Elect. Comm. in Probab. **9** (2004) 162–171

ELECTRONIC COMMUNICATIONS in PROBABILITY

ON STANDARD NORMAL CONVERGENCE OF THE MULTIVARIATE STUDENT t-STATISTIC FOR SYMMETRIC RANDOM VECTORS

EVARIST GINÉ[1]
Department of mathematics, U-3009, University of Connecticut, Storrs, CT, 06269, USA
Email: `gine@math.uconn.edu`

FRIEDRICH GÖTZE[2]
Facultät für Mathematik, Universität Bielefeld, 33501 Bielefeld, Germany
Email: `goetze@mathematik.uni-bielefeld.de`

Submitted 4 June 2004, *Final version accepted* 8 November 2004

AMS 2000 subject classifications: Primary 60F05; secondary 62E20.
Keywords: multivariate Student t-statistic, multivariate operator self-normalized sums, generalized normal domain of attraction

Abstract
It is proved that if the multivariate Student t-statistic based on i.i.d. symmetric random vectors is asymptotically standard normal, then these random vectors are in the generalized domain of attraction of the normal law. Uniform integrability is also considered, even in the absence of symmetry.

1. Introduction.

Let $X,\ X_i,\ i \in \mathbf{N}$, be i.i.d. $\mathbf{R}^d$ valued random variables that we assume are *full*, that is, if for each $u \in S^{d-1}$, the random variable $u^T X$ is not almost surely constant. X is said to be in the generalized domain of attraction of the normal law ($X \in GDOAN$ for short) if there exist (nonrandom) matrices A_n and constant vectors b_n such that

$$A_n(S_n - b_n) \to_d N(0, I),$$

where $S_n = \sum_{i=1}^n X_i$. See Hahn and Klass (1980), Maller (1993) and Sepanski (1994) for properties of GDOAN, including analytic characterizations. In particular, if $X \in GDOAN$ then $E|X| < \infty$ and the coordinates of X all belong to the domain of attraction of the normal law

[1] Research partially supported by the SFB 343 at Bielefeld and by NSA Grant No. H98230-04-1-0075.
[2] Research supported by the SFB 343 at Bielefeld.

in $\mathbf{R}$, actually in a uniform way (however X itself does not necessarily belong to the domain of attraction of the normal law in $\mathbf{R}^d$). Moreover, the matrices A_n can be taken to be symmetric and non-singular, and b_n to be nEX.
Let

$$C_n = \sum_{i=1}^{n} X_i X_i^T, \quad \overline{C}_n = \sum_{i=1}^{n} (X_i - \overline{X}_n)(X_i - \overline{X}_n)^T,$$

where $\overline{C}_n$ is the sample covariance matrix of $X_1, \dots, X_n$. If X is *full*, then (Maller (1993), Lemma 2.3) C_n and $\overline{C}_n$ are nonsingular on sets of probability tending to 1. So, for X full, in statements about convergence in law or in probability, we can assume without loss of generality that C_n^{-1} and $\overline{C}_n^{-1}$ exist, or, alternatively, we can replace $C_n^{-1}(\omega)$ and $\overline{C}_n^{-1}(\omega)$ by the zero matrix when they are not defined, and this will be our convention. The multidimensional version of the Student t-statistic for the random vector $X - EX$ is

$$\overline{T}_n := \overline{C}_n^{-1/2}(S_n - nEX),$$

and we will also consider

$$T_n := C_n^{-1/2} S_n,$$

the multivariate equivalent of *selfnormalized sums*. Here, $C_n^{-1/2}$ and $\overline{C}_n^{-1/2}$ are respectively the *symmetric* square roots of C_n^{-1} and $\overline{C}_n^{-1}$. Both, Sepanski (1996) and Vu, Maller, and Klass (1996) proved the following:

Theorem 1.1. $X \in GDOAN \implies \overline{C}_n^{-1/2}(S_n - nEX) \to_d N(0, I)$.

Giné, Götze and Mason (1997) proved the converse to this statement in $\mathbf{R}$. We believe (or at least we hope) that the converse to Theorem 1.1 is true as well in $\mathbf{R}^d$, and the main object of this note is to prove that this is indeed the case if the distribution of X is symmetric. This problem seems to have been first mentioned in Maller (1993), with the comment that it is more difficult than the direct implication. In fact, Roy Erickson (see Griffin and Mason (1991)) had a beautiful short proof of the result in $\mathbf{R}$ for symmetric variables, and here we will extend his proof to $\mathbf{R}^d$.
In order to handle the non-symmetric case in $\mathbf{R}$ a) we had to prove that stochastic boundedness of T_n, $n \in \mathbf{N}$, implies square exponential uniform integrability, meaning that there is $\lambda > 0$ such that

$$\sup_n E e^{\lambda T_n^2} < \infty,$$

(although uniform integrability of $\|T_n\|^6$ was enough for the proof) and b) we had to use the Mellin transform to break the dependence in T_n in order to get better estimates of certain moments. We will also show here that, even in the absence of symmetry, in $\mathbf{R}^d$, if the sequence T_n is stocastically bounded, then

$$\sup_n E\|T_n\|^m < \infty$$

for all $m > 0$, under the assumption that the law of X assigns mass zero to all hyperplanes through the origin. On the other hand, we do not know at present how to break the dependence in T_n for $d > 1$, and probably this requires some deep matrix analysis, even under this restricted continuity hypothesis on the law of X.
The following partial summary of results on $GDOAN$, taken from Maller (1993) will be useful in what follows:

Theorem 1.2. *For X full, the following conditions are equivalent to $X \in GDOAN$:*

i) $E\|X\|^p < \infty$ *for all* $0 < p < 2$ *and there exist (nonstochastic) symmetric nonsingular matrices* A_n *such that* $A_n(S_n - nEX) \to_d N(0, I)$*;*

ii) there exist matrices A_n *such that* $A_n \overline{C}_n A_n^T \to_{pr} I$*;*

iii) there exist A_n *and* b_n *such that*

$$A_n \sum_{i=1}^{n} (X_i - b_n)(X_i - b_n)^T A_n^T \to_{pr} I \quad \text{and} \quad n|A_n(b_n - \overline{X}_n)| \to_{pr} 0;$$

iv) $\max_{1 \le i \le n} X_i^T C_n^{-1} X_i \to_{pr} 0$*;*

v) $\max_{1 \le i \le n} (X_i - \overline{X}_n)^T \overline{C}_n^{-1} (X_i - \overline{X}_n) \to_{pr} 0$.

Moreover, the matrices A_n *can be taken to be the same in i)-iii) and* b_n *can be taken to be* EX *in iii).*

Vu, Maller and Klass (1980) also proved that if for A_n nonsingular and nonstochastic, and for symmetric stochastic V_n and for $\mathbf{R}^d$ random vectors Y_n, one has $A_n Y_n \to_d N(0, I)$ and $A_n V_n A_n^T \to_{pr} I$ then one also has $V_n^{-1/2} Y_n \to_d N(0, I)$. From this and from Theorem 1.2 ii) and iii) with $b_n = EX$, it follows that:

Theorem 1.3. *If* $X \in GDOAN$ *and* $EX = 0$*, then, both*

$$C_n^{-1/2} S_n \to_d N(0, I) \quad \text{and} \quad \overline{C}_n^{-1/2} S_n \to_d N(0, I).$$

2. Elementary reductions of the problem.

In Statistics there is more interest in $\overline{C}_n^{-1/2} S_n$, whereas $C_n^{-1/2} S_n$ may be easier to handle. The following remark shows that it does not mater whether one considers $\overline{C}_n$ or C_n in the problem we will treat here.

Lemma 2.1. *If either of the two sequences* $\{C_n^{-1/2} S_n\}$ *and* $\{\overline{C}_n^{-1/2} S_n\}$ *is stochastically bounded, then* $C_n^{-1/2} \overline{C}_n^{1/2} \to_{pr} I$ *and, in particular,*

$$\overline{C}_n^{-1/2} S_n - C_n^{-1/2} S_n \to_{pr} 0,$$

so that the two sequences are weak convergence equivalent.

Proof. (Since X is full, we can assume for this lemma that C_n and $\overline{C}_n$ are nonsingular.) Both directions are similar, so let us assume

$$\overline{C}_n^{-1/2} S_n \to_d N(0, I).$$

Then, $n\|\overline{C}_n^{-1/2} \overline{X}_n\|^2 = n^{-1} \|\overline{C}_n^{-1/2} S_n\|^2 \to_{pr} 0$. Therefore,

$$\overline{C}_n^{-1/2} C_n \overline{C}_n^{-1/2} = I + n \overline{C}_n^{-1/2} \overline{X}_n \overline{X}_n^T \overline{C}_n^{-1/2} \to_{pr} I.$$

Set $A_n = \overline{C}_n^{-1/2} C_n^{1/2}$, so that $A_n A_n^T \to_{pr} I$. Then, A_n diagonalizes and has strictly positive eigenvalues λ_{ni} (e.g. Horn and Johnson (1985), Theorem 7.6.3, p. 465). Moreover, the entries of A_n, a_{nij}, are stochastically bounded since, for each i, $\sum_j a_{nij}^2 \to_{pr} 1$, but then so are the eigenvalues $\lambda_{n,1} \le \ldots \le \lambda_{n,d}$ because $0 \le \lambda_{n,i} \le \sum_{j=1}^d a_{njj}$. So, the random vectors $(a_{nij}, 1 \le i,j \le d; \lambda_{n,i}, 1 \le i \le d)$ converge in law along a subsequence of every subsequence; pick up one convergence subsequence (still, for convenience, denoted by n), let $(a_{ij}, \lambda_i, 1 \le i,j \le d)$ be the limit in law, and set $A = \{a_{ij}\}$. Now, for each i, $\det(A_n - \lambda_{ni} I) \to_d \det(A - \lambda_i I)$, but $\det(A_n - \lambda_{ni} I) = 0$, so, $\det(A - \lambda_i I) = 0$ a.s. for each i, and the random variables $\lambda_i(\omega)$ are, for almost every ω, the eigenvalues of $A(\omega)$. Moreover, $AA^T = I$ *a.s.*, that is, A is unitary. Hence, the eigenvalues λ_i being nonnegative, we have $\lambda_i = 1$ a.s. The polynomial in λ, $\det(A_n - \lambda I)$, determines a continuous map from $\mathbf{R}^{d^2}$ into $C[-M, M]$ say for $M = d$, so that $\det(A_n - \lambda I) \to_d \det(A - \lambda I)$ in law in $C[-M, M]$. Since $\det(A_n - \lambda I) = (\lambda_{n1} - \lambda) \cdots (\lambda_{nd} - \lambda) \to_{pr} (1 - \lambda)^d$, we have $\det(A - \lambda I) = (1 - \lambda)^d$ a.s. on $[-M, M]$. But the only unitary matrix with this characteristic polynomial is the identity, that is, $A = I$ a.s. We have just proved that $A_n \to_{pr} I$. Hence also $C_n^{-1/2} \overline{C}_n^{1/2} \to_{pr} I$. This last limit and the hypothesis give

$$(I - C_n^{-1/2} \overline{C}_n^{1/2}) \overline{C}_n^{-1/2} S_n \to_{pr} 0,$$

that is, $\overline{C}_n^{-1/2} S_n - C_n^{-1/2} S_n \to_{pr} 0$. q.e.d.

So, when proving the converse of Theorem 1.3, which is our object, by Lemma 2.1, we only need to consider $C_n^{-1/2}$ (instead of $\overline{C}_n^{-1/2}$), that is, it suffices to prove that

$$C_n^{-1/2} S_n \to_d N(0, I) \implies X \in GDOAN \text{ and } EX = 0. \tag{2.1}$$

Next we make another remark to the effect that the centering question can be disposed of easily (obviously, this is not needed in the symmetric case).

Lemma 2.2. *If $X \in GDOAN$ and $C_n^{-1/2} S_n \to_d N(0, I)$ then $EX = 0$.*

Proof. The second hypothesis implies, by Lemma 2.1, that $C_n^{-1/2} \overline{C}_n^{1/2} \to_{pr} I$ and the first, by Theorem 1.1, that $\overline{C}_n^{-1/2}(S_n - nEX) \to_d N(0, I)$. Therefore, $C_n^{-1/2}(S_n - nEX) \to_d N(0, I)$. This, together with the second hypothesis gives that

$$\text{the sequence } \{nC_n^{-1/2} EX\}_{n=1}^\infty \text{ is tight.} \tag{2.2}$$

Let $\lambda_{\min}(A)$, $\lambda_{\max}(A)$ denote respectively the smallest and the largest eigenvalues of a positive definite symmetric matrix A (possibly random). Suppose that $\lambda_{\max}(C_n)/n^2 \to_{pr} 0$ and that $EX \ne 0$. Then,

$$\begin{aligned} \Pr\left\{ \|nC_n^{-1/2} EX\| > M \right\} &\ge \Pr\left\{ n\|EX\| \lambda_{\min}(C_n^{-1/2}) > M \right\} \\ &= \Pr\left\{ \lambda_{\max}(C_n) < \frac{n^2 \|EX\|^2}{M^2} \right\} \to 1 \end{aligned}$$

as $n \to \infty$, for all $M > 0$, which contradicts the tightness in (2.2). So, it suffices to prove that $\lambda_{\max}(C_n)/n^2 \to_{pr} 0$. Let $\xi^{(k)}$, $k = 1, \ldots, d$, denote the k-th coordinate of X, and $\xi_i^{(k)}$ that of X_i, $i = 1, \ldots, n$. If $X \in GDOAN$ then every coordinate $\xi^{(k)}$ is in the domain of attraction of the normal law in $\mathbf{R}$ and therefore, by Raikov's theorem, for each k, there exist sequences of constants $\{a_{n,k}\}_{n=1}^\infty$ such that $a_{n,k} = \sqrt{n}\ \ell_k(n)$ with either $\ell_k(n) = 1$ for all n or $\ell_k(n)$

slowly varying and increasing to infinity, such that $\sum_{i=1}^n (\xi_i^{(k)})^2/a_{n,k}^2 \to_{pr} 1$ if $\ell_k(n) \nearrow \infty$ and $\sum_{i=1}^n (\xi_i^{(k)})^2/a_{n,k}^2 \to_{pr} E(\xi^{(k)})^2$ if $\ell_k(n) \equiv 1$. Hence, there is $c < \infty$ such that

$$\Pr\left\{\frac{\text{trace}(C_n)}{\max_{1\le k\le d} a_{n,k}^2} > c\right\} = \Pr\left\{\frac{\sum_{k=1}^d \sum_{i=1}^n (\xi_i^{(k)})^2}{\max_{1\le k\le d} a_{n,k}^2} > c\right\} \to 0$$

and therefore, since $\max_{1\le k\le d} a_{n,k}^2/n^2 \to 0$ and the trace dominates the largest eigenvalue, we obtain $\lambda_{\max}(C_n)/n^2 \to_{pr} 0$. q.e.d.

So, as a consequence of the last two lemmas, in order to prove that

$$\overline{C}_n^{-1/2} S_n \to_d N(0, I) \implies X \in GDOAN \text{ and } EX = 0,$$

it suffices to prove:

$$C_n^{-1/2} S_n \to_d N(0, I) \implies X \in GDOAN, \tag{2.3}$$

a reduction of the problem that is trivial in the case $d = 1$.

3. Proof of statement (2.3) in the symmetric case.

We begin with two lemmas that hold in general (without assuming symmetry of X).

Lemma 3.1. $\sum_{i=1}^n X_i^T(\omega) C_n^{-1}(\omega) X_i(\omega) = d$ *whenever* $C_n(\omega)$ *is non-singular, and this sum is zero otherwise.*

Proof. Assuming C_n non-singular, if $C_n = (c_{i,j})$ and we let $M_{i,j}$ denote the (i,j)-th minor of C_n multiplied by $(-1)^{i+j}$, the above expression is just $\sum_{i,j} c_{i,j} M_{i,j}/(\det C_n)$, but $\sum_{i,j} c_{i,j} M_{i,j} = d(\det C_n)$. In the singular case, the lemma follows from the convention $C_n^{-1}(\omega) = \mathbf{0}$. q.e.d.

Lemma 3.2. $\sum_{1\le i,j\le n} \left[X_i^T C_n^{-1} X_j\right]^2 = \sum_{1\le i,j\le n} \left[X_i^T C_n^{-1} X_j\right]\left[X_j^T C_n^{-1} X_i\right] = d$ *at those* ω *for which* $C_n(\omega)$ *is non-singular, and this sum is zero otherwise.*

Proof. With the same notation as in the previous lemma, assuming $\det C_n \ne 0$, and setting $\xi_i^{(r)}$ to be the r-th coordinate of X_i, $r = 1, \dots, d$, we have:

$$\begin{aligned}
(\det C_n)^2 \sum_{i,j} \left(X_i^T C_n^{-1} X_j\right)^2 &= \sum_{i,j} \left(\sum_{r,s} \xi_i^{(r)} \xi_j^{(s)} M_{r,s}\right)^2 \\
&= \sum_{i,j,r,s,t,u} \xi_i^{(r)} \xi_j^{(s)} \xi_i^{(t)} \xi_j^{(u)} M_{r,s} M_{t,u} \\
&= \sum_{t,u} M_{t,u} \left(\sum_{r,s} \left(\sum_i \xi_i^{(r)} \xi_i^{(t)}\right) \left(\sum_j \xi_j^{(s)} \xi_j^{(u)}\right) M_{r,s}\right) \\
&= \sum_{t,u} M_{t,u} \left(\sum_{r,s} c_{r,t} c_{s,u} M_{r,s}\right) = \sum_{t,u} M_{t,u} \left(\sum_s c_{s,u} \left(\sum_r c_{r,t} M_{r,s}\right)\right) \\
&= \sum_{t,u} M_{t,u} \left(\sum_s c_{s,u} (\det C_n) \delta_{t,s}\right) \\
&= (\det C_n) \sum_{t,u} M_{t,u} c_{t,u} = d(\det C_n)^2.
\end{aligned}$$

The other identity has a similar proof. q.e.d.

Corollary 3.3. *If X is symmetric, then, for all $k \in \mathbf{N}$,*

$$\sup_n E \left\| C_n^{-1/2} S_n \right\|^{2k} \leq (2kd)^k.$$

Proof. Let ε_i, $i \in \mathbf{N}$, be i.i.d. with $\Pr\{\varepsilon_i = 1\} = \Pr\{\varepsilon_i = -1\} = 1/2$, independent of X_i, $i \in \mathbf{N}$. Then, since $X_i X_i^T = (\varepsilon_i X_i)(\varepsilon_i X_i)^T$ and $(X_i, i \leq n)$ has the same joint distribution as $(\varepsilon_i X_i, i \leq n)$, it follows that $C_n^{-1} S_n$ has the same probability law as $C_n^{-1}\left(\sum_{i=1}^n \varepsilon_i X_i\right)$. So, letting E_ε denote integration with respect to the Rademacher variables only, we get, by Khinchin's inequality (e.g., de la Peña and Giné (1999), p. 16) and Lemma 3.1, that

$$\begin{aligned} E \left\| C_n^{-1/2} S_n \right\|^{2k} &= E \left\| \sum_{i=1}^n \varepsilon_i C_n^{-1/2} X_i \right\|^{2k} = E E_\varepsilon \left\| \sum_{i=1}^n \varepsilon_i C_n^{-1/2} X_i \right\|^{2k} \\ &\leq (2k)^k E \left(\sum_{i=1}^n \left\| C_n^{-1/2} X_i \right\|^2 \right)^k \\ &= (2k)^k E \left(\sum_{i=1}^n X_i^T C_n^{-1} X_i \right)^k \leq (2kd)^k. \end{aligned}$$

q.e.d.

Theorem 3.4. *Let X be a symmetric full random variable in $\mathbf{R}^d$. Then*

$$C_n^{-1/2} S_n \to_d N(0, I) \iff X \in GDOAN$$

and

$$\overline{C}_n^{-1/2} S_n \to_d N(0, I) \iff X \in GDOAN.$$

Proof. By Theorem 1.3 and Lemma 2.1, it suffices to prove

$$C_n^{-1/2} S_n \to_d N(0, I) \implies X \in GDOAN,$$

and, by Theorem 1.2 iv, this further reduces to proving that

$$\max_{1 \leq i \leq n} X_i^T C_n^{-1} X_i \to_{pr} 0, \tag{3.1}$$

assuming the multivariate Student t-statistic is asymptotically standard normal. Under this assumption, Corollary 3.3 implies, by uniform integrability, that

$$E \left\| C_n^{-1/2} S_n \right\|^4 \to E\|Z\|^4 = d^2 + 2d, \tag{3.2}$$

where Z is $N(0, I)$. Now, X being full, by Lemma 3.1 and the randomization from the proof of Corollary 3.3, we have that, asymptotically,

$$\begin{aligned} E \left\| C_n^{-1/2} S_n \right\|^4 &= E \left[\sum_{1 \leq i, j \leq n} \varepsilon_i \varepsilon_j X_i^T C_n^{-1} X_j \right]^2 \\ &= E \left[d I(\det C_n \neq 0) + \sum_{1 \leq i \neq j \leq n} \varepsilon_i \varepsilon_j X_i^T C_n^{-1} X_j \right]^2 \\ &= d^2 I(\det C_n \neq 0) \\ &\quad + n(n-1) \left[E \left(X_1^T C_n^{-1} X_2 \right)^2 + E \left(X_1^T C_n^{-1} X_2 \right) \left(X_2^T C_n^{-1} X_1 \right) \right]. \end{aligned}$$

Adding and subtracting $2nE\left(X_1^T C_n^{-1} X_1\right)^2$ and applying Lemma 3.2, this becomes

$$E\left\|C_n^{-1/2} S_n\right\|^4 = (d^2+2d) I(\det C_n \neq 0) - 2nE\left(X_1^T C_n^{-1} X_1\right)^2. \tag{3.3}$$

Comparing (3.2) and (3.3) and noting that $\Pr\{\det C_n \neq 0\} \to 1$, we conclude

$$nE\left(X_1^T C_n^{-1} X_1\right)^2 \to 0,$$

and therefore,

$$E \max_{1\le i\le n} \left(X_i^T C_n^{-1} X_i\right)^2 \le nE\left(X_1^T C_n^{-1} X_1\right)^2 \to 0,$$

which implies (3.1). q.e.d.

4. Integrability of the Student t-statistic in the general case.

We recall the notation $T_n = C_n^{-1} S_n$. We also recall from matrix theory (e.g., Horn and Johnson (1985), Corollary 7.7.4, p. 471) that if C and D are two positive definite matrices such that $C \le D$, meaning that $D - C$ is positive semidefinite, then $D^{-1} \le C^{-1}$, and in particular then, for all x, $x^T D^{-1} x \le x^T C^{-1} x$. In this section X needs not be symmetric, however, our proof of integrability will require $C_n^{-1} \le C_m^{-1}$ a.s. for $m \le n$, at least for all m large enough; of course, $C_n \ge C_m$ and therefore, $C_n^{-1} \le C_m^{-1}$ on the event $\det C_m \neq 0$, but it seems impossible to define C_m on the set $\det C_m = 0$ only in terms of $X_1, \dots, X_m$ and in such a way that $C_m \le C_n$ a.s. for all $n \ge m$. This problem disappears if $\det C_n \neq 0$ a.s. from some non-random n on, and it is easy to check that this happens if and only if $\Pr\{u^T X = 0\} = 0$ for all $u \in S^{d-1}$, that is, iff the law of X assigns probability zero to any hyperplane through the origin. Moreover, in this case, $\det C_n \neq 0$ a.s. for all $n \ge d$. [If, for some $n \ge d$ and $\omega \in \Omega$, U is the unitary transformation that diagonalizes $C_n(\omega)$ and $v_i = UX_i(\omega)$ has coordinates v_i^k, $k = 1, \dots, d$, then $\det C_n(\omega) \neq 0$ iff the dual vectors $(v_1^k, \dots, v_n^k)$, $k = 1, \dots, d$, are orthogonal and non-zero, which happens iff the vectors $X_1(\omega), \dots, X_n(\omega)$ span $\mathbf{R}^d$; but, by Fubini, this holds a.s. iff $\Pr\{u^T X = 0\} = 0$ for all $u \in S^{d-1}$, and if this is the case, then also $X_1, \dots, X_d$ span $\mathbf{R}^d$ a.s.]

Note that the conditions 'X is full' and 'the law of X assigns mass zero to every hyperplane through the origin' do not imply each other. In this section we require the second condition but fullness of X is not assumed.

Along the steps in Giné, Götze and Mason (1997), we begin with the following lemma:

Lemma 4.1. *Let X be a not necessarily symmetric random vector in $\mathbf{R}^d$ such that $\Pr\{u^T X = 0\} = 0$, $u \in S^{d-1}$. Assume that $i_\lambda \neq j_\lambda$, $\lambda = 1, \dots, r$, and that exactly m indices occur among $i_1, \dots, i_r, j_1, \dots, j_r$, each with multiplicity r_ℓ, $\sum_{\ell=1}^m r_\ell = 2r$. Let $s = \mathrm{Card}\{\ell \le m : r_\ell = 1\}$. Then, if $n \ge md$,*

$$\left|E\left[\left(X_{i_1}^T C_n^{-1} X_{j_1}\right) \cdots \left(X_{i_r}^T C_n^{-1} X_{j_r}\right)\right]\right| \le d^{(2r-s)/2} \left[\frac{n}{m}\right]^{-m} \left(E\left\|T_{[n/m]}\right\|\right)^s.$$

In particular, for $n \ge md$ and setting

$$M_n := \max_{d\le \ell\le n} \left[1, E\|T_\ell\|\right],$$

we have

$$\left|E\left[\left(X_{i_1}^T C_n^{-1} X_{j_1}\right) \cdots \left(X_{i_r}^T C_n^{-1} X_{j_r}\right)\right]\right| \le \left(\frac{d+1}{d}\right)^m d^r \left(\frac{m}{n}\right)^m M_n^m.$$

Proof. For ease of notation, we consider a simple case that nevertheless shows how the proof works in general. Let

$$U = E\left[\left(X_1^T C_n^{-1} X_2\right)\left(X_2^T C_n^{-1} X_3\right)\left(X_4^T C_n^{-1} X_2\right)^3\right],$$

where the index 2 has multiplicity 5, 4 has multiplicity 3 and the other two, multiplicity 1 each, and $r = 5$, $m = 4$, $s = 2$. We set $S_{[n/4]} = S_{[n/4],1} = \sum_{i=1}^{[n/4]} X_i$, $S_{[n/4],2} = \sum_{i=[n/4]+1}^{2[n/4]} X_i$, $C_{[n/4]} = C_{[n/4],1} = \sum_{i=1}^{[n/4]} X_i X_i^T$ and $C_{[n/4],2} = \sum_{i=[n/4]+1}^{2[n/4]} X_i X_i^T$. Then, using Cauchy-Schwartz's inequality (to bound five scalar products of the type $X_\ell^T C_n^{-1} X_k$ in terns of $X_k^T C_n^{-1} X_k$), the fact that $C_n^{-1} \leq C_{[n/4],j}^{-1}$ a.s., Lemma 3.1 and independence and equidistribution of the vectors X_i, we obtain

$$\begin{aligned}
\left[\frac{n}{m}\right]^m |U| &= \left| E\left[\sum_{k=2[n/4]+1}^{3[n/4]} \sum_{\ell=3[n/4]+1}^{4[n/4]} \left(S_{[n/4],1}^T C_n^{-1} X_k\right)\left(X_k^T C_n^{-1} S_{[n/4],2}\right)\left(X_\ell^T C_n^{-1} X_k\right)^3\right]\right| \\
&\leq E\Bigg[\left(S_{[n/4],1}^T C_n^{-1} S_{[n/4],1}\right)^{1/2}\left(S_{[n/4],2}^T C_n^{-1} S_{[n/4],2}\right)^{1/2} \\
&\quad \times \sum_{k=2[n/4]+1}^{3[n/4]} \left(X_k^T C_n^{-1} X_k\right)^{5/2} \sum_{\ell=3[n/4]+1}^{4[n/4]} \left(X_\ell^T C_n^{-1} X_\ell\right)^{3/2}\Bigg] \\
&\leq d^{(5+3)/2} E\left[\left(S_{[n/4],1}^T C_{[n/4],1}^{-1} S_{[n/4],1}\right)^{1/2}\left(S_{[n/4],2}^T C_{[n/4],2}^{-1} S_{[n/4],2}\right)^{1/2}\right] \\
&\leq d^{(2r-s)/2}\left(E\left(S_{[n/m]}^T C_{[n/m]}^{-1} S_{[n/m]}\right)^{1/2}\right)^s \\
&= d^{(2r-s)/2}(E\|T_{[n/m]}\|)^s,
\end{aligned}$$

proving the lemma in this case. Obviously, this proof generalizes. q.e.d.

Corollary 4.2. *With X as in Lemma 4.1, for $k \in \mathbf{N}$ and $n \geq 2kd$, we have*

$$E\left\|C_n^{-1/2} S_n\right\|^{2k} \leq 4ed^k \sum_{r=0}^{k} \binom{k}{r} (2r)!\,(1+4e)^{2r} M_n^{2r} \leq 4e(2k)!\left(1+(1+4e)^2\right)^k d^k M_n^{2k}. \quad (4.1)$$

In particular, if $\sup_n E\left\|C_n^{-1/2} S_n\right\| = M < \infty$, *then*

$$\sup_n E\left\|C_n^{-1/2} S_n\right\|^m < \infty$$

for all $0 < m < \infty$.

Proof. First we note that, by Lemma 3.1, for $k \in \mathbf{N}$,

$$\begin{aligned} E\|T_n\|^{2k} &= E\left[S_n^T C_n^{-1} S_n\right]^k \\ &= E\left[\sum_i X_i^T C_n^{-1} X_i + \sum_{i\neq j} X_i^T C_n^{-1} X_j\right]^k \\ &= E\left[d + \sum_{i\neq j} X_i^T C_n^{-1} X_j\right]^k \\ &= \sum_{r=0}^k \binom{k}{r} d^{k-r} E\left(\sum_{i\neq j} X_i^T C_n^{-1} X_j\right)^r < \infty. \end{aligned}$$

Next, Lemma 4.1 and elementary combinatorics give that, for all $r \in \mathbf{N}$ and $n \geq 2rd$,

$$\begin{aligned} \left| E\left(\sum_{1\leq i\neq j\leq n} X_i^T C_n^{-1} X_j\right)^r \right| &\leq M_n^{2r} d^r \sum_{m=2}^{2r} \binom{2n}{m} \sum_{\substack{r_1+\cdots+r_m=2r \\ r_i\geq 1}} \binom{2r}{r_1,\ldots,r_m} \left(\frac{m}{n}\right)^m 2^m \\ &\leq M_n^{2r} d^r \sum_{m=2}^{2r} \binom{2n}{m}\binom{2r-1}{m-1}(2r)!\left(\frac{m}{n}\right)^m 2^m \\ &\leq M_n^{2r} d^r (2r)! \sum_{m=2}^{2r} \binom{2r-1}{m-1}(4e)^m \\ &\leq 4eM_n^{2r} d^r (2r)!\,(1+4e)^{2r-1}. \end{aligned}$$

Inequality (4.1) follows by combining these two estimates. The boundedness of moments is a consequence of (4.1) and the fact that, by Lemma 3.1, setting $a_i = \left\|C_n^{-1/2} X_i\right\|^2 = X_i^T C_n^{-1} X_i$, we have

$$\begin{aligned} \left\|C_n^{-1/2} S_n\right\|^2 &= \sum_{i,j} X_i^T C_n^{-1} X_j = \sum_{i,j} \langle C_n^{-1/2} X_i, C_n^{-1/2} X_j\rangle \\ &\leq \sum_{i,j} a_i^{1/2} a_j^{1/2} = \left(\sum_i a_i^{1/2}\right)^2 \leq n\sum_i a_i = nd \end{aligned} \tag{4.2}$$

(which controls $\max_{n\leq(m+1)d} \left\|C_n^{-1/2} S_n\right\|^m$). q.e.d.

This yields the main result of this section, namely that tightness of the sequence $\{T_n\}$ implies uniform integrability of $\|T_n\|^m$ for every m:

Theorem 4.3. *Assume* $\Pr\{u^T X = 0\} = 0$ *for all* $u \in S^{d-1}$. *If the sequence* $C_n^{-1/2} S_n$, $n \in \mathbf{N}$, *is stochastically bounded then*

$$\sup_n E\left\|C_n^{-1/2} S_n\right\| < \infty$$

and therefore

$$\sup_n E\left\|C_n^{-1/2} S_n\right\|^m < \infty$$

for $0 < m < \infty$.

Proof. Inequalities (4.1) and (4.2) allow to carry out the same Paley-Zygmund type argument as in Lemma 2.4 in Giné, Götze and Mason (1987), to conclude that if $\{\|T_n\|\}$ is stochastically bounded then $\sup_n E\|T_n\| < \infty$. Now the theorem follows from the second part of Corollary 4.2. q.e.d.

In particular this theorem implies that if T_n is asymptotically standard normal, then, as in the symmetric case, we have convergence of all its moments to the corresponding moments of the standard normal variable in $\mathbf{R}^d$. In particular, $E\|T_n\|^2 \to d$ and $E\|T_n\|^4 \to d^2 + 2d$. However, in the absence of symmetry $E\|T_n\|^4$ has an expression that contains terms other than $nE(X_1^T C_n^{-1} X_1)^2$, and they must be shown to tend to zero in order to conclude, as in the proof of Theorem 3.4, that $nE(X_1^T C_n^{-1} X_1)^2 \to 0$. Concretely, $E\|T_n\|^2 \to d$ and Lemma 3.1 imply $E(X_1^T C_n^{-1} X_2) = o(n^{-2})$, and using this and Lemma 3.2, $E\|T_n\|^4$ can be written as

$$\begin{aligned}
E\|T_n\|^4 &= E\left[d + \sum_{i\neq j} X_i^T C_n^{-1} X_j\right]^2 \\
&= d^2 + 2d + o(1) - 2nE\left(X_1^T C_n^{-1} X_1\right)^2 \\
&\qquad + n(n-1)(n-2)E\left[\left(X_1^T C_n^{-1} X_2\right)\left(X_1^T C_n^{-1} X_3\right) + \left(X_1^T C_n^{-1} X_2\right)\left(X_3^T C_n^{-1} X_2\right)\right. \\
&\qquad\qquad \left. + \left(X_1^T C_n^{-1} X_2\right)\left(X_2^T C_n^{-1} X_3\right) + \left(X_1^T C_n^{-1} X_2\right)\left(X_3^T C_n^{-1} X_1\right)\right] \\
&\qquad + n(n-1)(n-2)(n-3)E\left(X_1^T C_n^{-1} X_2\right)\left(X_3^T C_n^{-1} X_4\right),
\end{aligned}$$

so that one should prove that the last two expected values tend to zero. This was achieved in $\mathbf{R}$ by means of an integral representation for $(\sum_{i=1}^n X_i^2)^{-r}$ and a delicate argument involving the measure $[E\exp(\lambda X^2)]^r d\lambda/\lambda$ which do not directly generalize to the case in hand.

References

de la Peña, V. and Giné, E. (1999). *Decoupling: from Dependence to Independence.* Springer, New York.

Giné, E; Götze, F.; Mason, D. (1997). When is the Student t-statistic asymptotically standard normal? *Ann. Probab.* **25** 1514-1531.

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

Hahn, M.G. and Klass, M.J. (1980). Matrix normalization of sums of random vectors in the domain of attraction of the multivariate normal. *Ann. Probab.* **8** 262-280.

Horn, R.A. and Johnson, C.H. (1985). *Matrix Analysis.* Cambridge U. Press, Cambridge, England.

Maller, R. A. (1993) Quadratic negligibility and the asymptotic normality of operator normed sums. *J. Multivariate Analysis* **44**, 191-219.

Maller, R.A.; Klass, M.J.: Vu, H.T.V. (1996). On the Studentization of random vectors. *J. Multivariate Analysis* **57** 142-155.

Sepanski, S.J. (1994). Probabilistic characterizations of the generalized domain of attraction of the multivariate normal law. *J. Theoret. Probab.* **7** 857-866.

Sepanski, S.J. (1996). Asymptotics for multivariate t-statistic for random vectors in the generalized domain of atraction of the multivariate normal law. *Statist. and Probab. Letters* **30** 179-188.