JOURNAL OF MULTIVARIATE ANALYSIS 44, 191–219 (1993)

Quadratic Negligibility and the Asymptotic Normality of Operator Normed Sums

R. A. MALLER

University of Western Australia, Nedlands 6009, Australia

The condition $\max_{1 \leqslant i \leqslant n} X_i^T V_n^{-1} X_i \xrightarrow{P} 0$, where X_i are vectors in R^d and $V_n = \sum_{i=1}^n X_i X_i^T$, is important in the asymptotics of various linear and nonlinear regression models. We call it "quadratic negligibility." It is shown that, when X_i are independent and identically distributed random vectors in R^d, quadratic negligibility is equivalent to X_i being in the operator normed domain of attraction of the multivariate normal distribution, thereby generalising the one-dimensional case. Related results on the convergence of the matrix V_n, along with results on the centering and norming constants for operator-normed convergence, are also given.

1. INTRODUCTION AND SUMMARY

In this paper we investigate the condition

$$\max_{1 \leqslant i \leqslant n} X_i^T V_n^{-1} X_i \xrightarrow{P} 0, \tag{1.1}$$

where "$\xrightarrow{P}$" denotes convergence in probability, X_i is a sample of independent and identically distributed (i.i.d.) d-dimensional (column) vectors, and

$$V_n = \sum_{i=1}^n X_i X_i^T. \tag{1.2}$$

("T" denotes a vector or matrix transpose.) This condition is of interest in the theory of regression, where it is related to the requirement that the components of a weighted sum of random variables be uniformly asymptotically negligible. So we could dub (1.1) "quadratic negligibility."

In one dimension, condition (1.1) or its almost sure (a.s.) version has

Received September 25, 1990; revised March 1, 1992.

AMS 1991 subject classifications: primary 60F15; secondary 60F05, 62G30.

Key words and phrases: asymptotic normality, operator norming, sums and maxima of random variables, sum of squares and products matrix, covariance matrix, relative stability.

0047-259X/93 $5.00

also attracted the interest of probabilists because in one way it expresses the property of individual random variables being negligible with respect to their sum of squares. To date, (1.1) seems not to have been studied for $d>1$, and this, motivated by the above statistical and probabilistic applications, is the object of this paper. In view of known results for the one-dimensional case (see below), we might conjecture that (1.1) requires the asymptotic normality of the sample vectorial sum

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

after appropriate norming and centering. But what form should this take? It turns out that the right kind is precisely the operator norming studied by Hahn and Klass [9, 10] in which the existence of nonstochastic vectors A_n and square matrices B_n is required so that, as $n \to \infty$,

$$B_n(S_n - A_n) \xrightarrow{D} N(0, I).$$

Here $N(0, I)$ denotes a standard normal random vector in d dimensions, and "$\xrightarrow{D}$" denotes convergence in distribution. ($N(0, 1)$ will be the univariate standard normal.) This is the main result of Theorem 1.1.

Before stating the theorem, we introduce some further notation. Let

$$\bar{X}_n = n^{-1} \sum_{i=1}^{n} X_i$$

be the sample mean, and let

$$\bar{V}_n = \sum_{i=1}^{n} (X_i - \bar{X}_n)(X_i - \bar{X}_n)^{\mathrm{T}} \tag{1.3}$$

be the sample sum of squares and products matrix. It will follow from Lemma 2.3 below that V_n and $\bar{V}_n$ are non-singular on sets whose probabilities approach one as $n \to \infty$, so we can assume without loss of generality that V_n^{-1} and $\bar{V}_n^{-1}$ exist.

We assume X_i are full, i.e., $u^{\mathrm{T}}X$ is not degenerate for $u \in S^{d-1}$, where X is a random variable with distribution function F, the same as that of X_i, and S^{d-1} is the unit sphere in R^d. Thus for each $u \in S^{d-1}$, the quantity

$$V_u(x) = E[(u^{\mathrm{T}}X)^2 1_{\{|u^{\mathrm{T}}X| \leqslant x\}}], \tag{1.4}$$

is not zero for x large enough. (1_A denotes the indicator of an event A.) Throughout this paper, a supremum or infimum over u will be understood to mean a supremum or infimum over $u \in S^{d-1}$.

THEOREM 1.1. *The following are equivalent*:

there exist nonstochastic vectors A_n and square matrices B_n such that

$$B_n(S_n - A_n) \xrightarrow{D} N(0, I) \qquad (n \to \infty); \tag{1.5}$$

there exist nonstochastic square matrices B_n such that

$$B_n \bar{V}_n B_n^{\mathrm{T}} \xrightarrow{P} I \qquad (n \to \infty); \tag{1.6}$$

there exist nonstochastic vectors A_n and square matrices B_n such that

$$B_n \sum_{i=1}^{n} (X_i - A_n)(X_i - A_n)^{\mathrm{T}} B_n^{\mathrm{T}} \xrightarrow{P} I \qquad (n \to \infty) \tag{1.7a}$$

and

$$n\,|B_n(A_n - \bar{X}_n)|^2 \xrightarrow{P} 0 \qquad (n \to \infty); \tag{1.7b}$$

$$\max_{1 \leqslant i \leqslant n} X_i^{\mathrm{T}} V_n^{-1} X_i \xrightarrow{P} 0 \qquad (n \to \infty); \tag{1.8}$$

$$\sup_u \frac{x^2 P(|u^{\mathrm{T}}X| > x)}{V_u(x)} \longrightarrow 0 \quad (x \to \infty); \tag{1.9}$$

$$\max_{1 \leqslant i \leqslant n} (X_i - \bar{X}_n)^{\mathrm{T}} \bar{V}_n^{-1}(X_i - \bar{X}_n) \xrightarrow{P} 0 \qquad (n \to \infty). \tag{1.10}$$

Remarks. (i) We use "I" to denote the identity matrix in d dimensions, and by convergence in probability of $d \times d$ random matrices we mean that each component converges separately.

(ii) In the course of the proof of Theorem 1.1, it is shown that if (1.5)–(1.10) hold, then $E\,|X|^\alpha < \infty$ for $0 \leqslant \alpha < 2$ and A_n may be chosen as $n\mu = nE(X)$ in (1.5) or as μ in (1.7); thus centering occurs at the mean, as we would hope. The norming matrix B_n in (1.5), (1.6), or (1.7) is clearly nonsingular for large enough n (so we choose it to be so, in general) and may be chosen to be symmetric ($B_n = B_n^{\mathrm{T}}$). B_n shrinks to 0 as $n \to \infty$, and we show that the "size" of B_n, as measured by its largest eigenvalue, is at most of order $n^{-1/2}$. (See Theorem 2.1 for these.) We note that Hahn and Klass [9] use $E[(u^{\mathrm{T}}X)^2 \wedge x^2]$ rather than $V_u(x)$ in their formulation of (1.9), but these are equivalent since

$$E[(u^{\mathrm{T}}X)^2 \wedge x^2] = V_u(x) + x^2 P(|u^{\mathrm{T}}X| > x).$$

When (1.6) holds and $\mu = 0$ we also have

$$\frac{\sum_{i=1}^{n} |X_i|^2}{b_n^2} \xrightarrow{P} 1 \qquad (n \to \infty), \tag{1.11}$$

where $b_n^2 = \text{trace}(B_n^{-2}) \to \infty$ is a nonstochastic norming sequence. Equation (1.11) is "relative stability" of $\sum_{i=1}^n |X_i|^2$ and is equivalent to

$$\frac{x^2 P(|X| > x)}{E[|X|^2 1_{\{|X| \le x\}}]} \to 0 \qquad (x \to \infty) \tag{1.12}$$

(cf. Feller [4, p. 236], Rogozin [22], Maller [16]), which is Lévy's [15, p. 113] necessary and sufficient condition for $|X_i|$ to be in the domain of attraction of the normal. However, (1.12) does not imply (1.9) (see the remark following the proof of Theorem 1.1), so the conditions in Theorem 1.1 are genuinely d-dimensional. In general, B_n cannot be taken as diagonal in (1.5); diagonal special cases were studied earlier by Greenwood and Resnick [6]. Another way of expressing (1.11) (when $\mu = 0$) is

$$\frac{\text{trace}(V_n)}{\text{trace}(B_n^{-2})} \xrightarrow{P} 1 \qquad (n \to \infty).$$

(iii) In one dimension, Breiman [1] showed that (1.8) holds if and only if X_i^2 are relatively stable, i.e.,

$$B_n^2 \sum_{i=1}^n X_i^2 \xrightarrow{P} 1$$

for a nonstochastic sequence B_n with $B_n \to 0$ as $n \to \infty$, and this in turn is equivalent to the asymptotic normality of $B_n(\sum_{i=1}^n X_i - A_n)$ for a nonstochastic sequence A_n. Theorem 1.1 contains the d-dimensional versions of these results. Again in the one-dimensional case, Kesten [13] shows that (1.8) holds for a.s. convergence if and only if $EX_i^2 < \infty$; this was extended by O'Brien [19] to the case of pairwise independent X_i.

(iv) In statistical applications, condition (1.8) or its mean-centered equivalent (1.10) appears in the asymptotic theory of generalized linear models [3], in time series [17], in censored regression [23], in stochastic adaptive control models [14], and in the analysis of many other nonlinear regression models. In these situations it is used as a condition on stochastic or fixed covariates or regressor variables, usually to ensure asymptotic normality of regression coefficients. Equation (1.8) is sufficient and in some cases necessary for the consistency and asymptotic normality of the least squares estimator of the regression coefficient in the linear regression of a dependent variable on a vector X_i of covariates. This seems to have first been proved by Eicker [2], and there are also versions in Hajek [11] and Huber [12, p. 159]. The quantity in (1.8) is the diagonal term in the "hat" matrix [12, p. 156] and as such measures the "influence" of the ith covariate; a large influence is to be noted. Hence the importance in statistics of understanding (1.8).

(vi) Condition (1.7b) or something like it is necessary if we wish to allow the general centering sequence A_n in (1.7a), as an example at the end of Section 3 shows.

(vii) We may wish to apply (1.8) when X_i are not full; for example, a design matrix may contain a column of ones to allow estimation of an intercept. Suppose

$$\tilde{X}_i = [1 X_i^{\mathrm{T}}]^{\mathrm{T}} \in R^{d+1},$$

where X_i are full i.i.d. random vectors in R^d. By a formula for the inverse of a partitioned matrix (Rao [21, p. 29]) we can show that

$$\tilde{X}_i^{\mathrm{T}} \left(\sum_{j=1}^{n} \tilde{X}_j \tilde{X}_j^{\mathrm{T}} \right)^{-1} \tilde{X}_i = n^{-1} + (X_i - \bar{X}_n)^{\mathrm{T}} \bar{V}_n^{-1} (X_i - \bar{X}_n) \tag{1.13}$$

so

$$\max_{1 \leqslant i \leqslant n} \tilde{X}_i^{\mathrm{T}} \left(\sum_{j=1}^{n} \tilde{X}_j \tilde{X}_j^{\mathrm{T}} \right)^{-1} \tilde{X}_i \xrightarrow{P} 0 \qquad (n \to \infty)$$

if and only if (1.10) and hence (1.5) holds. Thus the desired asymptotic normality obtains simply by omitting the degenerate component.

(viii) The "self-normalisation" or "studentisation" of the sum S_n by the square root of an estimate of its variance is of great interest in statistics. From (1.6) we suspect that $\bar{V}_n^{-1/2}$ (a symmetric square root of $\bar{V}_n^{-1}$) should be of the "right" size to normalise S_n, and we ask whether (1.5) implies

$$\bar{V}_n^{-1/2}(S_n - A_n) \xrightarrow{D} N(0, I) \qquad (n \to \infty). \tag{1.14}$$

This is true in one dimension. Following the proof of Theorem 1.1 we show that if $B_n(S_n - n\mu) \xrightarrow{D} N(0, I)$ then

$$(S_n - n\mu)^{\mathrm{T}} \bar{V}_n^{-1} (S_n - n\mu) \xrightarrow{D} \chi_d^2 \qquad (n \to \infty), \tag{1.15}$$

where χ_d^2 is the chi-square random variable with d degrees of freedom. This expresses the convergence in distribution of the "Mahalanobis T^2" statistic, and shows that $\bar{V}_n^{-1/2}(S_n - n\mu)$ is stochastically compact under (1.5). If in addition X_i are spherically symmetric around μ then we do have $\bar{V}_n^{-1/2}(S_n - n\mu) \xrightarrow{D} N(0, I)$, as follows from (1.15). More difficult is the reverse implication from (1.14) to (1.5). In one dimension this is true for symmetric X_i, as recently shown by Griffin and Mason [8].

2. Preliminary Results

In this section we give some useful results on the forms of sequences A_n and B_n which generalise the "nice" properties which hold in one dimension. There, $B_n(S_n - A_n) \xrightarrow{D} N(0, 1)$ implies that $E|X| < \infty$ and that A_n may be chosen as nEX, while B_n satisfies

$$B_n^{-2} \sim n[E(X-EX)^2 1_{\{B_n|X| \leqslant 1\}}] \qquad (n \to \infty)$$

and $n^{1/2}B_n$ is bounded above as $n \to \infty$. The d-dimensional counterparts of these are given in Theorem 2.1. Lemma 2.1 contains a symmetry result on B_n. The section also contains two lemmas concerning the "size" of the matrices V_n and $\bar{V}_n$ and convergence of eigenvalues of a matrix.

Recall that X is assumed full throughout this paper. We will let $\lambda_{\min}(V)$, $\lambda_{\max}(V)$, and $\lambda_j(V)$ or $\lambda(V)$, denote the minimum, maximum, and an arbitrary eigenvalue of a square matrix V. We will say that a triangular array of scalar random variables $(Z_{in}, 1 \leqslant i \leqslant n)$ is "uniformly asymptotically negligible" (UAN) (or "infinitesimal," Gnedenko and Kolmogorov [5, p. 95]) if for any $\varepsilon > 0$,

$$\max_{1 \leqslant i \leqslant n} P(|Z_{in}| > \varepsilon) \to 0$$

as $n \to \infty$. We often use the Cramér–Wold device, that a sequence of random vectors $Z_n \xrightarrow{D} Z \in R^d$ if and only if $u^T Z_n \xrightarrow{D} u^T Z \in R$, for each $u \in S^{d-1}$.

Theorem 2.1. *Suppose there exist nonstochastic vectors A_n and square matrices B_n such that $B_n(S_n - A_n) \xrightarrow{D} N(0, I)$. Then*

$$\limsup_{n \to \infty} n^{1/2} \lambda_{\max}(B_n) < \infty; \tag{2.1}$$

$$B_n X \xrightarrow{P} 0 \textit{ and } u^T B_n X_i \textit{ are UAN for } u \in S^{d-1}; \tag{2.2}$$

$E|X|^\alpha < \infty$ *for* $0 \leqslant \alpha < 2$ *and for a nonstochastic matrix* $\tilde{B}_n$,

$$\tilde{B}_n(S_n - n\mu) \xrightarrow{D} N(0, I), \qquad \textit{where} \quad \mu = EX; \tag{2.3}$$

$$nB_n E\{(X-\mu)(X-\mu)^T 1_{\{|B_n X| \leqslant 1\}}\} B_n \to I. \tag{2.4}$$

Proof of Theorem 2.1. The proof requires some preliminary results which we give as the following three lemmas.

Lemma 2.1. (i) *Suppose for a sequence of random vectors $Z_n \in R^d$, we have $B_n Z_n \xrightarrow{D} G$ as $n \to \infty$, where G is a nondegenerate spherically sym-*

metric random vector in R^d *and* B_n *are square nonsingular nonstochastic matrices. Then* B_n *may be taken to be symmetric matrices.*

(ii) *Suppose for a sequence of symmetric nonsingular random* $d \times d$ *matrices* Z_n,

$$B_n Z_n B_n^{\mathrm{T}} \xrightarrow{P} I \qquad (n \to \infty),$$

where B_n *are square nonsingular nonstochastic matrices. Then* B_n *may be taken to be symmetric matrices.*

Proof. Let $(B_n^{\mathrm{T}} B_n)^{1/2}$ be a symmetric square root of the symmetric matrix $B_n^{\mathrm{T}} B_n$, and when $u \in S^{d-1}$ consider

$$u^{\mathrm{T}}(B_n^{\mathrm{T}} B_n)^{1/2} Z_n = [u^{\mathrm{T}}(B_n^{\mathrm{T}} B_n)^{1/2} B_n^{-1}]\, B_n Z_n = v_n^{\mathrm{T}} B_n Z_n,$$

where $v_n^{\mathrm{T}} = u^{\mathrm{T}}(B_n^{\mathrm{T}} B_n)^{1/2} B_n^{-1} \in S^{d-1}$. Use Helly's theorem to find a subsequence (denoted N) so that $v_n \to v (n \to \infty, n \in N)$, where the convergence is componentwise and $v \in S^{d-1}$. Thus

$$v_n^{\mathrm{T}} B_n Z_n = v^{\mathrm{T}} B_n Z_n + (v_n - v)^{\mathrm{T}} B_n Z_n \xrightarrow{D} v^{\mathrm{T}} G. \tag{2.5}$$

Now $v^{\mathrm{T}} G$ has the same distribution as G_1, where G_1 is the first (or any) component of G; this follows by the spherical symmetry of G. This in turn has the same distribution as $u^{\mathrm{T}} G$. Thus $u^{\mathrm{T}}(B_n^{\mathrm{T}} B_n)^{1/2} Z_n \xrightarrow{D} u^{\mathrm{T}} G$, the convergence in (2.5) is independent of the subsequence, and so $(B_n^{\mathrm{T}} B_n)^{1/2} Z_n \xrightarrow{D} G$. The second part is proved similarly. ∎

Remarks. Lemma 2.1 also follows when G is the spherically symmetric subsequential limit of stochastically compact S_n, from the results of Hahn and Klass [9, 10] and Griffin [7], who in effect *define* B_n by its eigenvector decomposition. Note that Lemma 2.1 also holds if $n \to \infty$ through a subsequence.

LEMMA 2.2. *Let* W_n *be a symmetric random matrix with eigenvalues* $\lambda_j(n)$, $1 \leq j \leq d$. *Then the following are equivalent as* $n \to \infty$:

$$W_n \xrightarrow{P} I; \tag{2.6}$$

$$\lambda_j(n) \xrightarrow{P} 1, \qquad 1 \leq j \leq d; \tag{2.7}$$

$$\inf_u (u^{\mathrm{T}} W_n u) \xrightarrow{P} 1, \qquad \sup_u (u^{\mathrm{T}} W_n u) \xrightarrow{P} 1; \tag{2.8}$$

$$u^{\mathrm{T}} W_n u \xrightarrow{P} 1 \qquad \textit{for each} \quad u \in S^{d-1}; \tag{2.9}$$

$$W_n^{-1} \textit{ exists with probability approaching 1 and } W_n^{-1} \xrightarrow{P} I. \tag{2.10}$$

Proof. (See also Tyler [24] for (2.7)). Let $\lambda(n)$ be any eigenvalue of W_n and $\theta(n)$ a corresponding (random) eigenvector. If (2.6) holds then

$$\begin{aligned}\lambda(n) &= \theta(n)^{\mathrm{T}} W_n \theta(n)\\ &= \theta(n)^{\mathrm{T}}(I+\Delta_n)\,\theta(n)\\ &= 1+\theta(n)^{\mathrm{T}}\,\Delta_n\theta(n),\end{aligned}$$

where Δ_n is a symmetric matrix each of whose elements converge in probability to 0. Let $\theta_j(n)$ and $\Delta_{jk}(n)$ be the elements of $\theta(n)$ and Δ_n. Since $|\theta(n)|=1$, $\theta_j(n)$ are bounded, and since $\Delta_n \xrightarrow{P} 0$, $\Delta_{jk}(n) \xrightarrow{P} 0$ as $n\to\infty$. Thus

$$|\theta(n)^{\mathrm{T}}\,\Delta_n\theta(n)| \leqslant \sum_{j=1}^{d}\sum_{k=1}^{d} |\theta_j(n)\,\theta_k(n)\,\Delta_{jk}(n)| \xrightarrow{P} 0,$$

showing that $\lambda(n)\xrightarrow{P} 1$ and proving (2.7).

We have for $u\in S^{d-1}$

$$\min_{1\leqslant j\leqslant d} \lambda_j(W_n) \leqslant u^{\mathrm{T}} W_n u \leqslant \max_{1\leqslant j\leqslant d} \lambda_j(W_n),$$

and if (2.7) holds the right and left extremes of this converge in probability to one as $n\to\infty$. Thus (2.7) implies (2.8). (2.9) follows immediately from (2.8); in fact the convergence in (2.9) is uniform in $u\in S^{d-1}$.

When (2.9) holds, let e_j be the coordinate vector in R^d, i.e., having a one in the jth position and zeros elsewhere. Then if $W_n=(w_{jk}(n))$, we have, as $n\to\infty$,

$$w_{jj}(n)=e_j^{\mathrm{T}} W_n e_j \xrightarrow{P} 1.$$

Also, if $j\neq k$, $(e_j+e_k)/2^{1/2}\in S^{d-1}$, so because of the symmetry of W_n,

$$(e_j+e_k)^{\mathrm{T}} W_n(e_j+e_k) = w_{jj}(n)+2w_{jk}(n)+w_{kk}(n) \xrightarrow{P} 2.$$

Thus $w_{jk}(n)\xrightarrow{P} 0$ when $j\neq k$. So $W_n\xrightarrow{P} I$, and (2.6) follows. So far we have shown that (2.6)–(2.9) are equivalent.

Finally, if (2.6) holds, we have that W_n is positive definite and therefore nonsingular on a set whose probability approaches one by (2.7). On this set,

$$1/\lambda_{\min}(W_n)=\lambda_{\max}(W_n^{-1})\geqslant u^{\mathrm{T}} W_n^{-1}u \geqslant \lambda_{\min}(W_n^{-1})=1/\lambda_{\max}(W_n)$$

if $u \in S^{d-1}$. By (2.7), the right and left extremes of this converge in probability to one as $n \to \infty$; thus

$$u^{\mathrm{T}} W_n^{-1} u \xrightarrow{P} 1 \qquad \text{for} \quad u \in S^{d-1}.$$

By (2.9) applied to W_n^{-1}, this implies $W_n^{-1} \xrightarrow{P} I$, which is (2.10). Similarly, (2.10) implies (2.6). ∎

Lemma 2.3. *When X is full, there is a $c>0$ such that*

$$P(\lambda_{\min}(V_n) \geqslant cn) \to 1 \qquad (n \to \infty) \tag{2.11}$$

while for every $c>0$,

$$P(\lambda_{\min}(\bar{V}_n) \geqslant c) \to 1 \qquad (n \to \infty). \tag{2.12}$$

Proof. When X is full, i.e., $u^{\mathrm{T}}X$ is not degenerate for $u \in S^{d-1}$, we can show that there are constants $c>0$ and $T>0$ such that

$$\inf_u u^{\mathrm{T}} E[XX^{\mathrm{T}} 1_{\{|X| \leqslant T\}}] u \geqslant 2c. \tag{2.13}$$

To see this, suppose for a sequence $T_i \to \infty$ there were $u_i \in S^{d-1}$ such that

$$u_i^{\mathrm{T}} E[XX^{\mathrm{T}} 1_{\{|X| \leqslant T_i\}}] u_i \to 0.$$

Choose a subsequence such that $u_i \to u \in S^{d-1}$ and given $T>0$, choose $T_i > T$. Then

$$\begin{aligned} u^{\mathrm{T}} E[XX^{\mathrm{T}} 1_{\{|X| \leqslant T\}}] u &= \lim_i u_i^{\mathrm{T}} E[XX^{\mathrm{T}} 1_{\{|X| \leqslant T\}}] u_i \\ &\leqslant \limsup_i u_i^{\mathrm{T}} E[XX^{\mathrm{T}} 1_{\{|X| \leqslant T_i\}}] u_i = 0. \end{aligned}$$

Letting $T \to \infty$ shows that $u^{\mathrm{T}}X = 0$ a.s., a contradiction. Thus (2.13) holds. To prove (2.11), we have for $T>0$

$$\begin{aligned} u^{\mathrm{T}} V_n u = \sum_{i=1}^{n} (u^{\mathrm{T}} X_i)^2 &\geqslant \sum_{i=1}^{n} (u^{\mathrm{T}} X_i)^2 1_{\{|X_i| \leqslant T\}} \\ &= u^{\mathrm{T}} \sum_{i=1}^{n} X_i X_i^{\mathrm{T}} 1_{\{|X_i| \leqslant T\}} u \end{aligned}$$

and by the weak law of large numbers

$$n^{-1} \sum_{i=1}^{n} X_i X_i^{\mathrm{T}} 1_{\{|X_i| \leqslant T\}} \xrightarrow{P} E[XX^{\mathrm{T}} 1_{\{|X| \leqslant T\}}].$$

By Lemma 2.2 of Tyler [24] and (2.13), if T is large enough,

$$\lambda_{\min}\left(n^{-1}\sum_{i=1}^{n} X_i X_i^{\mathrm{T}} 1_{\{|X_i|\leqslant T\}}\right) \xrightarrow{P} \lambda_{\min}(E[XX^{\mathrm{T}} 1_{\{|X|\leqslant T\}}]) \geqslant 2c.$$

This implies (2.11). To prove (2.12), by Pakes [20], $u^{\mathrm{T}}\bar{V}_n u$ is non-decreasing in n, so for any sample path, $u^{\mathrm{T}}\bar{V}_n u \to V$, say. Suppose $V<\infty$ on a set B which has probability 0 or 1 by the Hewitt–Savage 0–1 law. If $P(B)=1$, Pakes [20] shows that on the set B,

$$\Sigma(Z_i - A)^2 < \infty$$

(where $Z_i = u^{\mathrm{T}}X_i$) for some finite A. But then $Z_i - A \to 0$ a.s. as $i\to\infty$, which, since Z_i is an i.i.d. sequence, implies that Z_i is a constant. So $u^{\mathrm{T}}X_i$ is degenerate, contrary to assumption. Thus, in fact, $\lambda_{\min}(\bar{V}_n)\to\infty$ a.s., certainly implying (2.12). ∎

We now prove Theorem 2.1. Suppose $B_n(S_n - A_n) \xrightarrow{D} N(0, I)$. Then according to Lemma 2.1 we may take B_n to be symmetric. Define the eigenvalue decomposition (e.g., Rao [21, p. 39]) of B_n by

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

where $\Lambda(n)$ is a diagonal $d\times d$ matrix whose diagonal elements are the eigenvalues of B_n, and $\theta(n)$ is an orthogonal matrix, i.e., $\theta(n)^{\mathrm{T}}\theta(n)=I$. Use Helly's theorem to find a subsequence which we denote N so that $\theta(n)\to\theta$, an orthogonal matrix, when $n\to\infty$, $n\in N$. We have

$$\begin{aligned}\Lambda(n)\,\theta(n)^{\mathrm{T}}(S_n - A_n) &= \theta(n)^{\mathrm{T}} B_n(S_n - A_n)\\ &= \theta^{\mathrm{T}} B_n(S_n - A_n) + [\theta(n)-\theta]^{\mathrm{T}} B_n(S_n - A_n)\\ &\xrightarrow{D} N(0, I) \qquad (n\to\infty,\ n\in N).\end{aligned}$$

Thus since the limit does not depend on the choice of N,

$$\Lambda(n)\,\theta(n)^{\mathrm{T}}(S_n - A_n) \xrightarrow{D} N(0, I) \qquad (n\to\infty).$$

So if $\theta_j(n)$ are the columns of $\theta(n)$ and $\lambda_j(n)$ are the diagonal elements of $\Lambda(n)$, we have, for $1\leqslant j\leqslant d$,

$$\lambda_j(n)\,\theta_j(n)^{\mathrm{T}}(S_n - A_n) \xrightarrow{D} N(0, 1).$$

At this stage we do not have the UAN condition which is needed to apply criteria for the convergence of triangular arrays. Let $\tilde{X}_i$ be i.i.d. random variables with the same distribution as X_i but independent of them, and let

$X_i^s = \tilde{X}_i - X_i$. Then the symmetrised random variables $\theta_j(n)^T X_i^s$ satisfy, by a simple characteristic function argument,

$$\lambda_j(n) \sum_{i=1}^{n} \theta_j(n)^T X_i^s \xrightarrow{D} N(0, 2) \qquad (n \to \infty). \tag{2.14}$$

Let $\lambda_1(n)$ denote the largest of $\lambda_j(n)$ and suppose (2.1) fails to hold. Then $n^{1/2}\lambda_1(n) \to \infty$ as $n \to \infty$, $n \in N$, for a subsequence N of integers. Then by (2.14),

$$\sum_{i=1}^{n} Z_{in} \xrightarrow{P} 0 \qquad (n \to \infty, n \in N),$$

where we define

$$Z_{in} = [\theta_1(n)^T X_i^s]/n^{1/2}, \qquad Z_n = [\theta_1(n)^T X_1^s]/n^{1/2}.$$

Since Z_{in} are symmetric we then have for real t,

$$[E \cos tZ_n]^n \to 1 \qquad (n \to \infty, n \in N)$$

and letting $r_n = n(1 - E \cos tZ_n)$, this implies

$$(1 - n^{-1}r_n)^n = \exp[n \log(1 - n^{-1}r_n)] \to 1,$$

so $-n \log(1 - n^{-1}r_n) \to 0$, and $r_n \to 0$ $(n \to \infty,\ n \in N)$. But since $1 - \cos tx \geqslant t^2x^2/2$ for $|tx| \leqslant 1$ and since $|\theta_1(n)^T X_1^s| \leqslant |X_1^s|$,

$$\begin{aligned} r_n &= nE(1 - \cos tZ_n) \geqslant nt^2 E[Z_n^2 1_{\{|tZ_n| \leqslant 1\}}]/2 \\ &= t^2 E\{[\theta_1(n)^T X_1^s]^2 1_{\{|t\theta_1(n)^T X_1^s| \leqslant n\}}\}/2 \\ &\geqslant t^2 \theta_1(n)^T E\{X_1^s (X_1^s)^T 1_{\{|tX_1^s| \leqslant 1\}}\} \theta_1(n)/2 \\ &\geqslant t^2 \inf_u [u^T E\{X_1^s (X_1^s)^T 1_{\{|tX_1^s| \leqslant 1\}}\} u]/2. \end{aligned}$$

Now the last expression is bounded away from 0 for t small enough by (2.13) applied to X_1^s, so $r_n \to 0$ $(n \to \infty,\ n \in N)$ is impossible. This contradiction proves (2.1), and (2.1) implies $\lambda_{\max}(B_n) \to 0$, so (2.2) follows easily from (2.1).

To prove (2.3), apply Gnedenko and Kolmogorov [5, Theorem 5, p. 143] to (2.14). (Note that $\lambda_j(n)\theta_j(n)^T X_i^s$ are UAN for each j, since by (2.1), $\lambda_j(n) \to 0$.) This gives

$$\sum_{i=1}^{n} [\theta_j(n)^T X_i^s]^2 \overset{P}{\sim} 2/\lambda_j^2(n), \qquad 1 \leqslant j \leqslant d,$$

where $\overset{P}{\sim}$ signifies that the ratio of the two sides converges in probability to one as $n \to \infty$. Adding over $1 \leqslant j \leqslant d$ and recalling that $\sum_{j=1}^{d} \theta_j(n)\,\theta_j(n)^{\mathrm{T}} = I$ gives

$$\frac{\sum_{i=1}^{n} |X_i^s|^2}{b_n^2} \xrightarrow{P} 1,$$

where

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

This is relative stability of $\sum_{i=1}^{n} |X_i^s|^2$ and by (1.12) is equivalent to

$$\frac{x^2 P(|X_1^s| > x)}{E[|X_1^s|^2\, 1_{\{|X_1^s| \leqslant x\}}]} \to 0 \qquad (x \to \infty).$$

Thus $|X_1^s|$ is in the domain of attraction of the normal distribution, and this implies $E|X_1^s|^\alpha < \infty$, $0 < \alpha < 2$, and hence $E|X|^\alpha < \infty$, $0 < \alpha < 2$. Hence certainly $E|X| < \infty$. To complete the proof of (2.3), we need $\tilde{B}_n(S_n - n\mu) \xrightarrow{D} N(0, I)$ for some $\tilde{B}_n$. However, this is easily seen by symmetrising then using the equivalence lemma of Hahn and Klass [9, p. 264) and Theorem 5 of Hahn and Klass [9], so we omit the details.

To prove (2.4), by (2.3) we may replace $X - \mu$ by X; i.e., we may take $\mu = 0$. Then $u^{\mathrm{T}} B_n S_n \xrightarrow{D} N(0, 1)$ for $u \in S^{d-1}$. If $Z_n = u^{\mathrm{T}} B_n X$, then by Gnedenko and Kolmogorov [5, Theorem 2, p. 128], we have for $\varepsilon > 0$

$$nP\{|Z_n| > \varepsilon\} \to 0 \tag{2.15}$$

and

$$n\{E[Z_n^2 1_{\{|Z_n| \leqslant \varepsilon\}}] - E^2[Z_n 1_{\{|Z_n| \leqslant \varepsilon\}}]\} \to 1.$$

Theorem 5 of Hahn and Klass [9] applies so (1.9) holds and by their equivalence lemma we have, since $EZ_n = 0$,

$$\begin{aligned} E^2[Z_n 1_{\{|Z_n| \leqslant \varepsilon\}}] &\leqslant \varepsilon\, |E[Z_n 1_{\{|Z_n| \leqslant \varepsilon\}}]| \\ &= \varepsilon\, |E[Z_n 1_{\{|Z_n| > \varepsilon\}}]| \\ &= o\{E[Z_n^2 1_{\{|Z_n| \leqslant \varepsilon\}}]\} \qquad (n \to \infty). \end{aligned}$$

We thus obtain

$$nE[Z_n^2 1_{\{|Z_n| \leqslant \varepsilon\}}] \to 1,$$

and so

$$nE[(u^{\mathrm{T}}B_nX)^2 1_{\{|u^{\mathrm{T}}B_nX| \leq \varepsilon\}}] \to 1. \tag{2.16}$$

This is not quite what we want, since the indicator function depends on u. But (2.15) says

$$nP(|u^{\mathrm{T}}B_nX| > \varepsilon) \to 0 \qquad (n \to \infty)$$

from which we obtain, taking u to be the coordinate vector along each axis in turn,

$$nP(|B_nX| > \varepsilon) \to 0 \qquad (n \to \infty). \tag{2.17}$$

Finally, since $|u^{\mathrm{T}}B_nX| \leq |B_nX|$, we easily conclude from (2.16) with $\varepsilon = 1$, that

$$nE[(u^{\mathrm{T}}B_nX)^2 1_{\{|B_nX| \leq 1\}}] \to 1. \tag{2.18}$$

Thus, replacing X by $X - \mu$, we have by Lemma 2.2 that

$$nB_nE\{(X-\mu)(X-\mu)^{\mathrm{T}} 1_{\{|B_n(X-\mu)| \leq 1\}}\} B_n \to I. \tag{2.19}$$

We may then omit the μ from the indicator function in (2.19) by a similar argument to that by which we deduced (2.18) from (2.16). Thus (2.4) holds. ∎

Remarks. (i) Having proved (2.3) we do not distinguish between $\tilde{B}_n$ and B_n.

(ii) Following (2.4) it is of interest to ask whether we can *define* (uniquely) a matrix sequence by

$$B_n^{-2} = nE\{(X-\bar{X}_n)(X-\bar{X}_n)^{\mathrm{T}} 1_{\{|B_n(X-\bar{X}_n)| \leq 1\}}\};$$

the answer to a closely related question is yes, as shown by Theorem 2 of Maronna [18], when X is full. Currently we do not know whether this is the "right" norming sequence for the asymptotic normality (or, more generally, stochastic compactness) of S_n.

3. Proof of Theorem 1.1

The proof is divided into six sections denoted (a), ..., (f), giving the proofs of the implications (1.5) → (1.6), ..., (1.9) → (1.5), and (1.8) ↔ (1.10). The methods rely on a reduction to the one-dimensional case by projecting on a unit vector. The results concerning the sum of squares and products matrices V_n and $\bar{V}_n$ essentially use Raikov's theorem (Gnedenko and Kolmogorov [5, p. 143]) showing that, after centering and norming, a sum of squares is relatively stable if the sum is asymptotically normal. The major part of the proof is the implication from (1.8) to (1.9), and this is essentially an extension of Breiman's [1] arguments to the multivariate case, using some of the ideas now current in the theory of "trimmed sums." And, of course, important use is made of the Hahn–Klass analytic equivalence for the operator-normed asymptotic normality.

(a) Let (1.5) hold, so by Theorem 2.1, $\mu = EX$ exists and may be taken as zero, B_n may be taken as symmetric and nonsingular, and $u^T B_n X_i$ are UAN. We will show that (1.6) holds. We have $u^T B_n S_n \xrightarrow{D} N(0, 1)$ for $u \in S^{d-1}$ so (see (2.15) and (2.16)) as $n \to \infty$, for each $\varepsilon > 0$,

$$nP\{|u^T B_n X| > \varepsilon\} \to 0 \tag{3.1}$$

and

$$nE[(u^T B_n X)^2 1_{\{|u^T B_n X| \leqslant \varepsilon\}}] \to 1. \tag{3.2}$$

By Gnedenko and Kolmogorov [5, Theorem 2, p. 140], (3.1) and (3.2) imply

$$\sum_{i=1}^{n} (u^T B_n X_i)^2 \xrightarrow{P} 1$$

which implies by Lemma 2.2 that $B_n V_n B_n \xrightarrow{P} I$, since $B_n V_n B_n$ is symmetric. Also

$$n\,|B_n \bar{X}_n|^2 = n^{-1}\,|B_n S_n| \xrightarrow{P} 0$$

so we obtain

$$B_n \bar{V}_n B_n = B_n \sum_{i=1}^{n} X_i X_i^T B_n - nB_n \bar{X}_n \bar{X}_n^T B_n = I + o_p(1),$$

where $o_p(1)$ is a quantity converging to 0 in probabilty. Thus (1.5) implies (1.6).

(b) Next we show that (1.6) implies (1.7). Our first task is to show that $E|X| < \infty$. According to Lemma 2.1 we may take B_n to be symmetric. Let the eigenvalue decomposition of B_n be $B_n = \theta(n)\,\Lambda(n)\,\theta(n)^T$, where $\Lambda(n)$ is diagonal and $\theta(n)$ is orthogonal. Use Helly's theorem to find a subsequence N so that $\theta(n) \to \theta$, an orthogonal matrix, when $n \to \infty$, $n \in N$. Then

$$\Lambda(n)\,\theta^T(n)\,\bar{V}_n \theta(n)\,\Lambda(n) = \theta^T(n)\,B_n \bar{V}_n B_n \theta(n) \xrightarrow{P} I, \qquad \text{as} \quad n \to \infty, n \in N.$$

Since the limit does not depend on N this implies

$$\lambda_j^2(n)\,\theta_j(n)^{\mathrm{T}}\bar{V}_n\theta_j(n) \xrightarrow{P} 1, \qquad 1 \leqslant j \leqslant d \quad (n \to \infty),$$

where $\theta_j(n)$ are the columns of $\theta(n)$ and $\lambda_j(n)$ are the diagonal elements of $\Lambda(n)$. Thus we have $\lambda_j(n) > 0$ and

$$\sum_{i=1}^{n} [\theta_j(n)^{\mathrm{T}}(X_i - \bar{X}_n)]^2 \overset{P}{\sim} 1/\lambda_j^2(n), \qquad 1 \leqslant j \leqslant d. \tag{3.3}$$

Adding over $1 \leqslant j \leqslant d$ and recalling that $\sum_{j=1}^{d} \theta_j(n)\,\theta_j(n)^{\mathrm{T}} = I$ gives

$$\frac{\sum_{i=1}^{n} |X_i - \bar{X}_n|^2}{b_n^2} \xrightarrow{P} 1,$$

where

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

Now when $E\,|X|^2 = \infty$,

$$\sum_{i=1}^{n} |X_i - \bar{X}_n|^2 = \sum_{i=1}^{n} |X_i|^2 - n\,|\bar{X}_n|^2 \overset{P}{\sim} \sum_{i=1}^{n} |X_i|^2 \tag{3.4}$$

which we demonstrate as follows: Given $\varepsilon > 0$ fix T so large that $P(|X| > T) \leqslant \varepsilon^2$, and let $u \in S^{d-1}$. Then

$$n^{-1/2}\left|\sum_{i=1}^{n} X_i 1_{\{|X_i| \leqslant T\}}\right| \leqslant n^{1/2}T \leqslant o_P\left(\sum_{i=1}^{n} |X_i|^2\right)^{1/2},$$

since we assume $E\,|X|^2 = \infty$. Also by the Cauchy–Schwarz inequality,

$$\begin{aligned} n^{-1/2}\left|\sum_{i=1}^{n} X_i 1_{\{|X_i| > T\}}\right| &\leqslant n^{-1/2}\left[\sum_{i=1}^{n} 1_{\{|X_i| > T\}}\right]^{1/2}\left(\sum_{i=1}^{n} |X_i|^2\right)^{1/2} \\ &\overset{P}{\sim} (P(|X| > T))^{1/2}\left(\sum_{i=1}^{n} |X_i|^2\right)^{1/2} \\ &\leqslant \varepsilon\left(\sum_{i=1}^{n} |X_i|^2\right)^{1/2}, \end{aligned}$$

and adding these gives

$$n^{1/2}\,|\bar{X}_n| = o_P\left(\sum_{i=1}^{n} |X_i|^2\right)^{1/2},$$

so, indeed, (3.4) is true. Thus when $E|X|^2=\infty$,

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

this is relative stability of $\sum_{i=1}^{n}|X_i|^2$ and is equivalent to (1.12). This implies $E|X|<\infty$ and, of course, this is also true if $E|X|^2<\infty$.

Now we can show that (1.7) holds with $A_n=\mu=EX$. Since

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

for (1.7a) with $A_n=\mu$, it will suffice to prove (1.7b) with $A_n=\mu$. Now $\bar{V}_n$ is invariant under the transformation $X_i\to X_i-\mu$, so assume $\mu=0$. Then by the weak law of large numbers, $\bar{X}_n\xrightarrow{P}0$, and for (1.7b) with $A_n=0$ it will suffice to show that

$$\limsup_{n\to\infty} n^{1/2}\lambda_{\max}(B_n)<\infty, \tag{3.5}$$

because then

$$n\,|B_n\bar{X}_n|^2\leqslant n\,|\bar{X}_n|^2\,\lambda_{\max}(B_n^2)$$

converges to zero in probability as $n\to\infty$. Let $\lambda_1(n)$ be the largest eigenvalue of B_n and suppose (3.5) fails, so there is a sequence N of integers such that $n\lambda_1^2(n)\to\infty$ and $n\to\infty$, $n\in N$. By (3.3) this means

$$n^{-1}\sum_{i=1}^{n}[\theta_1(n)^{\mathrm{T}}(X_i-\bar{X}_n)]^2\xrightarrow{P}0 \qquad (n\to\infty,\, n\in N).$$

But

$$|\theta_1(n)^{\mathrm{T}}\bar{X}_n|^2\leqslant|\bar{X}_n|^2\xrightarrow{P}0,$$

so we have

$$\begin{aligned}n^{-1}\theta_1(n)^{\mathrm{T}}V_n\theta_1(n)&=n^{-1}\sum_{i=1}^{n}[\theta_1(n)^{\mathrm{T}}X_i]^2\\&=n^{-1}\sum_{i=1}^{n}[\theta_1(n)^{\mathrm{T}}(X_i-\bar{X}_n)]^2+|\theta_1(n)^{\mathrm{T}}\bar{X}_n|^2\\&\xrightarrow{P}0 \qquad (n\to\infty,\, n\in N)\end{aligned}$$

which contradicts (2.11). Thus (3.5) holds and we have shown that (1.6) implies (1.7), in which we may take $A_n=\mu$.

(c) Next, (1.7) implies (1.8). To prove this, first assume $A_n = 0$ in (1.7). Then we have $B_n V_n B_n \xrightarrow{P} I$. This implies for $\varepsilon > 0$ and $u \in S^{d-1}$ that

$$nP(|u^{\mathrm{T}} B_n X| > \varepsilon) \to 0$$

which we prove as follows. Let $Z_{in} = u^{\mathrm{T}} B_n X_i$ and $Z_n = u^{\mathrm{T}} B_n X$ so we have

$$\sum_{i=1}^{n} Z_{in}^2 \xrightarrow{P} 1. \tag{3.6}$$

For $\varepsilon > 0$, (3.6) obviously implies

$$P(\max_{1 \leqslant i \leqslant n} |Z_{in}| > 1 + \varepsilon) \to 0$$

and so for $\varepsilon > 1$

$$nP(|Z_n| > \varepsilon) \to 0. \tag{3.7}$$

At this stage we do not know that Z_{in} are UAN. But (3.6) implies that for any t

$$[E(\exp(itZ_n^2)]^n = e^{it} + \Delta_n(t),$$

where $\Delta_n(t) \to 0$ as $n \to \infty$. Thus, defining the log appropriately,

$$\begin{aligned} E(1 - \cos tZ_n^2) - iE(\sin tZ_n^2) &= 1 - [e^{it} + \Delta_n(t)]^{1/n} \\ &= 1 - \exp\{\log[e^{it} + \Delta_n(t)]/n\} \\ &= -\log[e^{it} + \Delta_n(t)]/n + O(n^{-2}) \\ &= -it/n + o(n^{-1}), \end{aligned}$$

so that

$$nE(1 - \cos tZ^2) \to 0 \qquad (n \to \infty).$$

Now keep $0 < t < 1$. We have

$$2nE(1 - \cos tZ_n^2) \geqslant nt^2 E[Z_n^4 1_{\{tZ_n^2 \leqslant 1\}}] = nt^2 \int_0^{t^{-1/2}} y^4 \, dP(|Z_n| \leqslant y)$$

in which the left-hand side $\to 0$. After integrating by parts the right-hand side is

$$-nP(|Z_n| > t^{-1/2}) + 4nt^2 \int_0^{t^{-1/2}} y^3 P(|Z_n| > y) \, dy$$

and by (3.7) the first here is $o(1)$. But the integral is bounded below by

$$4nt^2 \int_0^{\varepsilon} y^3 P(|Z_n| > y)\, dy \geqslant nt^2 \varepsilon^4 P(|Z_n| > \varepsilon),$$

where $0 < \varepsilon < t^{-1/2}$. Thus (3.7) holds in fact for all $\varepsilon > 0$.

From (3.7), taking each component of $B_n X$ in turn, we obtain $nP(|B_n X| > \varepsilon) \to 0$; thus

$$\max_{1 \leqslant i \leqslant n} |B_n X_i| \xrightarrow{P} 0,$$

and so

$$\max_{1 \leqslant i \leqslant n} \sup_u |u^{\mathrm{T}} B_n X_i| \xrightarrow{P} 0.$$

Also from $B_n V_n B_n \xrightarrow{P} I$, by Lemma 2.2,

$$\inf_u (u^{\mathrm{T}} B_n V_n B_n u) \xrightarrow{P} 1.$$

If $u \in S^{d-1}$, let

$$v_n^{\mathrm{T}} = u^{\mathrm{T}} B_n^{-1} (u^{\mathrm{T}} B_n^{-2} u)^{-1/2} \in S^{d-1}.$$

Then we have for $1 \leqslant i \leqslant n$

$$\begin{aligned} \frac{(u^{\mathrm{T}} X_i X_i^{\mathrm{T}} u)}{(u^{\mathrm{T}} V_n u)} &= \frac{(v_n^{\mathrm{T}} B_n X_i X_i^{\mathrm{T}} B_n v_n)}{(v_n^{\mathrm{T}} B_n V_n B_n v_n)} \\ &\leqslant \frac{\sup_v |v^{\mathrm{T}} B_n X_i|^2}{\inf_v (v^{\mathrm{T}} B_n V_n B_n v)} \\ &\leqslant \max_{1 \leqslant i \leqslant n} \frac{\sup_v |v^{\mathrm{T}} B_n X_i|^2}{(1 + o_p(1))} \xrightarrow{P} 0. \end{aligned}$$

This proves (1.8) using the identity (Rao [21, p. 60]):

$$X_i^{\mathrm{T}} V_n^{-1} X_i = \sup_u \frac{|u^{\mathrm{T}} X_i|^2}{u^{\mathrm{T}} V_n u}. \tag{3.8}$$

We assumed at first that $A_n = 0$ in (1.7) but we can remove this restriction by noting that (1.7) as stated implies (1.6) and, hence, by part (b) of the proof, that A_n may be taken as μ and, hence, as 0. To see that (1.7) as stated implies (1.6), we have

$$\begin{aligned} B_n \sum_{i=1}^{n} (X_i - A_n)(X_i - A_n)^{\mathrm{T}} B_n &= B_n \sum_{i=1}^{n} (X_i - \bar{X}_n)(X_i - \bar{X}_n)^{\mathrm{T}} B_n \\ &\quad + n B_n (\bar{X}_n - A_n)(\bar{X}_n - A_n)^{\mathrm{T}} B_n \\ &= B_n \bar{V}_n B_n + o_p(1) \end{aligned}$$

by (1.7b), so $B_n \bar{V}_n B_n \xrightarrow{P} I$ $(n \to \infty)$, proving (1.6).

(d) Now we show that (1.8) implies (1.9). For this part of the proof we will assume that $u^{\mathrm{T}} X$ has a continuous distribution for each $u \in S^{d-1}$ and remove this restriction later. For brevity, write $Y_i^u = (u^{\mathrm{T}} X_i)^2$. According to (3.8), we have

$$\max_{1 \leqslant i \leqslant n} \sup_u \frac{Y_i^u}{\sum_{j=1}^n Y_j^u} \xrightarrow{P} 0.$$

Thus

$$\sup_u \frac{\max_{1 \leqslant i \leqslant n} Y_i^u}{\sum_{i=1}^n Y_i^u} \xrightarrow{P} 0$$

or

$$\inf_u \frac{{}^{(1)}S_n^u}{Y_{n1}^u} \xrightarrow{P} \infty,$$

where

$${}^{(1)}S_n^u = \sum_{i=1}^{n} Y_i^u - Y_{n1}^u \qquad \text{and} \qquad Y_{n1}^u = \max_{1 \leqslant i \leqslant n} Y_i^u.$$

Thus we have for all $x > 0$

$$P({}^{(1)}S_n^u > x Y_{n1}^u, \text{ all } u \in S^{d-1}) \to 1 \qquad (n \to \infty)$$

or

$$P \bigcup_u ({}^{(1)}S_n^u \leqslant x Y_{n1}^u) \to 0 \qquad (n \to \infty).$$

This implies, for all $x > 0$, that

$$\sup_u P({}^{(1)}S_n^u \leqslant x Y_{n1}^u) \to 0 \qquad (n \to \infty). \tag{3.9}$$

Now let $h_u(x)$ be the distribution function of Y_1^u and note that (see (1.4))

$$V_u(x^{1/2}) = E[(u^{\mathrm{T}}X)^2 1_{\{|u^{\mathrm{T}}X| \leq x^{1/2}\}}]$$
$$= E[Y_1^u 1_{\{Y_1^u \leq x\}}] = \int_0^x y\, dh_u(y).$$

Following arguments used in Hahn and Klass [9] and in Griffin [7, Lemma 2.2] we can define for each u a nondecreasing sequence

$$B_n(u) = \inf\{x > 0 : V_u(x)/x^2 \leq n^{-1}\}, \tag{3.10}$$

so that, by continuity of F and of V_u,

$$B_u^2(u) = nV_u(B_n(u)). \tag{3.11}$$

Then, since X is full,

$$\inf_u h_u(x) \to 1 \quad (x \to \infty), \qquad \inf_u B_n(u) \to \infty \quad (n \to \infty). \tag{3.12}$$

Next, since Y_i^u has a continuous distribution, Y_{n1}^u is uniquely defined a.s., and we have

$$P({}^{(1)}S_n^u \leq xY_{n1}^u) = nP(Y_2^u + \cdots + Y_n^u < xY_1^u,\ Y_i^u < Y_1^u \text{ for } 2 \leq i \leq n)$$
$$= n\int_0^\infty P(S_{n-1}^u(y) \leq xy)\, h_u^{n-1}(y)\, dh_u(y), \tag{3.13}$$

where for $y > 0$

$$S_n^u(y) = \sum_{i=1}^n Y_i^u(y)$$

and $Y_i^u(y)$, $1 \leq i \leq n$, are i.i.d. random variables, each with the conditional distribution of Y_1^u, given that $Y_1^u \leq y$. If $x, y > 0$ we have by Markov's inequality

$$P(S_{n-1}^u(y) \leq xy) = 1 - P(S_{n-1}^u(y) > xy)$$
$$\geq 1 - E[S_{n-1}^u(y)]/(xy)$$
$$= 1 - (n-1)\, V_u(y^{1/2})/(xyh_u(y))$$
$$\geq 1 - 2(n-1)\, V_u(y^{1/2})/(xy),$$

where we fix x_0 so large that $\inf_u h_u(x) \geq \frac{1}{2}$ for $x \geq x_0$, as is possible by

(3.12), and keep $y \geqslant x_0$. By (3.12) we may also take $B_n(u) \geqslant x_0$ for $u \in S^{d-1}$. Suppose that $y^{1/2} \geqslant B_n(u)$; (3.10) and (3.11) show then that

$$\frac{V_u(y^{1/2})}{y} \leqslant \frac{V_u(B_n(u))}{B_n^2(u)} = \frac{1}{n}$$

which gives

$$P(S_{n-1}^u(y) \leqslant xy) \geqslant 1 - 2(n-1)/(nx) \geqslant \tfrac{1}{2}$$

if $y^{1/2} \geqslant B_n(u) \geqslant x_0$ and $x \geqslant 4$. Going back to (3.13) shows then that (recall the continuity of h)

$$2P({}^{(1)}S_n^u \leqslant xY_{n1}^u) \geqslant n \int_{B_n^2(u)}^{\infty} h_u^{n-1}(y)\, dh_u(y)$$

$$= 1 - h_u^n(B_n^2(u)).$$

By (3.9) we then deduce that

$$\sup_u [1 - h_u^n(B_n^2(u))] \to 0 \qquad (n \to \infty). \tag{3.14}$$

Now we claim that

$$\sup_u n[1 - h_u(B_n^2(u))] \to 0 \qquad (n \to \infty). \tag{3.15}$$

If not, there are sequences $n_i \to \infty$ and u_i such that

$$n_i [1 - h_{u_i}(B_{n_i}^2(u_i))] \to \delta \in (0, \infty] \qquad (i \to \infty).$$

Since

$$h_u^n(B_n^2(u)) = \exp[n \log h_u(B_n^2(u))] \leqslant \exp\{-n[1 - h_u(B_n^2(u))]\},$$

if $\delta = \infty$, this means $h_{u_i}^{n_i}(B_{n_i}^2(u_i)) \to 0$ $(i \to \infty)$ which contradicts (3.14). Likewise if $0 < \delta < \infty$ we obtain a contradiction, so (3.15), indeed, holds.

Together with (3.11) this means we have proved

$$\sup_u \left\{ \frac{B_n^2(u)[1 - h_u(B_n^2(u))]}{V_u(B_n(u))} \right\} \to 0 \qquad (n \to \infty). \tag{3.16}$$

Essentially, to obtain (1.9), we need to replace $B_n(u)$ by an arbitrary

$x \to \infty$ in (3.16). To do this, we show that $B_{n+1}(u) \sim B_n(u)$, uniformly in $u \in S^{d-1}$. We have by (3.11)

$$\begin{aligned} 0 &= (n+1)\, B_{n+1}^{-2}(u)\, V_u(B_{n+1}(u)) - nB_n^{-2}(u)\, V_u(B_n(u)) \\ &= n[B_{n+1}^{-2}(u)\, V_u(B_{n+1}(u)) - B_n^{-2}(u)\, V_u(B_n(u))] + B_{n+1}^{-2}(u)\, V_u(B_{n+1}(u)) \\ &= nB_n^{-2}(u)\, V_u(B_n(u))[B_{n+1}^{-2}(u)\, B_n^2(u) - 1] \\ &\quad + n[V_u(B_{n+1}(u)) - V_u(B_n(u))]\, B_{n+1}^{-2}(u) + (n+1)^{-1}. \end{aligned}$$

Since $nB_n^{-2}(u)\, V_u(B_n(u)) = 1$ and

$$\begin{aligned} \frac{n[V(B_{n+1}(u)) - V(B_n(u))]}{B_{n+1}^2(u)} &= \frac{n \int_{B_n^2(u)}^{B_{n+1}^2(u)} x\, dh_u(x)}{B_{n+1}^2(u)} \\ &\leqslant n[1 - h_u(B_n^2(u))] \to 0 \end{aligned}$$

uniformly in u, by (3.15), we have, indeed, shown that

$$\sup_u \left| 1 - \frac{B_n^2(u)}{B_{n+1}^2(u)} \right| \to 0 \qquad (n \to \infty). \tag{3.17}$$

Finally, we claim that

$$\sup_u \frac{x^2[1 - h_u(x^2)]}{V_u(x)} \to 0 \qquad (x \to \infty). \tag{3.18}$$

If not, there are sequences $x_i \to \infty$, $u_i \in S^{d-1}$, such that

$$\frac{x_i^2[1 - h_{u_i}(x_i^2)]}{V_{u_i}(x_i)} \geqslant \delta \in (0, \infty].$$

Since $\inf_u B_n(u) \to \infty$ as $n \to \infty$ we can define a sequence n_i of integers by

$$n_i = \sup\{n : B_n(u_i) \leqslant x_i\}$$

so that

$$B_{n_i}(u_i) \leqslant x_i < B_{n_i+1}(u_i).$$

Now $n_i \to \infty$; if not, $n_i + 1 \leqslant N$, say. Since $x_i \leqslant B_{n_i+1}(u_i) \leqslant B_N(u_i)$, this means $B_N(u_i) \to \infty$. Then by (3.11), (1.4), and integration by parts,

$$B_N^2(u_i) = NV_{u_i}(B_N(u_i)) = N\int_0^{B_N(u_i)} y^2\, dP(|u_i^{\mathrm{T}}X| \leqslant y)$$

$$\leqslant 2N\int_0^{B_N(u_i)} yP(|u_i^{\mathrm{T}}X| > y)\, dy$$

$$= 2NB_N^2(u_i)\int_0^1 yP(|u_i^{\mathrm{T}}X| > yB_N(u_i))\, dy.$$

But then

$$1 \leqslant 2N\int_0^1 yP(|u_i^{\mathrm{T}}X| > yB_N(u_i))\, dy$$

$$\leqslant 2N\int_0^1 yP(|X| > yB_N(u_i))\, dy,$$

whereas the last expression converges to zero as $i \to \infty$ by dominated convergence, since N is fixed and $P(|X| > x) \to 0$ as $x \to 0$. Thus indeed $n_i \to \infty$.

We now have

$$\frac{x_i^2[1 - h_{u_i}(x_i^2)]}{V_{u_i}(x_i)} \leqslant \frac{B_{n_i+1}^2(u_i)}{B_{n_i}^2(u_i)}\,\frac{B_{n_i}^2(u_i)[1 - h_{u_i}(B_{n_i}^2(u_i))]}{V_{u_i}(B_{n_i}(u_i))}$$

$$\leqslant \sup_u \frac{B_{n_i+1}^2(u)}{B_{n_i}^2(u)} \sup_u \left\{\frac{B_{n_i}^2(u)[1 - h_u(B_{n_i}^2(u))]}{V_u(B_{n_i}(u))}\right\}$$

$$\to 0 \; (i \to \infty)$$

by (3.17) and (3.16). This demonstrates (3.18), which is (1.9), and shows that (1.8) implies (1.9) when $u^{\mathrm{T}}X$ is continuous.

Now assume (1.8) holds for a general X. By Lemma 2.3 we can choose $c > 0$ so that

$$P(\lambda_{\min}(V_n) > 10cn) \to 1 \qquad (n \to \infty). \tag{3.19}$$

Let Z_i be i.i.d. random variables distributed as $N(0, cI)$, independent of X_i, and define

$$\tilde{X}_i = X_i + Z_i, \qquad \tilde{V}_n = \sum_{i=1}^{n} \tilde{X}_i \tilde{X}_i^{\mathrm{T}}.$$

Then $u^{\mathrm{T}}\tilde{X}_i$ are continuous i.i.d. random variables. To show that (1.8) holds for $\tilde{X}_i$, consider

$$
\begin{aligned}
u^{\mathrm{T}}\tilde{V}_n u &= u^{\mathrm{T}} V_n u + 2u^{\mathrm{T}} \sum_{i=1}^{n} X_i Z_i^{\mathrm{T}} u + u^{\mathrm{T}} \sum_{i=1}^{n} Z_i Z_i^{\mathrm{T}} u \\
&\geqslant u^{\mathrm{T}} V_n u + 2u^{\mathrm{T}} \sum_{i=1}^{n} X_i Z_i^{\mathrm{T}} u \\
&\geqslant u^{\mathrm{T}} V_n u - 2\left[(u^{\mathrm{T}} V_n u)\left(u^{\mathrm{T}} \sum_{i=1}^{n} Z_i Z_i^{\mathrm{T}} u\right)\right]^{1/2},
\end{aligned}
$$

the last step following from the Cauchy–Schwartz inequality. Let $A(u)$ be the set where

$$
u^{\mathrm{T}} \sum_{i=1}^{n} Z_i Z_i^{\mathrm{T}} u \leqslant u^{\mathrm{T}} V_n u/9;
$$

then on $A(u)$,

$$
u^{\mathrm{T}} \tilde{V}_n u \geqslant u^{\mathrm{T}} V_n u/3 \geqslant u^{\mathrm{T}} \sum_{i=1}^{n} Z_i Z_i^{\mathrm{T}} u
$$

and also

$$
\frac{|u^{\mathrm{T}} \tilde{X}_i|}{(u^{\mathrm{T}} \tilde{V}_n u)^{1/2}} \leqslant 3 \frac{|u^{\mathrm{T}} X_i|}{(u^{\mathrm{T}} V_n u)^{1/2}} + \frac{|u^{\mathrm{T}} Z_i|}{\sum_{i=1}^{n} (u^{\mathrm{T}} Z_i Z_i^{\mathrm{T}} u)^{1/2}}.
$$

The first term on the right-hand side converges to 0 in probability uniformly in u and $1 \leqslant i \leqslant n$ by assumption. The second term converges in probability to 0 uniformly in u and i because $n^{-1/2} \sum Z_i$ is asymptotically normal with mean 0 and covariance matrix cI, so the implication (1.5) to (1.8) holds for the continuous random variable Z. Also

$$
\begin{aligned}
&P(A(u) \text{ fails to hold for some } u) \\
&\quad = P\left(u^{\mathrm{T}} \sum_{i=1}^{n} Z_i Z_i^{\mathrm{T}} u > u^{\mathrm{T}} V_n u/9 \text{ for some } u\right) \\
&\quad \leqslant P\left(\sup_u u^{\mathrm{T}} \sum_{i=1}^{n} Z_i Z_i^{\mathrm{T}} u/n > 10c/9\right) + P(\inf_u u^{\mathrm{T}} V_n u/n \leqslant 10c) \\
&\quad \leqslant P\left(\sum_{i=1}^{n} |Z_i|^2/n > 10c/9\right) + o(1) \qquad \text{(see (3.19))} \\
&\quad = o(1)
\end{aligned}
$$

since $n^{-1}\sum_{i=1}^{n}|Z_i|^2 \xrightarrow{P} c < 10c/9$. Thus for $\varepsilon > 0$,

$$\begin{aligned}
&P\left\{\max_{1\leqslant i\leqslant n}\sup_{u}\frac{|u^{\mathrm{T}}\tilde{X}_i|^2}{u^{\mathrm{T}}\tilde{V}_n u} > \varepsilon\right\} \\
&\quad\leqslant P\{\max_{1\leqslant i\leqslant n}|u^{\mathrm{T}}\tilde{X}_i|^2 > \varepsilon u^{\mathrm{T}}\tilde{V}_n u \text{ for some } u\in S^{d-1}\} \\
&\quad\leqslant P\{\max_{1\leqslant i\leqslant n}|u^{\mathrm{T}}\tilde{X}_i|^2 > \varepsilon u^{\mathrm{T}}\tilde{V}_n u \text{ for some } u\in S^{d-1},\ A(u) \text{ holds for all } u\} \\
&\qquad + P\{A(u) \text{ fails to hold for some } u\} \\
&\quad\to 0.
\end{aligned}$$

Thus (cf. (3.8))

$$\max_{1\leqslant i\leqslant n}\tilde{X}_i^{\mathrm{T}}\tilde{V}_n^{-1}\tilde{X}_i \xrightarrow{P} 0$$

and so (1.8) holds for $\tilde{X}$. By the continuous part of the proof, then, (1.9) holds for $\tilde{X}$. Since

$$P(|Z_1| > x) = O(e^{-x^2/2c}) \qquad (x\to\infty),$$

it is not hard to see from this that (1.9) holds for X. This proves that (1.8) implies (1.9), in general.

(e) That (1.9) implies (1.5) follows from Theorem 5 of Hahn and Klass [9] provided we can show that $E|X| < \infty$. If e_j is a coordinate vector (one in the jth position, zeros elsewhere) and $x_j = e_j^{\mathrm{T}}X$ is the jth component of X, we have by (1.9) that

$$\frac{x^2P(|x_j| > x)}{\int_0^x yP(|x_j| > y)\,dy} \to 0 \qquad (x\to\infty),\ 1\leqslant j\leqslant d. \tag{3.20}$$

Thus the random variables x_j are each in the domain of attraction of the normal, so $E|x_j|^{\alpha} < \infty$, $0\leqslant\alpha<2$, $1\leqslant j\leqslant d$. But since

$$|X| = (x_1^2 + \cdots + x_d^2)^{1/2} \leqslant d\max_{1\leqslant j\leqslant d}|x_j|,$$

we have $E|X|^{\alpha} < \infty$ for $0\leqslant\alpha<2$. Now apply the above-mentioned result of Hahn and Klass to $X_i - \mu$ to obtain $B_n(S_n - n\mu) \xrightarrow{D} N(0, I)$, which is (1.5).

(f) Finally we show that (1.8) and (1.10) are equivalent. If (1.8) holds then it holds with X_i replaced by $X_i - A$, where A is any constant, i.e.,

$$\max_{1 \leqslant i \leqslant n} (X_i - A)^{\mathrm{T}} \times \left\{ \sum_{j=1}^{n} (X_j - A)(X_j - A)^{\mathrm{T}} \right\}^{-1} (X_i - A) \xrightarrow{P} 0, \qquad n \to \infty. \tag{3.21}$$

This is true because we showed above that (1.8) implies (1.5), and (1.5) is invariant under the transformation $X_i \to X_i - A$ (after redefining $A_n = A_n + nA$), and (1.5) in this form implies (3.21) via (1.8). Likewise (1.10) is obviously invariant under this transformation. Furthermore, when (1.5) holds, we know from Theorem 2.1 that $\mu = E(X)$ exists. So we can let $X_i = X_i - \mu$, or equivalently, take $\mu = 0$, without loss of generality. Then $\bar{X}_n \xrightarrow{P} \mu = 0$ and, since

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

we have

$$u^{\mathrm{T}} \bar{V}_n u = u^{\mathrm{T}} V_n u + o_p(n) = (1 + o_p(1))\, u^{\mathrm{T}} V_n u,$$

the last equality holding since, with probability approaching one, $u^{\mathrm{T}} V_n u/n$ is bounded away from zero uniformly in u by Lemma 2.3. Thus $u^{\mathrm{T}} \bar{V}_n u/n$ must be bounded away from zero uniformly in u with probability approaching one, and

$$\begin{aligned} &\sup_u \frac{|u^{\mathrm{T}}(X_i - \bar{X}_n)|}{(u^{\mathrm{T}} \bar{V}_n u)^{1/2}} \\ &\quad \leqslant \sup_u \frac{|u^{\mathrm{T}} X_i|}{[1 + o_p(1)](u^{\mathrm{T}} V_n u)^{1/2}} + \sup_u \frac{|u^{\mathrm{T}} \bar{X}_n|}{(u^{\mathrm{T}} \bar{V}_n u)^{1/2}}. \end{aligned}$$

The last expression converges in probability to zero uniformly in $1 \leqslant i \leqslant n$ if (1.8) holds (and $\bar{X}_n \xrightarrow{P} 0$). Thus (see (3.8)) (1.8) implies (1.10). Conversely, $u^{\mathrm{T}} \bar{V}_n u \leqslant u^{\mathrm{T}} V_n u$, so it is easy to show that (1.10) implies (1.8) (without invoking (1.5)). ∎

Remarks. (i) For the counterexample mentioned in Section 1, take $d = 2$, $E|X|^2 < \infty$, $EX = 0$, and let $B_n = \operatorname{diag}(n^{-1/2}, n^{-1})$, $\delta = [0\ \ 1]^{\mathrm{T}}$. Then we have

$$B_n S_n \xrightarrow{D} N(0, I - \delta\delta^{\mathrm{T}}) \qquad (n \to \infty),$$

i.e., convergence to a singular normal, so (1.5) fails. But

$$\sum_{i=1}^{n} (u^{\mathrm{T}} B_n X_i)^2 \xrightarrow{P} 1 - (u^{\mathrm{T}} \delta)^2 \qquad (n \to \infty),$$

so

$$\sum_{i=1}^{n} [u^{\mathrm{T}} B_n (X_i - A_n)]^2$$

$$= \sum_{i=1}^{n} (u^{\mathrm{T}} B_n X_i)^2 - 2(u^{\mathrm{T}} B_n S_n)(n^{-1/2} u^{\mathrm{T}} \delta) + (u^{\mathrm{T}} \delta)^2$$

$$\xrightarrow{P} 1 - (u^{\mathrm{T}} \delta)^2 + (u^{\mathrm{T}} \delta)^2 = 1 \qquad (n \to \infty),$$

and (1.7a) holds, if we let $A_n = n^{-1/2} B_n^{-1} \delta$. Note that

$$n\, |B_n (A_n - \bar{X}_n)|^2 = n\, |n^{-1/2} \delta - n^{-1} B_n S_n|^2 \xrightarrow{P} 1 \qquad (n \to \infty),$$

so (1.7b) fails.

(ii) We remark also that (1.11) does not imply (1.9); because (3.20) implies (1.11), but (3.20) cannot imply (1.9) by the example on page 238 of Hahn and Klass [9]. In fact, the uniformity in u is essential in (1.9), as they show. However, when X has a spherically symmetric distribution, each projection of X has the same distribution, so X being in the operator-normed domain of attraction of the normal in the sense of (1.5) is equivalent to each component of X or to $|X|$ being in the domain of attraction of the one-dimensional normal, in the usual sense.

(iii) Finally, we show that (1.5) implies (1.15). Let $B_n S_n \xrightarrow{D} N(0, I)$ and $\mu = 0$. Then from $B_n \bar{V}_n B_n \xrightarrow{P} I$ (see (1.6)) we obtain $B_n^{-1} \bar{V}_n^{-1} B_n^{-1} \xrightarrow{P} I$ (Lemma 2.2) and so

$$S_n^{\mathrm{T}} \bar{V}_n^{-1} S_n = Q_n^{\mathrm{T}} Q_n - Q_n^{\mathrm{T}} \Delta_n Q_n,$$

where

$$\Delta_n = I - B_n \bar{V}_n^{-1} B_n \xrightarrow{P} 0$$

and

$$Q_n = B_n S_n \xrightarrow{D} N(0, I).$$

Thus by the continuous mapping theorem $Q_n^T Q_n \xrightarrow{D} \chi_d^2$. Also

$$|Q_n^T \Delta_n Q_n| \leqslant \sum_{j=1}^{d} \sum_{k=1}^{d} |q_j(n)\, q_k(n)\, \Delta_{jk}(n)| \xrightarrow{P} 0,$$

where $q_j(n)$ and $\Delta_{jk}(n)$ are the components of Q_n and Δ_n. The former are stochastically bounded, since Q_n converges in distribution to $N(0, I)$; the latter converge in probability to 0, since $\Delta_n \xrightarrow{P} 0$. Replacing X by $X-\mu$ completes the proof of (1.15).

Acknowledgment

I am grateful to Professor Phil Griffin for a hint which helped with the proof of (3.4).

References

[1] Breiman, L. (1965). On some limit theorems similar to the arcsine law. *Theory Probab. Appl.* **10** 323–331.

[2] Eicker, F. (1963). Asymptotic normality and consistency of the least squares estimates for families of regressions. *Ann. Math. Statist.* **34** 447–456.

[3] Fahrmeir, L., and Kaufmann, H. (1985). Consistency and asymptotic normality of the maximum likelihood estimator in generalised linear models. *Ann. Statist.* **13** 342–368.

[4] Feller, W. (1976). *An Introduction to Probability Theory and Its Applications*, Vol. 2, 2nd ed., Wiley, New York.

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

[6] Greenwood, P., and Resnick, S. (1979). A bivariate stable characterisation and domains of attraction. *J. Multivariate Anal.* **9** 206–221.

[7] Griffin, P. S. (1986). Matrix normalised sums of independent identically distributed random vectors. *Ann. Probab.* **14** 224–246.

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

[9] 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. Probab.* **8** 262–280.

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

[11] Hajek, J. (1967). *Linear Rank Statistics*, Wiley, New York.

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

[13] Kesten, H. (1971). Sums of random variables with infinite expectation. *Amer. Math. Monthly* **78** 305–308.

[14] Lai, T. L. (1986). Asymptotically efficient adaptive control in stochastic regression models. *Adv. in Appl. Math.* **7** 23–45.

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

[16] Maller, R. A. (1978). Relative stability and the strong law of large numbers. *Z. Wahrsch. Verw. Gebiete* **43** 141–148.

[17] Maller, R. A. (1989). Regression with autoregressive errors—Some asymptotic results. *Statistics* **20** 23–39.

[18] MARONNA, R. A. (1976). Robust M-estimators of multivariate location and scatter. *Ann. Statist.* **1** 51–67.

[19] O'BRIEN, G. L. (1980). A limit theorem for sample maxima and heavy branches in Galton Watson trees. *J. Appl. Probab.* **17** 539–545.

[20] PAKES, A. G. (1982). Classical and inverse calibration for the simple linear regression model. Dept. of Math., University of Western Australia, research report.

[21] RAO, C. R. (1973). *Linear Statistical Inference and Its Applications.* 2nd ed., Wiley, New York.

[22] ROGOZIN, B. A. (1976). Relatively stable walks. *Theory Probab. Appl.* **21** 375 379.

[23] SMITH, P. J. (1988). Asymptotic properties of linear regression estimators under a fixed censorship model. *Austral. J. Statist.* **30** 52–66.

[24] TYLER, D. E. (1983). The asymptotic distribution of principal component roots under local alternatives to multiple roots. *Ann. Statist.* **11** 1232–1242.