Submitted to the Bernoulli

An Empirical Likelihood Approach To Goodness of Fit Testing

HANXIANG PENG and ANTON SCHICK

Indiana University Purdue University at Indianapolis, Department of Mathematical Sciences, Indianapolis, IN 46202, USA
E-mail: hpeng@math.iupui.edu

Binghamton University, Department of Mathematical Sciences, Binghamton, NY 13902, USA
E-mail: anton@math.binghamton.edu

Motivated by applications to goodness of fit testing, the empirical likelihood approach is generalized to allow for the number of constraints to grow with the sample size and for the constraints to use estimated criteria functions. The latter is needed to deal with nuisance parameters. The proposed empirical likelihood based goodness of fit tests are asymptotically distribution free. For univariate observations, tests for a specified distribution, for a distribution of parametric form, and for a symmetric distribution are presented. For bivariate observations, tests for independence are developed.

Keywords: infinitely many constraints; nuisance parameter; estimated constraint functions; regression model; testing for a specific distribution; testing for a parametric model; testing for symmetry; testing for independence.

1. Introduction

The empirical likelihood approach was introduced by Owen (1988, 1990) to construct confidence intervals in a nonparametric setting, see also Owen (2001). As a likelihood approach possessing nonparametric properties, it does not require us to specify a distribution for the data and often yields more efficient estimates of the parameters. It allows data to decide the shape of confidence regions and is Bartlett correctable (DiCiccio, Hall and Romano, 1991). The approach has been developed to various situations, e.g., to generalized linear models (Kolaczyk, 1994), local linear smoother (Chen and Qin, 2000), partially linear models (Shi and Lau, 2000; Wang and Jing, 2003), parametric and semiparametric models in multiresponse regression (Chen and Van Keilegom, 2009), linear regression with censored data (Zhou and Li, 2008), and plug-in estimates of nuisance parameters in estimating equations in the context of survival analysis (Qin and Jing, 2001; Wang and Jing, 2001; Li and Wang, 2003). Algorithms, calibration and higher-order precision of the approach can be found in Hall and La Scala (1990), Emerson and Owen (2009) and Liu and Chen (2010) among others. It is especially convenient to incorporate side information expressed through equality constraints. Qin and Lawless (1994) linked empirical likelihood with finitely many estimating equations. These estimating equations serve as finitely many equality constraints.

In semiparametric settings, information on the model can often be expressed by means of *infinitely* many constraints which may also depend on parameters of the model. In goodness of fit testing, the null hypothesis can typically be expressed by infinitely many such constraints. This is the case when testing for a fixed distribution (see Example 1 below), when testing for a given parametric model (Example 2), when testing for symmetry about a fixed point (Example 3), and when testing for independence (Example 4). Modeling conditional expectations can also be done by means of infinitely many constraints. This has applications to heteroscedastic regression models (Section 3) and to conditional moment restriction models treated by Tripathi and Kitamura (2003) using a smoothed empirical likelihood approach.

Recently Hjort, McKeague and Van Keilegom (2009) extended the scope of the empirical method. In particular, they developed a general theory for constraints with nuisance parameters and considered the case with infinitely many constraints. Their results for infinitely many constraints, however, do not allow for nuisance parameters. In this paper we will fill this gap and in the process improve on their results. Let us now discuss some of our results in the following special case.

Let $Z_1, \dots, Z_n$ be independent copies of a random vector Z with distribution Q. Let $u_1, u_2, \dots$ be orthonormal elements of

$$L_{2,0}(Q) = \{u \in L_2(Q) : \int u\, dQ = 0\}.$$

Then the random variables $u_1(Z), u_2(Z), \dots$ have mean zero, variance one and are uncorrelated. Now consider the empirical likelihood based on the first m of these functions,

$$\mathscr{R}_n = \sup\Big\{ \prod_{j=1}^n n\pi_j : \pi \in \mathscr{P}_n,\ \sum_{j=1}^n \pi_j u_k(Z_j) = 0, \quad k = 1, \dots, m\Big\},$$

where $\mathscr{P}_n = \{\pi = (\pi_1, \dots, \pi_n)^\top \in [0,1]^n : \pi_1 + \dots + \pi_n = 1\}$ denotes the closed probability simplex in dimension n. For fixed m, it follows from Owen's work that $-2\log \mathscr{R}_n$ has asymptotically a chi-square distribution with m degrees of freedom. In other words,

$$P(-2\log \mathscr{R}_n > \chi^2_{1-\alpha}(m)) \to \alpha, \quad 0 < \alpha < 1, \tag{1.1}$$

where $\chi^2_\beta(m)$ denotes the β-quantile of the chi-square distribution with m degrees of freedom. Hjort *et al* (2009) have shown that (1.1) holds under some additional assumptions even if m tends to infinity with n by proving the asymptotic normality result

$$(-2\log \mathscr{R}_n - m)/\sqrt{2m} \implies N(0,1). \tag{1.2}$$

This result requires higher moment assumptions on the functions $u_1, u_2, \dots$ and restrictions on the rate at which m can tend to infinity. For example, if the functions $u_1, u_2, \dots$ are uniformly bounded, then the rate $m^3 = o(n)$ suffices for (1.2). They also state in their

Theorem 4.1, that if $\sup_k \int |u_k|^q \, dQ$ is finite for some $q > 2$, then $m^{3+6/(q-2)} = o(n)$ suffices for (1.2). A gap in their argument was fixed by Peng and Schick (2012). We shall show that larger m are allowed in some cases. In particular, for $q = 4$, it suffices that $m^4 = o(n)$ holds (instead of their $m^6 = o(n)$) and if $q = 3$, then $m_n^6 = o(n)$ is enough (instead of their $m_n^9 = o(n)$), see our Theorems 7.2 and 7.3 below.

Our rate $m^4 = o(n)$ for $q = 4$ matches the rate given in Theorem 2 of Chen, Peng and Qin (2009). These authors obtain asymptotic normality for m larger than in Hjort *et al* (2009) by imposing additional structural assumptions. These assumptions, however, are typically not met in the applications we have in mind.

One of the key points in our proof is a simple condition for the convex hull of some vectors $x_1, \dots, x_n$ to have the origin as an interior point. Our condition is that the smallest eigen value of $\sum_{i=1}^n x_i x_i^\top$ exceeds $5|\sum_{i=1}^n x_i| \max_{1\le j\le n} |x_j|$. Here $|x|$ denotes the euclidean norm of a vector x. This sufficient condition ties in nicely with the other requirements used to establish the asymptotic behavior of the empirical likelihood and is typically implied by these. For example, conditions (A1)–(A3) in Theorem 2.1 of Hjort *et al* (2009) already imply their (A0). Thus the conclusion of their theorem is valid under (A1)–(A3) only, see our Theorem 6.1.

Let us now look at the case when the functions $u_1, u_2, \dots$ are unknown. Then we can work with the empirical likelihood

$$\hat{\mathscr{R}}_n = \sup\Big\{ \prod_{j=1}^n n\pi_j : \pi \in \mathscr{P}_n,\ \sum_{j=1}^n \pi_j \hat{u}_k(Z_j) = 0, \quad k = 1, \dots, m \Big\}$$

where $\hat{u}_k$ is an estimator of u_k such that

$$\sum_{k=1}^m \frac{1}{n} \sum_{j=1}^n (\hat{u}_k(Z_j) - u_k(Z_j))^2 = o_p(m^{-1}). \tag{1.3}$$

Now we have the conclusion $(-2\log \hat{\mathscr{R}}_n - m)/\sqrt{2m} \implies N(0,1)$ under the condition

$$\sum_{k=1}^m \Big(n^{-1/2} \sum_{j=1}^n (\hat{u}_k(Z_j) - u_k(Z_j)) \Big)^2 = o_p(1) \tag{1.4}$$

and mild additional conditions such as

(i) $|\hat{u}_k| + |u_k| \le B$ for some constant B and all k and $m^3 = o(n)$, or

(ii) $\sum_{k=1}^m \int u_k^4 \, dQ = O(m^2)$ and $m_n^4 = o(n)$.

Our results, however, go beyond this simple result. If (1.4) is replaced by

$$\sum_{k=1}^m \Big(n^{-1/2} \sum_{j=1}^n (\hat{u}_k(Z_j) - u_k(Z_j) + E[u_k(Z)\psi^\top(Z)]\psi(Z_j)) \Big)^2 = o_p(1) \tag{1.5}$$

with ψ a measurable function into $\mathbb{R}^q$ which is standardized under Q in the sense that $E[\psi(Z)] = 0$ and $E[\psi(Z)\psi^\top(Z)] = I_q$, the $q \times q$ identity matrix, then the conclusion $(-2\log \hat{\mathscr{R}}_n - (m-q))/\sqrt{2(m-q)} \implies N(0,1)$ holds under (i) or (ii).

Our paper is organized as follows. In Section 2, we give four examples that motivate our research. The emphasis in these examples is on goodness of fit testing. The proposed empirical likelihood based goodness of fit tests are asymptotically distribution free. For univariate observations, tests for a specified distribution, for a distribution of parametric form, and for a symmetric distribution are presented. For bivariate observations, tests for independence are discussed. Another example is given in Section 3 with a small simulation study. This example considers tests for the regression parameters in simple linear heteroscedastic regression. The simulations compare our new procedure based on infinitely many constraints with the classical empirical likelihood procedure and illustrate improvements by the new procedures. In Section 4 we introduce notation and recall some results on the spectral norm of matrices. In Section 5 we derive a lemma that extracts the essence from the proofs of Owen (2001, Chapter 11) and also obtains the aforementioned sufficient condition for a convex hull of vectors to contain the origin as interior point. The results are derived for non-stochastic vectors and formulated as inequalities. The inequalities are used in Section 6 to obtain the behavior of the empirical likelihood with random vectors whose dimension may increase. The results are formulated abstractly and do not require independence. In Section 7 we specialize our results to the case of independent observations with infinitely many constraints, both known and unknown. We also briefly discuss the behavior under contiguous alternatives. The details for our examples are given in Section 8.

2. Motivating examples

In this section, we give examples that motivated the research in this paper.

***Example* 1.** TESTING FOR A FIXED DISTRIBUTION. Let $X_1, \dots, X_n$ be independent copies of a random variable X. Suppose we want to test whether their common distribution function F equals a known *continuous* distribution function F_0. Under the null hypothesis, we have $E[h(X)] = 0$ for every $h \in L_{2,0}(F_0)$, and $F_0(X)$ has a uniform distribution. An orthonormal basis of $L_{2,0}(F_0)$ is thus given by $v_1 \circ F_0, v_2 \circ F_0, \dots$ for any orthonormal basis $v_1, v_2, \dots$ of $L_{2,0}(U)$, where U is the uniform distribution on $(0, 1)$. We shall work with the trigonometric basis $\phi_1, \phi_2, \dots$ defined by

$$\phi_k(x) = \sqrt{2}\cos(k\pi x), \quad x \in [0,1],\ k = 1, 2, \dots, \tag{2.1}$$

as these basis functions are uniformly bounded by $\sqrt{2}$. As test statistic we take

$$\mathscr{R}_n(F_0) = \sup\Big\{ \prod_{j=1}^n n\pi_j : \pi \in \mathscr{P}_n,\ \sum_{j=1}^n \pi_j \phi_k(F_0(X_j)) = 0, \quad k = 1, \dots, m\Big\}$$

which uses the first m of the trigonometric functions. Under the null hypothesis we have $P(-2\log \mathscr{R}_n(F_0) > \chi^2_{1-\alpha}(m)) \to \alpha$ for every $0 < \alpha < 1$ as both m and n tend to infinity and m^3/n tends to zero. Thus the test $\mathbf{1}[-2\log \mathscr{R}_n(F_0) > \chi^2_{1-\alpha}(m)]$ has asymptotic size α. Here we are still in the framework of Hjort *et al* (2009) with infinitely many *known* constraints.

Example 2. TESTING FOR A PARAMETRIC MODEL. Let $X_1, \dots, X_n$ be again independent and identically distributed random variables. But now suppose we want to test whether their common distribution function F belongs to a model $\mathscr{F} = \{F_\vartheta : \vartheta \in \Theta\}$ indexed by an open subset Θ of $\mathbb{R}^q$. Suppose that the distribution functions F_ϑ have densities f_ϑ such that the map $\vartheta \mapsto s_\vartheta = \sqrt{f_\vartheta}$ is continuously differentiable in L_2 with derivative $\vartheta \mapsto \dot{s}_\vartheta$ and the matrix $J(\vartheta) = 4 \int \dot{s}_\vartheta(x)\dot{s}_\vartheta(x)^\top \, dx$ is invertible for each $\vartheta \in \Theta$. In this case we set $\dot{\ell}_\vartheta = 2\dot{s}_\vartheta / s_\vartheta$. Let now $\hat{\theta}$ be an estimator of the parameter in the model. We require it to satisfy the stochastic expansion

$$\hat{\theta} = \theta + \frac{1}{n} \sum_{j=1}^{n} J(\theta)^{-1} \dot{\ell}_\theta(X_j) + o_{P_\theta}(n^{-1/2}) \tag{2.2}$$

for each $\theta \in \Theta$, where P_θ is the measure for which $F = F_\theta$. Such estimators are efficient in the parametric model. Candidates are maximum likelihood estimators. As test statistic we take $\mathscr{R}_n(F_{\hat{\theta}})$, the test statistic from the previous example with F_0 replaced by $F_{\hat{\theta}}$. Here we are no longer in the framework of Hjort *et al* (2009) as we now have infinitely many *unknown* constraints. We shall show that under the null hypothesis $P(-2 \log \mathscr{R}_n(F_{\hat{\theta}}) > \chi^2_{1-\alpha}(m-q)) \to \alpha$ for every $0 < \alpha < 1$ as both m and n tend to infinity and $\log n \; m^3/n$ tends to zero. In view of this result the test $\mathbf{1}[-2 \log \mathscr{R}_n(F_{\hat{\theta}}) > \chi^2_{1-\alpha}(m-q)]$ has asymptotic size α. It is crucial for our result that we have chosen an estimator $\hat{\theta}$ satisfying (2.2).

Example 3. TESTING FOR SYMMETRY. Let $X_1, \dots, X_n$ be independent copies of a random variable X with a continuous distribution function F. We want to test whether F is symmetric about zero in the sense that $F(t) = 1 - F(-t)$ for all real t. Under the null hypothesis of symmetry, the random variables $\mathrm{sign}(X)$ and $|X|$ are independent, and $\mathrm{sign}(X)$ takes values -1 and 1 with probability one half. This is equivalent to $E[\mathrm{sign}(X)v(|X|)] = 0$ for every $v \in L_2(H)$, where H is the distribution function of $|X|$. Since H is continuous, an orthonormal system of $L_2(H)$ is given by $\phi_0 \circ H, \phi_1 \circ H, \dots$ where $\phi_0 = 1$ and $\phi_1, \phi_2, \dots$ are given in (2.1). This suggests the test statistic

$$\mathscr{R}_n = \sup \Big\{ \prod_{j=1}^{n} n\pi_j : \pi \in \mathscr{P}_n, \; \sum_{j=1}^{n} \pi_j \, \mathrm{sign}(X_j)\phi_k(R_j) = 0, \; k = 0, \dots, m \Big\},$$

where $R_j = \mathbb{H}(|X_j|)$ and $\mathbb{H}$ is the empirical distribution function based on $|X_1|, \dots, |X_n|$. We shall show that under symmetry one has $P(-2 \log \mathscr{R}_n > \chi^2_{1-\alpha}(m+1)) \to \alpha$ for every $0 < \alpha < 1$ as m and n tend to infinity and m^3/n tends to zero. From this we derive that the test $\mathbf{1}[-2 \log \mathscr{R}_n > \chi^2_{1-\alpha}(m+1)]$ has asymptotic size α.

Example 4. TESTING FOR INDEPENDENCE. Let $(X_1, Y_1), \dots, (X_n, Y_n)$ be independent copies of a bivariate random vector (X, Y). We assume that the marginal distribution functions F and G are continuous. We want to test whether X and Y are independent. Independence is equivalent to $E[a(X)b(Y)] = 0$ for all $a \in L_{2,0}(F)$ and $b \in L_{2,0}(G)$ and thus equivalent to $E[\phi_k(F(X))\phi_l(G(Y))] = 0$ for all positive integers k and l.

(a) Assume first that F and G are known. This is for example the case in an actuarial setting where X and Y denote residual lifetimes and their distribution functions are available from life tables. Motivated by the above we take as test statistics

$$\mathscr{R}_n(F,G) = \sup\Big\{ \prod_{j=1}^n n\pi_j : \pi \in \mathscr{P}_n,\ \sum_{j=1}^n \pi_j \phi_k(F(X_j))\phi_l(G(Y_j)) = 0, \quad k,l = 1,\dots,r \Big\}.$$

Under the null hypothesis one has $P(-2\log \mathscr{R}_n(F,G) > \chi^2_{1-\alpha}(r^2)) \to \alpha$ for every $0 < \alpha < 1$ as r and n tend to infinity and r^6/n tends to zero. Here we are in the framework of Hjort, McKeague and Van Keilegom (2009). The above shows that the test $\mathbf{1}[-2\log \mathscr{R}_n(F,G) > \chi^2_{1-\alpha}(r^2)]$ has asymptotic size α.

(b) Now assume that F and G are unknown. In this case we replace both marginal distribution functions by their empirical distribution functions. The resulting test statistic is $\mathscr{R}_n(\mathbb{F},\mathbb{G})$, where $\mathbb{F}$ denotes the empirical distribution based on $X_1,\dots,X_n$ and $\mathbb{G}$ the one based on $Y_1,\dots,Y_n$. We shall show that under the null hypothesis $P(-2\log \mathscr{R}_n(\mathbb{F},\mathbb{G}) > \chi^2_{1-\alpha}(r^2)) \to \alpha$ for every $0 < \alpha < 1$ as r and n tend to infinity and r^6/n tends to zero. Thus the test $\mathbf{1}[-2\log \mathscr{R}_n(\mathbb{F},\mathbb{G}) > \chi^2_{1-\alpha}(r^2)]$ has asymptotic size α.

***Remark* 2.1.** Suppose that (X,Y) form a simple linear homoscedastic regression model, $Y = \beta_1 + \beta_2 X + \varepsilon$, with X and ε independent. We can use the test statistic from case (b) to test the hypothesis whether the slope parameter β_2 is zero. Indeed, $\beta_2 = 0$ is equivalent to the independence of X and Y.

***Remark* 2.2.** The asymptotic distributions of the above tests under contiguous alternatives are linked to non-central chi-square distributions; see Remark 7.3 for details. As the non-centrality parameters are bounded, the local asymptotic power along such a contiguous alternative coincides with the level. Our tests are asymptotically equivalent to Neyman's smooth tests with increasing dimensions. In view of the optimality results of Inglot and Ledwina (1996) for those tests under moderate deviations, we expect similar results for our tests. Of course, this needs to be explored more carefully.

3. Another example and simulations

Let $(X_1,Y_1),\dots,(X_n,Y_n)$ be independent copies of (X,Y), where $Y = \beta_1+\beta_2 X+\varepsilon$, with $E[\varepsilon|X] = 0$, $\sigma^2(X) = E[\varepsilon^2|X]$ bounded and bounded away from zero, and $E[\varepsilon^4] < \infty$. Assume that X has a finite variance and a continuous distribution function G. We are interested in testing whether the regression parameter $\beta = (\beta_1,\beta_2)^\top$ equals some specific value θ. We could proceed as in Owen (1991) and use the test $\delta_0 = \mathbf{1}[-2\log \mathscr{R}_{n0}(\theta) > \chi^2_{1-\alpha}(2)]$ based on the empirical likelihood

$$\mathscr{R}_{n0}(\theta) = \sup\Big\{ \prod_{j=1}^n n\pi_j : \pi \in \mathscr{P}_n,\ \sum_{j=1}^n \pi_j \begin{pmatrix} 1 \\ X_j \end{pmatrix} (Y_j - \theta_1 - \theta_2 X_j) = 0 \Big\}.$$

But this empirical likelihood does not use all the information of the model. Here we have $E[a(X)\varepsilon] = 0$ for every $a \in L_2(G)$. Since G is continuous (but unknown), we work with the empirical likelihood

$$\hat{\mathscr{R}}_{n1}(\theta) = \sup\Big\{ \prod_{j=1}^{n} n\pi_j : \pi \in \mathscr{P}_n,\ \sum_{j=1}^{n} \pi_j u_r(\mathbb{G}(X_j))(Y_j - \theta_1 - \theta_2 X_j) = 0 \Big\}$$

where $u_r = (1, \phi_1, \ldots, \phi_r)^\top$ and $\mathbb{G}$ is the empirical distribution function based on the covariate observations $X_1, \ldots, X_n$. It follows from Corollary 7.6 and Lemma 8.1 below that $P(-2\log \hat{\mathscr{R}}_{n1}(\beta) > \chi^2_{1-\alpha}(1+r)) \to \alpha$ if $r^4 = o(n)$. The resulting test is $\delta_1 = \mathbf{1}[-2\log \hat{\mathscr{R}}_{n1}(\theta) > \chi^2_{1-\alpha}(r+1)]$. Both tests have asymptotic size α.

We performed a small simulation study to compare the procedures. For our simulation we chose $\alpha = .05$ and $n = 100$ and took $\theta = (1,2)^\top$. We modeled the error ε as $s(X)\eta$, with $s(X) = \min(\sqrt{1+X^2}, 100)$ and η independent of X. As distributions for X we chose the exponential distribution with mean 5 (Ex(5)) and the t-distribution with three degrees of freedom (t(3)), while for η we chose the standard normal distribution (N(0,1)) and the double exponential distribution with location 0 and scale .5 (L(0,.5)).

Table 1. Simulated powers of the tests δ_0 and δ_1.

			$t(3)$					$Ex(5)$				
	β_1	β_2	0	2	3	4	5	0	2	3	4	5
$N(0,1)$	0.6	2.3	.71	.88	.86	.85	.84	.38	.37	.39	.40	.41
	0.8	1.5	.68	.82	.84	.83	.83	.95	.99	.99	.99	.99
	1.0	2.0	.13	.09	.10	.12	.13	.12	.07	.09	.12	.14
	1.2	2.2	.37	.42	.43	.43	.44	.51	.54	.52	.50	.52
	1.4	1.7	.71	.88	.87	.86	.86	.37	.34	.37	.40	.44
$L(0,.5)$	0.6	2.3	.89	.98	.99	.98	.98	.61	.64	.68	.71	.74
	0.8	1.5	.84	.96	.98	.98	.98	.93	1.00	1.00	1.00	1.00
	1.0	2.0	.14	.10	.14	.17	.21	.13	.10	.11	.14	.17
	1.2	2.2	.57	.70	.70	.70	.74	.68	.84	.84	.82	.83
	1.4	1.7	.89	.99	.99	.99	.99	.62	.67	.72	.73	.76

Table 1 reports simulated powers of the tests δ_0 and δ_1 (with several choices of r) and for some values of θ. The reported values are based on 1000 repetitions. The column labeled 0 corresponds to Owen's test δ_0, while the columns labeled 2,3,4,5 correspond to our tests δ_1 with $r = 2, 3, 4, 5$ respectively. Clearly our new test is more powerful than the traditional test. The values in the rows corresponding to the parameter values $(1.0, 2.0)$ are the observed significance levels of the nominal significance level .05. Our new test overall has closer observed significance levels than the traditional one except for $r = 5$.

4. Notation

In this section we introduce some of the notation we use throughout. We write $|A|$ for the euclidean norm and $|A|_o$ for the operator (or spectral) norm of a matrix A which are defined by

$$|A|^2 = \operatorname{trace}(A^\top A) = \sum_{i,j} A_{ij}^2 \quad \text{and} \quad |A|_o = \sup_{|u|=1} |Au| = \sup_{|u|=1} (u^\top A^\top A u)^{1/2}.$$

In other words, the squared euclidean norm $|A|^2$ equals the sum of the eigen values of $A^\top A$, while the squared operator norm $|A|_o^2$ equals the largest eigen value of $A^\top A$. Consequently, the inequality $|A|_o \le |A|$ holds. Thus we have

$$|Ax| \le |A|_o |x| \le |A||x|$$

for compatible vectors x. We should also point out the identity

$$|A|_o = \sup_{|u|=1} \sup_{|v|=1} u^\top A v.$$

If A is a nonnegative definite symmetric matrix, this simplifies to

$$|A|_o = \sup_{|u|=1} u^\top A u.$$

Using this and the Cauchy-Schwarz inequality we obtain

$$\Big| \int f g^\top \, d\mu \Big|_o^2 \le \Big| \int f f^\top \, d\mu \Big|_o \Big| \int g g^\top \, d\mu \Big|_o, \tag{4.1}$$

$$\Big| \int f f^\top \, d\mu \Big|_o \le \int |f|^2 \, d\mu, \tag{4.2}$$

whenever μ is a measure and f and g are measurable functions into $\mathbb{R}^s$ and $\mathbb{R}^t$ such that $\int |f|^2 \, d\mu$ and $\int |g|^2 \, d\mu$ are finite. As a special case we derive the inequality

$$|S_{x+y} - S_x|_o \le |S_y|_o + 2|S_x|_o^{1/2} |S_y|_o^{1/2}$$

and therefore

$$|S_{x+y} - S_x|_o \le \frac{1}{n} \sum_{j=1}^n |y_i|^2 + 2|S_x|_o^{1/2} \Big(\frac{1}{n} \sum_{j=1}^n |y_i|^2 \Big)^{1/2} \tag{4.3}$$

with

$$S_{x+y} = \frac{1}{n} \sum_{j=1}^n (x_j + y_j)(x_j + y_j)^\top, \quad S_x = \frac{1}{n} \sum_{j=1}^n x_j x_j^\top, \quad S_y = \frac{1}{n} \sum_{j=1}^n y_j y_j^\top$$

for vectors $x_1, y_1, \dots, x_n, y_n$ of the same dimension.

5. A maximization problem

Let $x_1, ..., x_n$ be m-dimensional vectors. Set $x_* = \max_{1\le j\le n} |x_j|$,

$$\bar{x} = \frac{1}{n}\sum_{j=1}^n x_j, \quad S = \frac{1}{n}\sum_{j=1}^n x_j x_j^\top, \quad x^{(\nu)} = \sup_{|u|=1}\Big|\frac{1}{n}\sum_{j=1}^n (u^\top x_j)^\nu\Big|, \quad \nu = 3, 4,$$

and let λ and Λ denote the smallest and largest eigen value of the matrix S,

$$\lambda = \inf_{|u|=1} u^\top S u \quad \text{and} \quad \Lambda = \sup_{|u|=1} u^\top S u.$$

Using Lagrange multipliers, Owen (1988,2001) obtained the identity

$$\mathscr{R} = \sup\Big\{\prod_{j=1}^n n\pi_j : \pi \in \mathscr{P}_n,\ \sum_{j=1}^n \pi_j x_j = 0\Big\} = \prod_{j=1}^n \frac{1}{1+\zeta^\top x_j}$$

if there exists a ζ in $\mathbb{R}^m$ such that $1 + \zeta^\top x_j > 0$, $j = 1, \dots, n$, and

$$\sum_{j=1}^n \frac{x_j}{1+\zeta^\top x_j} = 0. \tag{5.1}$$

He also showed that such a vector ζ exists and is unique if (i) the origin is an interior point of the convex hull of $x_1, \dots, x_n$ and (ii) the matrix S is invertible. Let us now show that the inequality $\lambda > 5x_*|\bar{x}|$ implies these two conditions. Indeed, the matrix S is then positive definite and hence invertible as its smallest eigen value λ is positive. To show (i) we will rely on the following lemma.

Lemma 5.1. *A random variable Y with $E[Y] = 0$ and $P(|Y| \le K) = 1$ for some positive K obeys the inequality*

$$P(Y > a) \ge \frac{E[Y^2] - 2Ka}{2K^2}, \quad 0 \le a < K.$$

Proof. Fix a in $[0, K)$. By the properties of Y, we obtain $2K^2 P(Y > a) \ge 2KE[Y\mathbf{1}[Y > a]] \ge 2KE[Y\mathbf{1}[Y > 0]] - 2Ka$ and $2KE[Y\mathbf{1}[Y > 0]] = KE[|Y|] \ge E[Y^2]$. □

The origin is an interior point of the convex hull of $x_1, \dots, x_n$ if for every unit vector $u \in \mathbb{R}^m$ there is at least one $j \in \{1, \dots, n\}$ such that $u^\top x_j > 0$. This latter condition is equivalent to

$$N = \inf_{|u|=1} \sum_{j=1}^n \mathbf{1}[u^\top x_j > 0] \ge 1.$$

For a unit vector u, we have $-u^\top \bar{x} \le |\bar{x}|$ and thus

$$\sum_{j=1}^n \mathbf{1}[u^\top x_j > 0] \ge \sum_{j=1}^n \mathbf{1}[u^\top (x_j - \bar{x}) > |\bar{x}|] = N(u).$$

It follows from the triangle inequality that $|x_j - \bar{x}| \le |x_j| + |\bar{x}| \le 2x_*$ for $j = 1, \dots, n$. Note that x_* is positive if S is positive definite. Thus Lemma 5.1 yields the lower bound $N(u)/n \ge (\sigma^2(u) - 4x_*|\bar{x}|)/(8x_*^2)$ with

$$\sigma^2(u) = \frac{1}{n}\sum_{j=1}^n (u^\top(x_j - \bar{x}))^2 = u^T S u - (u^\top \bar{x})^2 \ge \lambda - |\bar{x}|^2 \ge \lambda - x_*|\bar{x}|.$$

Thus we have $N \ge n(\lambda - 5|\bar{x}|x_*)/(8x_*^2)$. This shows that the inequality $\lambda > 5|\bar{x}|x_*$ implies $N \ge 1$ and hence the desired condition (i).

Assume now that the inequality $\lambda > 5x_*|\bar{x}|$ holds. We proceed as on page 220 of Owen (2001). Let u be a unit vector such that $\zeta = |\zeta|u$. Then we have the identity

$$0 = \frac{1}{n}\sum_{j=1}^n \frac{u^\top x_j(1 + \zeta^\top x_j - \zeta^\top x_j)}{1 + \zeta^\top x_j} = u^\top \bar{x} - |\zeta|\frac{1}{n}\sum_{j=1}^n \frac{(u^\top x_j)^2}{1 + \zeta^\top x_j}$$

and the inequality

$$\lambda \le u^\top S u = \frac{1}{n}\sum_{j=1}^n (u^\top x_j)^2 \le \frac{1}{n}\sum_{j=1}^n \frac{(u^\top x_j)^2(1 + |\zeta|x_*)}{1 + \zeta^\top x_j}.$$

Consequently, we find $\lambda|\zeta| \le (1 + |\zeta|x_*)u^\top \bar{x} \le (1 + |\zeta|x_*)|\bar{x}|$ and obtain the bound

$$|\zeta| \le \frac{|\bar{x}|}{\lambda - |\bar{x}|x_*}. \tag{5.2}$$

From this one immediately derives

$$|\zeta|x_* \le \frac{|\bar{x}|x_*}{\lambda - |\bar{x}|x_*} < \frac{1}{4}, \tag{5.3}$$

$$\max_{1 \le j \le n} \frac{1}{1 + \zeta^\top x_j} \le \frac{1}{1 - |\zeta|x_*} < \frac{4}{3}, \tag{5.4}$$

$$\frac{1}{n}\sum_{j=1}^n (\zeta^\top x_j)^2 = \zeta^\top S \zeta \le \Lambda|\zeta|^2 \le \frac{\Lambda|\bar{x}|^2}{(\lambda - |\bar{x}|x_*)^2}. \tag{5.5}$$

The identity $1/(1+d) - 1 + d = d^2 - d^3/(1+d)$ and (5.4) yield

$$\Big|\frac{1}{n}\sum_{j=1}^n \Big(\frac{r_j}{1 + \zeta^\top x_j} - r_j + r_j x_j^\top \zeta\Big)\Big| \le \Big|\frac{1}{n}\sum_{j=1}^n r_j(\zeta^\top x_j)^2\Big| + \frac{4}{3}\frac{1}{n}\sum_{j=1}^n |r_j|\,|\zeta^\top x_j|^3$$

for vectors $r_1, \dots, r_n$ of the same dimension. Taking $r_j = S^{-1}x_j$, we derive with the help of (5.1)

$$|\zeta - S^{-1}\bar{x}| \le \Big|\frac{1}{n}\sum_{j=1}^n S^{-1}x_j(\zeta^\top x_j)^2\Big| + \frac{4}{3}\frac{1}{n}\sum_{j=1}^n |S^{-1}x_j||\zeta^\top x_j|^3.$$

Using $|x| = \sup_{|v|=1} v^\top x$, the Cauchy-Schwarz inequality, (5.3) and (5.5) we bound the square of the first summand of the right-hand side by

$$\frac{1}{n}\sum_{j=1}^{n}(\zeta^\top x_j)^4 \sup_{|v|=1} v^\top S^{-1} v \le \frac{1}{\lambda}|\zeta|^4 x^{(4)}$$

and the square of the second summand by

$$\frac{16}{9\lambda^2} x_*^2 \zeta^\top S\zeta \frac{1}{n}\sum_{j=1}^{n}(\zeta^\top x_j)^4 \le \frac{\Lambda}{9\lambda^2}|\zeta|^4 x^{(4)}.$$

Combining the above we obtain

$$|\zeta - S^{-1}\bar{x}|^2 \le 2\Big(\frac{1}{\lambda} + \frac{\Lambda}{9\lambda^2}\Big)|\zeta|^4 x^{(4)}. \tag{5.6}$$

Using the inequality $|2\log(1+t) - 2t + t^2 - 2t^3/3| \le |t|^4/(2(1-|t|)^4)$ valid for $|t| < 1$, and then (5.4) we derive

$$\Big|2\sum_{j=1}^{n}\log(1+\zeta^\top x_j) - 2n\zeta^\top\bar{x} + n\zeta^\top S\zeta\Big| \le \frac{2}{3}\Big|\sum_{j=1}^{n}(\zeta^\top x_j)^3\Big| + \frac{1}{2}\Big(\frac{4}{3}\Big)^4\sum_{j=1}^{n}|\zeta^\top x_j|^4.$$

With $\Delta = \zeta - S^{-1}\bar{x}$, we can write $\zeta^\top S\zeta = \zeta^\top\bar{x} + \zeta^\top S\Delta$ and $\zeta^\top\bar{x} = \bar{x}^\top S^{-1}\bar{x} + \Delta^\top\bar{x}$, and obtain the identity $2\zeta^\top\bar{x} - \zeta^\top S\zeta = \bar{x}S^{-1}\bar{x} - \Delta^\top S\Delta$. Using this and (5.6) we arrive at the bound

$$\Big|2\sum_{j=1}^{n}\log(1+\zeta^\top x_j) - n\bar{x}^\top S^{-1}\bar{x}\Big| \le n|\zeta|^3 x^{(3)} + n\Big(\frac{16}{9} + \frac{2\Lambda}{\lambda} + \frac{2\Lambda^2}{9\lambda^2}\Big)|\zeta|^4 x^{(4)}.$$

In view of (5.2) and $\Lambda \ge \lambda$, this becomes

$$\Big|2\sum_{j=1}^{n}\log(1+\zeta^\top x_j) - n\bar{x}^\top S^{-1}\bar{x}\Big| \le \frac{n|\bar{x}|^3 x^{(3)}}{(\lambda - |\bar{x}|x_*)^3} + \frac{\Lambda^2}{\lambda^2}\frac{4n|\bar{x}|^4 x^{(4)}}{(\lambda - |\bar{x}|x_*)^4}. \tag{5.7}$$

If we bound $x^{(3)}$ by $x_*\Lambda$ and $x^{(4)}$ by $x_*^2\Lambda$ and use (5.3) we obtain the bound

$$\Big|2\sum_{j=1}^{n}\log(1+\zeta^\top x_j) - n\bar{x}^\top S^{-1}\bar{x}\Big| \le \Big(\Lambda + \frac{\Lambda^3}{\lambda^2}\Big)\frac{nx_*|\bar{x}|^3}{(\lambda - |\bar{x}|x_*)^3}. \tag{5.8}$$

Thus we have proved the following result.

Lemma 5.2. *The inequality $\lambda > 5|\bar{x}|x_*$ implies that there is a unique ζ in $\mathbb{R}^m$ satisfying $1+\zeta^\top x_j > 0$, $j = 1,\dots,n$, and (5.1) to (5.8).*

6. Applications with random vectors

We shall now discuss implications of Lemma 5.2 to the case when the vectors x_j are replaced by random vectors. We are interested in the case when the dimension of the random vectors increases with n.

Let $T_{n1}, \dots, T_{nn}$ be m_n-dimensional random vectors. With these random vectors we associate the empirical likelihood

$$\mathscr{R}_n = \sup\Big\{ \prod_{j=1}^n n\pi_j : \pi \in \mathscr{P}_n,\ \sum_{j=1}^n \pi_j T_{nj} = 0 \Big\}.$$

To study the asymptotic behavior of $\mathscr{R}_n$ we introduce

$$T_n^* = \max_{1\le j\le n} |T_{nj}|, \quad \bar{T}_n = \frac{1}{n}\sum_{j=1}^n T_{nj}, \quad T_n^{(\nu)} = \sup_{|u|=1} \Big| \frac{1}{n}\sum_{j=1}^n (u^\top T_{nj})^\nu \Big|,$$

and the matrix

$$S_n = \frac{1}{n}\sum_{j=1}^n T_{nj} T_{nj}^\top,$$

and let λ_n and Λ_n denote the smallest and largest eigen values of S_n,

$$\lambda_n = \inf_{|u|=1} u^\top S_n u \quad \text{and} \quad \Lambda_n = \sup_{|u|=1} u^\top S_n u.$$

We say a sequence W_n of $m_n \times m_n$ dispersion matrices is *regular* if the following condition holds,

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

We impose the following conditions.

(A1) $m_n^{1/2} T_n^* = o_p(n^{1/2})$.
(A2) $n|\bar{T}_n|^2 = O_p(m_n)$.
(A3) There is a regular sequence of dispersion matrices W_n such that $|S_n - W_n|_o = o_p(m_n^{-1/2})$.
(A4) $m_n T_n^{(3)} = o_p(n^{1/2})$ and $m_n^{3/2} T_n^{(4)} = o_p(n)$.

The first two conditions imply $T_n^* |\bar{T}_n| = o_p(1)$, the third condition implies that there are positive numbers $a < b$ such that $P(a \le \lambda_n \le \Lambda_n \le b) \to 1$. Thus all three conditions imply that the probability of the event $\{\lambda_n > 5T_n^*|\bar{T}_n|\}$ tends to one. Consequently, by Lemma 5.2, there exists an m_n-dimensional random vector $\hat{\zeta}_n$ which is uniquely determined on this event by the properties $1 + \hat{\zeta}_n^\top T_{nj} > 0$, $j = 1, \dots, n$, and

$$\frac{1}{n}\sum_{j=1}^n \frac{T_{nj}}{1 + \hat{\zeta}_n^\top T_{nj}} = 0. \tag{6.1}$$

On this event we have $-2\log\mathscr{R}_n = 2\sum_{j=1}^n \log(1+\hat\zeta_n^\top T_{nj})$. It follows from (A3) that S_n is invertible except on an event whose probability tends to zero. It follows from (A2) and (A4) that

$$n|\bar T_n|^3 T_n^{(3)} = o_p(m_n^{1/2}) \quad \text{and} \quad n|\bar T_n|^4 T_n^{(4)} = o_p(m_n^{1/2}).$$

Thus, under (A1)–(A4), the following expansion follows from (5.7)

$$-2\log\mathscr{R}_n = n\bar T_n^\top S_n^{-1}\bar T_n + o_p(m_n^{1/2}). \tag{6.2}$$

From (A3) we can also derive the rate $|S_n^{-1} - W_n^{-1}|_o = o_p(m_n^{-1/2})$. Thus, if (A1)–(A4) hold, then (6.2) holds with S_n replaced by W_n,

$$-2\log\mathscr{R}_n = n\bar T_n^\top W_n^{-1}\bar T_n + o_p(m_n^{1/2}). \tag{6.3}$$

In view of the inequalities $T_n^{(3)} \le \Lambda_n T_n^*$ and $T_n^{(4)} \le \Lambda_n (T_n^*)^2$, a sufficient condition for (A1) and (A4) is given by

$$m_n T_n^* = o_p(n^{1/2}). \tag{B1}$$

In view of the bound $(T_n^{(3)})^2 \le \Lambda_n T_n^{(4)}$, which is a consequence of the Cauchy-Schwarz inequality, a sufficient condition for (A4) is given by

$$m_n^2 T_n^{(4)} = o_p(n). \tag{B2}$$

We first treat the case when the dimension m_n does not increase with n. In this case (B1) and (A2) are implied by $T_n^* = o_p(n^{1/2})$ and $\bar T_n = O_p(n^{-1/2})$, and (A3) is implied by the condition: $S_n = W + o_p(1)$ for some positive definite matrix W. Thus we have the following result.

Theorem 6.1. *Let $m_n = m$ for all n. Suppose*

$$T_n^* = o_p(n^{1/2}), \quad n^{1/2}\bar T_n \implies N(0,V) \quad \text{and} \quad S_n = W + o_p(1),$$

for dispersion matrices V and W, with W positive definite. Then $-2\log\mathscr{R}_n$ converges in distribution to $Z^\top V^{1/2} W^{-1} V^{1/2} Z$, where the m-dimensional random vector Z is standard normal. For $V = W$ the limiting distribution is a chi-square distribution with m degrees of freedom.

If we replace $n^{1/2}\bar T_n \implies N(0,V)$ by $n^{1/2}\bar T_n \implies U$ for some random variable U, then the conclusion becomes $-2\log\mathscr{R}_n$ converges in distribution to $U^\top W^{-1} U$. This version of the theorem yields Theorem 2.1 of Hjort *et al* (2009) without their (A0).

Theorem 6.1 does not require the independence of the random vectors $T_{n,1},\dots,T_{n,n}$. This is important when dealing with estimated constraint functions as we shall see below.

Suppose the condition in the theorem hold with $V = W$. Under a contiguous alternative, one typically has $n^{1/2}\bar T_n \implies N(\mu, V)$ for some μ different from zero, but retains the other conditions. In this case, $-2\log\mathscr{R}_n$ has a limiting chi-square distribution with m degrees of freedom and non-centrality parameter $|V^{-1/2}\mu|$.

Let us address some applications of Theorem 6.1. For this discussion we let $Z_1, \dots, Z_n$ be independent copies of a k-dimensional random vector Z with distribution Q and let w be a measurable function from $\mathbb{R}^k$ into $\mathbb{R}^m$ such that $E[w(Z)] = \int w\,dQ = 0$ and $W = E[w(Z)w^\top(Z)] = \int ww^\top\,dQ$ is positive definite. Let us first look at the empirical likelihood

$$\mathscr{R}_{n1} = \sup\Big\{\prod_{j=1}^n n\pi_j : \pi \in \mathscr{P}_n,\ \sum_{j=1}^n \pi_j w(Z_j) = 0\Big\}.$$

It follows from Owen that $-2\log\mathscr{R}_{n1}$ has a limiting chi-square distribution with m degrees of freedom. This also follows from Theorem 6.1 applied with $T_{nj} = w(Z_j)$. Indeed, the first condition follows from the inequality

$$P(\max_{1\le j\le n}|w(Z_j)| > \epsilon n^{1/2}) \le \frac{1}{\epsilon^2}E[|w(Z)|^2\mathbf{1}[|w(Z)| > \epsilon n^{1/2}] \tag{6.4}$$

and the Lebesgue dominated convergence theorem; the central limit theorem yields the second condition with $V = W$; the third condition

$$\frac{1}{n}\sum_{j=1}^n w(Z_j)w^\top(Z_j) = W + o_p(1) \tag{6.5}$$

follows from the weak law of large numbers. This shows that Owen's result is a special case of our result.

Now consider the empirical likelihood

$$\hat{\mathscr{R}}_{n1} = \sup\Big\{\prod_{j=1}^n n\pi_j : \pi \in \mathscr{P}_n,\ \sum_{j=1}^n \pi_j \hat{w}(Z_j) = 0\Big\},$$

where $\hat{w}$ is an estimator of w based on the observations $Z_1, \dots, Z_n$ which is consistent in the following sense,

$$\frac{1}{n}\sum_{j=1}^n |\hat{w}(Z_j) - w(Z_j)|^2 = o_p(1). \tag{6.6}$$

Then $-2\log\hat{\mathscr{R}}_{n1}$ has a limiting chi-square distribution with m degrees of freedom if also

$$n^{-1/2}\sum_{j=1}^n \hat{w}(Z_j) = n^{-1/2}\sum_{j=1}^n w(Z_j) + o_p(1) \tag{6.7}$$

holds. To see this, we verify the assumptions of Theorem 6.1 with $T_{nj} = \hat{w}(Z_j)$. The first condition follows from (6.4), (6.6) and the inequality

$$T_n^* \le \max_{1\le j\le n}|w(Z_j)| + \Big(\sum_{j=1}^n |\hat{w}(Z_j) - w(Z_j)|^2\Big)^{1/2}.$$

The central limit theorem, Slutsky's theorem and (6.7) yield the second condition with $V = W$. The third condition follows from (6.5), (6.6) and the inequality (4.3).

The requirement (6.7) is rather strong. One often only derives

$$n^{-1/2}\sum_{j=1}^{n}\hat{w}(Z_j) = n^{-1/2}\sum_{j=1}^{n} v(Z_j) + o_p(1) \tag{6.8}$$

for some function v satisfying $E[v(Z)] = 0$ and $E[|v(Z)|^2] < \infty$. Under (6.6) and (6.8), $-2\log\hat{\mathscr{R}}_{n1}$ has limiting distribution as given in Theorem 6.1 with V the dispersion matrix of $v(Z)$. This follows from Theorem 6.1 whose assumptions are now verified as above.

In situations when $w(Z) = u(Z,\eta)$ for some q-dimensional nuisance parameter η and $\hat{w}(Z) = u(Z,\hat{\eta})$ for some estimator $\hat{\eta}$ of η, one typically has $v(Z) = w(Z) + D\psi(Z)$, where the $m \times q$ matrix D is the derivative of the map $t \mapsto E[u(Z,\eta+t)]$ at $t = 0$, and ψ is the influence function of $\hat{\eta}$.

We now address the case when m_n increases with the sample size.

Theorem 6.2. *Let (A1)–(A4) hold. Suppose that m_n increases with n to infinity and that there are $m_n \times m_n$ dispersion matrices V_n such that $m_n/\mathrm{trace}(V_n^2) = O(1)$ and*

$$\Big(n\bar{T}_n^\top W_n^{-1}\bar{T}_n - \mathrm{trace}(V_n)\Big)\Big/\sqrt{2\mathrm{trace}(V_n^2)} \implies N(0,1). \tag{6.9}$$

Then we have

$$\Big(-2\log\mathscr{R}_n - \mathrm{trace}(V_n)\Big)\Big/\sqrt{2\mathrm{trace}(V_n^2)} \implies N(0,1). \tag{6.10}$$

Proof. We have already seen that (A1)–(A4) imply (6.3). It follows from (6.3) and $m_n/\mathrm{trace}(V_n^2) = O(1)$ that the difference of the left-hand sides of (6.9) and (6.10) converge to zero in probability. Thus the desired (6.10) follows from (6.9) and Slutsky's Theorem. □

Of special interest is the case when V_n is the $m_n \times m_n$ identity matrix I_{m_n}. Then $\mathrm{trace}(V_n) = \mathrm{trace}(V_n^2) = m_n$ and (6.10) simplifies to (1.2). Sufficient conditions for (6.9) are given by Peng and Schick (2012).

7. Main results

In this section we assume that $(\mathscr{Z},\mathscr{S})$ is a measurable space, that $Z_1,\dots,Z_n$ are independent copies of the $\mathscr{Z}$-valued random variable Z with distribution Q, and that m_n is a positive integer that tends to infinity with n. We let w_n denote a measurable function from $\mathscr{Z}$ to $\mathbb{R}^{m_n}$ such that $\int w_n\,dQ = 0$ and $\int |w_n|^2\,dQ$ is finite.

We first study

$$\mathscr{R}_n = \sup\Big\{\prod_{j=1}^{n} n\pi_j : \pi \in \mathscr{P}_n,\ \sum_{j=1}^{n}\pi_j w_n(Z_j) = 0\Big\}.$$

Our goal is to show (1.2). To this end we set

$$\bar{w}_n = \frac{1}{n}\sum_{j=1}^{n} w_n(Z_j), \quad \bar{W}_n = \frac{1}{n}\sum_{j=1}^{n} w_n(Z_j)w_n^\top(Z_j), \quad W_n = \int w_n w_n^\top \, dQ$$

and introduce the following condition.

(C1) The sequence W_n is regular.

Motivated by the results in Peng and Schick (2012) we call a sequence v_n of measurable functions from $\mathscr{Z}$ to $\mathbb{R}$ *Lindeberg* if

$$\int |v_n|^2 \mathbf{1}[|v_n| > \epsilon\sqrt{n}]\, dQ \to 0, \quad \epsilon > 0. \tag{7.1}$$

The following are easy to check. If the sequences u_n and v_n are Lindeberg, so are the sequences $\max\{|u_n|, |v_n|\}$ and $u_n + v_n$. If the sequence v_n is Lindeberg and $|u_n| \le |v_n|$, then the sequence u_n is also Lindeberg. We also need the following properties.

(L1) If v_n is Lindeberg, then one has the rate $\max_{1\le j\le n} |v_n(Z_j)| = o_p(n^{1/2})$.
(L2) If $\int |v_n|^r \, dQ = o(n^{r/2-1})$ for some $r > 2$, then v_n is Lindeberg.

The first statement follows from an inequality similar to (6.4), the second from Remark 1 in Peng and Schick (2012).

To show (1.2) we apply Theorem 6.2 with $T_{nj} = w_n(Z_j)$. In the presence of (C1), the conditions (6.9) and (A1)–(A4) of this theorem are implied by

$$(n\bar{w}_n^\top W_n^{-1}\bar{w}_n - m_n)/\sqrt{2m_n} \implies N(0,1), \tag{D0}$$

$$\max_{1\le j\le n} m_n^{1/2}|w_n(Z_j)| = o_p(n^{1/2}), \tag{D1}$$

$$n|\bar{w}_n|^2 = O_p(m_n), \tag{D2}$$

$$|\bar{W}_n - W_n|_o = o_p(m_n^{-1/2}), \tag{D3}$$

$$\sup_{|u|=1} \frac{m_n^2}{n}\sum_{j=1}^{n} |u^\top w_n(Z_j)|^4 = o_p(n). \tag{D4}$$

By part (c) of Corollary 3 in Peng and Schick (2012), (D0) follows if the function $|W_n^{-1/2}w_n|$ is Lindeberg. In the presence of (C1), the latter condition is equivalent to $|w_n|$ being Lindeberg. By (L1), a sufficient condition for (D1) is that $m_n^{1/2}|w_n|$ is Lindeberg. It follows from (C1) that $\operatorname{trace}(W_n) \le Bm_n$ for some constant B. Thus (C1) implies $E[n|\bar{w}_n|^2] = \operatorname{trace}(W_n) = O(m_n)$ and hence (D2). In view of (C1), a sufficient condition for (D3) is that $m_n|w_n|$ is Lindeberg. To see this fix $\epsilon > 0$ and let $\bar{W}_{n,1}$ and $\bar{W}_{n,2}$ be the matrices obtained by replacing in the definition of $\bar{W}_n$ the function w_n by $v_n = w_n\mathbf{1}[|m_n w_n| \le \epsilon\sqrt{n}]$ and $w_n - v_n = w_n\mathbf{1}[|m_n w_n| > \epsilon\sqrt{n}]$, respectively. Then we find

$$nE[|\bar{W}_{n,1} - E[\bar{W}_{n,1}]|^2] \le E[|v_n|^4(Z)] \le \frac{\epsilon^2 n}{m_n^2}E[|w_n|^2(Z)] \le \frac{\epsilon^2 nBm_n}{m_n^2},$$

$$P(\bar{W}_{n,2} \neq 0) \leq P(\max_{1\leq j\leq n} |m_n w_n(Z_j)| > \epsilon\sqrt{n}) \to 0$$

and using (4.2)

$$|E[\bar{W}_{n,2}]|_o \leq E[|w_n|^2(Z)\mathbf{1}[|m_n w_n(Z)| > \epsilon\sqrt{n}]] = o(m_n^{-2}).$$

The above inequalities show that (C1) and $m_n|w_n|$ is Lindeberg imply statement (D3). The latter condition also implies (B1) and hence (D1) and (D4), the latter in the presence of (C1). Thus we have the following result.

Theorem 7.1. *Suppose (C1) holds and the sequence $m_n|w_n|$ is Lindeberg. Then (1.2) holds as m_n tends to infinity with n.*

From this, simple calculations and the property (L2) we immediately derive the following corollaries.

Corollary 7.1. *Suppose (C1) holds and $|w_n| \leq \sqrt{m_n}B$ for some constant B. Then (1.2) holds if $m_n^3 = o(n)$.*

Corollary 7.2. *Suppose (C1) holds and $\int |w_n|^r \, dQ = O(m_n^{r/2})$ for some $r > 2$. Then (1.2) holds if $m_n^{3r/(r-2)} = o(n)$.*

These two corollaries give the conclusions in Theorem 4.1 in Hjort *et al* (2009) under slightly weaker conditions in the case of Corollary 7.2. We now present some additional results that allow for larger m_n if r is small. For example, if $r = 4$, Corollary 7.2 requires $m_n^6 = o(n)$, while Theorem 7.2 below allows $m_n^4 = o(n)$. For $r = 3$, Corollary 7.2 requires $m_n^9 = o(n)$, while Theorem 7.3 below allows $m_n^6 = o(n)$.

Theorem 7.2. *Suppose (C1) holds and $\int |w_n|^4 \, dQ = O(m_n^2)$. Then (1.2) holds if $m_n^4 = o(n)$.*

Proof. Using (L2) and $m_n^4 = o(n)$ we derive that $m_n^{1/2}|w_n|$ is Lindeberg. This latter condition and (C1) imply (D1)–(D2) as shown prior to Theorem 7.1. Next we calculate $nE[|\bar{W}_n - W_n|^2] \leq E[|w_n|^4(Z)] = O(m_n^2)$. This yields (D3) in view of $|\bar{W}_n - W_n|_o \leq |\bar{W}_n - W_n| = O_p(m_n/\sqrt{n})$ and $m_n^4 = o(n)$. Finally we have (D4) as the left-hand side of (D4) is bounded by

$$\frac{m_n^2}{n} \sum_{j=1}^{n} |w(Z_j)|^4 = O_p(m_n^4) = o_p(n).$$

Thus (D0)–(D4) hold and we obtain the desired result from Theorem 6.2. □

Theorem 7.3. *Suppose (C1) holds and $\int |w_n|^r \, dQ = O(m_n^{r/2})$ for some $2 < r < 4$. Then (1.2) holds if $m_n^{2r/(r-2)} = o(n)$.*

Proof. There is a constant B such that $\int |w_n|^r \, dQ \le B m_n^{r/2}$. In view of (L2) and the properties of m_n we derive that $m_n^{1/2}|w_n|$ is Lindeberg. This condition and (C1) imply (D0)–(D2). It follows from (D1), the moment condition on w_n, and the properties of m_n that

$$\begin{aligned}\frac{m_n^2}{n}\sum_{j=1}^n |w(Z_j)|^4 &\le \frac{m_n^2}{n}\sum_{j=1}^n |w(Z_j)|^r \max_{1\le j\le n}|w_n(Z_j)|^{4-r}\\ &= o_p(m_n^2 m_n^{r/2}(n/m_n)^{(4-r)/2}) = o_p(m_n^r n^{(4-r)/2}) = o_p(n).\end{aligned}$$

This establishes (D4). Finally (D3) follows as we have $|\bar W_n - W_n|_o = o_p(m_n^{-1})$. To prove the latter we mimic the argument prior to Theorem 7.1 used to verify (D3) if $m_n|w_n|$ is Lindeberg. But now $|w_n|\mathbf{1}[m_n^{1/2}|w_n| \le \sqrt{n}]$ plays the role of v_n. For the corresponding matrices $\bar W_{n1}$ and $\bar W_{n2}$ we have

$$m_n^2 E[|\bar W_{n1} - E[\bar W_{n1}|^2] \le \frac{m_n^2}{n}\Big(\frac{n}{m_n}\Big)^{(4-r)/2} B m_n^{r/2} \le \frac{B m_n^r}{n^{r/2-1}} \to 0,$$

$$P(\bar W_{n2} \ne 0) \le P(\max_{1\le j\le n} m_n^{1/2}|w_n(Z_j)| > n^{1/2}) \to 0,$$

$$m_n|E[\bar W_{n,2}]|_o \le \int \frac{m_n^{r/2}|w_n|^r}{n^{(r-2)/2}}\, dQ \le \frac{B m_n^r}{n^{r/2-1}} \to 0.$$

Consequently (D0)–(D4) hold and the desired result follows. □

Now we study

$$\hat{\mathscr{R}}_n = \sup\Big\{\prod_{j=1}^n n\pi_j : \pi \in \mathscr{P}_n,\ \frac{1}{n}\sum_{j=1}^n \pi_j \hat w_n(Z_j) = 0\Big\},$$

where $\hat w_n$ is an estimator of w_n. Let us set

$$\hat W_n = \frac{1}{n}\sum_{j=1}^n \hat w_n(Z_j)\hat w_n^\top(Z_j).$$

Theorem 7.4. *Suppose (C1) holds and assume we have the expansions*

$$m_n \max_{1\le j\le n}|\hat w_n(Z_j)| = o_p(n^{1/2}), \tag{7.2}$$

$$|\hat W_n - W_n|_o = o_p(m_n^{-1/2}) \tag{7.3}$$

$$\frac{1}{n}\sum_{j=1}^n \hat w_n(Z_j) = \frac{1}{n}\sum_{j=1}^n v_n(Z_j) + o_p(n^{-1/2}) \tag{7.4}$$

for some measurable function v_n from S into $\mathbb{R}^{m_n}$ such that $\int v_n\,dQ = 0$ and $|v_n|$ is Lindeberg. Furthermore assume that the dispersion matrix

$$U_n = W_n^{-1/2} \int v_n v_n^\top \, dQ W_n^{-1/2}$$

of $W_n^{-1/2} v_n(Z)$ satisfies $|U_n|_o = O(1)$ and $m_n/\mathrm{trace}(U_n^2)$ is bounded. Then, as m_n tends to infinity with n, $(-2\log\hat{\mathscr{R}}_n - \mathrm{trace}(U_n))/\sqrt{2\mathrm{trace}(U_n^2)}$ is asymptotically standard normal.

Proof. Set $\xi_{nj} = W_n^{-1/2} v_n(Z_j)$, and introduce the averages $\bar{v}_n = \frac{1}{n}\sum_{j=1}^n v_n(Z_j)$ and $\bar{T}_n = \frac{1}{n}\sum_{j=1}^n \hat{w}_n(Z_j)$. It follows from (C1) that $|W_n^{1/2}|_o + |W_n^{-1/2}|_o = O(1)$. Using this and the Lindeberg property of $|v_n|$ we derive

$$L_n(\epsilon) = E[|\xi_{n,1}|^2 \mathbf{1}[|\xi_{n,1}| > \epsilon\sqrt{n}]] \to 0, \quad \epsilon > 0, \tag{7.5}$$

We have $\mathrm{trace}(U_n)/\mathrm{trace}(U_n^2) \le |U_n|_o m_n/\mathrm{trace}(U_n^2) = O(1)$. From $m_n/\mathrm{trace}(U_n^2) = O(1)$ we conclude $\mathrm{trace}(U_n^2) \to \infty$. Thus Theorem 2 in Peng and Schick (2012) yields that $\big(n\bar{v}_n W_n^{-1}\bar{v}_n - \mathrm{trace}(U_n)\big)/\sqrt{2\mathrm{trace}(U_n^2)}$ is asymptotically standard normal. From this, (C1), $\mathrm{trace}(U_n) = O(m_n)$ and $\mathrm{trace}(U_n^2) \le |U_n|_0^2 m_n$ we conclude $n|\bar{v}_n|^2 = O_p(m_n)$. With the help of (7.4) and the assumption $m_n/\mathrm{trace}(U_n^2) = O(1)$ we then derive $n|\bar{T}_n|^2 = O_p(m_n)$ and that $\big(n\bar{T}_n W_n^{-1}\bar{T}_n - \mathrm{trace}(U_n)\big)/\sqrt{2\mathrm{trace}(U_n^2)}$ is asymptotically standard normal. Thus in view of (B1), conditions (A1)–(A4) hold with $T_{nj} = \hat{w}_n(Z_j)$, and the desired result follows from Theorem 6.2. □

Let us first mention the special case when $v_n = w_n$. In this case U_n equals I_{m_n} and $\mathrm{trace}(U_n) = \mathrm{trace}(U_n^2) = m_n$.

Corollary 7.3. *Suppose (C1), (7.2) and (7.3) hold, $|w_n|$ is Lindeberg, and the following expansion is valid,*

$$\frac{1}{n}\sum_{j=1}^n \hat{w}_n(Z_j) = \frac{1}{n}\sum_{j=1}^n w_n(Z_j) + o_p(n^{-1/2}). \tag{7.6}$$

Then $(-2\log\hat{\mathscr{R}}_n - m_n)/\sqrt{2m_n}$ is asymptotically standard normal.

Next we treat $v_n = w_n - A_n\psi$ with A_n and ψ as in the next condition.

(C2) There is a measurable function ψ from $\mathscr{Z}$ into $\mathbb{R}^q$ satisfying $\int \psi\,dQ = 0$ and $\int \psi\psi^\top\,dQ = I_q$ such that, with $A_n = \int w_n\psi^\top\,dQ$, the expansion

$$\frac{1}{n}\sum_{j=1}^n \hat{w}_n(Z_j) = \frac{1}{n}\sum_{j=1}^n w_n(Z_j) - A_n\psi(Z_j) + o_p(n^{-1/2})$$

and the convergence $\mathrm{trace}(A_n^\top W_n^{-1} A_n) \to q$ hold.

Corollary 7.4. *Suppose (C1), (C2), (7.2) and (7.3) hold, and $|w_n|$ is Lindeberg. Then $(-2\log\hat{\mathscr{R}}_n - m_n + q)/\sqrt{2(m_n-q)}$ is asymptotically standard normal.*

***Remark* 7.1.** Suppose that w_n is the vector formed by the first m_n elements of an orthonormal basis $u_1, u_2, \dots$ for $L_{2,0}(Q)$. Then the ν-th column of the matrix A_n is formed by the first m_n Fourier coefficients of the ν-th component of ψ with respect to this basis. In this case we have the identity

$$\operatorname{trace}(A_n^\top W_n^{-1} A_n) = \operatorname{trace}(A_n^\top A_n) = \sum_{\nu=1}^{q}\sum_{k=1}^{m_n}\Big(\int \psi_\nu u_k\, dQ\Big)^2$$

and obtain under the assumptions $\int \psi\, dQ = 0$ and $\int \psi\psi^\top\, dQ = I_q$ the convergence $\operatorname{trace}(A_n^\top W_n^{-1} A_n) \to \int |\psi|^2\, dQ = q$.

In our goodness-of-fit examples the following condition holds.

(C3) There is a constant B such that $|w_n| \le B\sqrt{m_n}$ and $|\hat{w}_n| \le B\sqrt{m_n}$.

Under this condition, the rate $m_n^3/n \to 0$ implies (7.2), the Lindeberg property of $m_n|w_n|$, and (D3). Sufficient conditions for (7.3) can now be given directly or by verifying

$$|\hat{W}_n - \bar{W}_n|_o = o_p(m_n^{-1/2}). \tag{7.7}$$

In view of the inequality (4.3) a sufficient condition for the latter is

$$D_n = \frac{1}{n}\sum_{j=1}^{n} |\hat{w}_n(Z_j) - w_n(Z_j)|^2 = o_p(m_n^{-1}). \tag{7.8}$$

Thus we have the following results.

Corollary 7.5. *Suppose (C1), (C3), $m_n^3 = o(n)$, and one of (7.3), (7.7), (7.8) hold. Then (i) (7.6) implies that $(-2\log\hat{\mathscr{R}}_n - m_n)/\sqrt{2m_n}$ is asymptotically standard normal, while (ii) (C2) implies that $(-2\log\hat{\mathscr{R}}_n - m_n + q)/\sqrt{2(m_n-q)}$ is asymptotically standard normal.*

***Remark* 7.2.** The conditions in Theorem 7.4 are based on the sufficient condition (B1) for (A1) and (A4). Working with (A1) and (B2) instead, we see that (7.2) can be replaced by the conditions,

$$m_n^{1/2}\max_{1\le j\le n}|\hat{w}_n(Z_j)| = O_p(n^{1/2}) \quad \text{and} \quad \frac{m_n^2}{n}\sum_{j=1}^{n}|\hat{w}_n(Z_j)|^4 = o_p(n).$$

With D_n as in (7.8), we derive the bounds

$$\max_{1\le j\le n}|\hat{w}_n(Z_j)| \le \max_{1\le j\le n}|w_n(Z_j)| + (nD_n)^{1/2}$$
$$\frac{m_n^2}{n}\sum_{j=1}^{n}|\hat{w}_n(Z_j)|^4 \le \frac{8m_n^2}{n}\sum_{j=1}^{n}|w_n(Z_j)|^4 + 8m_n^2 n D_n^2.$$

Here we used that $(a+b)^4 \le 8(a^4+b^4)$ for non-negative a and b. Assume now that $\int |w_n|^4\,dQ = O(m_n^2)$ and that $m_n^4/n \to 0$. Then we have (D1) and (D3) as shown in the proof of Theorem 7.2 and obtain the above two conditions and (7.3) from (7.8).

Corollary 7.6. *Suppose (C1), (7.8), $\int |w_n|^4\,dQ = O(m_n^2)$ and $m_n^4 = o(n)$ hold. Then (i) (7.6) implies that $(-2\log\hat{\mathscr{R}}_n - m_n)/\sqrt{2m_n}$ is asymptotically standard normal, while (ii) (C2) implies that $(-2\log\hat{\mathscr{R}}_n - m_n + q)/\sqrt{2(m_n-q)}$ is asymptotically standard normal.*

***Remark* 7.3.** Let us now describe the behavior of $-2\log\hat{\mathscr{R}}_n$ under a local alternative. For this we follow Remarks 6 and 7 in Peng and Schick (2012). As there let h be a measurable function satisfying $\int h\,dQ = 0$ and $\int h^2\,dQ < \infty$ and let $Q_{n,h}$ be a distribution satisfying

$$\int |n^{1/2}(\sqrt{dQ_{n,h}} - \sqrt{dQ}) - (1/2)h\sqrt{dQ}|^2 \to 0. \tag{7.9}$$

Then the product measures $Q_{n,h}^n$ and Q^n are mutually contiguous. All results in this section obtain the expansion

$$-2\log\hat{\mathscr{R}}_n - |n^{-1/2}\sum_{j=1}^n u_n(Z_j)|^2 = o_p(m_n^{1/2}) \tag{7.10}$$

for some measurable function u_n from $\mathscr{Z}$ into $\mathbb{R}^{m_n}$ with the properties $\int u_n\,dQ = 0$, $\int |u_n|^2\,dQ = O(m_n)$, $|u_n|$ is Lindeberg, and the matrix $U_n = \int u_n u_n^\top\,dQ$ satisfies $|U_n|_o = O(1)$ and $m_n/\mathrm{trace}(U_n^2) = O(1)$. For example, in Theorem 7.4 one has $u_n = W_n^{-1/2}v_n$. By contiguity, one has the expansion (7.10) even if $Z_1,\dots,Z_n$ are independent with distribution $Q_{n,h}$. Under this distributional assumption one has

$$\big(|n^{-1/2}\sum_{j=1}^n u_n(Z_j)|^2 - |\mu_n(h)|^2 - \mathrm{trace}(U_n)\big)\big/\sqrt{2\mathrm{trace}(U_n^2)} \implies N(0,1)$$

with $\mu_n(h) = \int u_n h\,dQ$. Thus, under the local alternative $Q_{n,h}$ one has

$$\big(-2\log\hat{\mathscr{R}}_n - |\mu_n(h)|^2 - \mathrm{trace}(U_n)\big)\big/\sqrt{2\mathrm{trace}(U_n^2)} \implies N(0,1).$$

If $U_n = I_{m_n}$, this simplifies to $(-2\log\hat{\mathscr{R}}_n - |\mu_n(h)|^2 - m_n)/\sqrt{2m_n} \implies N(0,1)$ and may be interpreted as $-2\log\mathscr{R}_n$ being approximately a non-central chi-square random variable with m_n degrees of freedom and non-centrality parameter $|\mu_n(h)|$.

8. Details for the examples

In this section we use the results of the previous section to provide the details for the examples of Sections 2. In all examples, the components of w_n are orthonormal and uniformly bounded, so that (C1) and (C3) hold with $W_n = I_{m_n}$. We begin with a technical lemma.

Lemma 8.1. *Let $(S_1, T_1), \dots, (S_n, T_n)$ be independent copies of the bivariate random vector (S, T), where T has a continuous distribution function H and $E[S|T] = 0$ and $\sigma^2(T) = E[S^2|T]$ is bounded (by say B) and bounded away from zero (by say b), Let $\mathbb{H}$ denote the empirical distribution function based on $T_1, \dots, T_n$. Set $u_r = (1, \phi_1, \dots, \phi_r)^\top$, $D_j = u_r(\mathbb{H}(T_j)) - u_r(H(T_j))$, and $M = E[S^2 u_r(H(T)) u_r^\top(H(T)]$. Then we have the following inequalities*

$$b \le v^\top M v \le B, \quad v \in \mathbb{R}^{1+r}, |v| = 1, \tag{8.1}$$

$$\Big|\frac{1}{n}\sum_{j=1}^n u_r(\mathbb{H}(T_j)) u_r^\top(\mathbb{H}(T_j)) - I_{1+r}\Big|^2 \le \frac{16\pi^2 r^2 (1+r)^2}{n^2} \quad a.s., \tag{8.2}$$

$$\frac{1}{n}\sum_{j=1}^n E[|S_j D_j|^2] \le \frac{1}{n}\sum_{j=1}^n BE[|D_j|^2]\Big] \le \frac{B\pi^2 r^3}{n}, \tag{8.3}$$

$$E\Big[\Big|n^{-1/2}\sum_{j=1}^n S_j D_j\Big|^2\Big] \le \frac{2B\pi^2 r^3}{n}. \tag{8.4}$$

Moreover, if $E[S^4]$ is finite, then we have the bound

$$E[S^4 |u_r(H(T))|^4) \le (1+2r)^2 E[S^4].$$

Proof. The last inequality follows from the bound $|u_r|^2 \le 1 + 2r$. The inequality (8.1) is an easy consequence of $b \le \sigma^2(T) \le B$. Conditioning on $T_1, \dots, T_n$ shows that the left-hand side of (8.4) is bounded by the left-hand side of (8.3) and yields the first inequality in (8.3). Since $|\phi_k'| \le \sqrt{2}\pi k$, we obtain $|D_j|^2 \le 2\pi r^3 (\mathbb{H}(T_j) - H(T_j))^2$. It is easy to check that $E[(\mathbb{H}(T_j) - H(T_j))^2] \le 1/n$. This proves (8.3) and (8.4). Next, we have almost surely,

$$\frac{1}{n}\sum_{j=1}^n u_r(\mathbb{H}(T_j)) u_r^\top(\mathbb{H}(T_j)) = \frac{1}{n}\sum_{j=1}^n u_r(j/n) u_r^\top(j/n)).$$

For a function h defined on $[0,1]$ with Lipschitz constant L, we have

$$\Big|\frac{1}{n}\sum_{j=1}^n h(j/n) - \int_0^1 h(u)\,du\Big| \le \frac{1}{n}\sum_{j=1}^n \sup_{j-1 \le nu \le j} |h(j/n) - h(u)| \le L/n.$$

Since the function $\phi_k \phi_l$ is Lipschitz with Lipschitz constant $2\pi(k+l)$, we derive the desired bound (8.2). □

Details for Example 2. Let $X_1, \dots, X_n$ be independent copies of a random variable X that has distribution function F_θ and density f_θ for some θ in the open subset Θ of $\mathbb{R}^q$. Recall we assumed in Example 2 that the map $\vartheta \mapsto s_\vartheta = \sqrt{f_\vartheta}$ is continuously differentiable in L_2 with derivative $\vartheta \mapsto \dot{s}_\vartheta$ and that the information matrix $J(\vartheta) = 4\int \dot{s}_\vartheta(x)\dot{s}_\vartheta(x)^\top\,dx$ is invertible for each ϑ in Θ. Thus we have

$$\rho(\tau) = \int (s_{\theta+\tau}(x) - s_\theta(x) - \tau^\top \dot{s}_\theta(x))^2\,dx = o(|\tau|^2). \tag{8.5}$$

Recall also that $\dot\ell_\theta = 2\dot s_\theta / s_\theta$ denotes the score function. By the properties of the densities, there is a $\delta > 0$ and a constant K such that

$$\int |f_{\vartheta_1}(x) - f_{\vartheta_2}(x)|\,dx \le K|\vartheta_1 - \vartheta_2|, \quad |\vartheta_1 - \theta| < \delta, |\vartheta_2 - \theta| < \delta. \tag{8.6}$$

As a consequence we have

$$\sup_{x\in\mathbb{R}} |F_{\vartheta_1}(x) - F_{\vartheta_2}(x)| \le K|\vartheta_1 - \vartheta_2|, \quad |\vartheta_1 - \theta| < \delta, |\vartheta_2 - \theta| < \delta. \tag{8.7}$$

Let $m = m_n \to \infty$ and $\log(n) m_n^3 = o(n)$. It suffices to show

$$\big(-2\log \mathscr{R}_n(F_{\hat\theta}) - m_n + q\big)\big/\sqrt{2(m_n - q)} \implies N(0,1).$$

For this, we take $w_n = q_n \circ F_\theta$ and $\hat w_n = q_n \circ F_{\hat\theta}$ with $q_n = (\phi_1, \dots, \phi_{m_n})^\top$ and verify (7.7) and (C2) with $\psi = J(\theta)^{-1/2}\dot\ell_\theta$. The desired result then follows from (ii) of Corollary 7.5.

We have $W_n = I_{m_n} = \int \hat w_n \hat w_n^\top \, dF_{\hat\theta}$ and obtain

$$\Big|\int \hat w_n \hat w_n^\top \, dF_\theta - W_n\Big| \le 2m_n \int |f_{\hat\theta}(x) - f_\theta(x)|\,dx = o_p(m_n^{-1/2})$$

in view of (8.6) and (2.2). Thus (7.7) follows if we verify

$$\Big|\hat W_n - \bar W_n - \int \hat w_n \hat w_n^\top \, dF_\theta + W_n\Big|^2 = o_p(m_n^{-1}). \tag{8.8}$$

Note that ψ has mean 0 and identity dispersion matrix under F_θ and that $A_n\psi$ equals $D_n J(\theta)^{-1}\dot\ell_\theta$, with $D_n = \int w_n \dot\ell_\theta^\top \, dF_\theta$. Thus (C2) follows from Remark 7.1,

$$\frac{1}{n}\sum_{j=1}^n \hat w_n(X_j) - w_n(X_j) + D_n(\hat\theta - \theta) = o_p(n^{-1/2}), \tag{8.9}$$

the stochastic expansion (2.2), and the fact that $|D_n|_o$ is bounded.

We are left to verify (8.8) and (8.9). For this we set

$$U_{nk}(t) = \frac{1}{n}\sum_{j=1}^n [\phi_k(F_{\theta + n^{-1/2}t}(X_j)) - \phi_k(F_\theta(X_j))],$$

$$V_{nkl}(t) = \frac{1}{n}\sum_{j=1}^n [(\phi_k\phi_l)(F_{\theta + n^{-1/2}t}(X_j)) - (\phi_k\phi_l)(F_\theta(X_j))],$$

and note that $D_n^\top = (d_1, \dots, d_{m_n})$ with $d_k = \int \phi_k(F_\theta)\dot\ell_\theta \, dF_\theta$. The statements (8.8) and

(8.9) follow if we show that, for each finite C,

$$T_{n1}(C) = \sup_{|t|\le C} \sum_{k=1}^{m_n} \sum_{l=1}^{m_n} (V_{nkl}(t) - E[V_{nkl}(t)])^2 = o_p(m_n^{-1}),$$

$$T_{n2}(C) = \sup_{|t|\le C} \sum_{k=1}^{m_n} (U_{nk}(t) - E[U_{nk}(t)])^2 = o_p(n^{-1}),$$

$$T_{n3}(C) = \sup_{|t|\le C} \sum_{k=1}^{m_n} (E[U_{nk}(t)] + n^{-1/2} d_k^\top t)^2 = o(n^{-1}).$$

The first two statements can be verified using the exponential inequality given in Lemma 5.2 in Peng and Schick (2004). This requires the fact that $(\log n) m_n^3/n \to 0$.

The identity $f_{\theta+\tau} - f_\theta - \dot\ell_\theta^\top \tau f_\theta = 2 s_\theta (s_{\theta+\tau} - s_\theta - \dot s_\theta^\top \tau) + (s_{\theta+\tau} - s_\theta)^2$ and the definition of d_k yield the formula

$$\int \phi_k(F_\theta(x)))(f_{\theta+\tau}(x) - f_\theta(x))\,dx = d_k^\top \tau + \int \phi_k(F_\theta(x))(s_{\theta+\tau}(x) - s_\theta(x))^2\,dx + 2\int \phi_k(F_\theta(x)) s_\theta(x)(s_{\theta+\tau}(x) - s_\theta(x) - \dot s_\theta^\top(x)\tau)\,dx.$$

In view of this and the fact that $\int \phi_k(F_\vartheta)\,dF_\vartheta = 0$ for all ϑ, we have the identity

$$\begin{aligned} E[U_{nk}(t)] + d_k^\top t n^{-1/2} &= \int (\phi_k(F_\theta(x)) - \phi_k(F_{\theta+n^{-1/2}t}(x)))(f_{\theta+n^{-1/2}t}(x) - f_\theta(x))\,dx \\ &\quad - \int \phi_k(F_\theta(x))(s_{\theta+n^{-1/2}t}(x) - s_\theta(x))^2\,dx \\ &\quad - 2\int \phi_k(F_\theta(x)) s_\theta(x)(s_{\theta+n^{-1/2}t}(x) - s_\theta(x) - n^{-1/2} t^\top \dot s_\theta(x))\,dx. \end{aligned}$$

Using (8.6), (8.7) and the orthonormality of the the functions $s_\theta \phi_k \circ F_\theta$, $k = 1, 2, \dots$, in L_2, $T_{n3}(C)$ can be bounded by

$$\frac{6\pi^2 m_n^3 K^4 C^4}{n^2} + 6 m_n \Big(\int (s_{\theta+n^{-1/2}t}(x) - s_\theta(x))^2\,dx \Big)^2 + 12 \sup_{|t|\le C} \rho(n^{-1/2}t).$$

The desired statement $T_{n3}(C) = o(n^{-1})$ now follows from (8.5) and $m_n^3 = o(n)$. This completes the proof of (7.4).

Details for Example 3. Assume that the distribution function of X is symmetric and continuous. Then $S = \mathrm{sign}(X)$ and $T = |X|$ are independent, S has mean zero and variance 1, and T has a continuous distribution function H. Let $\mathscr{R}_n$ be defined as in Example 3 with as $r = r_n \to \infty$ and $r_n^3 = o(n)$. It suffices to show that $(-2\log \mathscr{R}_n - (1 + r_n))/\sqrt{2(1 + r_n)}$ is asymptotically standard normal. This follows from Corollary 7.5 if we verify (7.3) and (7.6). These conditions follow from Lemma 8.1 applied with $S_j = \mathrm{sign}(X_j)$ and $T_j = |X_j|$. Indeed, in view of the properties of r_n, (7.3) is a consequence of (8.2) and (7.6) of (8.4).

Details for Example 4. Assume that X and Y are independent. Part (a) is an immediate consequence of Corollary 7.1. Part (b) follows if we show $(-2\log \mathscr{R}_n(\mathbb{F},\mathbb{G}) - r_n^2)/\sqrt{2}r_n$ is asymptotically standard normal. We shall use Corollary 7.5 to conclude this. Here m_n equals r_n^2 and thus satisfies $m_n^3 = o(n)$. We shall now verify (7.8) and (7.6). Let us set

$$D_{klj} = \phi_k(\mathbb{F}(X_j))\phi_l(\mathbb{G}(Y_j)) - \phi_k(F(X_j))\phi_l(G(Y_j)),$$

$$\Phi_{kj} = \phi_k(\mathbb{F}(X_j)) - \phi_k(F(X_j)) \quad \text{and} \quad \Gamma_{lj} = \phi_l(\mathbb{G}(Y_j)) - \phi_l(G(Y_j)).$$

In view of the inequality $|D_{klj}| \le \sqrt{2}|\Phi_{kj}| + \sqrt{2}|\Gamma_{lj}|$, we obtain with the help of (8.3) the bound

$$\sum_{k=1}^{r_n}\sum_{l=1}^{r_n}\frac{1}{n}\sum_{j=1}^{n} E[|D_{klj}|^2] \le \frac{8\pi^2 r_n^4}{n}.$$

From this and $r_n^6 = o(n)$ we conclude (7.8).

In view of the identity $D_{klj} = \phi_k(F(X_j))\Gamma_{lj} + \phi_l(G(Y_j))\Phi_{kj} + \Phi_{kl}\Gamma_{jl}$, (7.6) follows if we verify

$$T_{n1} = \sum_{k=1}^{r_n}\sum_{l=1}^{r_n}\Big(n^{-1/2}\sum_{j=1}^{n}\phi_k(F(X_j))\Gamma_{lj}\Big)^2 = o_p(1),$$

$$T_{n2} = \sum_{k=1}^{r_n}\sum_{l=1}^{r_n}\Big(n^{-1/2}\sum_{j=1}^{n}\Phi_{kj}\phi_l(G(Y_j))\Big)^2 = o_p(1),$$

$$T_{n3} = \sum_{k=1}^{r_n}\sum_{l=1}^{r_n}\Big(n^{-1/2}\sum_{j=1}^{n}\Phi_{kj}\Gamma_{lj}\Big)^2 = o_p(1).$$

Applications of (8.4) with $S_j = \phi_k(F(X_j))$ yield the bound $E[T_{n1}] \le \pi^2 r_n^4/n$, and this proves $T_{n1} = o_p(1)$. The proof of $T_{n2} = o_p(1)$ is similar. To deal with T_{n3} we set

$$H(k,l) = \sum_{j=1}^{n}\Phi_{k,j}\Gamma_{lj}, \quad \bar{\Phi}_k = \frac{1}{n}\sum_{j=1}^{n}\Phi_{kj} \quad \text{and} \quad \bar{\Gamma}_l = \frac{1}{n}\sum_{j=1}^{n}\Gamma_{lj}.$$

Note that $R_j = n\mathbb{F}(X_j)$ is the rank of X_j. Given $Y_1,\dots,Y_n$ and the order statistics $X_{(1)},\dots,X_{(n)}$, the sum $H(k,l)$ is a simple linear rank statistic with scores $a(j) = \phi_k(j/n) - \phi_k(F(X_{(j)})$ and coefficients $\mathbb{G}_{lj}$ and consequently has (conditional) mean $n\bar{\Phi}_k\bar{\Gamma}_l$ and (conditional) variance

$$\frac{1}{n-1}\sum_{j=1}^{n}(\Phi_{kj} - \bar{\Phi}_k)^2\sum_{j=1}^{n}(\Gamma_{li} - \bar{\Gamma}_l)^2 \le \frac{n}{n-1}\frac{1}{n}\sum_{j=1}^{n}\Phi_{kj}^2\frac{1}{n}\sum_{j=1}^{n}\Gamma_{lj}^2.$$

In view of this bound we derive the inequality

$$E[T_{n3}] \le \frac{n}{n-1}\sum_{k=1}^{n}E[\frac{1}{n}\sum_{j=1}^{n}\Phi_{kj}^2]\sum_{l=1}^{r_n}E[\frac{1}{n}\sum_{j=1}^{n}\Gamma_{lj}^2] + n\sum_{k=1}^{r_n}E[\bar{\Phi}_k^2]\sum_{l=1}^{r_n}E[\bar{\Gamma}_l^2].$$

We have

$$E[\bar{\Gamma}_k^2) = E[\bar{\Phi}_l^2] = \frac{1}{n} + \Big(\frac{1}{n}\sum_{j=1}^n \phi_k(j/n)\Big)^2 \le \frac{1}{n} + \frac{2\pi^2 k^2}{n^2}.$$

Using this and (8.3), we obtain $E[T_{n3}] = O(r_n^6 n^{-2}) = o(1)$ and thus $T_{n3} = o_p(1)$.

Acknowledgements

This work was completed while Anton Schick was visiting the Department of Statistics at Texas A&M University. He wants to thank the members of the department for their extraordinary hospitality. Thanks go also to Ingrid Van Keilegom for discussions and for providing an important reference. Hanxiang Peng's research was supported in part by NSF Grant DMS 0940365. Anton Schick's research was supported in part by NSF Grant DMS 0906551.

References

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

[2] Chen, S.X. and Qin, Y.S. (2000). Empirical likelihood confidence intervals for local linear smoothers. *Biometrika* **87**, 946–953.

[3] Chen, S.X. and Van Keilegom, I. (2009). A goodness-of-fit test for parametric and semiparametric models in multiresponse regression. *Bernoulli*, **15**, 955–976.

[4] Hall, P. and La Scala, B. (1990). Methodology and algorithms of empirical likelihood. *Internat. Statist. Review*, **58**, 109–127.

[5] DiCiccio, T.J., Hall, P. and Romano, J.P. (1991). *Ann. Statist.* **19**, 1053–1061.

[6] Emerson, S. and Owen, A. (2009). Calibration of the empirical likelihood method for a vector mean. *Electron. J. Statist.* **3**, 1161–1192.

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

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

[9] Inglot, T. and Ledwina, R. (1996). Asymptotic optimality of data-driven Neyman's tests for uniformity. *Ann. Statist.* **24**, 1982–2019.

[10] Kolaczyk, E.D. (1994). Empirical likelihood for generalized linear models. *Statist. Sinica* **4**, 199–218.

[11] Li, G. and Wang, Q.-H. (2003). Empirical likelihood regression analysis for right censored data. *Statist. Sinica* **13**, 51–68.

[12] Liu, Y. and Chen, J. (2010). Adjusted empirical likelihood with high-order precision. *Ann. Statist.* **38**, 1341-1362.

[13] Neyman, J. (1937). "Smooth test" for goodness of fit. *Skand. Aktuarietidskr.* **20**, 149–199.

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

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

[16] Owen, A. (1991). Empirical likelihood for linear models. *Ann. Statist.* **19**, 1725–1747.

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

[18] Peng, H. and Schick, A. (2004). Estimation of linear functionals of bivariate distributions with parametric marginals. *Statist. Decisions* **22**, 61–77.

[19] 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.

[20] Peng, H. and Schick, A. (2012). Asymptotic normality of quadratic forms with random vectors of increasing dimension. Preprint.

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

[22] Qin, G. and Jing, B.-Y. (2001). Empirical likelihood for censored linear regression. *Scand. J. Statist.* **28**, 661–673.

[23] Qin, J. and Lawless, J. (1994). Empirical likelihood and general estimating equations. *Ann. Statist.* **22**, 300–325.

[24] Shi, J. and Lau, T.S. (2000). Empirical likelihood for partially linear models. *J. Multivariate Anal.* **72**, 132–148.

[25] Tripathi, G. and Kitamura, Y. (2003). Testing conditional moment restrictions *Ann. Statist.* **31**, 2059–2095.

[26] Wang, Q.-H. and Jing, B.-Y. (2001). Empirical likelihood for a class of functionals of survival distribution with censored data. *Ann. Inst. Statist. Math.* **53**, 517–527.

[27] Wang, Q.-H. and Jing, B.-Y. (2003). Empirical likelihood for partially linear models. *Ann. Inst. Statist. Meth.* **55**, 585–595.

[28] Zhou, M. and Li, G. (2008). Empirical likelihood analysis of the Buckley-James estimator. *J Multivariate Anal.* **99**, 649–664.