Asymptotic normality of quadratic forms with random vectors of increasing dimension

Hanxiang Peng and Anton Schick

Abstract. This paper provides sufficient conditions for the asymptotic normality of quadratic forms of averages of random vectors of increasing dimension and improves on conditions found in the literature. Such results are needed in applications of Owen's empirical likelihood when the number of constraints is allowed to grow with the sample size. In this connection we fix a gap in the proof of Theorem 4.1 of Hjort, McKeague and Van Keilegom (2009). We also demonstrate how our results can be used to obtain the asymptotic distribution of the empirical likelihood with an increasing number of constraints under contiguous alternatives. In addition, we discuss potential applications of our result. One example treats testing for the equality of the marginal distributions of a bivariate random vector. Another example treats a test for diagonality of a covariance matrix of a normal random vector with increasing dimension.

Key words: Martingale central limit theorem; Lindeberg condition; empirical likelihood; contiguous alternatives; chi-square of fit test with increasing number of cells; testing for equal marginals; testing for diagonality of a covariance matrix.

1. Introduction

Let r_n be positive integers that tend to infinity with n. Let $\xi_{n,1}, \dots, \xi_{n,n}$ be independent and identically distributed r_n-dimensional random vectors with mean $E[\xi_{n,1}] = 0$ and dispersion matrix $V_n = E[\xi_{n,1}\xi_{n,1}^\top]$. We assume throughout that the largest eigenvalue of V_n is bounded,

$$\text{(C1)} \qquad \rho_n = \sup_{|u|=1} u^\top V_n u = O(1),$$

and that the euclidean norm of V_n tends to infinity,

$$\text{(C2)} \qquad \operatorname{trace}(V_n^2) \to \infty.$$

Let $|x|$ denote the euclidean norm of a vector x. We are interested in the asymptotic behavior of $|\tilde{\xi}_n + \mu_n|^2$ with μ_n an r_n-dimensional vector satisfying

$$\text{(C3)} \qquad \frac{\mu_n^\top V_n \mu_n}{\operatorname{trace}(V_n^2)} \to 0$$

The research of Hanxiang Peng was partially supported by NSF Grant DMS 0940365.
The research of Anton Schick was partially supported by NSF Grant DMS 0906551.

and $\tilde{\xi}_n$ the r_n-dimensional random vector defined by

$$\tilde{\xi}_n = n^{-1/2} \sum_{j=1}^{n} \xi_{n,j}.$$

More precisely, we are looking for conditions that imply the asymptotic normality

$$\frac{|\tilde{\xi}_n + \mu_n|^2 - |\mu_n|^2 - \operatorname{trace}(V_n)}{\sqrt{2\operatorname{trace}(V_n^2)}} \implies N(0,1). \tag{1.1}$$

Of special interest is the case, when μ_n is the zero vector and V_n is idempotent with rank q_n tending to infinity. Then (1.1) simplifies to

$$\frac{|\tilde{\xi}_n|^2 - q_n}{\sqrt{2q_n}} \implies N(0,1). \tag{1.2}$$

In particular, if μ_n is the zero vector and V_n equals I_{r_n}, the $r_n \times r_n$ identity matrix, then (1.2) becomes

$$\frac{|\tilde{\xi}_n|^2 - r_n}{\sqrt{2r_n}} \implies N(0,1). \tag{1.3}$$

Such results are needed to obtain the asymptotic behavior of the likelihood ratio statistic in exponential families of increasing dimensions and to study the behavior of Owen's empirical likelihood when the data dimension is allowed to increase with the sample size. The former was done by Portnoy (1988) who proved (1.3) under the assumption that the sixth moments of the coordinates of $\xi_{n,1}$ are uniformly bounded. The latter has been recently studied by Hjort, McKeague and Van Keilegom (2009) who rely on Portnoy's result and by Chen, Peng and Chin (2009) who rely on results and structural assumptions of Bai and Saranadasa (1996). We are interested in verifying (1.1) under weaker moment assumptions than used by these authors. This allows us to fix a gap in the proof of Theorem 4.1 in Hjort, McKeague and Van Keilegom (2009), see Remark 4 below. Our results are used in Peng and Schick (2013) where the authors generalize the results in Hjort, McKeague and Van Keilegom (2009) to allow for infinitely many constraints that depend on nuisance parameters.

We achieve our goal by proving two central limit theorems. The first one uses the following growth conditions.

$$\operatorname{Var}[|\xi_{n,1}|^2] = o(n \operatorname{trace}(V_n^2)), \tag{1.4}$$

$$\operatorname{Var}[|V_n^{1/2}\xi_{n,1}|^2] = o(n \operatorname{trace}^2(V_n^2)), \tag{1.5}$$

$$E[(\xi_{n,1}^\top \xi_{n,2})^4] = o(n^2 \operatorname{trace}^2(V_n^2)). \tag{1.6}$$

Theorem 1. *Suppose (C1)–(C3) hold. Then (1.4)-(1.6) imply (1.1).*

In the presence of (C1) and (C2), the growth conditions (1.4)–(1.6) are implied by

$$E[|\xi_{n,1}|^4] = o(n \operatorname{trace}(V_n^2)). \tag{1.7}$$

Thus we have the following corollary.

Corollary 1. *Suppose (C1)–(C3) hold. Then (1.7) implies (1.1).*

The next theorem uses

$$\frac{\operatorname{trace}(V_n)}{\operatorname{trace}(V_n^2)} = O(1) \tag{C4}$$

and the Lindeberg condition,

$$L_n(\epsilon) = E[|\xi_{n,1}|^2 \mathbf{1}[|\xi_{n,1}| > \epsilon\sqrt{n}]] \to 0, \quad \epsilon > 0, \tag{L}$$

to obtain the desired result.

THEOREM 2. *Suppose (C1)–(C4) hold. Then (L) implies (1.1).*

The proofs of the theorems are given in Section 9. Section 8 gives technical details needed in the proofs. A first example with simulations is given in Section 2. There a chi-square goodness of fit test is discussed and shown to valid even if the number of cells increases almost as fast as the sample size. In Section 3 we discuss the results in more detail and compare our results with those in the literature. Section 4 illustrates how our results can be used to give the asymptotic behavior under contiguous alternatives. In Section 5 we discuss potential applications of our results. We illustrate such an application in Section 6 by presenting a test for the equality of the marginal distributions of a bivariate random vector. As another application a test for the diagonality of a covariance matrix of a normal random vector with increasing dimension is presented in Section 7.

2. A first example with simulations

Let $X_1, \dots, X_n$ be independent random variables with common distribution function F. To test the null hypothesis that F equals a specified continuous distribution function F_0, we can use the test statistic

$$T_{n,r} = \sum_{i=1}^{r} \frac{(N_i - n/r)^2}{n/r}$$

with

$$N_i = \sum_{j=1}^{n} \mathbf{1}\Big[\frac{i-1}{r} < F_0(X_j) \le \frac{i}{r}\Big], \quad i = 1, \dots, r,$$

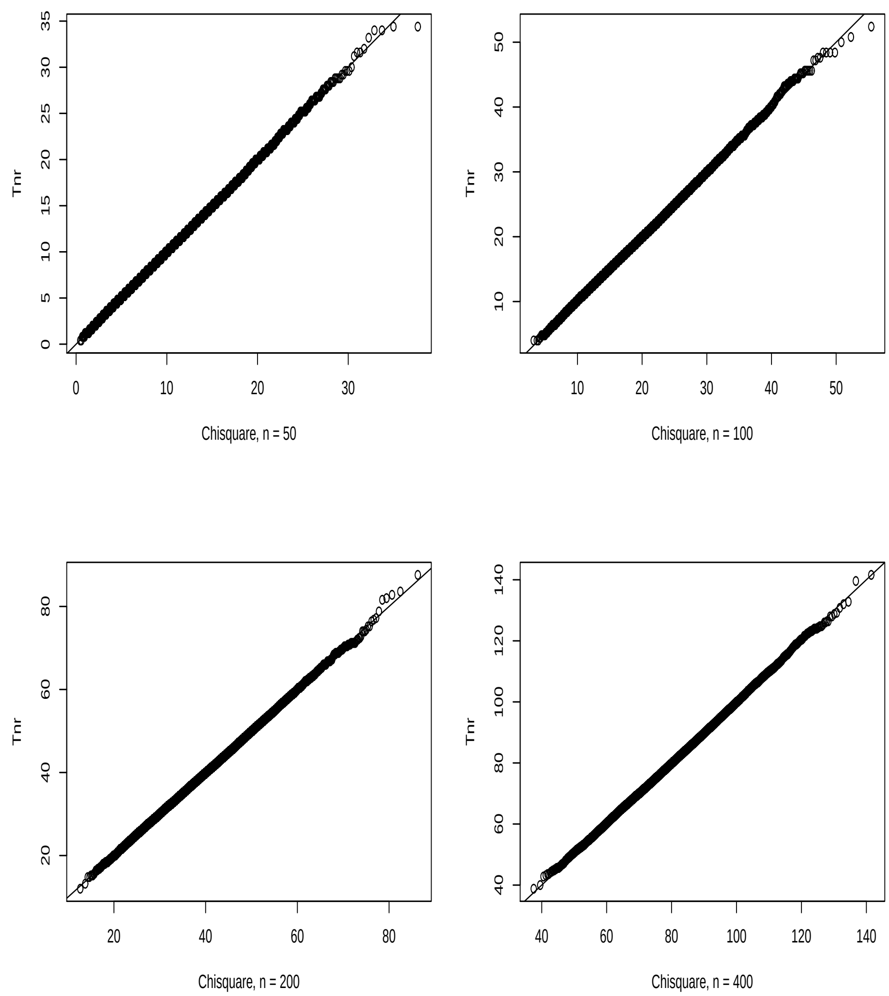
and reject the null hypothesis if $T_{n,r}$ exceeds $\chi_{1-\alpha}(r-1)$, the $(1-\alpha)$-quantile of the chi-square distribution with $r-1$ degrees of freedom. This test has asymptotic size α. This follows from the fact that under the null hypothesis the test statistic is asymptotically chi-square with r degrees of freedom. Keep in mind that under the null hypothesis the random variables $F_0(X_1), \dots, F_0(X_n)$ are uniformly distributed on $(0,1)$ and that $P((i-1)/r < U \le i/r) = 1/r$ holds for a uniform random variable U and $i = 1, \dots, r$.

Can we let r grow with n and still maintain the asymptotic size of this test? The answer is yes. More precisely, we have the following result. The test

$$\delta_{n,r_n} = \mathbf{1}[T_{n,r_n} > \chi_{1-\alpha}(r_n - 1)]$$

has asymptotic size α as long as r_n tends to infinity at a rate slower than n, i.e., $r_n = o(n)$. The proof of this claim is based on the observation that a chi-square random variable with m degrees of freedom is approximately normal with mean m and variance $2m$ for large m. This result is a consequence of the central limit theorem and the fact that a chi-square random variable with m degrees of freedom has the same distribution as a sum of m independent chi-square random

FIGURE 1. Q-Q plots of the simulated values of $T_{n,n/5}$ against the chi-square distribution with $(n/5-1)$ degrees of freedom for $n = 50, 100, 200, 400$.



variables with one degree of freedom. Thus our claim can be verified by showing the asymptotic normality result

$$\frac{T_{n,r_n}-(r_n-1)}{\sqrt{2(r_n-1)}} \implies N(0,1). \tag{2.1}$$

We note that T_{n,r_n} equals $n|\tilde{\xi}_n|^2$ if we take ξ_{nj} to be the r_n-dimensional random vector whose i-th coordinate is $\sqrt{r_n}\mathbf{1}[i-1 < r_nF_0(X_j) \le i] - 1/\sqrt{r_n}$. These random vectors are independent and identically distributed with mean vector 0 and dispersion matrix $V_n = I_{r_n} - 1/r_n J_{r_n}$, with J_r the $r \times r$ matrix with all its entries equal to one, and satisfy $|\xi_{nj}|^2 = r_n - 1$ almost surely. The matrix V_n is idempotent with trace $r_n - 1$ and $E[|\xi_{n1}|^4]$ equals $(r_n - 1)^2$. Consequently, the assumptions of Corollary 1 are met with $\rho_n = 1$ and $\mu_n = 0$ if $r_n \to \infty$ and $r_n = o(n)$ hold, and this corollary gives the desired (2.1).

TABLE 1. Simulated sizes of the test $\delta_{n,n/5}$ for selected values of n and α

$\alpha \backslash n$	40	60	80	100	120	140	160	180	200	400
.10	.0910	.0941	.0985	.0926	.0936	.1040	.0998	.0949	.0956	.1028
.05	.0449	.0464	.0460	.0476	.0503	.0498	.0501	.0495	.0466	.0525
.01	.0090	.0109	.0098	.0102	.0096	.0100	.0108	.0096	.0098	.0114

We have run some simulations to assess this result. In the simulations F_0 was taken to be the uniform distribution and $r_n = n/5$. We generated 25,000 independent copies of $T_{n,n/5}$ for several choices of n. Figure 1 gives quantile-quantile plots. These show that the chi-square approximation is quite good. Table 2 reports the simulated size of the test for three choices of α.

3. Discussion of the results

We begin by addressing sufficient conditions for the Lindeberg condition.

REMARK 1. In view of the inequality

$$L_n(\epsilon) \le E[|\xi_{n,1}|^2 \log(1 + |\xi_{n,1}|)]/\log(1 + \epsilon\sqrt{n})$$

the Lindeberg condition (L) is implied by

$$E[|\xi_{n,1}|^2 \log(1 + |\xi_{n,1}|)] = o(\log n). \tag{L1}$$

In view of the inequality

$$L_n(\epsilon) \le E[|\xi_{n,1}|^{2+\delta}](\epsilon\sqrt{n})^{-\delta}, \quad \delta > 0,$$

the Lindeberg condition (L) holds whenever

$$E[|\xi_{n,1}|^{2\alpha}] = o(n^{\alpha-1}), \quad \text{for some } \alpha > 1. \tag{L2}$$

In particular, if $E[|\xi_{n,1}|^{2\alpha}] = O(r_n^\alpha)$ holds for some $\alpha > 1$, then (L2) is implied by $r_n = o(n^{1-1/\alpha})$.

Let us now specialize our results to the case when μ_n is the zero vector and V_n is an idempotent matrix with rank q_n tending to infinity. In this case (C1) – (C4) hold and (1.1) simplifies to (1.2).

COROLLARY 2. *Suppose V_n is idempotent with rank q_n tending to infinity. Then the following are true.*

(a) *The growth conditions* $\mathrm{Var}[|\xi_{n,1}|^2] = o(nq_n)$, $\mathrm{Var}[|V_n^{1/2}\xi_{n,1}|^2] = o(nq_n^2)$ *and* $E[(\xi_{n,1}^\top \xi_{n,2})^4] = o(n^2 q_n^2)$ *imply (1.2).*
(b) *The moment condition* $E[|\xi_{n,1}|^4] = o(nq_n)$ *implies (1.2).*
(c) *The Lindeberg condition (L) implies (1.2).*
(d) *If* $E[|\xi_{n,1}|^{2\alpha}] = O(r_n^\alpha)$ *and* $r_n = o(n^{1-1/\alpha})$ *hold for some* $\alpha > 1$, *then (1.2) holds.*

For its importance we formulate the special case $V_n = I_{r_n}$.

COROLLARY 3. *Suppose V_n equals I_{r_n}. Then the following are true.*

(a) $\mathrm{Var}[|\xi_{n,1}|^2] = o(nr_n)$ *and* $E[(\xi_{n,1}^\top \xi_{n,2})^4] = o(n^2 r_n^2)$ *imply (1.3).*
(b) *The moment condition* $E[|\xi_{n,1}|^4] = o(nr_n)$ *implies (1.3). In particular,* $E[|\xi_{n,1}|^4] = O(r_n^2)$ *and* $r_n = o(n)$ *imply (1.3).*
(c) *The Lindeberg condition (L) implies (1.3).*
(d) *If* $E[|\xi_{n,1}|^{2\alpha}] = O(r_n^\alpha)$ *and* $r_n = o(n^{1-1/\alpha})$ *hold for some* $\alpha > 1$*, then (1.3) holds.*

Note that in the case $E[|\xi_{n,1}|^4] = r_n^2$ part (b) allows for larger r_n than part (d). More precisely, part (b) requires $r_n = o(n)$, while part (d) requires $r_n = o(n^{1/2})$.

REMARK 2. Portnoy (1988, Theorem 4.1) obtains the conclusion (1.3) in the case $V_n = I_{r_n}$ under the growth condition $r_n/n \to 0$ and the assumption that the coordinates $\xi_{n,1,i}$ of $\xi_{n,1}$ have a uniformly bounded sixth moment,

$$\max_{1\le i\le r_n} E[\xi_{n,1,i}^6] = O(1).$$

His last condition implies

$$\max_{1\le i\le r_n} E[\xi_{n,1,i}^4] = O(1), \tag{3.1}$$

and the latter implies

$$E[|\xi_{n,1}|^4] = E[(\sum_{i=1}^{r_n} \xi_{n,1,i}^2)^2] \le r_n \sum_{i=1}^{r_n} E[\xi_{n,1,i}^4] = O(r_n^2).$$

Thus his result is a special case of part (b) of Corollary 3.

REMARK 3. Assume that $\xi_{n,1} = V_n Z_n$ for some symmetric idempotent matrix V_n with rank q_n tending to infinity and some random vector Z_n satisfying $E[Z_n] = 0$, $E[Z_n Z_n^\top] = I_{r_n}$,

$$\zeta_n = \max_{1\le i\le r_n} E[|Z_{n,i}^4] = o(n), \tag{3.2}$$

and

$$E[Z_{n,i}^{\alpha_1} Z_{n,j}^{\alpha_2} Z_{n,k}^{\alpha_3} Z_{n,l}^{\alpha_4}] = E[Z_{n,i}^{\alpha_1}]E[Z_{n,j}^{\alpha_2}]E[Z_{n,k}^{\alpha_3}]E[Z_{n,l}^{\alpha_4}] \tag{3.3}$$

for distinct indices i, j, k, l and non-negative integers $\alpha_1, \ldots, \alpha_4$ that sum to 4. The above conditions generalize those in Chen, Peng and Qin (2009) with our V_n equal to their $\Gamma_n^\top(\Gamma_n\Gamma_n^\top)^{-1}\Gamma_n$. These authors require instead of (3.2) the stronger $E[Z_{n,1}^4] = \cdots = E[Z_{n,r_n}^4] = \beta$ for some β. Relying on results of Bai and Saranadasa (1996), they obtain (1.2) under the condition that $q_n = O(n)$. We shall show

$$\mathrm{Var}(|V_n^{1/2}\xi_{n,1}|^2) = \mathrm{Var}(|\xi_{n,1}|^2) \le (2+\zeta_n)q_n = o(nq_n) \tag{3.4}$$

and

$$E[(\xi_{n,1}^\top \xi_{n,2})^4] \le 3(q_n^2 + 2q_n + \zeta_n q_n) + \zeta_n(3+\zeta_n)q_n = o(n^2 q_n^2). \tag{3.5}$$

Thus we obtain (1.2) from part (a) of Corollary 2 without their restrictions.

Note that the right-hand side in (3.3) equals zero if at least one of $\alpha_1, \dots, \alpha_4$ equals one and that (3.3) yields $E[Z_{n,i}^2 Z_{n,j}^2] = E[Z_{n,i}^2]E[Z_{n,j}^2] = 1$ for $i \neq j$. Thus we calculate

$$
\begin{aligned}
E[|\xi_{n,1}|^4] = E[(Z_n^\top V_n Z_n)^2] &= \sum_{i=1}^{r_n}\sum_{j=1}^{r_n}\sum_{k=1}^{r_n}\sum_{l=1}^{r_n} E[Z_{n,i}V_{n,i,j}Z_{n,j}Z_{n,k}V_{n,k,l}Z_{n,l}] \\
&= \sum_{i\neq k} V_{n,i,i}V_{n,k,k} + \sum_{i\neq j} 2V_{n,i,j}V_{n,i,j} + \sum_i E[Z_{n,i}^4]V_{n,i,i}^2 \\
&= (\operatorname{trace}(V_n))^2 + 2\operatorname{trace}(V_n) + \sum_{i=1}^{r_n}(E[Z_{n,i}^4]-3)V_{n,i,i}^2,
\end{aligned}
$$

$$
\begin{aligned}
E[(Z_n^\top \xi_{n,2})^4] &= \sum_{i=1}^{r_n}\sum_{j=1}^{r_n}\sum_{k=1}^{r_n}\sum_{l=1}^{r_n} E[Z_{n,i}\xi_{n,2,i}Z_{n,j}\xi_{n,2,j}Z_{n,k}\xi_{n,2,k}Z_{n,l}\xi_{n,2,l}] \\
&\leq 3E[|\xi_{n,2}|^4] + \zeta_n \sum_{i=1}^{r_n} E[\xi_{n,2,i}^4],
\end{aligned}
$$

$$
\begin{aligned}
\sum_{\nu=1}^{r_n} E[\xi_{n,1,\nu}^4] &= \sum_{\nu=1}^{r_n}\sum_{i=1}^{r_n}\sum_{j=1}^{r_n}\sum_{k=1}^{r_n}\sum_{l=1}^{r_n} E[V_{n,\nu,i}Z_{n,i}V_{n,\nu,j}Z_{n,j}V_{n,\nu,k}Z_{n,k}V_{n,\nu,l}Z_{n,l}] \\
&\leq \sum_{\nu=1}^{r_n}\Big[3\Big(\sum_{i=1}^{r_n} V_{n,\nu,i}^2\Big)^2 + \zeta_n \sum_{i=1}^{r_n} V_{n,\nu,i}^4\Big] \leq 3\operatorname{trace}(V_n) + \zeta_n \operatorname{trace}(V_n).
\end{aligned}
$$

Here we used the identity

$$\sum_{j=1}^{r_n} V_{n,i,j}^2 = \sum_{j=1}^{r_n} V_{n,i,j}V_{n,j,i} = V_{n,i,i}$$

and the inequalities

$$0 \leq V_{n,i,i} \leq 1 \quad \text{and} \quad V_{n,i,j}^2 \leq 1, \quad i,j = 1,\dots,r_n.$$

Using the identities $|\xi_{n,1}|^2 = |V_n^{1/2}\xi_{n,1}|^2$ and $\xi_{n,1}^\top \xi_{n,2} = Z_n^\top \xi_{n,2}$ we obtain (3.4) and (3.5). Note also that $E[|\xi_{n,1}|^2]$ equals $\operatorname{trace}(V_n)$.

REMARK 4. Our results are motivated by recent results on extending Owen's (1988, 1990, 2001) empirical likelihood approach to allow for an increasing number of constraints, see Hjort, McKeague and Van Keilegom (2009) and Chen, Peng and Qin (2009). The empirical likelihood for this case is given by

$$\mathscr{R}_n = \sup\{\prod_{j=1}^{n} n\pi_j : 0 \leq \pi_j, \sum_{j=1}^{n}\pi_j = 1, \sum_{j=1}^{n}\pi_j X_{n,j} = 0\}$$

where $X_{n,1},\dots,X_{n,n}$ are independent and identically distributed r_n-dimensional random variables with mean $E[X_{n,1}] = 0$ and invertible dispersion matrix W_n. It is equivalent to

$$\mathscr{R}_n = \sup\{\prod_{j=1}^{n} n\pi_j : 0 \leq \pi_j, \sum_{j=1}^{n}\pi_j = 1, \sum_{j=1}^{n}\pi_j \xi_{n,j} = 0\}$$

with $\xi_{n,j} = W_n^{-1/2} X_{n,j}$. The goal is to show that $-2\log\mathscr{R}_n$ is approximately a chi-square random variable with r_n degrees of freedom. This is done by showing the asymptotic normality result

$$\frac{-2\log\mathscr{R}_n - r_n}{\sqrt{2r_n}} \implies N(0,1). \tag{3.6}$$

This result is typically achieved in two steps. The first step establishes the approximation

$$-2\log\mathscr{R}_n - |\tilde{\xi}_n|^2 = o_p(r_n^{1/2}), \tag{3.7}$$

and the second step obtains the asymptotic normality result (1.3).

In their Theorem 4.1, Hjort, McKeague and Van Keilegom (2009) claim (3.6) under the assumptions that the q-th moments of the coordinates of $X_{n,1}$ are uniformly bounded for some $q > 2$, that the eigenvalues of W_n are bounded and bounded away from zero, and that the dimension r_n satisfies

$$r_n^{3+6/(q-2)} = r_n^{3q/(q-2)} = o(n). \tag{3.8}$$

Their proof, however, is valid for the case $q \ge 6$ only, as they rely on Portnoy's (1988) asymptotic normality result mentioned in Remark 2 above. With C a bound on the largest eigenvalue of $W_n^{-1/2}$ and B a bound on the q-th moments of the coordinates of $X_{n,1}$, their assumptions imply

$$\begin{aligned} E[|\xi_{n,1}|^q] \le C^q E[|X_{n,1}|^q] &= C^q r_n^{q/2} E\Big[\Big(\frac{1}{r_n}\sum_{j=1}^{n} X_{n,1,j}^2\Big)^{q/2}\Big] \\ &\le C^q r_n^{q/2-1} \sum_{j=1}^{r_n} E[|X_{n,1,j}|^q] = C^q B r_n^{q/2}. \end{aligned}$$

Thus the required asymptotic normality follows from part (d) of Corollary 3 with $\alpha = q/2$. Note that their requirement (3.8) on r_n implies

$$r_n = o(n^{(q-2)/(3q)}) = o(n^{2(\alpha-1)/6\alpha}) = o(n^{1-1/\alpha})$$

as needed. This closes the gap in Theorem 4.1 of Hjort, McKeague and Van Keilegom (2009).

Remark 5. Suppose $E[|\xi_{n,1}|^{2\alpha}] = O(r_n^{\alpha})$ holds for some $\alpha \ge 2$. Then the moment inequality yields $E[|\xi_{n,1}|^4] \le E[|\xi_{n,1}|^{2\alpha}]^{2/\alpha} = O(r_n^2)$. In this case, part (b) of Corollary 3 allows for larger r_n than part (d).

4. Asymptotic behavior under local alternatives

Let $(\mathscr{X}, \mathscr{S}, Q)$ be a probability space and w_n be a function from $\mathscr{X}$ into $\mathbb{R}^{r_n}$ satisfying $\int w_n\,dQ = 0$, $\int |w_n|^2\,dQ < \infty$ and

$$\Lambda_n(\epsilon) = \int |w_n|^2 \mathbf{1}[|w_n| > \epsilon\sqrt{n}]\,dQ \to 0, \quad \epsilon > 0. \tag{4.1}$$

Assume also that the matrix $W_n = \int w_n w_n^\top\,dQ$ satisfies

$$\lambda_n = \sup_{|u|=1} u^\top W_n u = \sup_{|u|=1} \int (u^\top w_n)^2\,dQ = O(1),$$

$$\operatorname{trace}(W_n^2) \to \infty \quad \text{and} \quad \operatorname{trace}(W_n) = O(\operatorname{trace}(W_n^2)).$$

It then follows from Theorem 2 that

$$\frac{|\frac{1}{\sqrt{n}}\sum_{j=1}^n w_n(X_j)|^2 - \operatorname{trace}(W_n)}{\sqrt{2\operatorname{trace}(W_n^2)}} \implies N(0,1)$$

if $X_1, \dots, X_n$ are independent $\mathscr{X}$-valued random variables with distribution Q.

The next theorem answers the question of what happens if we slightly perturb the distribution Q. Let h denote a measurable function satisfying $\int h\,dQ = 0$ and $\int h^2\,dQ < \infty$ and set

$$h_n = h\mathbf{1}[|h| < c_n n^{1/2}/s_n] - \int h\mathbf{1}[|h| < c_n n^{1/2}/s_n]\,dQ$$

with $0 < c_n < 1/2$, $1 \le s_n$, $c_n \to 0$, $s_n^2 = o(\operatorname{trace}(W_n^2))$ and $c_n n^{1/2}/s_n \to \infty$. Let $Q_{n,h}$ denote the probability measure with density $1 + n^{-1/2}s_n h_n$ with respect to Q. By construction, we have

$$\int |n^{1/2}/s_n(\sqrt{1 + n^{-1/2}s_n h_n} - 1) - h/2|^2\,dQ \to 0. \tag{4.2}$$

If $s_n = 1$, this implies that the product measures $Q_{n,h}^n$ and Q^n are mutually contiguous. Set

$$\mu_n(h) = \int w_n h\,dQ \quad \text{and} \quad \Delta_n = n^{-1/2}s_n \int |w_n|^2 h_n\,dQ.$$

THEOREM 3. *Let $X_{n,1}, \dots, X_{n,n}$ be independent $\mathscr{X}$-valued random variables with distribution $Q_{n,h}$. Then we have the asymptotic normality result*

$$\frac{|\frac{1}{\sqrt{n}}\sum_{j=1}^n w_n(X_{n,j})|^2 - s_n^2|\mu_n(h)|^2 - \operatorname{trace}(W_n) - \Delta_n}{\sqrt{2\operatorname{trace}(W_n^2)}} \implies N(0,1). \tag{4.3}$$

In the case $s_n = 1$, this simplifies to

$$\frac{|\frac{1}{\sqrt{n}}\sum_{j=1}^n w_n(X_{n,j})|^2 - |\mu_n(h)|^2 - \operatorname{trace}(W_n)}{\sqrt{2\operatorname{trace}(W_n^2)}} \implies N(0,1). \tag{4.4}$$

PROOF. Taking $\nu_n = \mu_n(h_n)$ and $\xi_{n,j} = w_n(X_{n,j}) - n^{-1/2}s_n\nu_n$, we can write

$$\Big|\frac{1}{\sqrt{n}}\sum_{j=1}^n w_n(X_{n,j})\Big|^2 = |\tilde{\xi}_n + s_n\nu_n|^2.$$

The dispersion matrix of $\xi_{n,1}$ is given by $V_n = \tilde{W}_n - n^{-1}s_n^2\nu_n\nu_n^\top$ where

$$\tilde{W}_n = \int w_n w_n^\top\,dQ_{n,h} = W_n + n^{-1/2}s_n \int w_n w_n^\top h_n\,dQ.$$

By construction, $|n^{-1/2}s_n h_n|$ is bounded by $2c_n$. Thus, for $k = 1, 2$, we have the inequality

$$(1 - 2c_n)^k \operatorname{trace}(W_n^k) \le \operatorname{trace}(\tilde{W}_n^k) \le (1 + 2c_n)^k \operatorname{trace}(W_n^k)$$

and obtain

$$\frac{\operatorname{trace}(\tilde{W}_n)}{\operatorname{trace}(W_n)} \to 1 \quad \text{and} \quad \frac{\operatorname{trace}(\tilde{W}_n^2)}{\operatorname{trace}(W_n^2)} \to 1.$$

Since $\operatorname{trace}(W_n^2) \le \lambda_n \operatorname{trace}(W_n)$, we also have $\operatorname{trace}(W_n) \to \infty$.

The requirements on the sequences c_n and s_n imply $n^{-1/2}s_n = o(c_n) = o(1)$. Using this and the above, we find

$$\sup_{|u|=1} u^\top V_n u \le \sup_{|u|=1} u^\top \tilde{W}_n u \le (1+2c_n) \sup_{|u|=1} u^\top W_n u = O(\lambda_n) = O(1) \tag{4.5}$$

$$|\nu_n|^2 = \sup_{|u|=1} (u^\top \nu_n)^2 \le \int h_n^2\, dQ \sup_{|u|=1} \int (u^\top w_n)^2\, dQ \le \lambda_n \int h^2\, dQ = O(1), \tag{4.6}$$

$$|\nu_n - \mu_n(h)|^2 \le \lambda_n \int (h_n - h)^2\, dQ \to 0, \tag{4.7}$$

$$\operatorname{trace}(V_n) = \operatorname{trace}(\tilde{W}_n) - n^{-1}s_n^2|\nu_n|^2 = \operatorname{trace}(W_n) + o(\operatorname{trace}(W_n)). \tag{4.8}$$

$$\begin{aligned} \operatorname{trace}(V_n^2) &= \operatorname{trace}(\tilde{W}_n^2) - 2n^{-1}s_n^2\nu_n^\top \tilde{W}_n \nu_n + n^{-2}s_n^4|\nu_n|^4 \\ &= \operatorname{trace}(W_n^2) + o(\operatorname{trace}(W_n^2)). \end{aligned} \tag{4.9}$$

Thus the conditions (C1)–(C4) hold with $\mu_n = s_n\nu_n$. For (C3) note that $\nu_n^\top V_n \nu_n$ is bounded by (4.5) and (4.6). Finally, using (4.1), $|n^{-1/2}s_n\nu_n| = o(1)$ and the bound $n^{-1/2}s_n|h_n| \le 1$, we derive the Lindeberg condition (L). Thus Theorem 2 yields

$$\frac{|\frac{1}{\sqrt{n}}\sum_{j=1}^n w_n(X_{n,j})|^2 - s_n^2|\nu_n|^2 - \operatorname{trace}(V_n)}{\sqrt{2\operatorname{trace}(V_n^2)}} \implies N(0,1). \tag{4.10}$$

The desired result (4.3) follows from this, (4.7), (4.9) and the fact that $\operatorname{trace}(V_n) = \operatorname{trace}(W_n) + \Delta_n + o(1)$.

In the case $s_n = 1$, we have the bound

$$\begin{aligned} \int |w_n|^2|n^{-1/2}h_n|\, dQ &\le 2c_n\Lambda_n(\epsilon) + \epsilon \int |w_n||h_n|\mathbf{1}[|w_n| \le \epsilon\sqrt{n}]\, dQ \\ &\le 2c_n\Lambda_n(\epsilon) + \epsilon\Big(\int h_n^2\, dQ \int |w_n|^2\, dQ\Big)^{1/2}, \quad \epsilon > 0. \end{aligned}$$

This bound and $\operatorname{trace}(W_n) = O(\operatorname{trace}(W_n^2))$ yield $\Delta_n = o((\operatorname{trace}(W_n^2))^{1/2})$ and hence (4.4). □

Remark 6. Let $X_{n,1}, \dots, X_{n,n}$ be independent $\mathscr{X}$-valued random variables with distribution $Q_{n,h}$ for $s_n = 1$. Consider the empirical likelihood

$$\mathscr{R}_n = \sup\{\prod_{j=1}^n n\pi_j : 0 \le \pi_j, \sum_{j=1}^n \pi_j = 1, \sum_{j=1}^n \pi_j v_n(X_{n,j}, X_{n,1}, \dots, X_{n,n}) = 0\}$$

with v_n a measurable function from $\mathscr{X}^{n+1}$ into $\mathbb{R}^{r_n}$. Suppose that

$$-2\log\mathscr{R}_n - \Big|\frac{1}{\sqrt{n}}\sum_{j=1}^n w_n(X_{n,j})\Big|^2 = o_p(\sqrt{\operatorname{trace}(W_n^2)}) \tag{4.11}$$

when $h = 0$. By contiguity, this then also holds if $h \ne 0$ and we obtain

$$\frac{-2\log\mathscr{R}_n - |\mu_n(h)|^2 - \operatorname{trace}(W_n)}{\sqrt{2\operatorname{trace}(W_n^2)}} \implies N(0,1).$$

If W_n is idempotent with rank q_n tending to infinity, this simplifies to

$$\frac{-2\log\mathscr{R}_n - |\mu_n(h)|^2 - q_n}{\sqrt{2q_n}} \implies N(0,1)$$

and may be interpreted as $-2\log\mathscr{R}_n$ being approximately a non-central chi-square random variable with q_n degrees of freedom and non-centrality parameter $|\mu_n(h)|$.

REMARK 7. In the previous remark $Q_{n,h}$ was chosen to have density $1 + n^{-1/2}h_n$. By (4.2) this implies that

$$\int |n^{1/2}(\sqrt{dQ_{n,h}} - \sqrt{dQ}) - h/2\sqrt{dQ}|^2 \to 0. \tag{4.12}$$

The results of the previous remark remain true under the more general condition (4.12).

5. Applications

In applications, the quadratic form $|\tilde{\xi}_n|^2$ will often serve as an approximation to a more complicated statistic S_n. More precisely, suppose that we have the expansion

$$S_n = |\tilde{\xi}_n|^2 + o_p(r_n^{1/2}), \tag{5.1}$$

then the asymptotic normality result (1.3) implies the same asymptotic normality result for S_n,

$$\frac{S_n - r_n}{\sqrt{2r_n}} \implies N(0,1). \tag{5.2}$$

We have already encountered this concept in Remark 4.

Of special interest is the case $\xi_{nj} = W_n^{-1/2}w_n(Z_j)$, where $Z_1,\dots,Z_n$ are k-dimensional random vectors with common distribution Q and w_n is a measurable function from $\mathbb{R}^k$ into $\mathbb{R}^{r_n}$ such that $w_n(Z_1)$ has mean $\int w_n\,dQ = 0$ and dispersion matrix $W_n = \int w_n w_n^\top\,dQ$ which satisfies

$$0 < \inf_n \inf_{|u|=1} u^\top W_n u \le \sup_n \sup_{|u|=1} u^\top W_n u < \infty. \tag{5.3}$$

Suppose also that $E[|w_n(Z_1)|^4] = O(r_n^2)$ and $r_n = o(n)$. It then follows from part (b) of Corollary 3 that (1.3) holds.

Now let $\hat{w}_n$ denote an estimator of w_n and consider the statistic

$$S_n = \hat{T}_n^\top \hat{W}_n^{-1}\hat{T}_n$$

with

$$\hat{T}_n = \frac{1}{\sqrt{n}}\sum_{j=1}^n \hat{w}_n(Z_j) \quad \text{and} \quad \hat{W}_n = \frac{1}{n}\sum_{j=1}^n \hat{w}_n(Z_j)\hat{w}_n^\top(Z_j).$$

In this setting, (5.1) follows from the statements

$$\Big|\frac{1}{\sqrt{n}}\sum_{j=1}^n(\hat{w}_n(Z_j) - w_n(Z_j))\Big|^2 = o_p(1) \tag{5.4}$$

and

$$|\hat{W}_n - W_n|_o = \sup_{|u|=1}|u^\top(\hat{W}_n - W_n)u| = o_p(r_n^{-1/2}). \tag{5.5}$$

These statements typically require additional restrictions on the rate of growth of r_n. Let

$$\bar{W}_n = \frac{1}{n}\sum_{j=1}^n w_n(Z_j)w_n^\top(Z_j).$$

Then we have $|\bar{W}_n - W_n|_o = O_p(r_n/n^{1/2})$ as $E[|\bar{W}_n - W_n|_0^2] \le E[|\bar{W}_n - W_n|^2] \le E[|w_n(Z_1)|^4]/n = O(r_n^2/n)$ and

$$|\hat{W}_n - \bar{W}_n|_o \le D_n + 2|\bar{W}_n|_o^{1/2} D_n^{1/2}$$

with

$$D_n = \frac{1}{n}\sum_{j=1}^n |\hat{w}_n(Z_j) - w_n(Z_j)|^2.$$

Hence (5.5) is implied by $r_n^3 = o(n)$ and $D_n = o_p(r_n^{-1})$. Let us summarize our findings.

Proposition 1. *Suppose w_n is given as above and $r_n^3 = o(n)$, $D_n = o_p(r_n^{-1})$ and (5.4) hold. Then we have the asymptotic normality result (5.2) with $S_n = \hat{T}_n \hat{W}_n^{-1} \hat{T}_n$.*

6. Testing for equal marginals

Let us illustrate the result of the previous section by means of an example, namely testing for the equality of the marginal distributions of a bivariate random vector. Let the observations $(X_1, Y_1), \dots, (X_n, Y_n)$ be independent copies of a bivariate random vector (X, Y). We want to test whether the marginal distributions are the same. This is of importance when X denotes pre-treatment and Y post-treatment measurement. Equality of the marginal distributions indicates that there is no treatment effect. Assume that the marginal distribution functions F (of X) and G (of Y) are continuous.

Let us set $H = (F + G)/2$. We can estimate H by the pooled empirical distribution function,

$$\mathbb{H}(x) = \frac{1}{n}\sum_{j=1}^n (\mathbf{1}[X_j \le x] + \mathbf{1}[Y_j \le x])/2, \quad x \in \mathbb{R}.$$

Assume from now on that F equals G so that the null hypothesis holds. Then we have $H = F = G$ and $E[a(X) - a(Y)] = 0$ for every $a \in L_{2,0}(H)$ where

$$L_{2,0}(H) = \{a \in L_2(H) : \int a\, dH = 0\}.$$

We also impose the condition

$$\inf_{a \in A} E[(a(X) - a(Y))^2] > 0 \tag{6.1}$$

with $A = \{a \in L_{2,0}(H) : \int a^2\, dH = 1\}$ the unit sphere in $L_{2,0}(H)$.

Let $\psi_1, \psi_2, \dots$ denote an orthonormal basis of $L_{2,0}(U)$, where U is the uniform distribution on $[0, 1]$. Since H is continuous, the functions $\psi_1 \circ H, \psi_2 \circ H, \dots$ form an orthonormal basis of $L_{2,0}(H)$. We shall work with the trigonometric basis defined by

$$\psi_k(x) = \sqrt{2}\cos(\pi k x), \quad 0 \le x \le 1, k = 1, 2, \dots,$$

because these functions are bounded and have bounded derivatives. Let $v_n = (\psi_1, \dots, \psi_{r_n})^\top$ and set

$$w_n(x, y) = v_n(H(x)) - v_n(H(y))$$

and

$$\hat{w}_n(x, y) = v_n(\mathbb{H}(x)) - v_n(\mathbb{H}(y)), \quad x, y \in \mathbb{R}.$$

It follows from (6.1) that the dispersion matrix $W_n = E[w_n(X_1, Y_1)w_n^\top(X_1, Y_1)]$ of $w_n(X_1, Y_1)$ satisfies

$$0 < \inf_{a \in A} E[(a(X) - a(Y))^2] \le u^\top W_n u \le 4, \quad |u| = 1.$$

To see this use the fact that $u^\top v_n$ belongs to A for each unit vector u. Thus (5.3) holds. Since $|w_n| \le 2|v_n| \le 2\sqrt{2r_n}$, we obtain $E[|w_n(X,Y)|^4] = O(r_n^2)$.

If H and W_n were known, we could use the test statistic $|W_n^{-1/2}T_n|^2 = T_n^\top W_n^{-1} T_n$, where

$$T_n = \frac{1}{\sqrt{n}} \sum_{j=1}^n w_n(X_j, Y_j)).$$

Since H and W_n are unknown, we work instead with $\hat{T}_n^\top \hat{W}_n^{-1} \hat{T}_n$ where

$$\hat{T}_n = \frac{1}{\sqrt{n}} \sum_{j=1}^n \hat{w}_n(X_j, Y_j) \quad \text{and} \quad \hat{W}_n = \frac{1}{n} \sum_{j=1}^n \hat{w}_n(X_j, Y_j)\hat{w}_n^\top(X_j, Y_j)$$

Using $|v_n'|^2 \le 2\pi^2 r_n^3$, we have

$$\frac{1}{n} \sum_{j=1}^n |v_n(\mathbb{H}(X_j)) - v_n(H(X_j))|^2 \le 2\pi^2 r_n^3 \sup_{t \in \mathbb{R}} |\mathbb{H}(t) - H(t)|^2 = O_p(r_n^3/n),$$

$$\frac{1}{n} \sum_{j=1}^n |v_n(\mathbb{H}(Y_j)) - v_n(H(Y_j))|^2 \le 2\pi^2 r_n^3 \sup_{t \in \mathbb{R}} |\mathbb{H}(t) - H(t)|^2 = O_p(r_n^3/n).$$

This implies $D_n = o_p(r_n^{-1})$ if $r_n^4 = o(n)$. Finally (5.4) holds as shown in Peng and Schick (2005, pages 403–404) if $r_n^3 = o(n)$. Indeed it follows from there that

$$\sum_{k=1}^{r_n} E\Big[\Big(\frac{1}{n}\sum_{j=1}^n \hat{w}_n(X_j, Y_j) - w_n(X_j, Y_j)\Big)^2\Big] \le \frac{48\pi^2 r_n^3}{(n-1)^2} + 3\Big[\frac{8\pi^2 r_n^3}{n(n-1)} + \frac{32\pi^2 r_n^3 (n-1)}{n(n-1)^2}\Big].$$

Thus we have proved the following result.

COROLLARY 4. *Suppose F equals G and (6.1) holds. Then we have the asymptotic normality result*

$$\frac{\hat{T}_n^\top \hat{W}_n^{-1} \hat{T}_n - r_n}{\sqrt{2r_n}} \implies N(0,1)$$

provided r_n tends to infinity and r_n^4/n tends to zero.

This result shows that the test which rejects the null hypothesis if $\hat{T}_n^\top \hat{W}_n^{-1} \hat{T}_n$ exceeds the $(1-\alpha)$-quantile of the chi-square distribution with r_n degrees of freedom has asymptotic size α.

We conducted a small simulation study to investigate the power of this test. We first looked at data from a bivariate normal distribution with parameters $(0, \theta, 1, \sigma^2, \rho)$ where the first two coordinates refer to the means, the third and fourth to the variances, and the fifth to the correlation coefficient. By (a) (page 12) of Peng and Schick (2004) the bivariate normal model satisfies the condition (6.1). We simulated the power for some choices of θ, σ^2 and ρ, namely $\theta = .0, .2, .4$, $\sigma^2 = .7, 1, 1.3$ and $\rho = .5, .8$, for the sample sizes $n = 50, 100, 150$ and for the values $r_n = 1, 2, 3, 4$. In each case the power was estimated based on 10,000 repetitions

TABLE 2. Simulated power for bivariate normal data with $\alpha = .05$ for selected values of θ, σ^2, ρ, n and $r_n = 1, \ldots, 4$.

			$\rho = .5$				$\rho = .8$			
n	θ	σ^2	1	2	3	4	1	2	3	4
50	0.0	0.7	0.051	0.127	0.105	0.103	0.060	0.186	0.145	0.161
	0.0	1.0	0.049	0.044	0.047	0.042	0.050	0.045	0.042	0.037
	0.0	1.4	0.051	0.117	0.097	0.095	0.052	0.164	0.130	0.144
	0.2	0.7	0.276	0.290	0.240	0.224	0.531	0.539	0.484	0.453
	0.2	1.0	0.244	0.180	0.161	0.134	0.468	0.341	0.319	0.261
	0.2	1.4	0.210	0.225	0.187	0.171	0.409	0.433	0.380	0.352
	0.4	0.7	0.778	0.724	0.684	0.629	0.982	0.971	0.972	0.956
	0.4	1.0	0.714	0.590	0.554	0.484	0.969	0.925	0.928	0.888
	0.4	1.4	0.613	0.562	0.517	0.471	0.927	0.895	0.886	0.859
100	0.0	0.7	0.052	0.227	0.188	0.206	0.052	0.353	0.291	0.359
	0.0	1.0	0.048	0.049	0.047	0.044	0.053	0.049	0.050	0.048
	0.0	1.4	0.054	0.199	0.160	0.180	0.057	0.313	0.259	0.313
	0.2	0.7	0.505	0.564	0.514	0.499	0.836	0.876	0.861	0.862
	0.2	1.0	0.440	0.341	0.314	0.274	0.778	0.676	0.671	0.610
	0.2	1.4	0.370	0.450	0.403	0.400	0.691	0.756	0.727	0.730
	0.4	0.7	0.972	0.967	0.965	0.954	1.000	1.000	1.000	1.000
	0.4	1.0	0.950	0.899	0.891	0.856	1.000	1.000	1.000	0.999
	0.4	1.4	0.908	0.897	0.879	0.859	0.998	0.998	0.998	0.997
150	0.0	0.7	0.050	0.330	0.280	0.328	0.055	0.506	0.431	0.542
	0.0	1.0	0.048	0.052	0.049	0.047	0.052	0.052	0.051	0.051
	0.0	1.4	0.052	0.296	0.249	0.283	0.052	0.456	0.386	0.490
	0.2	0.7	0.677	0.758	0.728	0.727	0.945	0.973	0.974	0.974
	0.2	1.0	0.598	0.486	0.465	0.418	0.918	0.859	0.863	0.823
	0.2	1.4	0.516	0.629	0.578	0.589	0.856	0.919	0.908	0.915
	0.4	0.7	0.997	0.997	0.997	0.996	1.000	1.000	1.000	1.000
	0.4	1.0	0.993	0.983	0.982	0.972	1.000	1.000	1.000	1.000
	0.4	1.4	0.982	0.983	0.981	0.974	1.000	1.000	1.000	1.000

using a significance level of $\alpha = .05$. The results are reported in Table 1 for the above mentioned values of θ, σ^2, ρ, n and r_n. The rows corresponding to the value $(\theta, \sigma^2) = (0, 1)$ refer to the null hypothesis. We see from the table that the power is larger for the larger value of ρ.

We also generated data from the Farlie-Gumbel-Morgenstern copula model with marginals F and G possessing densities f and g, respectively. The density for this model is given by

$$p_{\gamma,F,G}(x,y) = \big(1 + \gamma(1 - 2F(x))(1 - 2G(y))\big) f(x)g(y), \quad x, y \in \mathbb{R},$$

where γ is a number in the interval $(-1, 1)$. As shown in Peng and Schick (2004), the density $p_{\gamma,F,G}$ satisfies the condition (6.1). For our simulation we took $\gamma = .5, .8$, F to be the logistic distribution function, $F(x) = 1/(1 + \exp(-x))$, and G of the form $G(x) = F((x - \theta)/\sigma)$ for some selected values of θ and σ, namely $\theta = 0, .2, .4$ and $\sigma = .8, 1, 1.2$. We again estimated the powers using 10,000 repetitions. Table 2 reports the simulated powers of the test for the above combinations of values of θ,

Table 3. Simulated power in the FGM copula with logistic marginals with $\alpha = .05$ for selected values of θ, σ, γ, n, and $r_n = 1, \ldots, 4$.

			$\gamma = .5$				$\gamma = .8$			
n	θ	σ	1	2	3	4	1	2	3	4
50	0.0	0.8	0.048	0.138	0.112	0.109	0.052	0.142	0.116	0.112
	0.0	1.0	0.050	0.050	0.049	0.047	0.052	0.049	0.046	0.042
	0.0	1.2	0.056	0.110	0.092	0.084	0.056	0.111	0.093	0.086
	0.2	0.8	0.110	0.179	0.143	0.134	0.116	0.189	0.155	0.140
	0.2	1.0	0.094	0.078	0.070	0.062	0.101	0.075	0.070	0.063
	0.2	1.2	0.080	0.127	0.106	0.100	0.086	0.134	0.112	0.102
	0.4	0.8	0.274	0.309	0.253	0.225	0.305	0.330	0.268	0.244
	0.4	1.0	0.231	0.169	0.137	0.118	0.257	0.185	0.153	0.131
	0.4	1.2	0.206	0.219	0.177	0.161	0.215	0.227	0.182	0.164
100	0.0	0.8	0.051	0.252	0.212	0.218	0.055	0.257	0.214	0.217
	0.0	1.0	0.050	0.051	0.048	0.046	0.050	0.049	0.047	0.046
	0.0	1.2	0.050	0.178	0.149	0.154	0.053	0.183	0.153	0.151
	0.2	0.8	0.166	0.350	0.296	0.290	0.186	0.361	0.307	0.298
	0.2	1.0	0.145	0.110	0.093	0.083	0.148	0.115	0.097	0.087
	0.2	1.2	0.121	0.239	0.202	0.194	0.136	0.249	0.204	0.199
	0.4	0.8	0.493	0.586	0.517	0.491	0.543	0.612	0.548	0.523
	0.4	1.0	0.417	0.321	0.270	0.237	0.468	0.354	0.309	0.267
	0.4	1.2	0.356	0.416	0.353	0.326	0.388	0.450	0.388	0.364
150	0.0	0.8	0.055	0.377	0.318	0.333	0.054	0.366	0.308	0.322
	0.0	1.0	0.050	0.049	0.052	0.050	0.051	0.054	0.049	0.048
	0.0	1.2	0.052	0.267	0.221	0.228	0.049	0.255	0.222	0.227
	0.2	0.8	0.226	0.498	0.433	0.434	0.255	0.511	0.446	0.452
	0.2	1.0	0.187	0.151	0.121	0.112	0.211	0.157	0.139	0.123
	0.2	1.2	0.163	0.351	0.294	0.289	0.181	0.364	0.310	0.310
	0.4	0.8	0.661	0.778	0.718	0.702	0.722	0.814	0.758	0.740
	0.4	1.0	0.577	0.463	0.407	0.354	0.635	0.524	0.460	0.412
	0.4	1.2	0.497	0.598	0.527	0.508	0.545	0.631	0.569	0.543

σ and $\gamma = .5, .8$, for sample sizes $n = 50, 100, 150$ and nominal level of significance .05. The rows with $(\theta, \sigma) = (0, 1)$ in Table 2 correspond to the null hypothesis. From the table it appears that the power is slightly larger for the larger value of γ.

7. Testing diagonality of a covariance matrix of a normal random vector with increasing dimension

Let $\{\mathbb{X}_{n1}(a) : a \in A_n\}, \ldots, \{\mathbb{X}_{n,n}(a) : a \in A_n\}$ be independent and identically distributed centered second order processes indexed by a finite set A_n with r_n elements. We denote the covariance function by

$$K_n(a, b) = E[\mathbb{X}_{n1}(a)\mathbb{X}_{n1}(b)], \quad a, b \in A_n,$$

and set

$$\tilde{\mathbb{X}}_n(a) = \frac{1}{\sqrt{n}} \sum_{j=1}^{n} \mathbb{X}_{nj}(a), \quad a \in A_n.$$

The results in this paper now apply with $\xi_{nj} = (\mathbb{X}_{nj}(a_1), \dots, \mathbb{X}_{nj}(a_{r_n}))^\top$ for any enumeration $(a_1, \dots, a_{r_n})$ of the elements of A_n. For notational considerations it might be more convenient to avoid this enumeration and work with the given parametrization. In the original parametrization, the analog of (1.1) with $\mu_n = 0$ is

$$\frac{\sum_{a\in A_n}(\tilde{\mathbb{X}}_n(a))^2 - \sum_{a\in A_n} K_n(a,a)}{\sqrt{2\sum_{a,b\in A_n} K_n(a,b)K_n(b,a)}} \implies N(0,1). \tag{7.1}$$

Simple sufficient conditions for this are

$$\max_{a\in A_n} \sum_{b\in A_n} |K_n(a,b)| = O(1), \tag{7.2}$$

$$k_n = \sum_{a,b\in A_n} K_n(a,b)K_n(b,a) \to \infty, \tag{7.3}$$

$$r_n \sum_{a\in A_n} E[|\mathbb{X}_{n1}(a)|^4] = o(nk_n). \tag{7.4}$$

Indeed, the first condition implies (C1) as the operator norm of a symmetric matrix is bounded by the maximal ℓ_1-norm of its rows, the second condition is equivalent to (C2), and the third condition implies (1.7). Thus (7.1) follows from Corollary 1.

Let us now illustrate this result. Suppose $Z_{n1}, \dots, Z_{nn}$ are independent and identically distributed centered p_n-dimensional random vectors. We are interested in testing whether their dispersion matrix Σ_n is diagonal. Since the random vectors are centered, we estimate Σ_n by $\hat{\Sigma}_n = \frac{1}{n}\sum_{j=1}^n Z_{nj}Z_{nj}^\top$. An estimator of Σ under the null hypothesis is $\mathrm{diag}(\hat{\Sigma}_n)$, the diagonal matrix formed by the diagonal entries of $\hat{\Sigma}_n$. As test statistic we can then take

$$T_n = (n/2)\|\hat{\Sigma}_n - \mathrm{diag}(\hat{\Sigma}_n)\|_2^2.$$

We can express T_n as

$$T_n = \sum_{1\le i<j\le p_n} (\tilde{\mathbb{X}}_n(i,j))^2$$

with

$$\tilde{\mathbb{X}}_n(i,j) = \frac{1}{\sqrt{n}} \sum_{k=1}^n Z_{n,k,i} Z_{n,k,j}$$

where $Z_{n,k,i}$ denotes the i-th coordinate of Z_{nk}. Here A_n equals $\{(i,j) : 1 \le i < j \le p_n\}$ and has $r_n = p_n(p_n-1)/2$ elements, while $\mathbb{X}_{nk}(a) = Z_{n,k,i}Z_{nk,j}$ for $a = (i,j) \in A_n$. Let us now derive the asymptotic behavior of T_n under the null hypothesis. For this we assume that Σ_n is a diagonal matrix whose diagonal entries fall into a compact subinterval $[\lambda, \Lambda]$ of $(0,\infty)$ for each n. Under this assumption we calculate

$$K_n((i,j),(k,l)) = \begin{cases} \Sigma_{n,i,i}\Sigma_{n,jj}, & (k,l) = (i,j), \\ 0, & \text{otherwise,} \end{cases}$$

and find

$$\max_{a\in A_n} \sum_{b\in A_n} |K_n(a,b)| = \max_{1\le i<j\le p_n} \Sigma_{n,i,i}\Sigma_{n,j,j} \le \Lambda^2,$$

$$k_n = \sum_{a,b\in A_n} K_n(a,b)K_n(b,a) = \sum_{1\le i<j\le p_n} \Sigma^2_{n,i,i}\Sigma^2_{n,j,j} \ge \lambda^4 p_n(p_n-1)/2 \to \infty$$

and

$$\sum_{1\le i<j\le p_n} E[Z^4_{n,1,i}Z^4_{n,1,j}] = \sum_{1\le i<j\le p_n} 9\Sigma^2_{n,i,i}\Sigma^2_{n,j,j} = 9k_n.$$

Thus, the sufficient conditions (7.2)–(7.4) are met if $p_n^2 = o(n)$. In this case we have

$$\frac{T_n - \sum_{1\le i<j\le p_n}\Sigma_{n,i,i}\Sigma_{n,j,j}}{\sqrt{2\sum_{1\le i<j\le p_n}\Sigma^2_{n,i,i,}\Sigma^2_{n,j,j}}} \implies N(0,1).$$

This result, however, is of limited practical use as the quantities $\Sigma_{n,i,i}$ are unknown. We claim that under the above assumptions

$$\frac{T_n - \sum_{1\le i<j\le p_n}\hat\Sigma_{n,i,i}\hat\Sigma_{n,j,j}}{\sqrt{2\sum_{1\le i<j\le p_n}\hat\Sigma^2_{n,i,i}\hat\Sigma^2_{n,j,j}}} \implies N(0,1).$$

This follows if we show

$$S_{n,1} = \sum_{1\le i<j\le p_n} (\hat\Sigma_{n,i,i}\hat\Sigma_{n,j,j} - \Sigma_{n,i,i}\Sigma_{n,j,j}) = o_p(k_n^{1/2})$$

and

$$S_{n,2} = \sum_{1\le i<j\le p_n} (\hat\Sigma^2_{n,i,i}\hat\Sigma^2_{n,j,j} - \Sigma^2_{n,i,i}\Sigma^2_{n,j,j}) = o_p(k_n).$$

It is easy to verify that the summands in $S_{n,1}$ are of the form

$$\Sigma_{n,i,i}\Sigma_{n,j,j}(Y_{ni} + Y_{nj} + Y_{ni}Y_{nj})$$

with

$$Y_{ni} = \frac{1}{n}\sum_{k=1}^n \Big(\frac{Z^2_{n,k,i}}{\Sigma_{n,i,i}} - 1\Big).$$

The random variables $Y_{n1},\dots,Y_{np_n}$ are independent with zero mean and variance $2/n$. Using this and the identity $a^2-b^2 = (a-b)^2 + 2b(a-b)$, we verify

$$E[|S_{n,1}|] \le \sum_{1\le i<j\le p_n} \Lambda^2 E[|Y_{ni} + Y_{nj} + Y_{ni}Y_{nj}| = O(p_n^2/\sqrt{n}) = o(p_n)$$

and

$$\begin{aligned} E[|S_{,2}|] &\le \sum_{1\le i<j\le p_n} \Lambda^4(E[|Y_{ni} + Y_{nj} + Y_{ni}Y_{nj}|^2] + 2E[|Y_{ni} + Y_{nj} + Y_{ni}Y_{nj}|]) \\ &= O(p_n^2/\sqrt{n}) = o(p_n). \end{aligned}$$

These yield the desired results in view of $k_n \ge \lambda^2 p_n(p_n-1)/2$.

8. An auxiliary lemma

Our proofs of the theorems will rely on the following simple lemma.

Lemma 1. *Let $X_1,\dots,X_m$ be independent and identically distributed random vectors with zero mean and dispersion matrix V and set $S_k = X_1 + \cdots + X_k$, $k = 1,\dots,m$. Then one has*

$$E[|S_k|^2] = kE[|X_1|^2] = k\,\mathrm{trace}[V], \quad k = 1,\dots,m,$$

and

$$\epsilon^2 P(\max_{1\le k\le m} |S_k| > \epsilon) \le E[|S_m|^2] = m\,\mathrm{trace}(V), \quad \epsilon > 0.$$

If also $E[|X_1|^4]$ *is finite, then one has*

$$\mathrm{Var}(|S_k|^2) = 2k(k-1)\,\mathrm{trace}(V^2) + k\,\mathrm{Var}(|X_1|^2), \quad k = 1, \dots, m,$$

and

$$\mathrm{Var}(\sum_{k=1}^m |S_k|^2) \le 2m^4 \mathrm{trace}(V^2) + 2m^3\,\mathrm{Var}(|X_1|^2).$$

Proof. The first inequality is the Kolmogorov inequality for random vectors. Let $X = X_1$ and $Y = X_2$. Then

$$E[|X|^2] = E[\mathrm{trace}[XX^\top)] = \mathrm{trace}(E[XX^\top]) = \mathrm{trace}(V)$$

and

$$\begin{aligned} E[(X^\top Y)^2] &= E[\mathrm{trace}(X^\top YY^\top X)] = E[\mathrm{trace}(YY^\top XX^\top)] \\ &= \mathrm{trace}(E[YY^\top XX^\top]) = \mathrm{trace}(E[YY^\top]E[XX^\top]) \\ &= \mathrm{trace}(V^2). \end{aligned}$$

Using independence we calculate

$$E[|S_k|^2] = \sum_{i=1}^k \sum_{j=1}^k E[X_i^\top X_j] = \sum_{i=1}^k E[|X_i|^2] = kE[|X|^2] = k\,\mathrm{trace}(V),$$

$$\begin{aligned} E[|S_k|^4] &= \sum_{i=1}^k \sum_{j=1}^k \sum_{l=1}^k \sum_{p=1}^k E[X_i^\top X_j X_l^\top X_p] \\ &= 4 \sum_{1\le i<j\le k} E[(X_i^\top X_j)^2] + \sum_{i=1}^k E[|X_i|^4] + 2 \sum_{1\le i<p\le k} E[|X_i|^2]E[|X_p|^2] \\ &= 2k(k-1)\mathrm{trace}(V^2) + kE[|X|^4] + k(k-1)(\mathrm{trace}(V))^2, \end{aligned}$$

and hence obtain the desired form of $\mathrm{Var}(|S_k|^2)$. It is easy to see that the covariance of $|S_i|^2$ and $|S_j|^2$ equals the variance of $|S_{\min(i,j)}|^2$. Thus we obtain

$$\begin{aligned} \mathrm{Var}(\sum_{k=1}^m |S_k|^2) &= \sum_{k=1}^m \mathrm{Var}(|S_k|^2)(1 + 2(m-k)) \\ &= \sum_{k=1}^m (1 + 2(m-k)k(2(k-1)\mathrm{trace}(V^2) + \mathrm{Var}(|X|^2)) \end{aligned}$$

and hence the desired bound on the variance of $\sum_{k=1}^m |S_k|^2$. □

9. Proof of the theorems

To simplify notation we abbreviate $\xi_{n,j}$ by ξ_j and $(\mathrm{trace}(V_n^2))^{1/2}$ by σ_n and introduce r_n-dimensional random vectors $D_0 = 0$ and

$$D_j = \frac{\sqrt{2}}{n\sigma_n} \sum_{i=1}^j \xi_i, \quad j = 1, \dots, n.$$

In view of the identity

$$|\tilde{\xi}_n + \mu_n|^2 = \frac{1}{n}\sum_{j=1}^n |\xi_j|^2 + \frac{2}{n}\sum_{1\le i<j\le n} \xi_i^\top \xi_j + 2\mu_n^\top \tilde{\xi}_n + |\mu_n|^2$$

we can write the left-hand side of (1.1) as $Q_n + R_n + T_n$ where

$$Q_n = \sum_{j=1}^n D_{j-1}^\top \xi_j, \quad R_n = \frac{1}{n\sigma_n}\sum_{j=1}^n [|\xi_j|^2 - E[|\xi_j|^2]] \quad \text{and} \quad T_n = \frac{\sqrt{2}}{\sigma_n\sqrt{n}}\sum_{j=1}^n \mu_n^\top \xi_j.$$

We have $E[T_n^2] = 2\mu_n^\top V_n \mu_n/\sigma_n^2 \to 0$. Thus the desired result follow if we show that R_n converges to zero in probability and that Q_n is asymptotically standard normal. The latter follows from the Martingale Central Limit Theorem (see e.g. part (a) of Theorem 2.5 of Helland (1982), or Corollary 3.1 in Hall and Heyde (1980) and the ensuing remarks) if we verify that

$$E_{j-1}(D_{j-1}^\top \xi_j) = 0, \quad j = 1, \dots, n, \tag{9.1}$$

$$\sum_{j=1}^n E_{j-1}((D_{j-1}^\top \xi_j)^2) = 1 + o_p(1) \tag{9.2}$$

and, for $\epsilon > 0$,

$$\sum_{j=1}^n E_{j-1}(|D_{j-1}^\top \xi_j|^2 \mathbf{1}[|D_{j-1}^\top \xi_j| > \epsilon]) = o_p(1), \tag{9.3}$$

with E_{j-1} the conditional expectation given $\xi_1, \dots, \xi_{j-1}$. Of course, (9.1) is a simple consequence of the independence of the random vectors $\xi_1, \dots, \xi_n$.

Proof of Theorem 1. Assume now (1.4)–(1.6) hold. We have $R_n = o_p(1)$ in view of (1.4) and the identity

$$E[R_n^2] = \frac{\mathrm{Var}(|\xi_1|^2)}{n\sigma_n^2}.$$

The left-hand side of (9.2) equals

$$S_n = \sum_{j=1}^n D_{j-1}^\top V_n D_{j-1} = \frac{2}{n^2\sigma_n^2}\sum_{j=1}^{n-1}\Big|\sum_{i=1}^j V_n^{1/2}\xi_i\Big|^2.$$

Note that the random vector $V_n^{1/2}\xi_1$ is centered and has dispersion matrix V_n^2. We have $\mathrm{trace}(V_n^4) \le \rho_n^2 \mathrm{trace}(V_n^2) = \rho_n^2\sigma_n^2$. Thus, with the aid of Lemma 1 and (1.5), we find

$$E[S_n] = \sum_{j=1}^{n-1}\frac{2j}{n^2} = \frac{n-1}{n} \to 1$$

and

$$\mathrm{Var}(S_n) \le \frac{8\rho_n^2}{\sigma_n^2} + \frac{8\,\mathrm{Var}(|V^{1/2}\xi_1|^2)}{n\sigma_n^4} \to 0.$$

This shows that $S_n = 1 + o_p(1)$. Finally, the expected value of the left-hand side of (9.3) is bounded by U_n/ϵ^2 with

$$U_n = \sum_{j=1}^n E[|D_{j-1}^\top \xi_j|^4] = \frac{4}{n^4\sigma_n^4}\sum_{j=2}^n E[(\sum_{i=1}^{j-1}\xi_i^\top \xi_j)^4].$$

Conditioning in the expectation with index j on ξ_j, we obtain with the aid of Lemma 1,

$$\begin{aligned} U_n &\le \frac{4}{n^4\sigma_n^4}\sum_{j=2}^n [3(j-1)(j-2)E[(\xi_1^\top V_n \xi_1)^2] + (j-1)E[(\xi_1^\top \xi_2)^4] \\ &\le \frac{4}{n^4\sigma_n^4}\big[3n^3(\mathrm{Var}(|V_n^{1/2}\xi_1|^2) + \mathrm{trace}^2(V_n^2)) + n^2E[|\xi_1^\top\xi_2|^4]\big]. \end{aligned}$$

It follows from (1.5) and (1.6) that U_n converges to zero. This proves (9.3) and completes the proof of Theorem 1.

Proof of Theorem 2. For an arbitrary positive ϵ, we can write $R_n = R_{n,1}+R_{n,2}$, where

$$R_{n,1} = \frac{1}{n\sigma_n}\sum_{j=1}^n \big(|\xi_j|^2\mathbf{1}[|\xi_j| \le \epsilon\sqrt{n}] - E[|\xi_j|^2\mathbf{1}[|\xi_{n,j}| \le \epsilon\sqrt{n}]]\big),$$

$$R_{n,2} = \frac{1}{n\sigma_n}\sum_{j=1}^n \big(|\xi_j|^2\mathbf{1}[|\xi_j| > \epsilon\sqrt{n}] - E[|\xi_j|^2[|\xi_j| > \epsilon\sqrt{n}]]\big),$$

and calculate $E[|R_{n,2}|] \le 2L_n(\epsilon)/\sigma_n$ and

$$E[R_{n,1}^2] \le \frac{E[|\xi_1|^4\mathbf{1}[|\xi_1| \le \epsilon\sqrt{n}]]}{n\sigma_n^2} \le \frac{\epsilon^2 E[|\xi_1|^2]}{\sigma_n^2} = \epsilon^2\frac{\mathrm{trace}(V_n)}{\mathrm{trace}(V_n^2)}.$$

This shows that $R_n = o_p(1)$.

Next, we show

$$D_n^* = \max_{1\le j\le n}\sqrt{n}|D_{j-1}| = O_p(1) \quad \text{and} \quad \sum_{j=1}^n |D_{j-1}|^2 = O_p(1). \tag{9.4}$$

Indeed, with the help of Lemma 1 we obtain

$$P(D_n^* > \sqrt{2}K) = P(\max_{1\le j\le n}|\sum_{i=1}^{j-1}\xi_i| > K\sigma_n\sqrt{n}) \le \frac{\mathrm{trace}(V_n)}{\sigma_n^2K^2} = \frac{1}{K^2}, \quad K>0,$$

and

$$\sum_{j=1}^n E[|D_{j-1}|^2] = \frac{2}{n^2\sigma_n^2}\sum_{j=1}^n (j-1)\mathrm{trace}(V_n) \le \frac{\mathrm{trace}(V_n)}{\sigma_n^2}.$$

The statements (9.4) imply (9.3), since the left-hand side of (9.3) is bounded by

$$\sum_{j=1}^n |D_{j-1}|^2\int y^2\mathbf{1}[|D_{j-1}|y > \epsilon]\,dF_n(y) \le \int y^2\mathbf{1}[D_n^*y > \epsilon\sqrt{n}]\,dF_n(y)\sum_{j=1}^n |D_{j-1}|^2,$$

where F_n is the distribution of $|\xi_1|$.

Finally, we obtain (9.2) by verifying

$$S_n = \frac{2}{n^2\sigma_n^2}\sum_{j=2}^n \Big|\sum_{i=1}^{j-1} V_n^{1/2}\xi_i\Big|^2 = 1 + o_p(1). \tag{9.5}$$

For this we write $\xi_j = X_j + Y_j$ with

$$\begin{aligned} X_j &= \xi_j\mathbf{1}[|\xi_j| \le \sqrt{n}] - E[\xi_j\mathbf{1}[|\xi_j| \le \sqrt{n}]], \\ Y_j &= \xi_j\mathbf{1}[|\xi_j| > \sqrt{n}] - E[\xi_j\mathbf{1}[|\xi_j| > \sqrt{n}]]. \end{aligned}$$

In view of the Cauchy–Schwarz inequality, the desired (9.5) follows from the statements

$$S_{n,1} = \frac{2}{n^2\sigma_n^2}\sum_{i=1}^{n}\Big|\sum_{i=1}^{j-1} V_n^{1/2}X_j\Big|^2 = 1 + o_p(1) \tag{9.6}$$

and

$$S_{n,2} = \frac{2}{n^2\sigma_n^2}\sum_{j=1}^{n}\Big|\sum_{i=1}^{j-1} V_n^{1/2}Y_j\Big|^2 = o_p(1). \tag{9.7}$$

The latter follows from the bound

$$E[S_{n,2}] = \frac{2}{n^2\sigma_n^2}\sum_{j=1}^{n}(j-1)E[|V_n^{1/2}Y_1|^2] \le \frac{\rho_n L_n(1)}{\sigma_n^2}.$$

The former follows if we show $E[S_{n,1}] \to 1$ and $\mathrm{Var}(S_{n,1}) \to 0$. We calculate

$$E[S_{n,1}] = \frac{(n-1)}{n\sigma_n^2}|E[V_n^{1/2}X_1|^2] = \frac{(n-1)}{n}\frac{\mathrm{trace}(W_n)}{\sigma_n^2}$$

with $W_n = V_n^{1/2}E[X_1X_1^\top]V_n^{1/2}$ the dispersion matrix of $V_n^{1/2}X_1$. We have the identity

$$W_n = V_n^{1/2}(V_n - E[Z_nZ_n^\top] - E[Z_n]E[Z_n]^\top)V_n^{1/2}$$

with $Z_n = \xi_1[\mathbf{1}|\xi_1| > \sqrt{n}]$ and obtain the inequality

$$\mathrm{trace}(V_n^2) - 2\rho_n L_n(1) \le \mathrm{trace}(W_n) \le \mathrm{trace}(V_n^2).$$

This lets us conclude $E[S_{n,1}] \to 1$. Lemma 1 and the inequalities $|V_n^{1/2}X_1|^2 \le 4n\rho_n$ and $\mathrm{trace}(W_n^2) \le \mathrm{trace}(V_n^4) \le \rho_n^2\mathrm{trace}(V_n^2)$ yield

$$\mathrm{Var}(S_{n,1}) \le \frac{8}{n^4\sigma_n^4}(n^4\mathrm{trace}(W_n^2) + n^3E(|V^{1/2}X_1|^4]) \le \frac{8\rho_n^2 + 32\rho_n}{\sigma_n^2} \to 0.$$

This completes the proof of Theorem 2.

References

[1] Bai, Z. and Saranadasa, H. (1996). Effect of high dimension: by an example of a two sample problem. *Statist. Sinica*, **6**, 311-329.

[2] Chen, S.X., Peng, L. and Qin, Y.-L. (2009). Effects of data dimension on empirical likelihood. *Biometrika* **96**, 711–722.

[3] Hall, P. and Heyde, C.C. (1980). *Martingale limit theory and its application.* Academic Press.

[4] Helland, I. (1982). Central limit theorems for martingales with discrete or continuous time. *Scand. J. Statist.* **9**, 79–94.

[5] Hjort, N.L., McKeague, I.W. and Van Keilegom, I. (2009). Extending the scope of empirical likelihood. *Ann. Statist.* **37**, 1079–1111.

[6] Owen, A. (1988). Empirical likelihood ratio confidence intervals for a single functional. *Biometrika* **75**, 237–249.

[7] Owen, A. (1990). Empirical likelihood ratio confidence regions. *Ann. Statist.* **18**, 90–120.

[8] Owen, A. (2001). *Empirical Likelihood.* Chapman & Hall/CRC, London.

[9] Peng, H. and Schick, A. (2005). Efficient estimation of linear functionals of a bivariate distribution with equal, but unknown, marginals: The least squares approach. *J. Multivariate Anal.* **95**, 385-409.

[10] Peng, H. and Schick, A. (2013). An empirical likelihood approach to goodness of fit testing. *Bernoulli*, **19**, 2250–2276.

[11] Portnoy, S. (1988). Behavior of likelihood methods for exponential families when the number of parameters tends to infinity. *Ann. Statist.* **16**, 356–366.

Hanxiang Peng, Indiana University Purdue University, at Indianapolis, Department of Mathematical Sciences, Indianapolis, IN 46202-3267, USA

Anton Schick, Department of Mathematical Sciences, Binghamton University, Binghamton, NY 13902-6000, USA