not result in asymptotic normality of the trimmed sample mean unless the distribution F from which the sample was drawn is itself in the domain of attraction of the normal distribution. Here we generalize this result to a multivariate distribution F and to a much more general deletion operation; the Maller–Mori result holds if any $r = r(n)$ points are removed from a sample of size n, provided $r(n)$ is bounded in probability as $n \to \infty$ in the sense of (1.4) below.

To specify our setup, let X, X_i, be independent and identically distributed (i.i.d.) random vectors in $\mathbb{R}^d$ with distribution F, $i \geqslant 1$. To allow for randomized rules (e.g., to break ties when equal observations are possible), we assume without loss of generality that some other random variables $W_1, W_2, \ldots$ are defined on the same probability space as the X_i and that these are independent of the X_i. These W's take values in some measurable space. The deletion scheme is determined by a random integer $r = r(n)$ and the random sets

$$I(n) = \{i_n(1), \ldots, i_n(r(n))\} \tag{1.1}$$

of the indices of the observations to be removed. The $I(n)$ are measurable with respect to the σ-field on our probability space. The random number r of observations deleted may depend on the sample size n, and as a special case it may be nonstochastic. In particular, we allow $r = 0$ (in which case $I(n) = \phi$ and no observations are deleted.)

We define the *deleted sum* as

$$^{D}S_n = \sum_{i \notin I(n)} X_i = X_1 + \cdots + X_n - X_{i_n(1)} - \cdots - X_{i_n(r)}. \tag{1.2}$$

When no points are deleted, we have the usual sample sum

$$S_n = \sum_{i=1}^{n} X_i. \tag{1.3}$$

The only restriction on the deletion scheme is the following:

Condition 1. $I(n) \subseteq \{1, 2, ..., n\}$ and for some nonstochastic integer s

$$P\{r(n) \leqslant s\} \to 1 \qquad \text{as} \quad n \to \infty. \tag{1.4}$$

Then we show in Theorems 2.1 and 2.2 of the next section that ${}^{D}S_n$ is asymptotically normal in a very general sense due to Hahn and Klass [10], if and only if the same is true of S_n (with the same norming matrix and centering vector).

The second strand we bring in relates to the "quadratic negligibility" of the X_i. This term was coined by Maller [18] to describe the following behaviour. Suppose from now on that the X_i are "full." Here a random vector X, or its distribution F, is called *full* if $P\{X \in A + H\} < 1$ for any vector $A \in \mathbb{R}^d$ and subspace H of $\mathbb{R}^d$ of dimension $d-1$ or lower. (Note that a full distribution cannot be concentrated on one point, even if $d=1$). Define

$$V_n = \sum_{i=1}^{n} X_i X_i^{\mathrm{T}}, \tag{1.5}$$

the sample sum of squares and products matrix, where the superscript T denotes vector or matrix transpose. Lemma 2.3 of Maller [18] shows that V_n is nonsingular with probability approaching 1 as $n \to \infty$. We say that the X_i are *quadratically negligible* if, as $n \to \infty$,

$$\max_{1 \leqslant i \leqslant n} X_i^{\mathrm{T}} V_n^{-1} X_i \xrightarrow{P} 0. \tag{1.6}$$

This concept can be motivated by thinking of the X_i as being values of regressor variables. The quantity in (1.6) is then the diagonal element of the "hat" matrix and measures the "leverage" or "influence" of the ith regressor. In certain regression models, (1.6) is necessary and sufficient for the asymptotic normality of the least squares regression coefficients (Huber [13, p. 159]) and, as discussed by Huber [13, p. 162] and Belsley, Kuh, and Welsch [2, Chap. 2], a large influence is to be avoided or at least noted. Common practice, in fact, is to delete an observation with a large influence from the sample and to recalculate the regression coefficients from the deleted sample; see, e.g., Belsley, Kuh, and Welsch [2, p. 11]. We may then iterate the procedure and delete, 2, 3, ..., r points. But are the remaining (undeleted) X_i still quadratically negligible?

This question provided the motivation for our second formulation. Define the deleted sum of squares and products matrix

$${}^{D}V_n = \sum_{i \notin I(n)} X_i X_i^{\mathrm{T}}. \tag{1.7}$$

(Summations as in (1.7) will be over values of i, $1 \leqslant i \leqslant n$, not including the indices in $I(n)$.) Under the above assumptions (including (1.4)), we will show that ${}^{D}V_n^{-1}$ exists with probability approaching 1 as $n \to \infty$, that

$$\max_{i \notin I(n)} X_i^{\mathrm{T}}\, {}^{D}V_n^{-1} X_i \xrightarrow{P} 0 \tag{1.8}$$

if and only if (1.6) holds and, furthermore, that these are both equivalent to the asymptotic normality, in the sense of Hahn and Klass [10] of ${}^{D}S_n$ and of S_n. Thus we answer the above questions and tie the two strands together in a satisfying way.

An attractive feature of our use of "deletions" is that the results apply in particular for extremes of a multivariate sample, however defined, as long as Condition 1 is satisfied. This suggests an application of our results to the treatment of outliers in multivariate (or even univariate) data. There is a large literature on this. See Barnett and Lewis [1, especially Chaps. 6, 7], where many procedures for detecting and dealing with outliers are summarized. A general theme is to take as null hypothesis, H_0, that the observed data represent observations on i.i.d. random vectors X_i. One is interested in inference on some aspect of the distribution of the X_i, say on its mean μ (which is assumed to exist). Let the sample mean and the mean of the deleted sample be

$$\bar{X}_n = \frac{S_n}{n}, \qquad {}^{D}\bar{X}_n = \frac{{}^{D}S_n}{n-r}, \tag{1.9}$$

and let the corresponding mean-corrected sum of squares and products matrices be

$$\bar{V}_n = \sum_{i=1}^{n} (X_i - \bar{X}_n)(X_i - \bar{X}_n)^{\mathrm{T}},$$

$${}^{D}\bar{V}_n = \sum_{i \notin I(n)} (X_i - {}^{D}\bar{X}_n)(X_i - {}^{D}\bar{X}_n)^{\mathrm{T}}. \tag{1.10}$$

One may reject H_0 if, for example, the test statistic

$$R_n = (X_{(n)} - \bar{X}_n)^{\mathrm{T}}\, \bar{V}_n^{-1} (X_{(n)} - \bar{X}_n) \tag{1.11}$$

is too large, where $X_{(n)}$ is an X_i for which $(X_i - \bar{X}_n)^{\mathrm{T}}\, \bar{V}_n^{-1}(X_i - \bar{X}_n)$ is maximized, and declare $X_{(n)}$ *discordant* in this case (see Barnett and Lewis [1, Chap. 7]) which also lists a variety of other test statistics). $X_{(n)}$ is then deleted from the sample. This specifies a deletion scheme according to our formulation. If H_0 is rejected, Barnett and Lewis, and practitioners, use a quantity such as

$$ {}^{D}M_n = ({}^{D}\bar{X}_n - \mu)^{\mathrm{T}}\, {}^{D}\bar{V}_n^{-1}({}^{D}\bar{X}_n - \mu), \tag{1.12} $$

based on the deleted sample, for inference on μ. (See, e.g., Barnett and Lewis [1, p. 274]). Commonly it is assumed that the deletion of discordant observations renders ${}^{D}\bar{X}_n$ approximately normally distributed, and ${}^{D}M_n$ approximately distributed as chi-squared with d degrees of freedom (χ_d^2), at least in large samples. Hampel *et al.* [12, p. 59] criticize this procedure on the grounds that "It appears as if subsequent statistical analysis is to proceed as if the parametric model would hold perfectly for the remaining data." In other words, practitioners usually act as if, after deleting outliers, the deleted sample can be treated as a sample of i.i.d. random variables with the same distribution as X.

This cannot be exactly true, of course, but it may be approximately so, and our results partially justify the procedure. Suppose H_0, in fact, specifies that the X_i are i.i.d. in the domain of attraction of the multivariate normal distribution (which is more general than assuming them normally distributed) and suppose that, in deleting $X_{(n)}$, a Type I error occurs, so that H_0 is erroneously rejected. Theorem 2.1 below can then be applied to deduce that ${}^{D}M_n$ is still approximately χ_d^2 in large samples. Consequently, no damage has been done, at least asymptotically. On the other hand, again as an application of Theorem 2.1, the converse is also true: the deleted mean ${}^{D}\bar{X}_n$ of the i.i.d. sample is asymptotically normal, after norming and centering, only if the X_i are in the domain of attraction of the normal. If they are not, we cannot achieve asymptotic normality of ${}^{D}\bar{X}_n$ by any deletion scheme satisfying Condition 1. This answers to some extent the question implicitly posed by Barnett and Lewis [1, p. 86], as to what range of distributions one can allow, while still obtaining asymptotic normality of a deleted quantity such as ${}^{D}\bar{X}_n$.

Alternatively one can think of (part of) Theorem 2.1 as a statement about the distribution of the trimmed sample mean when a few observations are discarded because they are *erroneously* believed to come from another distribution than the remaining observations.

We will also show that (1.8) remains equivalent to (1.6) if the X_i are centered at ${}^{D}\bar{X}_n$ and ${}^{D}V_n$ is replaced by ${}^{D}\bar{V}_n$, as is commonly done in practice.

2. RESULTS

Hahn and Klass [10] showed that S_n is asymptotically normal in the sense that *there exist nonstochastic vectors A_n and square matrices B_n such that*

$$B_n(S_n - A_n) \Rightarrow N(0, I) \tag{2.1}$$

(converges in distribution to the standard d-dimensional normal distribution), if and only if a certain analytic criterion ((2.7) below) is satisfied. For further background and discussion see also Hahn and Klass [11], Maller [18], and Meerschaert [19, 20]. Let us just mention here that (2.1) is truly d-dimensional, in that, in general, B_n may not be taken as diagonal; we may need to rotate S_n, after centering by A_n, in a direction which depends on the sample size n. Thus, even after centering, S_n need not "settle down" in an approximately fixed direction. The same will be seen to be true of the deleted sum ${}^DS_n$. (See also Remark (iii) below.)

THEOREM 2.1. *Suppose that F is full. For any deletion scheme D satisfying Condition* 1, *the following are equivalent*:

there are nonstochastic vectors A_n and square matrices B_n such that

$$B_n({}^DS_n - A_n) \Rightarrow N(0, I); \tag{2.2}$$

there are nonstochastic vectors A_n and square matrices B_n such that

$$B_n\left({}^DS_n - \frac{n-r}{n} A_n\right) \Rightarrow N(0, I); \tag{2.3}$$

there are nonstochastic square matrices B_n such that

$$B_n {}^D\bar{V}_n B_n^{\mathrm{T}} \xrightarrow{P} I; \tag{2.4}$$

${}^DV_n$ *is invertible with probability approaching* 1 *and*

$$\max_{i \notin I(n)} X_i^{\mathrm{T}} {}^DV_n^{-1} X_i \xrightarrow{P} 0; \tag{2.5}$$

${}^D\bar{V}_n$ *is invertible with probability approaching* 1 *and*

$$\max_{i \notin I(n)} (X_i - {}^D\bar{X}_n)^{\mathrm{T}} {}^D\bar{V}_n^{-1}(X_i - {}^D\bar{X}_n) \xrightarrow{P} 0; \tag{2.6}$$

$$\sup_u \frac{x^2 P\{|u^TX| > x\}}{E((u^TX)^2 \mathbb{1}\{|u^TX| \leqslant x\})} \to 0 \qquad \textit{as} \quad x \to \infty. \tag{2.7}$$

A supremum over directions u, as in (2.7), will be over all unit vectors in S^{d-1}, the unit sphere in $\mathbb{R}^d$. I is the $d\times d$ identity matrix in (2.2)–(2.4), and we use the notation $\mathbb{1}\{A\}$ for the indicator of an event A. The convergence in (2.2)–(2.6), and similar convergences elsewhere, is as $n\to\infty$. We say that a sequence of events A_n occurs with probability approaching 1 if $P(A_n)\to 1$. $N(\mu, \Sigma)$ will denote a normal random vector in d dimensions with mean vector μ and covariance matrix Σ.

Since (2.1) and (2.7) are equivalent, by Hahn and Klass [10], all conditions of Theorem 2.1 are equivalent to any of the conditions (1.5)–(1.10) of Maller [18], which, except for (1.7) of Maller [18], are the versions of (2.2)–(2.6) corresponding to the deletion of no points. Maller [18, Theorem 2.1] shows that $E|X|<\infty$ under (2.1), or, equivalently, under (2.7), and we can use this to show that A_n in (2.2)–(2.3) may be taken as nEX. Recall that r is a random variable, in general, so the centering in (2.3) is random. Nevertheless (2.3) is equivalent to (2.2) (and to (2.1)).

An important component in the proof of Theorem 2.1 will be the following result. Consider a triangular array

$$X_{n,1}, X_{n,2}, ..., X_{n,t(n)}, \qquad n\geqslant 1,$$

of d-vectors with $t(n)\to\infty$ and $\{X_{n,i}, 1\leqslant i\leqslant t(n)\}$ being i.i.d. for each n, but not necessarily having the same distribution for different values of n. $X_{n,i}$ need not be full. Define

$$T_n=\sum_{i=1}^{t(n)} X_{n,i}. \tag{2.8}$$

Let $W_1, W_2, ...$ be some random variables which are independent of the $\{X_{n,i}, 1\leqslant i\leqslant t(n)\}$ and let $I(n)=\{i_n(1), ..., i_n(r(n))\}$ be a set of random indices which is defined on the same probability space as $\{X_{n,i}, 1\leqslant i\leqslant t(n)\}$ and $W_1, W_2, ...$ Assume that

$$I(n)\subset\{1, 2, ..., t(n)\} \tag{2.9}$$

and that for some nonstochastic integer s

$$P\{r(n)\leqslant s\}\to 1. \tag{2.10}$$

Analogously to the preceding we define the deleted sum

$${}^{D}T_n=\sum_{i\notin I(n),\, i\leqslant t(n)} X_{n,i}=T_n-X_{n,i_n(1)}-\cdots-X_{n,i_n(r)}. \tag{2.11}$$

THEOREM 2.2. *Assume that the deletion scheme satisfies* (2.9) *and* (2.10) *and that for some sequence of integers* $k_1 < k_2 < \cdots$ *and nonstochastic* $C_1, C_2, ..., \in \mathbb{R}^d$,

$$ {}^{D}T_{k_n} - C_{k_n} \Rightarrow N(\mu, \Sigma), \tag{2.12}$$

where the covariance matrix Σ *may be singular. Then there exist nonstochastic vectors* $F_1, F_2, ..., \in \mathbb{R}^d$ *so that*

$$\max_{1 \leqslant i \leqslant t(k_n)} |X_{k_n, i} - F_{k_n}| \xrightarrow{P} 0. \tag{2.13}$$

If $C_{k_n} = 0$ *or if* $X_{n,i} = B_n X_i$ *for some sequence* $X_1, X_2, ...$ *of i.i.d. random d-vectors with a full distribution and nonstochastic* $d \times d$ *matrices* B_n, *then we may take* $F_{k_n} = 0$ *and in this case also*

$$ {}^{\tilde{D}}T_{k_n} - C_{k_n} \Rightarrow N(\mu, \Sigma) \tag{2.14}$$

for any other deletion scheme $\tilde{D}$ *which satisfies* (2.9) *and* (2.10).

Remarks. (i) Theorem 2.2 provides the basic tools we need for our analysis of deletion schemes. Note that (2.14) is immediate from (2.12) and (2.13) with $F_{k_n} = 0$, when (2.10) holds. (2.13) is related to, in fact is stronger than, "(uniform) asymptotic constancy" of the sequence $\{X_{k_n, i}\}_{1 \leqslant i \leqslant t(k_n)}$ (see, e.g., Gnedenko and Kolmogorov [7, p. 95]). When $F_{k_n} = 0$ it implies "uniform asymptotic negligibility" of the sequence. Thus these conditions are consequences of the convergence of the centered, deleted sum to normality as specified in (2.12). Uniform asymptotic negligibility of summands is an important ingredient in the theory of convergence of sums of triangular arrays to infinitely divisible laws. (Gnedenko and Kolmogorov [7, p. 95] use the terminology "infinitesimal" for this.)

(ii) In the situation where (2.14) applies we can take for $\tilde{D}$ the deletion scheme which deletes no observations. In this case we see that convergence of the deleted sum as in (2.12) is equivalent to that of the undeleted sum, after centering, to normality. Also, if we let $C_n = 0$ and

$$X_{n,i} = B_n \left(X_i - \frac{1}{n} A_n \right), \qquad 1 \leqslant i \leqslant n,$$

where B_n are nonsingular $d \times d$ matrices, A_n are constant vectors, and X_i are i.i.d. vectors in $\mathbb{R}^d$, then (2.14) of Theorem 2.2 asserts that when

$$B_{k_n} \left({}^{D}S_{k_n} - \left(\frac{k_n - r}{k_n} \right) A_{k_n} \right) \Rightarrow N(0, \Sigma) \tag{2.15}$$

holds for some sequence $k_n \uparrow \infty$, then we also have

$$B_{k_n}(S_{k_n} - A_{k_n}) \Rightarrow N(0, \Sigma). \tag{2.16}$$

Here ${}^{D}S_n$ and S_n are defined by (1.2) and (1.3). In this form Theorem 2.2 states that the attraction or "partial attraction" (subsequential convergence) of S_n, after norming and centering, to normality, is not affected by deletion of a bounded number of points. See Kesten [14] for an analysis in one dimension of trimming in relation to domains of attraction of nonnormal stable laws.

(iii) If (2.15) holds with $B_{k_n} = \gamma_{k_n} B$ for some fixed, nonsingular matrix B and real constants γ_{k_n}, then the same is true for (2.16). Similarly if (2.2) holds with $B_n = \gamma_n B$ for some deletion scheme D, then it holds for these same B_n for all D which satisfy Condition 1 (see next remark). Thus Theorems 2.1 and 2.2 include as a special case the case of vectors in the classical domain of attraction of the d-dimensional normal distribution, which do not call for the general operator norming of Hahn and Klass [10] with its allowance for (possibly) different rotations for different n.

(iv) (2.7) does not depend on D. Thus, if any of (2.2)–(2.6) hold for one deletion scheme satisfying Condition 1 then (2.2)–(2.6) hold for all such deletion schemes *with the same choice for* A_n *and* B_n *for all such deletion schemes.* The fact that the same A_n and B_n can be used for different schemes in (2.2) and (2.3) follows from the implication from (2.12) to (2.14) when $X_{n,i} = B_n X_i$ for full X_i. The lack of dependence of B_n on D in (2.4) can be obtained from (4.2) and (4.9) below, but we shall not give details.

(v) As an application of Theorem 2.1 we have that, when Condition 1 and any of (2.2)–(2.7) hold,

$${}^{D}\bar{V}_n^{-1/2}({}^{D}S_n - n\mu) \Rightarrow N(0, I). \tag{2.17}$$

Here ${}^{D}\bar{V}_n^{-1/2}$ is either the Cholesky or the symmetric square root of ${}^{D}\bar{V}_n^{-1}$. The convergence in (2.17) follows from (2.2), (2.4), and Theorem 2.1 of Vu, Maller, and Klass [23] and the observation made above that we may replace A_n by nEX in (2.2). (Vu, Maller, and Klass [23] define and briefly discuss the merits of the Cholesky and symmetric square roots of a matrix.) The "studentised" result (2.17) answers the question raised in Remark (viii) of Maller [18], and generalizes it to deletion schemes satisfying (1.4). It is immediate from (2.17) that

$$({}^{D}S_n - n\mu)^{\mathrm{T}}\,({}^{D}\bar{V}_n^{-1})({}^{D}S_n - n\mu) \Rightarrow \chi_d^2, \tag{2.18}$$

generalizing (1.15) of Maller [18] to a “deleted” version of the Hotelling T^2 statistic. It would be interesting to know whether (2.17) or (2.18) implies (2.2) and hence (2.1). In one dimension, and when there is no deletion, this is true and shown by Griffin and Mason [8] and Giné, Götze, and Mason [6].

We conclude this section with two examples. The first shows that ordering the X_i by their moduli, perhaps for the purpose of locating outliers, may have quite different implications from ordering them by their Mahalanobis distance from 0. The second example, in response to a question from a referee, shows that (2.10) cannot be weakened to requiring only that $r=r(n)$ be tight, rather than that $r(n)$ be bounded in probability, in Theorem 2.2.

Suppose $X_n^{(1)}$ denotes any observation whose length is $\max_{1\leqslant i\leqslant n}|X_i|$. By taking for D the deletion scheme which removes no observations, (2.5) immediately implies

$$(X_n^{(1)})^{\mathrm{T}}\, V_n^{-1} X_n^{(1)} \xrightarrow{P} 0. \tag{2.19}$$

Since $X_n^{(1)}$ is the sample point most distant from 0, we might expect (2.19) to imply (2.5). This, however, is not true.

Example 2.3. We may have (2.19) holding, even though

$$\max_{1\leqslant i\leqslant n} X_i^{\mathrm{T}} V_n^{-1} X_i \text{ does not converge to 0 in probability.} \tag{2.20}$$

We must of course have $d\geqslant 2$ in Example 2.3.

Example 2.4. When the X_i are i.i.d. symmetric random variables in $\mathbb{R}$, we may have ${}^{D}S_{k_n}/B_{k_n} \Rightarrow N(0,1)$ for some nonstochastic sequences $k_n\to\infty$ and B_{k_n}, and for some deletion scheme satisfying (2.9), with $r=r(n)$ tight rather than satisfying (2.10), but with S_{k_n}/B_{k_n} not converging to normality as $k_n\to\infty$.

The proofs of Examples 2.3 and 2.4 are in Section 5. Sections 3 and 4 give the proofs of Theorems 2.2 and 2.1, in that order, since some of the proof of Theorem 2.1 uses the results of Theorem 2.2. Some heuristics for the proofs are given at the beginnings of Sections 3 and 4.

3. PROOF OF THEOREM 2.2.

We begin with some heuristics. The idea is that if (2.13) fails, then some tail of the distribution of ${}^{D}T_{k_n}-C_{k_n}$ will be fatter than the corresponding tail of a normal distribution. This is best understood by considering the one-dimensional case with $C_n = 0$ and $k_n = n$. In this case we wish to prove (2.13) with $F_n = 0$, that is,

$$\max_{1 \leqslant i \leqslant n} |X_{n,i}| \xrightarrow{P} 0. \tag{3.1}$$

Assume that this fails. Some simple reductions then lead to the existence of an L_n and $\varepsilon > 0$, $\eta > 0$ for which

$$L_n \geqslant \varepsilon, \qquad P\{|X_{n,i}| > L_n\} \leqslant \frac{\eta}{t(n)}, \qquad \frac{\eta}{2t(n)} \leqslant P\{X_{n,i} > L_n\} \leqslant \frac{\eta}{t(n)}.$$

But then also, for large n,

$$P\{X_{n,i} > L_n \text{ for exactly } m \text{ values of } i \leqslant t(n) \text{ and } X_{n,i} \geqslant -L_n$$
$$\text{for all other } i \leqslant t(n)\} \geqslant \frac{1}{2m!} e^{-\eta} \left(\frac{\eta}{2}\right)^m. \tag{3.2}$$

Now consider a sample point for which the event on the left-hand side of (3.2) occurs and from which we delete at most s observations. For such a sample point

$$\begin{aligned} {}^D T_n &\geqslant \sum_{i \notin I(n)} X_{n,i} \mathbb{1}\{|X_{n,i}| \leqslant L_n\} + (m-s) L_n \\ &\geqslant \sum_{i \leqslant t(n)} X_{n,i} \mathbb{1}\{|X_{n,i}| \leqslant L_n\} + (m-2s) L_n. \end{aligned}$$

A separate argument will show that

$$\sum_{i \leqslant t(n)} X_{n,i} \mathbb{1}\{|X_{n,i}| \leqslant L_n\} \tag{3.3}$$

is with a probability tending to 1 not too large with respect to L_n; roughly speaking this is done by comparing the sum in (3.3) with ${}^D T_n$ when $|X_{n,i}| \leqslant L_n$ for all $i \leqslant t(n)$. This finally leads to

$$\liminf_{n \to \infty} P\left\{{}^D T_n \geqslant \frac{1}{4} m\varepsilon\right\} \geqslant \liminf_{n \to \infty} P\left\{{}^D T_n \geqslant \frac{1}{4} m L_n\right\} \geqslant \frac{1}{4} \frac{(\frac{1}{4}\eta)^m}{m!} e^{-\eta}. \tag{3.4}$$

This contradicts the asymptotic normality of ${}^D T_n$, because

$$\frac{1}{4} \frac{(\frac{1}{4}\eta)^m}{m!} e^{-\eta} \text{ is much bigger than } \frac{1}{\sigma\sqrt{2\pi}} \int_{(1/4)m\varepsilon}^{\infty} e^{-t^2/2\sigma^2}\, dt$$

for any fixed σ, as m becomes large. This contradiction will show that (3.1) cannot fail.

We break the proof up into a number of steps. After some technical reductions, the first step proves the variance estimate (3.20). The second step proves the Chebyshev-type estimate (3.29). The third step, which is the principal one, proves the property (3.14) of the quantiles of $|X_{n,i}|$. When $C_n = 0$ this completes the proof. For general C_n we complete the proof of (2.13) in step (iv) by proving (3.15). The final statement of the theorem is proven in step (v).

We now start the proof proper with some reductions. Let (2.12) hold. By ignoring the rows with numbers other than $k_1, k_2, \ldots$, we may, and shall, assume that $k_n = n$. We further may assume without loss of generality that the deletion scheme satisfies

$$P\{r(n) \leqslant s\} = 1 \tag{3.5}$$

which is stronger than (2.10). This is so because (2.10) allows us to replace $I(n)$ by $\{i_n(1), i_n(2), \ldots, i_n(r \wedge s)\}$, which contains at most s indices. For the remainder of this proof we shall assume that (3.5) holds.

By choosing a further subset of the rows we may further assume that either

$$\lim_{n \to \infty} P\{r(n) = s_0\} = 1 \qquad \text{for some fixed} \quad s_0 \in \{0, 1, 2, \ldots, \}, \tag{3.6}$$

or

$$\liminf_{n \to \infty} P\{r(n) = s_1\} \wedge P\{r(n) = s_2\} > 0$$
$$\text{for two distinct} \quad s_1, s_2 \in \{0, 1, 2, \ldots, s\}. \tag{3.7}$$

(We allow $s_0 = 0$ in (3.6), corresponding to the case of no deletion.) If (3.6) holds then

$$P\left\{{}^D T_n - C_n \neq \sum_{i \notin I(n)} \left(X_{n,i} - \frac{C_n}{t(n) - s_0}\right)\right\} \to 0.$$

We can then follow the proof below with $X_{n,i}$ replaced by $X_{n,i} - C_n/(t(n) - s_0)$ and C_n replaced by 0. We therefore concentrate on the case where (3.7) holds.

We shall then work with the shifted variables

$$\hat{X}_{n,i} = X_{n,i} - \frac{1}{t(n)} C_n. \tag{3.8}$$

Let $X_{n,i}$ and C_n have components $X_{n,i}(1), ..., X_{n,i}(d)$, and $C_n(1), ..., C_n(d)$, respectively. Then the qth component of $\hat{X}_{n,i}$ is

$$\hat{X}_{n,i}(q) = X_{n,i}(q) - \frac{1}{t(n)} C_n(q).$$

An important part in the proof is played by the $(1-\beta)$-quantiles of the $|\hat{X}_{n,i}(q)|$, $1 \leqslant q \leqslant d$. These are defined as any numbers $L_n(\beta, q)$ satisfying

$$P\{|\hat{X}_{n,1}(q)| < L_n(\beta, q)\} \leqslant 1-\beta \leqslant P\{|\hat{X}_{n,1}(q)| \leqslant L_n(\beta, q)\}, \tag{3.9}$$

where $0 < \beta < 1$. Suppose we can prove that for each $1 \leqslant q \leqslant d$ and each $\eta > 0$

$$L_n\left(\frac{\eta}{t(n)}, q\right) \to 0 \qquad (n \to \infty). \tag{3.10}$$

We then have for any $\varepsilon > 0$

$$\begin{aligned}
&\limsup_{n\to\infty} P\left\{\max_{1\leqslant i\leqslant t(n)} \left|X_{n,i} - \frac{C_n}{t(n)}\right| > \varepsilon\right\} \\
&\quad\leqslant \limsup_{n\to\infty} t(n)\, P\left\{\left|X_{n,1} - \frac{C_n}{t(n)}\right| > \varepsilon\right\} \\
&\quad\leqslant \limsup_{n\to\infty} \sum_{q=1}^{d} t(n)\, P\left\{\left|X_{n,1}(q) - \frac{C_n(q)}{t(n)}\right| > \frac{\varepsilon}{d}\right\} \\
&\quad\leqslant \limsup_{n\to\infty} \sum_{q=1}^{d} t(n)\, P\left\{\left|X_{n,1}(q) - \frac{C_n(q)}{t(n)}\right| > L_n\left(\frac{\eta}{t(n)}, q\right)\right\} \\
&\qquad\qquad \text{(by (3.10))} \\
&\quad\leqslant d\eta.
\end{aligned}$$

Thus (3.10) will imply (2.13) with $F_n = C_n/t(n)$.

It therefore suffices for (2.13) to prove (3.10). This will be proven for each component separately. We therefore suppress the q in our notation and treat $X_{n,i}$ as a real valued random variable, and C_n as a real number. We set

$$\hat{T}_n = \sum_{i=1}^{t(n)} \hat{X}_{n,i} = T_n - C_n$$

and

$$^{D}\hat{T}_n = \sum_{i \notin I(n)} \hat{X}_{n,i} = \hat{T}_n - \hat{X}_{n,i_n(1)} - \cdots - \hat{X}_{n,i_n(r)}. \tag{3.11}$$

Then

$$^{D}\hat{T}_n = {}^{D}T_n - \left(\frac{t(n)-r}{t(n)}\right) C_n, \tag{3.12}$$

and the hypothesis (2.12) of our theorem is that

$$^{D}\hat{T}_n - \frac{r}{t(n)} C_n \Rightarrow N(\tau, \sigma^2) \tag{3.13}$$

for some τ and $\sigma^2 \geqslant 0$, with N a one-dimensional normal random variable. (If $\sigma^2 = 0$, $N(\tau, 0)$ is degenerate at τ.)

To prove (3.10) we shall first prove that

$$\frac{L_n(\eta/t(n))}{1+|C_n|/t(n)} \to 0 \qquad \text{as} \quad n \to \infty. \tag{3.14}$$

Once we have (3.14), it will be easy to see that along any subsequence with $|C_n|/t(n) \to \infty$, $|C_n|^{-1}\, \hat{T}_n$ has to converge in distribution to an infinitely divisible distribution with compact support. It is known that such a distribution must be concentrated on one point. A simple argument shows that this is not possible if (3.7) holds. From this we shall obtain

$$\limsup_{n\to\infty} \frac{|C_n|}{t(n)} < \infty, \tag{3.15}$$

and then (3.14) yields (3.10). For (3.14) itself we give an indirect proof, basically by deriving (3.4).

Step i. Assume that (3.14) fails, so that for some $\varepsilon > 0$, $\eta > 0$, and an infinite sequence $n_1 < n_2 < \cdots$

$$L_{n_i}\left(\frac{\eta}{t(n_i)}\right) \geqslant \varepsilon \left(1 + \frac{|C_{n_i}|}{t(n_i)}\right). \tag{3.16}$$

To simplify notation we again assume that this holds along the full sequence of integers, so that

$$L_n\left(\frac{\eta}{t(n)}\right) \geqslant \varepsilon \left(1 + \frac{|C_n|}{t(n)}\right). \tag{3.17}$$

Esseen's concentration function inequality (Esseen [4, Theorem 3.1]) states that for $L>0$ and any $\lambda \in (0, L)$,

$$P\{|\hat{T}_n| \leqslant L\} \leqslant \frac{K_0 L}{\sqrt{t(n)\, U_n(\lambda)}}. \tag{3.18}$$

Here K_0 is a universal constant and

$$\begin{aligned} U_n(\lambda) &= E\{(\hat{X}^s_{n,1})^2 \wedge \lambda^2\} \\ &= E[(\hat{X}^s_{n,1})^2\, \mathbb{1}\{|\hat{X}^s_{n,1}| \leqslant \lambda\}] + \lambda^2 P\{|\hat{X}^s_{n,1}| > \lambda\}, \end{aligned} \tag{3.19}$$

where $\hat{X}^s_{n,1}$ has the distribution of $\hat{X}_{n,1} - \hat{X}_{n,2}$. We will show that (3.17) and (3.18) imply that for each $\gamma > 0$ there exists some $K(\gamma) < \infty$ for which

$$t(n)\, U_n\left(\gamma L_n\left(\frac{\eta}{t(n)}\right)\right) \leqslant K(\gamma)\, L_n^2\left(\frac{\eta}{t(n)}\right) \qquad \text{for all large } n. \tag{3.20}$$

We will need the following standard randomization device. Let $\hat{W}_i$, $i \geqslant 1$, be i.i.d. uniform $[0, 1]$ random variables which are also independent of the $\hat{X}_{n,i}$. Suppose $t(n) > \eta$, and if $P\{|\hat{X}_{n,1}| = L_n(\eta/t(n)\} > 0$, define

$$E(n,i) = \left\{ |\hat{X}_{n,i}| < L_n\left(\frac{\eta}{t(n)}\right), \text{ or } |\hat{X}_{n,i}| = L_n\left(\frac{\eta}{t(n)}\right) \right.$$

$$\left. \text{and } \hat{W}_i \leqslant \frac{1 - \dfrac{\eta}{t(n)} - P\left\{|\hat{X}_{n,i}| < L_n\left(\dfrac{\eta}{t(n)}\right)\right\}}{P\left\{|\hat{X}_{n,i}| = L_n\left(\dfrac{\eta}{t(n)}\right)\right\}} \right\}. \tag{3.21}$$

If, on the other hand, $P\{|\hat{X}_{n,1}| = L_n(\eta/t(n))\} = 0$, define

$$E(n,i) = \left\{ |\hat{X}_{n,i}| < L_n\left(\frac{\eta}{t(n)}\right) \right\}. \tag{3.22}$$

Then it is easy to see that

$$P\{E(n,i)\} = 1 - \eta/t(n). \tag{3.23}$$

Also for large n and each $L>0$

$$\begin{aligned}
&P\{E(n,i)\text{ occurs for all } i\leqslant t(n)\text{ and } |{}^D\hat{T}_n|\leqslant L\}\\
&\quad\geqslant P\{E(n,i)\text{ occurs for all } i\leqslant t(n)\}-P\{|{}^D\hat{T}_n|>L\}\\
&\quad=\left(1-\frac{\eta}{t(n)}\right)^{t(n)}-P\{|{}^D\hat{T}_n|>L\}\geqslant\frac{3}{4}e^{-\eta}-P\{|{}^D\hat{T}_n|>L.\}
\end{aligned}$$

By virtue of (3.5) and (3.17)

$$\begin{aligned}
P\{|{}^D\hat{T}_n|>L\}&\leqslant P\left\{\left|{}^D\hat{T}_n-\frac{r}{t(n)}C_n\right|+\frac{s}{t(n)}|C_n|>L\right\}\\
&\leqslant P\left\{\left|{}^D\hat{T}_n-\frac{r}{t(n)}C_n\right|>L-\frac{s}{\varepsilon}L_n\left(\frac{\eta}{t(n)}\right)\right\}.
\end{aligned}$$

Thus, by (3.13) and (3.17), there exists a constant $K_1=K_1(\varepsilon,\eta)$ such that

$$P\left\{|{}^D\hat{T}_n|>K_1L_n\left(\frac{\eta}{t(n)}\right)\right\}\leqslant\frac{1}{8}e^{-\eta}. \tag{3.24}$$

It follows that for all large n

$$P\left\{E(n,i)\text{ occurs for all } i\leqslant t(n)\text{ and } |{}^D\hat{T}_n|\leqslant K_1L_n\left(\frac{\eta}{t(n)}\right)\right\}\geqslant\frac{1}{2}e^{-\eta}. \tag{3.25}$$

Furthermore, if the event in braces in (3.25) occurs, then

$$\begin{aligned}
|\hat{T}_n|&\leqslant|{}^D\hat{T}_n|+\sum_{j=1}^{r}|\hat{X}_{n,i_n(j)}|\\
&\leqslant K_1L_n\left(\frac{\eta}{t(n)}\right)+rL_n\left(\frac{\eta}{t(n)}\right)\leqslant(K_1+s)\,L_n\left(\frac{\eta}{t(n)}\right)\qquad\text{a.s.}
\end{aligned}$$

This means that

$$P\left\{|\hat{T}_n|\leqslant(K_1+s)\,L_n\left(\frac{\eta}{t(n)}\right)\right\}\geqslant\frac{1}{2}e^{-\eta}. \tag{3.26}$$

On the other hand, by (3.18),

$$P\{|\hat{T}_n|\leqslant\tilde{L}\}\leqslant\frac{K_0\tilde{L}}{\sqrt{t(n)\,U_n(\gamma L_n(\eta/t(n)))}}$$

for any $\tilde{L} \geqslant \gamma L_n(\eta/t(n))$. Applying this to

$$\tilde{L} = \gamma L_n\left(\frac{\eta}{t(n)}\right) \vee \left((K_1+s)\, L_n\left(\frac{\eta}{t(n)}\right)\right),$$

we obtain from (3.26) that

$$\frac{(\gamma \vee (K_1+s))\, L_n(\eta/t(n))}{\sqrt{t(n)\, U_n(\gamma L_n(\eta/t(n)))}} \geqslant \frac{1}{2K_0}\, e^{-\eta}.$$

This implies (3.20) with $K(\gamma) = [2K_0 e^{\eta}(\gamma \vee (K_1+s))]^2$.

Step ii. Next define the truncated variables

$$Y_{n,i} = \left(\hat{X}_{n,i} \wedge L_n\left(\frac{\eta}{t(n)}\right)\right) \vee \left(-L_n\left(\frac{\eta}{t(n)}\right)\right)$$

and notice that

$$\begin{aligned} 2\,\mathrm{Var}(Y_{n,1}) &= \mathrm{Var}(Y_{n,1} - Y_{n,2}) = E(Y_{n,1} - Y_{n,2})^2 \\ &\leqslant E\left\{(\hat{X}_{n,1} - \hat{X}_{n,2})^2\, \mathbb{1}\left\{|\hat{X}_{n,1}|\right.\right. \\ &\qquad \left.\left.\leqslant 2L_n\left(\frac{\eta}{t(n)}\right),\ |\hat{X}_{n,2}| \leqslant 2L_n\left(\frac{\eta}{t(n)}\right)\right\}\right\} \\ &\qquad + 4L_n^2\left(\frac{\eta}{t(n)}\right) P\left\{|\hat{X}_{n,1}| > 2L_n\left(\frac{\eta}{t(n)}\right)\right. \\ &\qquad \left.\text{or } |\hat{X}_{n,2}| > 2L_n\left(\frac{\eta}{t(n)}\right)\right\} \end{aligned}$$

because $|Y_{n,1} - Y_{n,2}| \leqslant |\hat{X}_{n,1} - \hat{X}_{n,2}|$ (as an easy examination of the various cases shows) and $|Y_{n,1}|$ and $|Y_{n,2}|$ do not exceed $L_n(\eta/t(n))$. Since $\hat{X}_{n,1} - \hat{X}_{n,2}$ has the distribution of $\hat{X}^s_{n,1}$, we have

$$\begin{aligned} 2\,\mathrm{Var}(Y_{n,1}) &\leqslant E\left\{(\hat{X}^s_{n,1})^2\, \mathbb{1}\left\{|\hat{X}^s_{n,1}| \leqslant 4L_n\left(\frac{\eta}{t(n)}\right)\right\}\right\} \\ &\qquad + 8L_n^2\left(\frac{\eta}{t(n)}\right) P\left\{|\hat{X}_{n,1}| > 2L_n\left(\frac{\eta}{t(n)}\right)\right\} \\ &\leqslant U_n\left(4L_n\left(\frac{\eta}{t(n)}\right)\right) + 8L_n^2\left(\frac{\eta}{t(n)}\right) \\ &\qquad \times P\left\{|\hat{X}_{n,1}| > 2L_n\left(\frac{\eta}{t(n)}\right)\right\}. \end{aligned} \tag{3.27}$$

Furthermore, if m_n is a median of $\hat{X}_{n,1}$, then for $\eta/t(n) < \frac{1}{2}$ we have $|m_n| \leqslant L_n(\eta/t(n))$. A standard symmetrization inequality (see Lemma V.5.1 in Feller [5]) then gives

$$P\left\{|\hat{X}^s_{n,1}| > L_n\left(\frac{\eta}{t(n)}\right)\right\} \geqslant \frac{1}{2} P\left\{|\hat{X}_{n,1} - m_n| > L_n\left(\frac{\eta}{t(n)}\right)\right\}$$
$$\geqslant \frac{1}{2} P\left\{|\hat{X}_{n,1}| > 2L_n\left(\frac{\eta}{t(n)}\right)\right\}.$$

Consequently, from (3.27),

$$\operatorname{Var}(Y_{n,1}) \leqslant U_n\left(4L_n\left(\frac{\eta}{t(n)}\right)\right) + 16L_n^2\left(\frac{\eta}{t(n)}\right)$$
$$\times P\left\{|\hat{X}^s_{n,1}| > L_n\left(\frac{\eta}{t(n)}\right)\right\}$$
$$\leqslant 17U_n\left(4L_n\left(\frac{\eta}{t(n)}\right)\right). \tag{3.28}$$

It will also be useful to take

$$a_n = E(Y_{n,1}).$$

Our task in this step is to deduce that for suitable $x = x(\eta)$ and all $m = 0, 1, 2, ...,$

$$P\left\{E(n, i) \text{ occurs for all } i \leqslant t(n) \text{ and}\right.$$
$$\left.\left|\sum_{i=1}^{t(n)-m} (\hat{X}_{n,i} - a_n)\right| \leqslant xL_n\left(\frac{\eta}{t(n)}\right)\right\} \geqslant \frac{1}{4} e^{-\eta}, \tag{3.29}$$

provided n is large enough. Recall from (3.21) and (3.22) that $|\hat{X}_{n,i}| \leqslant L_n(\eta/t(n))$ when $E(n, i)$ occurs. Thus on the event in braces in (3.29), $\hat{X}_{n,i} = Y_{n,i}$, and so the probability in (3.29) is the same as

$$P\left\{E(n, i) \text{ occurs for all } i \leqslant t(n) \text{ and}\right.$$
$$\left.\left|\sum_{i=1}^{t(n)-m} (Y_{n,i} - a_n)\right| \leqslant xL_n\left(\frac{\eta}{t(n)}\right)\right\}.$$

By Chebyshev's inequality, (3.25), and (3.28), this is no smaller than

$$P\{E(n,i) \text{ occurs for all } i \leqslant t(n)\} - P\left\{\left|\sum_{i=1}^{t(n)-m} (Y_{n,i}-a_n)\right| > xL_n\left(\frac{\eta}{t(n)}\right)\right\}$$

$$\geqslant \frac{1}{2}e^{-\eta} - \frac{t(n)\operatorname{Var}(Y_{n,1})}{x^2L_n^2(\eta/t(n))}$$

$$\geqslant \frac{1}{2}e^{-\eta} - \frac{17t(n)\,U_n\{4L_n(\eta/t(n))\}}{x^2L_n^2(\eta/t(n))}.$$

In addition, by (3.20), the last expression is no smaller than

$$\frac{1}{2}e^{-\eta} - \frac{17K(4)}{x^2}$$

and this will exceed $\frac{1}{4}e^{-\eta}$, provided x^2 is chosen greater than $68K(4)\,e^{\eta}$. This establishes (3.29).

Step iii. To complete the proof of (3.14), we observe that, by virtue of (3.21) and (3.22), $|\hat{X}_{n,i}| \leqslant L_n(\eta/t(n))$ on $E(n,i)$, and $|\hat{X}_{n,i}| \geqslant L_n(\eta/t(n))$ on $E^c(n,i)$. Thus for each i at least one of the following two inequalities holds:

$$\frac{\eta}{t(n)} \geqslant P\left\{E^c(n,i) \text{ and } \hat{X}_{n,i} \geqslant L_n\left(\frac{\eta}{t(n)}\right)\right\} \geqslant \frac{\eta}{2t(n)}, \tag{3.30}$$

or

$$\frac{\eta}{t(n)} \geqslant P\left\{E^c(n,i) \text{ and } \hat{X}_{n,i} \leqslant -L_n\left(\frac{\eta}{t(n)}\right)\right\} \geqslant \frac{\eta}{2t(n)}. \tag{3.31}$$

For the sake of argument let us assume that (3.30) holds for infinitely many values of n and, then by selecting a further subsequence (if necessary), we may assume that (3.30) holds for all n. Now define events $G(n,m)$ as

$$\Big\{\hat{X}_{n,i} \geqslant L_n(\eta/t(n)) \text{ and } E^c(n,i) \text{ occurs for exactly } m \text{ values of } i \leqslant t(n),$$

$$\text{say for } i \in J_n = \{j_1, ..., j_m\}, \text{ whereas } E(n,i) \text{ occurs for } i \notin J_n,$$

$$\text{and also } \left|\sum_{i\notin J_n} (\hat{X}_{n,i} - a_n)\right| \leqslant xL_n\,(\eta/t(n))\Big\}. \tag{3.32}$$

Then for fixed m, $m > s$,

$$P\{G(n, m)\} \geqslant \binom{t(n)}{m}\left(\frac{\eta}{2t(n)}\right)^m P\left\{E(n, i) \text{ occurs for all } i \leqslant t(n) \text{ and } \left|\sum_{i=1}^{t(n)-m} (\hat{X}_{n,i} - a_n)\right| \leqslant xL_n\left(\frac{\eta}{t(n)}\right)\right\}$$

$$\geqslant \binom{t(n)}{m}\left(\frac{\eta}{2t(n)}\right)^m \frac{1}{4} e^{-\eta} \qquad \text{(by (3.29))}$$

$$\geqslant \frac{1}{4}\frac{(\frac{1}{4}\eta)^m}{m!} e^{-\eta}, \tag{3.33}$$

provided n is so large that (3.29) applies and $t(n) > 2m$.

Suppose the event $G(n, m)$ in (3.32) occurs. Write

$$^D\hat{T}_n = \sum_{\substack{i \notin \{i_n(1), ..., i_n(r)\} \\ i \notin J_n}} \hat{X}_{n,i} + \sum_{\substack{i \notin \{i_n(1), ..., i_n(r)\} \\ i \in J_n}} \hat{X}_{n,i}$$

$$= \sum_{i \notin J_n} \hat{X}_{n,i} - \sum_{i \in \{i_n(1), ..., i_n(r)\} \setminus J_n} \hat{X}_{n,i} + \sum_{i \in J_n \setminus \{i_n(1), ..., i_n(r)\}} \hat{X}_{n,i}.$$

Since $G(n, m)$ occurs,

$$\sum_{i \notin J_n} \hat{X}_{n,i} \geqslant (t(n) - m)\, a_n - xL_n\left(\frac{\eta}{t(n)}\right)$$

and also $E(n, i)$ occurs for $i \notin J_n$. For these i, $-\hat{X}_{n,i} \geqslant -L_n(\eta/t(n))$, so

$$- \sum_{i \in \{i_n(1), ..., i_n(r)\} \setminus J_n} \hat{X}_{n,i} \geqslant -sL_n\left(\frac{\eta}{t(n)}\right).$$

Lastly, if $i \in J_n$, then $\hat{X}_{n,i} \geqslant L(\eta/t(n))$, so

$$\sum_{i \in J_n \setminus \{i_n(1), ..., i_n(r)\}} \hat{X}_{n,i} \geqslant (m - s)\, L_n\left(\frac{\eta}{t(n)}\right).$$

Putting these three bounds together shows that, when $G(n, m)$ occurs,

$$^D\hat{T}_n \geqslant t(n)\, a_n + m\left(L_n\left(\frac{\eta}{t(n)}\right) - a_n\right) - (2s + x)\, L_n\left(\frac{\eta}{t(n)}\right). \tag{3.34}$$

We need one more estimate. It must be the case that

$$\limsup_{n\to\infty} \frac{t(n)\,|a_n|}{L_n(\eta/t(n))} < \infty. \tag{3.35}$$

Suppose in fact that (3.35) fails, so that for arbitrary $T > 0$,

$$t(n)\,|a_n| \geqslant T L_n\left(\frac{\eta}{t(n)}\right) \tag{3.36}$$

for infinitely many n. Suppose the event in braces on the left-hand side of (3.29) occurs with $m = 0$. Then $E(n, i)$ occurs for $1 \leqslant i \leqslant t(n)$ and

$$\begin{aligned} |{}^D\hat{T}_n - t(n)\,a_n| &= \left|\sum_{i=1}^{t(n)} (\hat{X}_{n,i} - a_n) - \sum_{i\in\{i_n(1),\ldots,i_n(r)\}} \hat{X}_{n,i}\right| \\ &\leqslant (x+s)\,L_n\left(\frac{\eta}{t(n)}\right). \end{aligned}$$

Thus by (3.36)

$$|{}^D\hat{T}_n| \geqslant t(n)\,|a_n| - (x+s)\,L_n\left(\frac{\eta}{t(n)}\right) > (T - x - s)\,L_n\left(\frac{\eta}{t(n)}\right).$$

From (3.29) it now follows that

$$\limsup_{n\to\infty} P\left\{|{}^D\hat{T}_n| > (T-x-s)\,L_n\left(\frac{\eta}{t(n)}\right)\right\} \geqslant \frac{1}{4}\,e^{-\eta}.$$

But this is not possible for large T because of (3.24). Thus, indeed, (3.35) holds, and so there is a $K_2 > 0$ such that

$$K_2 L_n\left(\frac{\eta}{t(n)}\right) \geqslant t(n)\,|a_n| \geqslant -t(n)\,a_n.$$

(3.35) also means that we can choose n large enough for

$$a_n \leqslant \frac{1}{2}\,L_n\left(\frac{\eta}{t(n)}\right).$$

Finally, we deduce from these inequalities and (3.34) that on $G(n, m)$

$${}^D\hat{T}_n \geqslant \left(-K_2 + \frac{1}{2}\,m - 2s - x\right) L_n\left(\frac{\eta}{t(n)}\right),$$

and if we choose $m > 4(K_2 + 2s + x)$ this implies

$$^{D}\hat{T}_n \geqslant \frac{1}{4} m L_n \left(\frac{\eta}{t(n)} \right).$$

Consequently, by (3.33),

$$\liminf_{n \to \infty} P \left\{ ^{D}\hat{T}_n \geqslant \frac{1}{4} m L_n \left(\frac{\eta}{t(n)} \right) \right\} \geqslant \frac{1}{4} \frac{(\frac{1}{4}\eta)^m}{m!} e^{-\eta}, \tag{3.37}$$

which is true for all sufficiently large m. On the other hand, by (3.17) and (3.13), for $m \geqslant 8(s+\tau)/\varepsilon$,

$$\begin{aligned} &\limsup_{n \to \infty} P \left\{ ^{D}\hat{T}_n \geqslant \frac{1}{4} m L_n \left(\frac{\eta}{t(n)} \right) \right\} \\ &\quad \leqslant \limsup_{n \to \infty} P \left\{ ^{D}\hat{T}_n - \frac{rC_n}{t(n)} \geqslant \frac{1}{8} m\varepsilon \right\} \\ &\quad = \frac{1}{\sqrt{2\pi}} \int_{(m\varepsilon/8 - \tau)/\sigma}^{\infty} e^{-(1/2)t^2} \, dt \\ &\quad \leqslant \frac{\sigma}{\sqrt{2\pi}\,(m\varepsilon/8 - \tau)} \exp \left\{ -\frac{1}{2\sigma^2} \left(\frac{1}{8} m\varepsilon - \tau \right)^2 \right\}. \end{aligned} \tag{3.38}$$

(Interpret the final expression in (3.38) as 0 if $\sigma = 0$.) (3.38) is clearly incompatible with (3.37) as $m \to \infty$, and this contradiction completes the proof of (3.14).

Step iv. In the case where (3.6) holds we are done, since we can then go through the preceding argument with $X_{n,i}$ and C_n replaced by $X_{n,i} - (n - s_0)^{-1} C_n$ and 0, respectively. (3.14) with $C_n = 0$ then gives the desired (3.10). Next consider the case when (3.7) holds. To obtain Theorem 2.2 from (3.14) in this case, we use a simple property of infinitely divisible distributions to show that (3.15) holds.

Assume, to the contrary, that there exists a sequence $n_1 < n_2 < \cdots$ such that

$$\frac{|C_{n_j}|}{t(n_j)} \to \infty, \qquad j \to \infty. \tag{3.39}$$

Then first we use (3.11) to write

$$\frac{t(n)\,\hat{T}_n}{|C_n|} = \frac{t(n)^{D}\hat{T}_n}{|C_n|} + \frac{t(n) \sum_{i \in I(n)} \hat{X}_{n,i}}{|C_n|}. \tag{3.40}$$

By definition of L_n we have

$$P\left\{\left|X_{n_j,i}-\frac{C_{n_j}}{t(n_j)}\right|>L_{n_j}\left(\frac{\eta}{t(n_j)}\right)\text{ for some } i\leqslant t(n_j)\right\}\leqslant\eta. \tag{3.41}$$

If the event in braces in (3.41) does not occur, then by (3.5), (3.14), and (3.39),

$$\left|\sum_{i\in I(n_j)}\hat{X}_{n_j,i}\right|\leqslant sL_{n_j}\left(\frac{\eta}{t(n_j)}\right)=o\left(\frac{|C_{n_j}|}{t(n_j)}\right).$$

Since (3.14) holds for all $\eta>0$, it follows from (3.40) that

$$\frac{t(n_j)\,\hat{T}_{n_j}}{|C_{n_j}|}-\frac{t(n_j)\,{}^D\hat{T}_{n_j}}{|C_{n_j}|}\xrightarrow{P}0. \tag{3.42}$$

Also (3.41) and $L_{n_j}(\eta/t(n_j))=o(|C_{n_j}|/t(n_j))$ imply that $\{t(n_j)\,\hat{X}_{n_j,i}/|C_{n_j}|\}$, $1\leqslant i\leqslant t(n_j)$, is a uniformly asymptotically negligible sequence. Now (3.42), (3.39), and (3.13) imply

$$\frac{t(n_j)\,\hat{T}_{n_j}}{|C_{n_j}|}-r(n_j)\operatorname{sgn}(C_{n_j})\xrightarrow{P}0 \tag{3.43}$$

and so

$$\lim_{j\to\infty}P\left\{t(n_j)\frac{|\hat{T}_{n_j}|}{|C_{n_j}|}\geqslant s+1\right\}=0.$$

Take a further subsequence, if necessary, so that $t(n_j)\,\hat{T}_{n_j}/|C_{n_j}|$ has a limiting distribution. Since $\{t(n_j)\,\hat{X}_{n_j,i}/|C_{n_j}|\}$, $1\leqslant i\leqslant t(n_j)$, is uniformly asymptotically negligible, this must be an infinitely divisible distribution (cf. Gnedenko and Kolmogorov [7, Theorem 24.2], or Feller [5, Chap. XVII.7]) and the only infinitely divisible distributions with compact support are the one-point distributions (cf. Feller [5, p. 177]). On the other hand, (3.43) and (3.7) would show that any limit distribution of $t(n_j)\,T_{n_j}/|C_{n_j}|$ has support on s_1 and s_2 or on $-s_1$ and $-s_2$. This contradiction shows that (3.39) is impossible, so that (3.15) holds. As noted before, (3.14) plus (3.15) yield (3.10), and therefore (2.13) for $F_n=C_n/t(n)$.

Step v. In this step we go back to considering the $X_{n,i}$ as vectors. Again we take $k_n=n$ for ease of notation. To prove the last statement of Theorem 2.2, consider first the case that (2.12) holds with $C_n=0$. Then $\hat{X}_{n,i}=X_{n,i}$ in the preceding proof, irrespective of the distribution of $r(n)$. Then (3.14) with $C_n=0$ immediately gives (3.10) and this in turn gives (2.13) with $F_n=0$ (because $\hat{X}_{n,i}=X_{n,i}$).

Next assume that $X_{n,i} = B_n X_i$. We must show that $\max_{1 \leqslant i \leqslant t(n)} |B_n X_i| \xrightarrow{P} 0$. Define

$$\|B_n\| = \sum_{k,\ell=1}^{d} |B_n(k,\ell)|.$$

We may assume that $\|B_n\| \neq 0$, since $B_n X_i = 0$ for those n with $\|B_n\| = 0$, and there is nothing to prove for them. For the remaining n we have

$$X_{n,i} = B_n X_i = \|B_n\| \frac{B_n}{\|B_n\|} X_i. \tag{3.44}$$

Now assume that (2.13) holds, but that for some subsequence $n_1 < n_2 < \cdots$

$$F_{n_j} \to F_0 \neq 0 \qquad \text{as} \quad j \to \infty. \tag{3.45}$$

We include in (3.45) the possibility that $|F_{n_j}| \to \infty$. By taking a further subsequence if necessary, we may assume that

$$\frac{B_{n_j}(k,\ell)}{\|B_{n_j}\|} \to M(k,\ell), \qquad 1 \leqslant k, \ell \leqslant d, \tag{3.46}$$

for some finite nonzero $d \times d$ matrix $M = (M(k,\ell))$. We may further assume that $\|B_{n_j}\|$ is bounded away from zero, for otherwise $X_{n_j,i} \xrightarrow{P} 0$ along some subsequence (by (3.44)) and this is ruled out by (2.13) and (3.45). Therefore (2.13) and (3.46) imply that

$$MX_1 - \frac{1}{\|B_{n_j}\|} F_{n_j} \xrightarrow{P} 0 \qquad (j \to \infty). \tag{3.47}$$

By taking still a further subsequence, if necessary, we may assume that $F_{n_j}/\|B_{n_j}\| \to A$ for some vector $A \in \mathbb{R}^d$. A must be finite because MX_1 is a finite (random) vector which does not depend on n_j any longer. We must then have $MX_1 = A$ a.s. This equation, however, is impossible when the distribution of X_1 is full. For if $MX_1 = A$ a.s., then $X_1 - A_0 \in N$ a.s. for some $A_0 \in \mathbb{R}^d$ satisfying $MA_0 = A$ and $N =$ null space of M (note that $N \neq \mathbb{R}^d$ when M is not the zero matrix). Thus (3.45) is impossible and we must have $F_n \to 0$.

If $F_n \to 0$, then we may take $F_n = 0$ and still preserve (2.13). In this case

$$\sum_{i \in I(k_n)} X_{k_n,i} \xrightarrow{P} 0$$

by virtue of (2.13) with $F_n = 0$ and (3.5). Then by (2.12)

$$T_{k_n} - C_{k_n} = {}^{D}T_{k_n} - C_{k_n} + \sum_{i \in I(k_n)} X_{k_n, i} \Rightarrow N(\mu, \Sigma).$$

Now

$${}^{\tilde{D}}T_{k_n} - C_{k_n} = T_{k_n} - C_{k_n} - \sum_{i \in \tilde{I}(k_n)} X_{k_n, i},$$

where $\tilde{I}$ is the index set of the variables removed in the deletion scheme $\tilde{D}$. Thus, if $\tilde{D}$ satisfies (2.9) and (2.10) and $F_n = 0$, then (2.14) follows. This completes the proof of Theorem 2.2. ∎

4. PROOF OF THEOREM 2.1

We shall prove separately that (2.i) ⇔ (2.7) for $i = 2, 3, 4, 5$, and that (2.5) ⇔ (2.6). If we use the deletion scheme which deletes no observations at all, then (2.2) and (2.3) both reduce to

$$B_n(S_n - A_n) \Rightarrow N(0, I). \tag{4.1}$$

It is the principal result of Hahn and Klass [10] that this is equivalent to (2.7); in the one-dimensional case this is of course the famous characterization of Lévy [16, Theorem 36.3], of the domain of attraction of the normal law. The equivalence of (2.2) and (2.3) to (4.1) for any of our permitted deletion schemes will follow easily from Theorem 2.2. Similarly, the other equivalences, in the case when no observations are deleted, were proven in Maller [18] and are related to Lévy [16, Theorem 38], which roughly speaking states that in one dimension (4.1) holds if and only if the maximal summand is small with respect to the fluctuations of S_n. Our proofs here are imitations of Maller's or reductions by simple algebra to the proofs of Maller [18]. We begin with

(2.2) ⇔ (2.7) *and* (2.3) ⇔ (2.7)

Let (2.2) hold for some deletion scheme D satisfying Condition 1. Then we can apply Theorem 2.2 to the random vectors

$$X_{n, i} = B_n X_i, \qquad 1 \leqslant i \leqslant n.$$

(2.2) then implies (2.12) with $k_n = n = t(n)$ and $C_n = B_n A_n$. By taking for $\tilde{D}$ the deletion scheme which deletes no variables at all, we obtain from (2.14) that (4.1) holds.

In the same way, (4.1) implies (2.2) for any deletion scheme which satisfies Condition 1.

If (2.3) holds, then we apply Theorem 2.2 with

$$X_{n,i} = B_n\left(X_i - \frac{1}{n} A_n\right)$$

and $C_n = 0$. Then, in the notation of (2.11), (2.3) says

$$^D T_n \Rightarrow N(0, I)$$

and, hence, by (2.14)

$$T_n = B_n(S_n - A_n) \Rightarrow N(0, I).$$

Thus, again (2.3) implies (4.1), and the converse is proved in the same way. Hence, both (2.2) and (2.3) are equivalent to (4.1).

Next, (4.1) is equivalent to (2.7) by Hahn and Klass [10, Theorem 5] and Maller [18, Theorem 1.1].

(2.4) ⇒ (2.7)

Next assume that (2.4) holds. We will show that (2.7) must hold. Our first task is to replace $^D\bar{X}_n$ by EX in $^D\bar{V}_n$ (see (1.10)). We follow Maller [18, Section 3(b)] to show that $E|X| < \infty$ and that, with $\mu = E(X)$,

$$n\,|B_n(^D\bar{X}_n - \mu)|^2 \xrightarrow{P} 0 \tag{4.2}$$

and

$$B_n\left[\sum_{i \notin I(n)} (X_i - \mu)(X_i - \mu)^{\mathrm{T}}\right] B_n^{\mathrm{T}} \xrightarrow{P} I. \tag{4.3}$$

This proof runs as follows. As in Maller [18, Lemma 2.1], we may replace B_n by the symmetric matrix $(B_n^{\mathrm{T}} B_n)^{1/2}$ in (2.4) (simply multiply by the orthogonal matrix $(B_n^{\mathrm{T}} B_n)^{1/2} B_n^{-1}$ on the left and by its transpose on the right; $A^{1/2}$ denotes the symmetric square root of a positive definite matrix A). Then, if B_n is taken symmetric, let

$$B_n = \theta(n)\, \Lambda(n)\, \theta(n)^{\mathrm{T}}$$

for some orthogonal matrix $\theta(n)$ and diagonal matrix $\Lambda(n)$ with $\lambda_j(n)$ as the (j, j) entry; the $\lambda_j(n)$ are the eigenvalues of B_n. Then, as in Section 3(b) of Maller [18], if $\theta_j(n)$ denotes the jth column of $\theta(n)$, (2.4) implies

$$\lambda_j^2(n)\,\theta_j(n)^{\mathrm{T}}\,{}^D\bar{V}_n\theta_j(n)$$
$$=\lambda_j^2(n)\,\theta_j(n)^{\mathrm{T}}\left[\sum_{i\notin I(n)} X_iX_i^{\mathrm{T}}-(n-r)\,{}^D\bar{X}_n\,{}^D\bar{X}_n^{\mathrm{T}}\right]\theta_j(n)\xrightarrow{P}1, \quad (4.4)$$
$$1\leqslant j\leqslant d.$$

This in turn implies

$$\frac{\sum_{i\notin I(n)}|X_i-{}^D\bar{X}_n|^2}{b_n^2}=\frac{\sum_{i\notin I(n)}|X_i|^2-(n-r)\,|{}^D\bar{X}_n|^2}{b_n^2}\xrightarrow{P}1, \quad (4.5)$$

where

$$b_n^2=\sum_{j=1}^{d}\lambda_j^{-2}(n)=\operatorname{trace}(B_n^{-2}).$$

To show that $E\,|X|<\infty$, first assume $E\,|X|^2=\infty$. Then, still following Maller [18], we have for any $T>0$

$$(n-r)\,|{}^D\bar{X}_n|^2\leqslant\frac{2}{n-r}\left|\sum_{i\notin I(n)}X_i\,\mathbb{1}\{|X_i|\leqslant T\}\right|^2$$
$$+\frac{2}{n-r}\left|\sum_{i\notin I(n)}X_i\,\mathbb{1}\{|X_i|>T\}\right|^2$$
$$\leqslant 2nT^2+\frac{2}{n-s}\sum_{i\notin I(n)}|X_i|^2\sum_{i=1}^{n}\mathbb{1}\{|X_i|>T\}\qquad\text{a.s.} \quad (4.6)$$

By choosing T large, this is with high probability small with respect to $\sum_{i\notin I(n)}|X_i|^2$. Indeed, if $r(n)\leqslant s$, then

$$\sum_{i\notin I(n)}|X_i|^2\geqslant\sum_{i=1}^{n}|X_i|^2-(s\text{ largest values of }|X_i|^2,\,1\leqslant i\leqslant n)$$
$$\geqslant\sum_{i=1}^{n}|X_i|^2\,\mathbb{1}\{|X_i|\leqslant T\} \quad (4.7)$$

as soon as there are more than s values $i\leqslant n$ with $|X_i|>T$, and this occurs with probability approaching 1. Thus, since

$$\frac{1}{n}\sum_{i=1}^{n}|X_i|^2\,\mathbb{1}\{|X_i|\leqslant T\}\xrightarrow{P}E(|X_1|^2\,\mathbb{1}\{|X_1|\leqslant T\}),$$

we have

$$P\left\{\frac{1}{n}\sum_{i\notin I(n)}|X_i|^2\geqslant\frac{1}{2}E(|X_1|^2\,\mathbb{1}\{|X_1|\leqslant T\})\right\}\to 1. \tag{4.8}$$

By letting $T\to\infty$ in this, we see that

$$\frac{1}{n}\sum_{i\notin I(n)}|X_i|^2\xrightarrow{P}\infty$$

for any deletion scheme satisfying (1.4) (and when $E|X|^2=\infty$). Thus by (4.6) and the weak law of large numbers

$$\frac{(n-r)\,|{}^D\bar{X}_n|^2}{\sum_{i\notin I(n)}|X_i|^2}\leqslant o_p(1)+2P\{|X|>T\}$$

so that, on letting $T\to\infty$,

$$(n-r)\,|{}^D\bar{X}_n|^2=o_p\left(\sum_{i\notin I(n)}|X_i|^2\right).$$

(Here $o_p(Z_n)$ denotes a sequence of random variables such that $o_p(Z_n)/Z_n$ converges in probability to 0. Also $O_p(Z_n)$ will denote a sequence such that $O_p(Z_n)/Z_n$ is stochastically bounded (tight).) Therefore, if $E\,|X|^2=\infty$, then (4.5) implies

$$\frac{1}{b_n^2}\sum_{i\notin I(n)}|X_i|^2\xrightarrow{P}1.$$

By Theorem 2.2 with $d=1$ and $X_{n,i}=b_n^{-2}\,|X_i|^2$ we then also have

$$\frac{1}{b_n^2}\sum_{i=1}^{n}|X_i|^2\xrightarrow{P}1. \tag{4.9}$$

This implies $E\,|X|<\infty$ as in Maller [18, Section 3(b)]. If, on the other hand, $E\,|X|^2<\infty$, then of course $E\,|X|<\infty$. Thus $\mu=EX$ certainly exists.

Now we wish to deduce from (4.4) that the $\lambda_j^2(n)$ are $O(n^{-1})$ as $n\to\infty$. This we do as follows. Since the distribution of the X_i is full, we have

$$\alpha:=\inf_{|u|=1}P\{|u^{\mathrm{T}}(X_i-\mu)|\geqslant 2\varepsilon\}>0 \tag{4.10}$$

for some $\varepsilon > 0$. Also, by the weak law of large numbers, we have ${}^D\bar{X}_n = \bar{X}_n + O_p(\max_{1 \leqslant i \leqslant n} |X_i|/n) \xrightarrow{P} \mu$. Thus the events

$$A_{in} = \{ |\theta_j(n)^{\mathrm{T}} (X_i - \mu)| \geqslant 2\varepsilon, |{}^D\bar{X}_n - \mu| \leqslant \varepsilon \}$$

have probability at least $\alpha/2$ for large n. But on A_{in}

$$|\theta_j(n)^{\mathrm{T}} (X_i - {}^D\bar{X}_n)| \geqslant |\theta_j(n)^{\mathrm{T}} (X_i - \mu)| - \varepsilon \geqslant \varepsilon$$

and so

$$\theta_j^{\mathrm{T}}(n) {}^D\bar{V}_n \theta_j(n) = \sum_{i \notin I(n)} |\theta_j(n)^{\mathrm{T}} (X_i - {}^D\bar{X}_n)|^2$$

$$\geqslant \varepsilon^2 \sum_{i \notin I(n)} \mathbb{1}\{A_{in}\} \geqslant \varepsilon^2 \left(\sum_{i=1}^{n} \mathbb{1}\{A_{in}\} - r \right).$$

By the weak law of large numbers we therefore have for ε satisfying (4.10)

$$P\{\theta_j(n)^{\mathrm{T}} \, {}^D\bar{V}_n \theta_j(n) \geqslant \tfrac{1}{4} \alpha \varepsilon^2 n\} \to 1. \tag{4.11}$$

With this, (4.4) implies the required relation:

$$\lambda^2_{\max}(B_n) := \max_{1 \leqslant j \leqslant d} \lambda_j^2(n) = O(n^{-1}). \tag{4.12}$$

(If in the previous paragraph we replace $\theta_j(n)$ throughout by $\rho_{\min}(n)$, where $\rho_{\min}(n)$ is an eigenvector corresponding to the minimum eigenvalue of ${}^D\bar{V}_n$, we obtain instead of (4.11) the relation

$$P\{\lambda_{\min}({}^D\bar{V}_n) \geqslant \tfrac{1}{4} \alpha \varepsilon^2 n\} \to 1, \tag{4.13}$$

where $\lambda_{\min}({}^D\bar{V}_n)$ is the minimum eigenvalue of ${}^D\bar{V}_n$. We will need this relationship later.)

To prove (4.2), use ${}^D\bar{X}_n \xrightarrow{P} \mu$ and (4.12) to write

$$n\, |B_n({}^D\bar{X}_n - \mu)|^2 \leqslant n \lambda^2_{\max}(B_n)\, |{}^D\bar{X}_n - \mu|^2 = O_p(|{}^D\bar{X}_n - \mu|^2) \xrightarrow{P} 0.$$

This gives (4.2). Using $\bar{X}_n \xrightarrow{P} \mu$ and ${}^D\bar{X}_n \xrightarrow{P} \mu$, thus $\bar{X}_n - {}^D\bar{X}_n \xrightarrow{P} 0$, together with (4.2), we then obtain

$$n^{1/2}\, |B_n(\bar{X}_n - \mu)| \leqslant n^{1/2}\, |B_n({}^D\bar{X}_n - \mu)| + o_p(1) \xrightarrow{P} 0. \tag{4.14}$$

That is, (4.2) also holds if no observations are deleted. Finally (4.3) then follows from (2.4) by some straightforward algebra.

Now, having proved that (2.4) implies (4.3) and (4.2), we apply Theorem 2.2 with d replaced by d^2 and $X_{n,i}$ equal to the matrix $B_n(X_i-\mu)(X_i-\mu)^{\mathrm{T}} B_n$ stacked column-wise into a $d^2\times 1$ vector; in the usual notation,

$$X_{n,i}=\operatorname{vec}\{B_n(X_i-\mu)(X_i-\mu)^{\mathrm{T}} B_n\}.$$

Then (4.3) asserts that after r summands have been deleted from $\sum_{i=1}^n X_{n,i}$ according to deletion scheme D, the resulting sum converges in distribution to a $d^2\times 1$ normal random variable degenerate at the point $v=\operatorname{vec}(I)$. Theorem 2.2 then shows that the full sum $\sum_{i=1}^n X_{n,i}$ converges to v, i.e., that (4.3) holds for the deletion scheme which deletes no summands. Finally, Theorem 1.1 of Maller [18] (use (1.7a) and (1.7b) with $A_n=\mu$; recall (4.14)) then shows that (2.7) holds.

$(2.7)\Rightarrow(2.4)$

Conversely, assume that (2.7) holds. Again by Theorems 1.1 and 2.1 of Maller [18], $E|X|^\alpha<\infty$ for $0\leqslant\alpha<2$ and, with $\mu=E(X)$,

$$B_n(S_n-n\mu)\Rightarrow N(0,I)$$

for some matrices $B_n\to 0$. Moreover, by the conditions for convergence to normality (cf. Gnedenko and Kolmogorov [7, Theorem 25.1]) for each $u\in S^{d-1}$ and $\varepsilon>0$

$$nP\{|u^{\mathrm{T}}B_nX|>\varepsilon\}\to 0 \qquad \text{as} \quad n\to\infty.$$

Consequently, for each $\varepsilon>0$, $nP\{|B_nX|>\varepsilon\}\to 0$ and $\max_{1\leqslant i\leqslant n}|B_nX_i| \xrightarrow{P} 0$. Again with $\bar{X}_n$ and ${}^D\bar{X}_n$ as in (1.9) we then also have

$$\begin{aligned} &n|B_n({}^D\bar{X}_n-\bar{X}_n)|^2\\ &\quad\leqslant 3n\left|B_n\left(\frac{1}{n-r}-\frac{1}{n}\right)\sum_{i\notin I(n)}X_i\right|^2+3n\left|\frac{B_n}{n}\sum_{i\in I(n)}X_i\right|^2\\ &\quad\leqslant\frac{3r^2}{n}\left|\frac{B_n}{n-r}\sum_{i\notin I(n)}X_i\right|^2+\frac{3r^2}{n}\max_{1\leqslant i\leqslant n}|B_nX_i|^2=o_p(1), \end{aligned} \tag{4.15}$$

because $|(n-r)^{-1}\sum_{i\notin I(n)}B_nX_i|=|B_n{}^D\bar{X}_n|\leqslant\max_{1\leqslant i\leqslant n}|B_nX_i|\xrightarrow{P}0$. Then also

$$\begin{aligned} B_n{}^D\bar{V}_nB_n^{\mathrm{T}}&=B_n\left[\sum_{i=1}^n X_iX_i^{\mathrm{T}}-(n-r)\,{}^D\bar{X}_n{}^D\bar{X}_n^{\mathrm{T}}-\sum_{i\in I(n)}X_iX_i^{\mathrm{T}}\right]B_n^{\mathrm{T}}\\ &=B_n\left[\sum_{i=1}^n X_iX_i^{\mathrm{T}}-n\bar{X}_n\bar{X}_n^{\mathrm{T}}\right]B_n^{\mathrm{T}}+o_p(1). \end{aligned}$$

Finally, as shown in Section 3(a) of Maller [18], under (2.7) we have

$$B_n\left[\sum_{i=1}^{n}(X_i-\bar{X}_n)(X_i-\bar{X}_n)^{\mathrm{T}}\right]B_n^{\mathrm{T}}$$

$$=B_n\left[\sum_{i=1}^{n}X_iX_i^{\mathrm{T}}-n\bar{X}_n\bar{X}_n^{\mathrm{T}}\right]B_n^{\mathrm{T}}\xrightarrow{P} I.$$

Hence (2.4) holds and we have shown that (2.4) and (2.7) are equivalent.

For the next equivalence the following lemma is useful. Let $Y_i^u=(u^{\mathrm{T}}X_i)^2$, $1\leqslant i\leqslant n$, and let $M_n^{(n)}(u)\leqslant\cdots\leqslant M_n^{(1)}(u)$ denote $Y_1^u, ..., Y_n^u$ arranged in increasing order. More precisely, let $m_n(j)$, $n\geqslant 1$, $1\leqslant j\leqslant n$, be the number of Y_i^u satisfying $Y_i^u>Y_j^u$, $1\leqslant i\leqslant n$, or $Y_i^u=Y_j^u$, $1\leqslant i\leqslant j$, and let $M_n^{(t)}(u)=Y_j^u$ if $m_n(j)=t$. Let

$$^{(0)}S_n^u=S_n^u=Y_1^u+Y_2^u+\cdots+Y_n^u, \tag{4.16}$$

and if $1\leqslant t<n$, let

$$^{(t)}S_n^u=S_n^u-M_n^{(1)}(u)-\cdots-M_n^{(t)}(u). \tag{4.17}$$

LEMMA 4.1. *If* $r(n)\leqslant s$, $0<\varepsilon<1/(s+1)$, *and*

$$\max_{i\notin I(n)} Y_i^u\leqslant\varepsilon\sum_{i\notin I(n)}Y_i^u \tag{4.18}$$

for some $u\in S^{d-1}$, *then for that* u,

$$(1-(s+1)\,\varepsilon)\,M_n^{(s+1)}(u)\leqslant\varepsilon\,{}^{(s+1)}S_n^u. \tag{4.19}$$

Remark. Note that Lemma 4.1 holds trivially if $r=s=0$, in the sense that

$$\max_{1\leqslant i\leqslant n} Y_i^u\leqslant\varepsilon\sum_{i=1}^{n}Y_i^u$$

for any $0<\varepsilon<1$ obviously implies $(1-\varepsilon)\,M_n^{(1)}(u)\leqslant\varepsilon\,{}^{(1)}S_n^u$.

Proof of Lemma 4.1. Choose $j_1=j_1(u), ..., j_r=j_r(u)$ such that $Y_{j_k}^u=M_n^{(k)}(u)$. Let

$$m=\inf\{k: j_k\notin\{i_n(1), ..., i_n(r)\}\},$$

so that $M_n^{(1)}(u), ..., M_n^{(m-1)}(u)$ are among the r deleted points $Y^u_{i_n(1)}, ..., Y^u_{i_n(r)}$, but $M_n^{(m)}(u)$ is not. Since $j_m \notin \{i_n(1), ..., i_n(r)\}$, we have $M_n^{(m)}(u) = \max_{i \notin I(n)} Y_i^u$. Thus by (4.18) and the fact that $Y_i^u \geqslant 0$ a.s.,

$$\begin{aligned} M_n^{(m)}(u) &\leqslant \varepsilon \left\{ \sum_{i=1}^{n} Y_i^u - Y^u_{i_n(1)} - \cdots - Y^u_{i_n(r)} \right\} \\ &\leqslant \varepsilon \{ S_n^u - Y^u_{j_1} - \cdots - Y^u_{j_{m-1}} \} \\ &= \varepsilon \{ S_n^u - Y^u_{j_1} - \cdots - Y^u_{j_m} \} + \varepsilon Y^u_{j_m} \\ &= \varepsilon\, {}^{(m)}S_n^u + \varepsilon M_n^{(m)}(u). \end{aligned}$$

It follows that

$$(1-\varepsilon)\, M_n^{(m)}(u) \leqslant \varepsilon\, {}^{(m)}S_n^u. \tag{4.20}$$

We must have $m \leqslant r+1$. If $m = r+1$ this ends the first step of the proof. If $m \leqslant r$, we continue as follows:

$$\begin{aligned} M_n^{(m+1)}(u) \leqslant M_n^{(m)}(u) &\leqslant \left(\frac{\varepsilon}{1-\varepsilon}\right) {}^{(m)}S_n^u \\ &= \left(\frac{\varepsilon}{1-\varepsilon}\right) \sum_{i \neq j_1, ..., j_{m+1}} Y_i^u + \left(\frac{\varepsilon}{1-\varepsilon}\right) Y^u_{j_{m+1}}. \end{aligned}$$

Since $Y^u_{j_{m+1}} = M_n^{(m+1)}(u)$ this implies

$$\left(\frac{1-2\varepsilon}{1-\varepsilon}\right) M_n^{(m+1)}(u) \leqslant \left(\frac{\varepsilon}{1-\varepsilon}\right) \sum_{i \neq j_1, ..., j_{m+1}} Y_i^u$$

or

$$M_n^{(m+1)}(u) \leqslant \left(\frac{\varepsilon}{1-2\varepsilon}\right) {}^{(m+1)}S_n^u.$$

Note that $m-1 < r$ and if we iterate the above procedure t times, where $m+t = r+1$, we obtain

$$M_n^{(r+1)}(u) \leqslant \left(\frac{\varepsilon}{1-(t+1)\,\varepsilon}\right) {}^{(r+1)}S_n^u \leqslant \left(\frac{\varepsilon}{1-(r+1)\,\varepsilon}\right) {}^{(r+1)}S_n^u. \tag{4.21}$$

By (4.20) this also holds when $m = r+1$.

We now carry out the second step. (4.21) shows that

$$M_n^{(r+1)}(u) \leqslant \left(\frac{\varepsilon}{1-(r+1)\,\varepsilon}\right) ({}^{(s+1)}S_n^u + M_n^{(r+2)}(u) + \cdots + M_n^{(s+1)}(u))$$

$$\leqslant \left(\frac{\varepsilon}{1-(r+1)\,\varepsilon}\right) ({}^{(s+1)}S_n^u + (s-r)\, M_n^{(r+1)}(u)).$$

Thus, on $\{r \leqslant s\}$,

$$M_n^{(s+1)}(u) \leqslant M_n^{(r+1)}(u) \leqslant \left(\frac{\varepsilon}{1-(s+1)\,\varepsilon}\right) {}^{(s+1)}S_n^u,$$

as desired. ∎

(2.5) ⇒ (2.7)

Now suppose (2.5) holds. We will prove (2.7). By a matrix identity (Rao [22, p. 60]) we have for any positive definite $d \times d$ symmetric matrix A and d-vector Y that

$$Y^{\mathrm{T}} A^{-1} Y = \sup_{u \in S^{d-1}} \frac{(u^{\mathrm{T}} Y)^2}{u^{\mathrm{T}} A u}. \tag{4.22}$$

In particular, for $i \notin I(n)$,

$$X_i^{\mathrm{T}} \left(\sum_{i \notin I(n)} X_i X_i^{\mathrm{T}} \right)^{-1} X_i = \sup_{u \in S^{d-1}} \frac{(u^{\mathrm{T}} X_i)^2}{\sum_{i \notin I(n)} (u^{\mathrm{T}} X_i)^2}. \tag{4.23}$$

Thus, under (2.5), (4.18) of Lemma 4.1 holds uniformly in $u \in S^{d-1}$ with probability approaching 1 for each $\varepsilon \in (0, 1/(s+1))$. By (4.19) of Lemma 4.1 we can then conclude that for $x > 0$

$$\sup_u P\{{}^{(s+1)}S_n^u \leqslant x M_n^{(s+1)}(u)\} \to 0 \qquad \text{as} \quad n \to \infty.$$

We can use this instead of (3.9) of Maller [18], as follows. If $y > 0$, let $Y_i^u(y)$, $1 \leqslant i \leqslant n$, be i.i.d. random variables each with the distribution of Y_1^u, conditional on $Y_1^u < y$. Let $S_n^u(y) = Y_1^u(y) + \cdots + Y_n^u(y)$. Then for $n > s+1$ and $x > 0$

$$P\{{}^{(s+1)}S_n^u \leqslant xM_n^{(s+1)}(u)\}$$

$$\geqslant P\{{}^{(s+1)}S_n^u \leqslant xM_n^{(s+1)}(u),\ M_n^{(s+2)}(u) < M_n^{(s+1)}(u)\}$$

$$\geqslant \binom{n}{s+1} \int_{[0,\infty)} P\{S_{n-s-1}^u(y) \leqslant xy\}\ P^{n-s-1}\{Y_1^u < y\}$$

$$\times P\{\min_{1 \leqslant i \leqslant s+1} Y_i^u \in dy\}.$$

Now follow virtually the same proof leading from (3.13) to (3.18) of Maller [18], and this is (2.7). (As a minor correction, replace "inf" by "sup" and "$\leqslant$" by "$\geqslant$" in (3.10) of Maller [18] for a correct proof. Note that we have not needed to assume continuity of the distribution of Y_1^u above, as was done in Maller [18].)

(2.7) $\Rightarrow$ (2.5)

Conversely let (2.7) hold, so by Lemma 2.3 of Maller [18], V_n is invertible with probability approaching 1 and, by (1.8) of Maller [18], (2.5) holds for the deletion scheme which deletes no variables, i.e.,

$$\max_{1 \leqslant i \leqslant n} X_i^{\mathrm{T}} \left(\sum_{i=1}^{n} X_i X_i^{\mathrm{T}} \right)^{-1} X_i \xrightarrow{P} 0.$$

By (4.23) this implies

$$\inf_u \frac{\sum_{j=1}^n (u^{\mathrm{T}} X_j)^2}{\max_{1 \leqslant i \leqslant n} (u^{\mathrm{T}} X_i)^2} \xrightarrow{P} \infty. \tag{4.24}$$

But

$$\frac{\sum_{i \notin I(n)} (u^{\mathrm{T}} X_i)^2}{\max_{i \notin I(n)} (u^{\mathrm{T}} X_i)^2} \geqslant \frac{\sum_{j=1}^n (u^{\mathrm{T}} X_j)^2 - r \max_{1 \leqslant i \leqslant n} (u^{\mathrm{T}} X_i)^2}{\max_{1 \leqslant i \leqslant n} (u^{\mathrm{T}} X_i)^2}$$

$$= \frac{\sum_{j=1}^n (u^{\mathrm{T}} X_j)^2}{\max_{1 \leqslant i \leqslant n} (u^{\mathrm{T}} X_i)^2} - r.$$

Since $P\{r \leqslant s\} \to 1$, (4.24) shows that

$$\sup_u \frac{\max_{i \notin I(n)} (u^{\mathrm{T}} X_i)^2}{\sum_{i \notin I(n)} (u^{\mathrm{T}} X_i)^2} \xrightarrow{P} 0.$$

This means that ${}^D V_n$ is invertible with probability approaching 1, and, via (4.23), that

$$\max_{i \notin I(n)} X_i^{\mathrm{T}} \left(\sum_{i \notin I(n)} X_i X_i^{\mathrm{T}} \right)^{-1} X_i \xrightarrow{P} 0,$$

which is (2.5). So we have shown that (2.5) and (2.7) are equivalent.

The following lemma is useful. If A and B are symmetric $d \times d$ matrices, write $A \geqslant B$ if $u^{\mathrm{T}} A u \geqslant u^{\mathrm{T}} B u$ for all $u \in \mathbb{R}^d$.

LEMMA 4.2. *For any deletion scheme* D, ${}^D V_n \geqslant {}^D \bar{V}_n$ *and* ${}^D \bar{V}_n^{-1} \geqslant {}^D V_n^{-1}$, *if the inverses exist.*

Proof. If A and B are symmetric positive definite $d \times d$ matrices and

$$A = B + x x^{\mathrm{T}}$$

for some $x \in \mathbb{R}^d$, then it is easy to verify that

$$A^{-1} = B^{-1} - \frac{B^{-1} x x^{\mathrm{T}} B^{-1}}{1 + x^{\mathrm{T}} B^{-1} x}.$$

Thus $B^{-1} \geqslant A^{-1}$. Applying this to

$$\begin{aligned} {}^D \bar{V}_n &= \sum_{i \notin I(n)} (X_i - {}^D \bar{X}_n)(X_i - {}^D \bar{X}_n)^{\mathrm{T}} \\ &= {}^D V_n - (n - r)\, {}^D \bar{X}_n {}^D \bar{X}_n^{\mathrm{T}} \end{aligned} \tag{4.25}$$

we obtain ${}^D \bar{V}_n^{-1} \geqslant {}^D V_n^{-1}$. That ${}^D V_n \geqslant {}^D \bar{V}_n$ follows from (4.25). ∎

$(2.6) \Rightarrow (2.5)$

Now suppose (2.6) holds. Thus, for each fixed $\varepsilon > 0$, with probability approaching 1,

$$\max_{i \notin I(n)} (X_i - {}^D \bar{X}_n)^{\mathrm{T}}\, {}^D \bar{V}_n^{-1} (X_i - {}^D \bar{X}_n) \leqslant \varepsilon^2. \tag{4.26}$$

If ${}^D \bar{V}_n$ is invertible, then ${}^D V_n$ is also invertible (see (4.25)), so when (4.26) holds we have

$$\max_{i \notin I(n)} (X_i - {}^D \bar{X}_n)^{\mathrm{T}}\, {}^D V_n^{-1} (X_i - {}^D \bar{X}_n) \leqslant \varepsilon^2 \tag{4.27}$$

because ${}^D \bar{V}_n^{-1} \geqslant {}^D V_n^{-1}$ by Lemma 4.2. By virtue of (4.22), (4.27) implies

$$|u^{\mathrm{T}} (X_i - {}^D \bar{X}_n)| \leqslant \varepsilon \sqrt{u^{\mathrm{T}}\, {}^D V_n u} \tag{4.28}$$

uniformly in $u \in S^{d-1}$ and $1 \leqslant i \leqslant n$, $i \notin I(n)$. From (4.28) we obtain

$$|u^{\mathrm{T}} X_i| \leqslant \varepsilon \sqrt{u^{\mathrm{T}}\, {}^{D}V_n u} + |u^{\mathrm{T}}\, {}^{D}\bar{X}_n|. \tag{4.29}$$

Now by the Cauchy–Schwarz inequality,

$$\begin{aligned} |u^{\mathrm{T} D}\bar{X}_n| &= \frac{|\sum_{i \notin I(n)} u^{\mathrm{T}} X_i|}{n-r} \leqslant \sqrt{\frac{\sum_{i \notin I(n)} (u^{\mathrm{T}} X_i)^2}{n-r}} \\ &= \sqrt{\frac{u^{\mathrm{T}}\, {}^{D}V_n u}{n-r}} \leqslant \varepsilon \sqrt{u^{\mathrm{T}}\, {}^{D}V_n u}, \end{aligned} \tag{4.30}$$

on $\{n > r + 1/\varepsilon^2\}$. (4.29) and (4.30) give

$$|u^{\mathrm{T}} X_i| \leqslant 2\varepsilon \sqrt{u^{\mathrm{T}}\, {}^{D}V_n u} \tag{4.31}$$

and, since this is uniform in $u \in S^{d-1}$ and $1 \leqslant i \leqslant n$, $i \notin I(n)$, we have, once again by (4.22)

$$\max_{i \notin I(n)} X_i^{\mathrm{T}}\, {}^{D}V_n^{-1} X_i \leqslant 4\varepsilon^2.$$

Since this holds with probability approaching 1, we see that (2.5) holds.

(2.5) $\Rightarrow$ (2.6)

Conversely let (2.5) hold. Then (again by (4.22)) for every $\varepsilon > 0$, (4.31) holds with probability approaching 1, uniformly in $u \in S^{d-1}$ and $1 \leqslant i \leqslant n$, $i \notin I(n)$. Thus on $\{n > r + 1/\varepsilon^2\}$

$$|u^{\mathrm{T}}(X_i - {}^{D}\bar{X}_n)| \leqslant 2\varepsilon \sqrt{u^{\mathrm{T}}\, {}^{D}V_n u} + |u^{\mathrm{T}}\, {}^{D}\bar{X}_n| \leqslant 3\varepsilon \sqrt{u^{\mathrm{T}}\, {}^{D}V_n u} \tag{4.32}$$

by (4.30). We showed above that (2.4), (2.5), and (2.7) are equivalent, and (2.7) implies $E|X| < \infty$ by Maller [18]. Replacing X_i by $X_i - E(X_1)$, we can assume that $E(X_1) = 0$. Then ${}^{D}\bar{X}_n \xrightarrow{P} 0$. Also (4.13), which holds under (2.4), hence under (2.5), and Lemma 4.2 imply that

$$P\{\inf_u u^{\mathrm{T}}\, {}^{D}V_n u > cn\} \to 1 \tag{4.33}$$

for some constant $c > 0$. Thus $(u^{\mathrm{T}}\, {}^{D}V_n u)/n$ is bounded away from 0, uniformly in $u \in S^{d-1}$, with probability approaching 1, and consequently, uniformly in $u \in S^{d-1}$,

$$\begin{aligned} u^{\mathrm{T}}\, {}^{D}\bar{V}_n u &= u^{\mathrm{T}}\, {}^{D}V_n u - (n-r)\, u^{\mathrm{T}}\, {}^{D}\bar{X}_n\, {}^{D}\bar{X}_n^{\mathrm{T}} u \\ &= (1 + o_p(1))\, u^{\mathrm{T}}\, {}^{D}V_n u. \end{aligned}$$

Hence ${}^{D}\bar{V}_n$ is invertible with probability approaching 1, and also (4.32) gives

$$|u^{\mathrm{T}}(X_i - {}^{D}\bar{X}_n)|^2 \leqslant 10\varepsilon^2 u^{\mathrm{T}}\, {}^{D}\bar{V}_n u$$

with probability approaching 1. Since this holds uniformly in $u \in S^{d-1}$ and $1 \leqslant i \leqslant n$, $i \notin I(n)$, we get

$$\max_{i \notin I(n)} (X_i - {}^{D}\bar{X}_n)^{\mathrm{T}}\, {}^{D}\bar{V}_n^{-1}(X_i - {}^{D}\bar{X}_n) \leqslant 10\varepsilon^2.$$

This holds with probability approaching 1, and $\varepsilon > 0$ is arbitrary, so (2.6) holds. We have shown that (2.5) and (2.6) are equivalent and completed the proof of Theorem 2.1. ▮

5. PROOFS OF EXAMPLES

Proof of Example 2.3. The idea behind this proof is to define an i.i.d. sequence X_i in $\mathbb{R}^2$ which takes a random value $X_i(1) \geqslant 0$ on the x-axis (say) on average 50% of the time, and a random value $X_i(2) \geqslant 0$ on the y-axis the remaining time. $X_i(1)$ will dominate $X_i(2)$ in a certain sense, so that $X_n^{(1)}$, the largest in modulus of the $X_i = (X_i(1), X_i(2))$, will lie on the x-axis, taking values no larger than $\bigvee_{i=1}^{n} X_i(1)$, with probability approaching 1. Then $(X_n^{(1)})^{\mathrm{T}}\, V_n^{-1} X_n^{(1)}$ will be of order of magnitude

$$\bigvee_{i=1}^{n} X_i^2(1) \Big/ \sum_{j=1}^{n} X_j^2(1) \tag{5.1}$$

and we can make this converge to 0 in probability by choosing the distribution of $X_i(1)$ appropriately. Then (2.19) will hold. On the other hand, $\max_{1 \leqslant i \leqslant n} X_i^{\mathrm{T}}\, V_n^{-1} X_i$ will be bounded below by an analogous expression for the $X_i(2)$ only (cf. (5.5) below)), and we can ensure that

$$\bigvee_{i=1}^{n} X_i^2(2) \Big/ \sum_{j=1}^{n} X_j^2(2) \tag{5.2}$$

does *not* converge to 0 in probability, so that (2.20) will hold.

To carry out this program, for $1 \leqslant i \leqslant n$ let $\{X_i(1), X_i(2), b_i\}$ be i.i.d. random variables, with the b_i independent of the $(X_i(1), X_i(2))$, $i \geqslant 1$, such that

$$P(b_i = 0) = \tfrac{1}{2} = P(b_i = 1).$$

Define an i.i.d. sequence of vectors in $\mathbb{R}^2$ by

$$X_i = \begin{pmatrix} X_i(1) \\ 0 \end{pmatrix} \mathbb{1}\{b_i = 0\} + \begin{pmatrix} 0 \\ X_i(2) \end{pmatrix} \mathbb{1}\{b_i = 1\}, \tag{5.3}$$

so that

$$V_n = \sum_{i=1}^{n} X_i X_i^{\mathrm{T}} = \begin{bmatrix} \sum_{i=1}^{n} X_i^2(1)\, \mathbb{1}\{b_i = 0\} & 0 \\ 0 & \sum_{i=1} X_i^2(2)\, \mathbb{1}\{b_i = 1\} \end{bmatrix}. \tag{5.4}$$

Thus

$$X_i^{\mathrm{T}} V_n^{-1} X_i = \frac{X_i^2(1)\, \mathbb{1}\{b_i = 0\}}{\sum_{j=1}^{n} X_j^2(1)\, \mathbb{1}\{b_j = 0\}} + \frac{X_i^2(2)\, \mathbb{1}\{b_i = 1\}}{\sum_{j=1}^{n} X_j^2(2)\, \mathbb{1}\{b_j = 1\}},$$

giving

$$\max_{1 \leqslant i \leqslant n} X_i^{\mathrm{T}} V_n^{-1} X_i \geqslant \frac{\max_{1 \leqslant i \leqslant n} X_i^2(2)\, \mathbb{1}\{b_i = 1\}}{\sum_{j=1}^{n} X_j^2(2)\, \mathbb{1}\{b_j = 1\}}. \tag{5.5}$$

We will choose the distribution of $X_i(2)$ so that the last expression does not converge to 0 in probability.

Let F_1 and F_2 be the distribution functions of $X_i^2(1)$ and $X_i^2(2)$, and suppose that

$$F_1(x) = 1 - \frac{\log x}{x}, \qquad x \geqslant e, \tag{5.6}$$

so that

$$A_1(x) := \int_0^x P\{X_1^2(1) > y\}\, dy \tag{5.7}$$

is slowly varying. Suppose also for simplicity that F_1 is continuous on $[0, e]$ and $F_1(0) = 0$. About F_2, we assume that $F_2(0) = 0$,

$$1 - F_2(x) \leqslant \frac{\sqrt{\log x}}{x} \qquad \text{for large } x \tag{5.8}$$

and

$$A_2(x) := \int_0^x P\{X_1^2(2) > y\}\, dy \qquad \text{is } \textit{not} \text{ slowly varying.} \tag{5.9}$$

This means that

$$\tilde{A}_2(x) := \int_0^x P\{X_1^2(2)\,\mathbb{1}\{b_1=1\} > y\}\,dy$$

$$= \tfrac{1}{2}\int_0^x P\{X_1^2(2) > y\}\,dy = \tfrac{1}{2}A_2(x)$$

also is not slowly varying. Thus it is *not* true that

$$\frac{\max_{1\leqslant i\leqslant n} X_i^2(2)\,\mathbb{1}\{b_i=1\}}{\sum_{j=1}^n X_j^2(2)\,\mathbb{1}\{b_j=1\}} \xrightarrow{P} 0,$$

because this is equivalent, by Theorem 2.1 of Kesten and Maller [15], to the condition $x(1-F_2(x))/\tilde{A}_2(x)\to 0$, and thus, via Feller [5, p. 283], to the slow variation of $\tilde{A}_2(x)$. Then (5.5) proves (2.20).

On the other hand, the slow variation of

$$\tilde{A}_1(x) := \int_0^x P\{X_1^2(1)\,\mathbb{1}\{b_1=0\} > y\}\,dy = \tfrac{1}{2}A_1(x)$$

shows that (5.1) converges to 0 in probability, and this will suffice for (2.19), as we now demonstrate. $X_i(1)$ dominates $X_i(2)$ in the sense that

$$\frac{\bigvee_{i=1}^n X_i^2(2)}{\bigvee_{i=1}^n X_i^2(1)\,\mathbb{1}\{b_i=0\}} \xrightarrow{P} 0, \tag{5.10}$$

which we prove as follows. Fix $T>0$ and for $n>T$ define $L_n = L_n(T)$ so that

$$n[1-F_1(L_n)] = T,$$

as we may by continuity of F_1. This means by (5.6) that $n\log L_n = TL_n$ for large n, so $L_n \sim n\log n/T$. If $\delta>0$ we have by (5.8) that

$$P\left\{\bigvee_{i=1}^n X_i^2(2) > \delta L_n\right\} \leqslant nP\{X_1^2(2) > \delta L_n\} = n[1-F_2(\delta L_n)]$$

$$\leqslant \frac{n\sqrt{\log(\delta L_n)}}{\delta L_n} = O\left\{\frac{n\sqrt{\log n}}{n\log n}\right\} \to 0.$$

On the other hand,

$$P\left\{\bigvee_{i=1}^{n} X_i^2(1)\,\mathbb{1}\{b_i=0\}\leqslant L_n\right\}=P^n\{X_1^2(1)\,\mathbb{1}\{b_1=0\}\leqslant L_n\}$$
$$=\left\{\frac{1}{2}\,[1+F_1(L_n)]\right\}^n$$
$$=\left\{\frac{1}{2}\left(2-\frac{T}{n}\right)\right\}^n\to e^{-T/2}.$$

It follows that

$$P\left\{\bigvee_{i=1}^{n} X_i^2(2)>\delta\bigvee\nolimits_{i=1}^{n} X_i^2(1)\,\mathbb{1}\{b_i=0\}\right\}$$
$$\leqslant P\left\{\bigvee_{i=1}^{n} X_i^2(2)>\delta L_n\right\}+P\left\{\bigvee_{i=1}^{n} X_i^2(1)\,\mathbb{1}\{b_i=0\}\leqslant L_n\right\}$$
$$=o(1)+e^{-T/2}.$$

This can be made arbitrarily small by choosing T large. Thus (5.10) holds. This in turn means that the event

$$E_n=\left\{X_i^2(1)>\bigvee_{j=1}^{n} X_j^2(2)\text{ and }b_i=0\text{ for some }i\leqslant n\right\}$$

occurs with probability approaching 1 as $n\to\infty$. When E_n occurs, $X_n^{(1)}$ must lie in the x-direction. In fact

$$X_n^{(1)}=\begin{pmatrix}X_{i_n}(1)\\ 0\end{pmatrix}, \tag{5.11}$$

where i_n is the i for which $\bigvee_{i=1}^{n} X_i(1)\,\mathbb{1}\{b_i=0\}$ is achieved. This is a well-defined random variable when E_n occurs and is even unique with probability 1 on E_n, since F_1 is continuous. But when (5.11) holds, (5.4) shows that

$$(X_n^{(1)})^{\mathrm{T}}\,V_n^{-1}X_n^{(1)}=\frac{X_{i_n}^2(1)}{\sum_{j=1}^{n} X_j^2(1)\,\mathbb{1}\{b_j=0\}}$$
$$=\frac{\max_{1\leqslant i\leqslant n}(X_i^2(1)\,\mathbb{1}\{b_i=0\})}{\sum_{j=1}^{n} X_j^2(1)\,\mathbb{1}\{b_j=0\}}.$$

If $m=m(n)=\sum_{i=1}^{n} \mathbb{1}\{b_i=0\}$, then the last ratio has the same distribution as $\max_{1\leqslant i\leqslant m} X_i^2(1)/\sum_{j=1}^{m} X_j^2(1)$. This converges to 0 in probability, because $m(n)\xrightarrow{P}\infty$, $m(n)$ is independent of the $X_i(1)$, and (5.1) converges to 0 in probability (as seen above). Thus (2.19) holds.

This completes the proof, subject to finding a distribution F_2 satisfying (5.8) and (5.9). Note that

$$1-F_2(x)=\left(\frac{d}{dx}\right)^{+} A_2(x) \tag{5.12}$$

must decrease to 0. Thus $A_2(x)$ has to be concave, continuous, with a derivative tending to 0 as $x\to\infty$ and with $0\leqslant(d/dx)^{+}A(x)\leqslant 1$. Conversely, given such an $A_2(x)$, we can define $F_2(x)$ to satisfy (5.12). We will choose $A_2(x)$ as a piecewise linear function. Define sequences a_n and b_n by $a_0=0$, $b_1=e^{-2}$, and for $n\geqslant 1$,

$$a_n=e^{2(n!)^2}, \qquad b_{n+1}=\frac{(n+1)!-n!}{a_{n+1}-a_n}.$$

Then define $A_2(a_0)=0$, $A_2(a_n)=n!$ for $n\geqslant 1$, and

$$A_2(x)=A_2(a_{n-1})+b_n(x-a_{n-1}) \qquad \text{for} \quad x\in(a_{n-1},a_n),\ n\geqslant 1.$$

It is easy to check that b_n is positive and decreasing, thus the continuous, piecewise linear function $A_2(x)$ is increasing and concave. Also for $n\geqslant 1$,

$$b_n=\frac{A_2(a_n)-A_2(a_{n-1})}{a_n-a_{n-1}},$$

and, since $A_2(a_n)/A_2(a_{n-1})\to\infty$ and $a_n/a_{n-1}\to\infty$, we have

$$b_n\sim\frac{A_2(a_n)}{a_n}=\frac{\sqrt{\log(a_n)}}{\sqrt{2}\,a_n}<\frac{\sqrt{\log(a_n)}}{a_n}$$

and

$$\frac{A_2(a_n/2)}{A_2(a_n)}=\frac{(n-1)!+b_n(a_n/2-a_{n-1})}{n!}\to 1/2$$

as $n\to\infty$. These show that $A_2(x)$ is not slowly varying, that

$$0\leqslant\left(\frac{d}{dx}\right)^{+} A_2(x)\leqslant b_1\leqslant 1$$

for all x, and that

$$\left(\frac{d}{dx}\right)^{+} A_2(x) \leqslant \frac{\sqrt{\log(x)}}{x}$$

for large x. We therefore need only take

$$1 - F_2(x) = \left(\frac{d}{dx}\right)^{+} A_2(x)$$

to complete the construction. ∎

Proof of Example 2.4. Let the X_i be symmetric i.i.d. random variables with mass p_j on $\pm a_j$, where the parameters are chosen such that $0 < a_j \uparrow \infty$, $\sum_{j \geqslant 1} p_j = 1/2$, $p_{j+1}/p_j \to 0$,

$$p_j a_j^2 = 2^j, \tag{5.13}$$

and

$$\frac{[p_k]^{1/2} \sum_1^k p_j a_j^3}{[\sum_1^k p_j a_j^2]^{3/2}} \leqslant 2, \qquad k \geqslant 1. \tag{5.14}$$

It is easy to see that one can choose such p_j and a_j inductively. The deletion scheme is defined separately for different blocks of values of n. Let $n_k := \lfloor 1/p_k \rfloor$. For $n_k < n \leqslant n_{k+1}$ the scheme is to remove from the sample all X_i with $|X_i| \geqslant a_{k+1}$, $i \leqslant n$. Thus for such n,

$${}^D S_n = \sum_{i=1}^{n} X_i I[\,|X_i| \leqslant a_k].$$

Then, with

$$\sigma_k^2 := \mathrm{Var}(X_i I[\,|X_i| \leqslant a_k]) = 2 \sum_{j=1}^{k} p_j a_j^2 = 2(2^{k+1} - 1) \qquad \text{(by (5.13))},$$

we have

$$\frac{{}^D S_{n_{k+1}}}{\sigma_k \sqrt{n_{k+1}}} \sim \frac{{}^D S_{n_{k+1}}}{\sqrt{2}\, a_{k+1}} \Rightarrow N(0, 1) \tag{5.15}$$

as $k \to \infty$. This follows from Liapunov's theorem; Liapunov's condition for (5.15) is immediate from (5.14) and $p_{j+1}/p_j \to 0$ (see Feller [5, p. 286]). In fact, more is true;

$$\frac{{}^D S_n}{\sigma_k \sqrt{n}} \Rightarrow N(0, 1)$$

whenever $n, k \to \infty$ in such a way that $n_k < n \leqslant n_{k+1}$ and $n/n_k \to \infty$.

On the other hand,

$$S_{n_{k+1}} = {}^D S_{n_{k+1}} + \sum_{i=1}^{n_{k+1}} X_i I[\,|X_i| \geqslant a_{k+1}], \tag{5.16}$$

and the numbers of X_i with $i \leqslant n_{k+1}$ which equal a_{k+1} and $-a_{k+1}$—which are each binomial (n_{k+1}, p_{k+1})—converge in distribution to N_+ and N_-, respectively, where N_+ and N_- are two independent Poisson variables with mean 1. Since $p_{j+1}/p_j \to 0$, we further have that

$$P\{\exists i \leqslant n_{k+1} \text{ with } |X_i| \geqslant a_{k+2}\}$$
$$\leqslant 2n_{k+1} \sum_{j \geqslant k+2} p_j \sim 2 \sum_{j \geqslant k+2} p_j/p_{k+1} \to 0.$$

It follows from (5.15) and (5.16) that

$$\frac{S_{n_{k+1}}}{\sigma_k \sqrt{n_{k+1}}} \sim \frac{S_{n_{k+1}}}{\sqrt{2}\, a_{k+1}} \Rightarrow N(0, 1) * G,$$

where G is the distribution of $2^{-1/2}(N_+ - N_-)$, i.e. of $2^{-1/2}$ times the difference of two independent mean 1 Poisson variables. Consequently, $S_{n_{k+1}}$ is not in the domain of attraction of the normal distribution. The number of summands trimmed at time n_{k+1} is asymptotically $N_+ + N_-$; hence it is tight. ∎

ACKNOWLEDGMENTS

Much of this work was done during a visit by the first author to the University of Western Australia in 1992–1993. He thanks the Mathematics Department of that University for its hospitality. We both thank two referees for their help, especially one who simplified Example 2.3, and we thank Jenny Harris for her expert typing.

REFERENCES

[1] Barnett, V., and Lewis, T. (1994). *Outliers in Statistical Data*, 3rd ed. Wiley, New York.

[2] Belsley, D. A., Kuh, E., and Welsch, R. E. (1980). *Regression Diagnostics*. Wiley, York.

[3] Csörgő, S., Horváth, L., and Mason, D. M. (1986). What portion of the sample makes a partial sum asymptotically stable or normal? *Probab. Theor. Rel. Fields* **2** 1–16.

[4] Esseen, C. G. (1968). On the concentration function of a sum of independent random variables. *Z. Wahrsch. Verw. Geb.* **9** 290–308.

[5] Feller, W. (1971). *An Introduction to Probability Theory and its Applications*, Vol. II, 2nd ed., Wiley, New York.

[6] Giné, E., Götze, F., and Mason, D. M. (1997). When is the Student t-statistic asymptotically standard normal? *Ann. Prob.* **25** 1514–1531.

[7] Gnedenko, B. V., and Kolmogorov, A. N. (1968). *Limit Theorems for Sums of Independent Random Variables*. 2nd ed., Addison–Wesley, Reading, MA.

[8] Griffin, P. S., and Mason, D. M. (1991). On the asymptotic normality of self-normalized sums. *Proc. Camb. Phil. Soc.* **109** 597–610.

[9] Griffin, P. S., and Pruitt, W. E. (1990). Weak convergence of trimmed sums. In *Sums Trimmed Sums, and Extremes* (M. G. Hahn, D. M. Mason, and D. C. Weiner, Eds.), Birkhäuser, Basel.

[10] Hahn, M. G., and Klass, M. J. (1980). Matrix normalisation of sums of random vectors in the domain of attraction of the multivariate normal. *Ann. Prob.* **8** 262–280.

[11] Hahn, M. G., and Klass, M. J. (1981). The multidimensional central limit theorem for arrays normed by affine transformations. *Ann. Prob.* **9** 611–623.

[12] Hampel, F. R., Ronchetti, E. M., Rousseeuw, P. J., and Stahel, W. A. (1986). *Robust Statistics: the Approach based on Influence Functions*. Wiley, New York.

[13] Huber, P. J. (1981). *Robust Statistics*. Wiley, New York.

[14] Kesten, Harry (1993). Convergence in distribution of lightly trimmed and untrimmed sums are equivalent. *Math. Proc. Camb. Phil. Soc.* **113** 615–638.

[15] Kesten, Harry, and Maller, R. A. (1992). Ratios of trimmed sums and order statistics. *Ann. Prob.* **20** 1805–1842.

[16] Lévy, P. (1937). *Théorie de l'Addition des Variables Aléatoires*. Gauthier-Villars.

[17] Maller, R. A. (1982). Asymptotic normality of lightly trimmed means-a converse. *Math. Proc. Camb. Phil. Soc.* **92** 535–545.

[18] Maller, R. A. (1993). Quadratic negligibility and the asymptotic normality of operator normed sums. *J. Mult. Analysis* **44** 191–219.

[19] Meerschaert, Mark M. (1993). Regular variation and generalized domains of attraction in $\mathbb{R}^k$. *Statist. Probab. Lett.* **18** 233–239.

[20] Meerschaert, Mark M. (1994). Norming operators for generalized domains of attraction. *J. Theoret. Probab.* **7** 793–798.

[21] Mori, T. (1984). On the limit distributions of lightly trimmed sums. *Proc. Camb. Math. Phil. Soc.* **96** 507–516.

[22] Rao, C. R. (1973). *Linear Statistical Inference and its Applications*. 2nd ed., Wiley, New York.

[23] Vu, H. T. V., Maller, R. A., and Klass, M. J. (1996). On the Studentisation of random vectors. *J. Mult. Analysis* **57** 142–155.