

Empirical phi-divergence test statistics for testing simple and composite null hypotheses

N. Balakrishnan¹, N. Martín² and L. Pardo³

¹Department of Mathematics and Statistics, McMaster University, Hamilton, Canada

²Department of Statistics, Carlos III University of Madrid, Getafe (Madrid), Spain

³Department of Statistics and O.R., Complutense University of Madrid, Madrid, Spain

Abstract

The main purpose of this paper is to introduce first a new family of empirical test statistics for testing a simple null hypothesis when the vector of parameters of interest are defined through a specific set of unbiased estimating functions. This family of test statistics is based on a distance between two probability vectors, with the first probability vector obtained by maximizing the empirical likelihood on the vector of parameters, and the second vector defined from the fixed vector of parameters under the simple null hypothesis. The distance considered for this purpose is the phi-divergence measure. The asymptotic distribution is then derived for this family of test statistics. The proposed methodology is illustrated through the well-known data of Newcomb's measurements on the passage time for light. A simulation study is carried out to compare its performance with respect to the empirical likelihood ratio test when confidence intervals are constructed based on the respective statistics for small sample sizes. The results suggest that the "empirical modified likelihood ratio test statistic" provides a competitive alternative to the empirical likelihood ratio test statistic, and is also more robust than the empirical likelihood ratio test statistic in the presence of contamination in the data. Finally, we propose empirical phi-divergence test statistics for testing a composite null hypothesis and present some asymptotic as well as simulation results to study the performance of these test procedures.

AMS 2001 Subject Classification: 62E20

Keywords and phrases: Empirical likelihood, Empirical phi-divergence test statistics, Influence function, Phi-divergence measures, Power function, Empirical likelihood ratio, Empirical modified likelihood ratio.

1 Introduction

Empirical likelihood (EL) is a powerful and currently widely used approach for developing efficient nonparametric statistics. The EL was introduced by Owen (1988) and since then many papers have appeared on this topic making varied contributions to different inferential problems. In this approach, the parameters are usually defined as functionals of the unknown population distribution.

The first purpose of this paper is to introduce a new family of empirical test statistics as an alternative to the likelihood ratio test statistic proposed by Qin and Lawless (1994) for testing a simple null hypothesis. As an extension of the empirical likelihood ratio test of Qin and Lawless (1995), a new family of empirical test statistics is also considered here for composite hypothesis. This new family of empirical test statistics is based on divergence measures.

Consider k -variate i.i.d. random vectors $\mathbf{X}_1, \dots, \mathbf{X}_n$ with unknown distribution function F , and a p -dimensional parameter, $\boldsymbol{\theta} \in \Theta$, associated with F having finite mean and non-singular covariance matrix. We assume that all the information about $\boldsymbol{\theta}$ and F is available in the form of $r \geq p$ functionally independent unbiased estimating

functions, through the functions $g_j(\mathbf{X}, \boldsymbol{\theta})$, $j = 1, 2, \dots, r$, such that $E_F [g_j(\mathbf{X}, \boldsymbol{\theta})] = 0$. In vector notation, we have $\mathbf{g}(\mathbf{X}, \boldsymbol{\theta}) = (g_1(\mathbf{X}, \boldsymbol{\theta}), \dots, g_r(\mathbf{X}, \boldsymbol{\theta}))^T$, such that

$$E_F [\mathbf{g}(\mathbf{X}, \boldsymbol{\theta})] = \mathbf{0}_r. \quad (1)$$

We shall assume that for each realization \mathbf{x} of \mathbf{X} , $\mathbf{g} : \mathbb{R}^p \rightarrow \mathbb{R}^r$ is a vector-valued function and the $r \times p$ matrix

$$\mathbf{G}_{\mathbf{x}}(\boldsymbol{\theta}) = \frac{\partial \mathbf{g}(\mathbf{x}, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}^T} \quad (2)$$

exists. This formulation is as in Qin and Lawless (1994), but a little different from that of Owen (1988, 1990). The essential difference is that Owen considered $r = p$ instead of $r \geq p$. For example, by taking into account that $E_F [X] = \theta$, we can adopt steps similar to the method of moments to obtain estimators through the estimating function $g_1(X, \theta) = X - \theta$ of a univariate distribution. In addition, if we assume $E_F [X^2] = 2\theta^2 + 1$, the other estimating function is $g_2(X, \theta) = X^2 - 2\theta^2 - 1$, and in this case we have $r = 2 > p = k = 1$.

Let $\mathbf{x}_1, \dots, \mathbf{x}_n$ be a realization of $\mathbf{X}_1, \dots, \mathbf{X}_n$. The empirical likelihood function is then given by

$$\mathcal{L}_n(F) = \prod_{i=1}^n dF(\mathbf{x}_i) = \prod_{i=1}^n p_i,$$

where $p_i = dF(\mathbf{x}_i) = P(\mathbf{X} = \mathbf{x}_i)$. Only distributions with an atom of probability at each \mathbf{x}_i have non-zero likelihood, and without consideration of estimating functions, the empirical likelihood function $\mathcal{L}_n(F)$ is seen to be maximized, at $\mathbf{X}_1 = \mathbf{x}_1, \dots, \mathbf{X}_n = \mathbf{x}_n$, by the empirical distribution function

$$F_n(\mathbf{x}) = \sum_{i=1}^n u_i I(\mathbf{X}_i < \mathbf{x}),$$

which is associated with the n -dimensional discrete uniform distribution

$$\mathbf{u} = (u_1, \dots, u_n)^T = \left(\frac{1}{n}, \dots, \frac{1}{n}\right)^T.$$

Now, let

$$F_{n,\boldsymbol{\theta}}(\mathbf{x}) = \sum_{i=1}^n p_i(\boldsymbol{\theta}) I(\mathbf{X}_i < \mathbf{x}),$$

be an empirical distribution function associated with the probability vector

$$\mathbf{p}(\boldsymbol{\theta}) = (p_1(\boldsymbol{\theta}), \dots, p_n(\boldsymbol{\theta}))^T, \quad p_i(\boldsymbol{\theta}) \geq 0, \quad \sum_{i=1}^n p_i(\boldsymbol{\theta}) = 1, \quad (3)$$

and $\ell_{E,n}(\boldsymbol{\theta})$ be the kernel of the empirical log-likelihood function, $\sum_{i=1}^n \log p_i(\boldsymbol{\theta})$. If we are interested in maximizing $\ell_{E,n}(\boldsymbol{\theta})$ subject to the restrictions defined by the estimating functions based on $F_{n,\boldsymbol{\theta}}(\mathbf{x})$ given by

$$\sum_{i=1}^n p_i(\boldsymbol{\theta}) \mathbf{g}(\mathbf{X}_i, \boldsymbol{\theta}) = \mathbf{0}_r, \quad (4)$$

we obtain, by applying the Lagrange multipliers method,

$$p_i(\boldsymbol{\theta}) = \frac{1}{n} \frac{1}{1 + \mathbf{t}(\boldsymbol{\theta})^T \mathbf{g}(\mathbf{X}_i, \boldsymbol{\theta})}, \quad i = 1, \dots, n, \quad (5)$$

where $\mathbf{t}(\boldsymbol{\theta})$ is an r -dimensional vector to be determined by solving the non-linear system of r equations,

$$\frac{1}{n} \sum_{i=1}^n \frac{1}{1 + \mathbf{t}(\boldsymbol{\theta})^T \mathbf{g}(\mathbf{x}_i, \boldsymbol{\theta})} \mathbf{g}(\mathbf{x}_i, \boldsymbol{\theta}) = \mathbf{0}_r, \quad (6)$$

subject to (3) and (5). Thus, the kernel of the empirical log-likelihood function is

$$\ell_{E,n}(\boldsymbol{\theta}) = - \sum_{i=1}^n \log [1 + \mathbf{t}(\boldsymbol{\theta})^T \mathbf{g}(\mathbf{x}_i, \boldsymbol{\theta})]. \quad (7)$$

One of the important results of Qin and Lawless (1994) is that the empirical likelihood ratio test statistic for testing

$$H_0: \boldsymbol{\theta} = \boldsymbol{\theta}_0 \text{ vs. } H_1: \boldsymbol{\theta} \neq \boldsymbol{\theta}_0 \quad (8)$$

is given by

$$L_{E,n}(\widehat{\boldsymbol{\theta}}_{E,n}, \boldsymbol{\theta}_0) = 2\ell_{E,n}(\widehat{\boldsymbol{\theta}}_{E,n}) - 2\ell_{E,n}(\boldsymbol{\theta}_0), \quad (9)$$

where $\widehat{\boldsymbol{\theta}}_{E,n}$ is the empirical maximum likelihood estimator of the parameter $\boldsymbol{\theta}$ obtained by maximizing $\ell_{E,n}(\boldsymbol{\theta})$ in (7). In particular, if $r = p$, it can be seen that $\mathbf{t}(\widehat{\boldsymbol{\theta}}_{E,n}) = \mathbf{0}_r$, $\mathbf{p}(\widehat{\boldsymbol{\theta}}_{E,n}) = \mathbf{u}$ and $\widehat{\boldsymbol{\theta}}_{E,n}$ is the solution of the system of r equations $\widehat{\mathbf{g}}_n(\boldsymbol{\theta}) = \mathbf{0}_r$, where

$$\widehat{\mathbf{g}}_n(\boldsymbol{\theta}) = \frac{1}{n} \sum_{i=1}^n \mathbf{g}(\mathbf{x}_i, \boldsymbol{\theta}), \quad (10)$$

subject to (3) and (5). Furthermore, if $r = p$, $\ell_{E,n}(\widehat{\boldsymbol{\theta}}_{E,n}) = 0$ and $L_{E,n}(\widehat{\boldsymbol{\theta}}_{E,n}, \boldsymbol{\theta}_0) = -2\ell_{E,n}(\boldsymbol{\theta}_0)$. The asymptotic properties of $\widehat{\boldsymbol{\theta}}_{E,n}$, when $r \geq p$, were studied by Qin and Lawless (1994). In fact, the asymptotic distribution of $L_{E,n}(\widehat{\boldsymbol{\theta}}_{E,n}, \boldsymbol{\theta}_0)$ in (9) is chi-square with p degrees of freedom.

Here, we propose a new family of test statistics for testing the hypotheses in (8), based on ϕ -divergence measures, and then derive their asymptotic distribution. This new family of empirical test statistics is referred to hereafter as “empirical ϕ -divergence test statistics”. In Section 3, the asymptotic null distribution of the empirical ϕ -divergence test statistics is derived. Then, two power approximations of the empirical ϕ -divergence test statistics are presented in Section 4. An illustrative example is presented in Section 5. In Section 6, a Monte Carlo simulation study is carried out to compare its performance with respect to the empirical likelihood ratio test when confidence intervals are constructed based on the respective statistics for small simple sizes. The results show that the empirical ϕ -divergence test statistic is competitive in terms of power when compared to the empirical likelihood ratio test statistic, and moreover is more robust than the empirical likelihood ratio test statistic in the presence of contamination in the data. Next, in Section 7, we propose empirical phi-divergence test statistics for testing a composite null hypothesis and present some asymptotic as well as simulation results to evaluate the performance of these test procedures. Finally, in Section 8, we make some concluding remarks.

2 New family of empirical phi-divergence test statistics

The Kullback-Leibler divergence measure between the probability vectors, \mathbf{u} and $\mathbf{p}(\boldsymbol{\theta})$, is given by

$$\sum_{i=1}^n u_i \log \frac{u_i}{p_i(\boldsymbol{\theta})} = \sum_{i=1}^n \frac{1}{n} \log \frac{\frac{1}{n}}{\frac{1}{n} \frac{1}{1 + \mathbf{t}(\boldsymbol{\theta})^T \mathbf{g}(\mathbf{x}_i, \boldsymbol{\theta})}} = \frac{1}{n} \sum_{i=1}^n \log (1 + \mathbf{t}(\boldsymbol{\theta})^T \mathbf{g}(\mathbf{x}_i, \boldsymbol{\theta})).$$

In the sequel, we shall denote it by $D_{\text{Kullback}}(F_n, F_{n,\boldsymbol{\theta}})$, since the above expression is the distance, in the sense of the Kullback-Leibler divergence, between the distributions functions F_n and $F_{n,\boldsymbol{\theta}}$. Using this notation, it is clear that the empirical likelihood ratio test statistic in (9) can be expressed as

$$L_{E,n}(\widehat{\boldsymbol{\theta}}_{E,n}, \boldsymbol{\theta}_0) = 2n \left(D_{\text{Kullback}}(F_n, F_{n,\boldsymbol{\theta}_0}) - D_{\text{Kullback}}(F_n, F_{n,\widehat{\boldsymbol{\theta}}_{E,n}}) \right),$$

or equivalently

$$L_{E,n}(\widehat{\boldsymbol{\theta}}_{E,n}, \boldsymbol{\theta}_0) = \frac{2n}{\phi''(1)} \left(D_\phi(F_n, F_{n,\boldsymbol{\theta}_0}) - D_\phi(F_n, F_{n,\widehat{\boldsymbol{\theta}}_{E,n}}) \right),$$

where

$$D_\phi(F_n, F_{n,\boldsymbol{\theta}}) = \frac{1}{n} \sum_{i=1}^n \frac{1}{1 + \mathbf{t}(\boldsymbol{\theta})^T \mathbf{g}(\mathbf{x}_i, \boldsymbol{\theta})} \phi(1 + \mathbf{t}(\boldsymbol{\theta})^T \mathbf{g}(\mathbf{x}_i, \boldsymbol{\theta})), \quad (11)$$

with $\phi(x) = x \log x - x + 1$, $x \in \mathbb{R}$.

Now, we shall denote by Φ^* the class of all convex functions $\phi : \mathbb{R}^+ \rightarrow \mathbb{R}$ such that at $x = 1$, $\phi(1) = 0$, $\phi''(1) > 0$, and at $x = 0$, $0\phi(0/0) = 0$ and $0\phi(p/0) = p \lim_{u \rightarrow \infty} \frac{\phi(u)}{u}$. Instead of $\phi(x) = x \log x - x + 1$, if we consider a function ϕ belonging to Φ^* , we obtain a new family of test statistics for testing (8) given by

$$T_n^\phi(\widehat{\boldsymbol{\theta}}_{E,n}, \boldsymbol{\theta}_0) = \frac{2n}{\phi''(1)} \left(D_\phi(F_n, F_{n,\boldsymbol{\theta}_0}) - D_\phi(F_n, F_{n,\widehat{\boldsymbol{\theta}}_{E,n}}) \right). \quad (12)$$

We will refer to this family of test statistics as empirical ϕ -divergence test statistics.

For every $\phi \in \Phi^*$ that is differentiable at $x = 1$, the function $\psi(x) \equiv \phi(x) - \phi'(1)(x - 1)$ also belongs to Φ^* . Then, we have

$$T_n^\phi(\widehat{\boldsymbol{\theta}}_{E,n}, \boldsymbol{\theta}_0) = T_n^\psi(\widehat{\boldsymbol{\theta}}_{E,n}, \boldsymbol{\theta}_0)$$

and ψ has the additional property that $\psi'(1) = 0$. Since the two divergence measures are equivalent, we can consider the set Φ^* to be equivalent to the set $\Phi = \Phi^* \cap \{\phi : \phi'(1) = 0\}$. In what follows, we shall assume that $\phi \in \Phi$.

The statistics in (12) have been considered recently by Broniatowski and Keziou (2012) to give some empirical test statistics, generalizing an important subfamily of test statistics introduced by Baggerly (1998). We present more details in Section 6 in this regard.

The main purpose of this paper is to present a new family of test statistics for testing the hypotheses in (8) based on the ϕ -divergence measure between $F_{n,\boldsymbol{\theta}_0}$ and $F_{n,\widehat{\boldsymbol{\theta}}_{E,n}}$, namely, $D_\phi(F_{n,\widehat{\boldsymbol{\theta}}_{E,n}}, F_{n,\boldsymbol{\theta}_0})$. We shall consider the empirical family of ϕ -divergence test statistics given by

$$S_n^\phi(\widehat{\boldsymbol{\theta}}_{E,n}, \boldsymbol{\theta}_0) = \frac{2n}{\phi''(1)} D_\phi(F_{n,\widehat{\boldsymbol{\theta}}_{E,n}}, F_{n,\boldsymbol{\theta}_0}) = \frac{2}{\phi''(1)} \sum_{i=1}^n \frac{1}{1 + \mathbf{t}(\boldsymbol{\theta}_0)^T \mathbf{g}(\mathbf{x}_i, \boldsymbol{\theta}_0)} \phi\left(\frac{1 + \mathbf{t}(\boldsymbol{\theta}_0)^T \mathbf{g}(\mathbf{x}_i, \boldsymbol{\theta}_0)}{1 + \mathbf{t}(\widehat{\boldsymbol{\theta}}_{E,n})^T \mathbf{g}(\mathbf{x}_i, \widehat{\boldsymbol{\theta}}_{E,n})}\right), \quad (13)$$

where ϕ is a function satisfying the same conditions as function ϕ used to construct $T_n^\phi(\widehat{\boldsymbol{\theta}}_{E,n}, \boldsymbol{\theta}_0)$. Observe that (12) and (13) are equivalent only when $r = p$. It is well-known that the family of test statistics based on ϕ -divergence has some nice and optimal properties for different inferential problems, and especially in relation to robustness; see Pardo (2006) and Basu et al. (2011).

3 Asymptotic null distribution

In this section, we derive the asymptotic distribution of $S_n^\phi(\widehat{\boldsymbol{\theta}}_{E,n}, \boldsymbol{\theta}_0)$. For this purpose, the asymptotic distribution of the maximum empirical likelihood estimator of the parameter $\boldsymbol{\theta}$, $\widehat{\boldsymbol{\theta}}_{E,n}$, as well as the asymptotic distribution of $\widehat{\mathbf{t}} = \mathbf{t}(\widehat{\boldsymbol{\theta}}_{E,n})$ are important. These asymptotic distributions are given in Qin and Lawless (1994), for example. Under some regularity assumptions (see Lemma 1 and Theorem 1 of Qin and Lawless (1994)), we have

$$\sqrt{n} \begin{pmatrix} \widehat{\boldsymbol{\theta}}_{E,n} - \boldsymbol{\theta}_0 \\ \widehat{\mathbf{t}}(\widehat{\boldsymbol{\theta}}_{E,n}) \end{pmatrix} \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \mathcal{N} \left(\mathbf{0}_{p+r}, \begin{pmatrix} \mathbf{V}(\boldsymbol{\theta}_0) & \mathbf{0}_{p \times r} \\ \mathbf{0}_{r \times p} & \mathbf{R}(\boldsymbol{\theta}_0) \end{pmatrix} \right), \quad (14)$$

where $\xrightarrow[n \rightarrow \infty]{\mathcal{L}}$ denotes convergence in law and

$$\mathbf{V}(\boldsymbol{\theta}_0) = \left(\mathbf{S}_{12}(\boldsymbol{\theta}_0)^T \mathbf{S}_{11}^{-1}(\boldsymbol{\theta}_0) \mathbf{S}_{12}(\boldsymbol{\theta}_0) \right)^{-1}, \quad (15)$$

$$\mathbf{R}(\boldsymbol{\theta}_0) = \mathbf{S}_{11}^{-1}(\boldsymbol{\theta}_0) - \mathbf{S}_{11}^{-1}(\boldsymbol{\theta}_0) \mathbf{S}_{12}(\boldsymbol{\theta}_0) \mathbf{V}(\boldsymbol{\theta}_0) \mathbf{S}_{21}(\boldsymbol{\theta}_0) \mathbf{S}_{11}^{-1}(\boldsymbol{\theta}_0), \quad (16)$$

with

$$\mathbf{S}_{11}(\boldsymbol{\theta}_0) = E_F \left[\mathbf{g}(\mathbf{X}, \boldsymbol{\theta}_0) \mathbf{g}(\mathbf{X}, \boldsymbol{\theta}_0)^T \right], \quad (17)$$

$$\mathbf{S}_{12}(\boldsymbol{\theta}_0) = E_F [\mathbf{G}_X(\boldsymbol{\theta}_0)], \quad \mathbf{S}_{21}(\boldsymbol{\theta}_0) = \mathbf{S}_{12}(\boldsymbol{\theta}_0)^T. \quad (18)$$

This result is derived from:

a)

$$\sqrt{n}(\widehat{\boldsymbol{\theta}}_{E,n} - \boldsymbol{\theta}_0) = \mathbf{V}(\boldsymbol{\theta}_0) \mathbf{S}_{12}(\boldsymbol{\theta}_0)^T \mathbf{S}_{11}^{-1}(\boldsymbol{\theta}_0) \sqrt{n}\bar{\mathbf{g}}_n(\boldsymbol{\theta}_0) + o_p(\mathbf{1}_p), \quad (19)$$

where $\bar{\mathbf{g}}_n(\boldsymbol{\theta})$ is given by (10). It is clear from the Central Limit Theorem that

$$\sqrt{n}\bar{\mathbf{g}}_n(\boldsymbol{\theta}_0) \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \mathcal{N}(\mathbf{0}_r, \mathbf{S}_{11}(\boldsymbol{\theta}_0)),$$

and so $\sqrt{n}(\widehat{\boldsymbol{\theta}}_{E,n} - \boldsymbol{\theta}_0) \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \mathcal{N}(\mathbf{0}_p, \mathbf{V}(\boldsymbol{\theta}_0))$;

b) $\sqrt{n}\mathbf{t}(\widehat{\boldsymbol{\theta}}_{E,n}) = -\mathbf{R}(\boldsymbol{\theta}_0) \sqrt{n}\bar{\mathbf{g}}_n(\boldsymbol{\theta}_0) + o_p(\mathbf{1}_r)$, where $\mathbf{R}(\boldsymbol{\theta}_0)$ is given by (16) and so $\sqrt{n}\mathbf{t}(\widehat{\boldsymbol{\theta}}_{E,n}) \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \mathcal{N}(\mathbf{0}_r, \mathbf{R}(\boldsymbol{\theta}_0))$.

In addition, $\widehat{\boldsymbol{\theta}}_{E,n}$ and $\mathbf{t}(\widehat{\boldsymbol{\theta}}_{E,n})$ are asymptotically uncorrelated.

Lemma 1 *The influence function of the empirical maximum likelihood estimator of parameter $\boldsymbol{\theta}$, $\widehat{\boldsymbol{\theta}}_{E,n}$, is given by*

$$\mathcal{IF}(\mathbf{x}, \widehat{\boldsymbol{\theta}}_{E,n}, F_{n,\boldsymbol{\theta}_0}) = \mathbf{V}(\boldsymbol{\theta}_0) \mathbf{S}_{12}(\boldsymbol{\theta}_0)^T \mathbf{S}_{11}^{-1}(\boldsymbol{\theta}_0) \mathbf{g}(\mathbf{x}, \boldsymbol{\theta}_0). \quad (20)$$

Proof. *It follows from the expression given in (19) and taking into account the definition of the influence function given in formula (20.1) of van der Vaart (2000, page 292). ■*

Remark 2 *The empirical maximum likelihood estimator of parameter $\boldsymbol{\theta}$, $\widehat{\boldsymbol{\theta}}_{E,n}$, is obtained maximizing the kernel of empirical log-likelihood function given in (7) or equivalently minimizing the function $-\frac{1}{n} \sum_{i=1}^n \log(np_i(\boldsymbol{\theta}))$, subject to the restrictions given in (4). This expression can be written as the ϕ -divergence measure between the probability vectors \mathbf{u} and $\mathbf{p}(\boldsymbol{\theta})$, i.e. $D_\phi(F_n, F_{n,\boldsymbol{\theta}}) = \sum_{i=1}^n p_i(\boldsymbol{\theta}) \phi\left(\frac{1}{np_i(\boldsymbol{\theta})}\right)$ with $\phi(x) = x \log x - x + 1$. Therefore,*

$$\widehat{\boldsymbol{\theta}}_{E,n} = \arg \min_{\boldsymbol{\theta}} D_\phi(F_n, F_{n,\boldsymbol{\theta}}), \quad \phi(x) = x \log x - x + 1,$$

subject to the restrictions given in (4). If we consider a general function $\phi \in \Phi$, defined in Section 2, instead of considering $\phi(x) = x \log x - x + 1$, then we can define the empirical minimum ϕ -divergence estimator by

$$\widehat{\boldsymbol{\theta}}_{E\phi,n} = \arg \min_{\boldsymbol{\theta}} D_\phi(F_n, F_{n,\boldsymbol{\theta}}), \quad \phi \in \Phi,$$

subject to the restrictions given in (4). The empirical exponential tilting estimator (ET), considered for instance in Schennach (2007), is defined by $\widehat{\boldsymbol{\theta}}_{ET,n} = \arg \min_{\boldsymbol{\theta}} \sum_{i=1}^n p_i(\boldsymbol{\theta}) \log(np_i(\boldsymbol{\theta}))$, subject to the restrictions given in (4). The ET is another member of this family of estimators since $\widehat{\boldsymbol{\theta}}_{ET,n} = \widehat{\boldsymbol{\theta}}_{E\phi,n}$ with $\phi(x) = -\log x + x - 1$. The asymptotic properties of $\widehat{\boldsymbol{\theta}}_{E\phi,n}$ and $\widehat{\boldsymbol{\theta}}_{E,n}$ are the same (for more details, see Ragusa (2011), Broniatowski and Keziou (2012) and Schennach (2007)). The Fisher consistency of $\widehat{\boldsymbol{\theta}}_{E\phi,n}$ was established in Ragusa (2011). Therefore, all the asymptotic results obtained for the test statistics considered in this paper are valid replacing

$\widehat{\boldsymbol{\theta}}_{E,n}$ by $\widehat{\boldsymbol{\theta}}_{E\phi,n}$. The expression given in (20), for the influence function of the empirical maximum likelihood estimator of parameter $\boldsymbol{\theta}$, can be found for the empirical minimum ϕ -divergence estimators, $\widehat{\boldsymbol{\theta}}_{E\phi,n}$, in Proposition 2.3 of Toma (2013). Hence, all the estimators based on ϕ -divergence measures, independently of the ϕ function, share the same influence function.

Let $\|\cdot\|$ denote any vector or matrix norm. We shall assume the following regularity conditions:

- i) $\mathbf{S}_{11}(\boldsymbol{\theta}_0)$ in (17) is positive definite, and for $\mathbf{S}_{12}(\boldsymbol{\theta}_0)$ in (18), $\text{rank}(\mathbf{S}_{12}(\boldsymbol{\theta}_0)) = p$.
- ii) There exists a neighbourhood of $\boldsymbol{\theta}_0$, in which $\|\mathbf{g}(\mathbf{x}, \boldsymbol{\theta})\|^3$ is bounded by some integrable function.
- iii) There exists a neighbourhood of $\boldsymbol{\theta}_0$, in which $\mathbf{G}_{\mathbf{x}}(\boldsymbol{\theta})$, given in (2), is continuous and $\|\mathbf{G}_{\mathbf{x}}(\boldsymbol{\theta})\|$ is bounded by some integrable function.
- iv) There exists a neighbourhood of $\boldsymbol{\theta}_0$ in which $\frac{\partial \mathbf{G}_{\mathbf{x}}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}$ is continuous and $\left\| \frac{\partial \mathbf{G}_{\mathbf{x}}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right\|$ is bounded by some integrable function.

The asymptotic distribution of the empirical ϕ -divergence test statistics, $S_n^\phi(\widehat{\boldsymbol{\theta}}_{E,n}, \boldsymbol{\theta}_0)$, is given in the following theorem.

Theorem 3 Under H_0 in (8) and the assumptions i)-iv) above, we have

$$S_n^\phi(\widehat{\boldsymbol{\theta}}_{E,n}, \boldsymbol{\theta}_0) \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \chi_p^2.$$

Proof. Based on the Taylor expansions of (13), as well as the asymptotic properties of $\widehat{\boldsymbol{\theta}}_{E,n}$, given in Qin and Lawless (1994), it holds

$$S_n^\phi(\widehat{\boldsymbol{\theta}}_{E,n}, \boldsymbol{\theta}_0) = \mathbf{U}_n(\widehat{\boldsymbol{\theta}}_{E,n}, \boldsymbol{\theta}_0)^T \mathbf{U}_n(\widehat{\boldsymbol{\theta}}_{E,n}, \boldsymbol{\theta}_0) + o_p(1), \quad (21)$$

with $\mathbf{U}_n(\widehat{\boldsymbol{\theta}}_{E,n}, \boldsymbol{\theta}_0) = \sqrt{n} \mathbf{V}^{-1/2}(\boldsymbol{\theta}_0) (\widehat{\boldsymbol{\theta}}_{E,n} - \boldsymbol{\theta}_0) \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \mathcal{N}(\mathbf{0}_p, \mathbf{I}_p)$, according to (14). ■

Remark 4 There are some measures of divergence which can not be expressed as a ϕ -divergence measure such as the divergence measures of Bhattacharya (1943), Rényi (1961), and Sharma and Mittal (1977). However, such measures can be written in the form

$$D_\phi^h(F_{n,\theta}, F_{n,\boldsymbol{\theta}_0}) = h(D_\phi(F_{n,\theta}, F_{n,\boldsymbol{\theta}_0})),$$

where h is a differentiable increasing function mapping from $[0, \infty)$ onto $[0, \infty)$, with $h(0) = 0$, $h'(0) > 0$, and $\phi \in \Phi$. In Table 1, these divergence measures are presented, along with the corresponding expressions of h and ϕ .

Divergence	$h(x)$	$\phi(x)$
Rényi	$\frac{1}{a(a-1)} \log(a(a-1)x + 1)$, $a \neq 0, 1$	$\frac{x^a - a(x-1) - 1}{a(a-1)}$, $a \neq 0, 1$
Sharma-Mittal	$\frac{1}{b-1} \left\{ [1 + a(a-1)x]^{\frac{b-1}{a-1}} - 1 \right\}$, $b, a \neq 1$	$\frac{x^a - a(x-1) - 1}{a(a-1)}$, $a \neq 0, 1$
Battacharya	$-\log(-x + 1)$	$-x^{1/2} + \frac{1}{2}(x + 1)$

Table 1: Some specific (h, ϕ) -divergence measures.

In the case of Rényi's divergence, we have

$$D_{Rényi}^a(F_{n,\theta}, F_{n,\boldsymbol{\theta}_0}) = \frac{1}{a(a-1)} \log \sum_{i=1}^n \frac{(1 + \mathbf{t}(\boldsymbol{\theta}_0)^T \mathbf{g}(\mathbf{x}_i, \boldsymbol{\theta}_0))^{a-1}}{n(1 + \mathbf{t}(\boldsymbol{\theta})^T \mathbf{g}(\mathbf{x}_i, \boldsymbol{\theta}))^a}, \text{ if } a \neq 0, 1, \quad (22)$$

and its “limit” cases corresponding to $a = 0$ and $a = 1$ as

$$\begin{aligned} D_{Rényi}^0(F_{n,\theta}, F_{n,\theta_0}) &= \lim_{a \rightarrow 0} D_{Rényi}(F_{n,\theta}, F_{n,\theta_0}) = D_{Kullback}(F_{n,\theta_0}, F_{n,\theta}) \\ &= \sum_{i=1}^n \frac{1}{n(1 + \mathbf{t}(\theta)^T \mathbf{g}(\mathbf{x}_i, \theta))} \log \left(\frac{1 + \mathbf{t}(\theta_0)^T \mathbf{g}(\mathbf{x}_i, \theta_0)}{1 + \mathbf{t}(\theta)^T \mathbf{g}(\mathbf{x}_i, \theta)} \right) \end{aligned}$$

and

$$\begin{aligned} D_{Rényi}^1(F_{n,\theta}, F_{n,\theta_0}) &= \lim_{a \rightarrow 1} D_{Rényi}(F_{n,\theta}, F_{n,\theta_0}) = D_{Kullback}(F_{n,\theta}, F_{n,\theta_0}) \\ &= \sum_{i=1}^n \frac{1}{n(1 + \mathbf{t}(\theta_0)^T \mathbf{g}(\mathbf{x}_i, \theta_0))} \log \left(\frac{1 + \mathbf{t}(\theta)^T \mathbf{g}(\mathbf{x}_i, \theta)}{1 + \mathbf{t}(\theta_0)^T \mathbf{g}(\mathbf{x}_i, \theta_0)} \right), \end{aligned}$$

respectively.

The (h, ϕ) -divergence measures were introduced in Menéndez et al. (1995) and some associated asymptotic results were established in Menéndez et al. (1997).

Theorem 5 Under the assumptions of Theorem 3, the asymptotic null distribution of the family of empirical test statistics

$$S_n^{\phi,h}(\hat{\theta}_{E,n}, \theta_0) = \frac{2n}{\phi''(1)h'(0)} h \left(D_\phi(F_{n,\hat{\theta}_{E,n}}, F_{n,\theta_0}) \right)$$

is chi-squared with p degrees of freedom.

Remark 6 Note that we can also consider the family of empirical (h, ϕ) -divergence test statistics given by

$$T_n^{\phi,h}(\hat{\theta}_{E,n}, \theta_0) = \frac{2n}{\phi''(1)h'(0)} \left(h(D_\phi(F_n, F_{n,\theta_0})) - h(D_\phi(F_n, F_{\hat{\theta}_{E,n}})) \right),$$

and their asymptotic distribution is chi-squared with p degrees of freedom as well. Test-statistics $T_n^{\phi,h}(\hat{\theta}_{E,n}, \theta_0)$ and $S_n^{\phi,h}(\hat{\theta}_{E,n}, \theta_0)$, include as particular cases $T_n^\phi(\hat{\theta}_{E,n}, \theta_0)$ and $S_n^\phi(\hat{\theta}_{E,n}, \theta_0)$, taking $h(x) = x$. In the same way done for $S_n^\phi(\hat{\theta}_{E,n}, \theta_0)$ in (21), it may be concluded that all of them have a Taylor expansion of second order which does not depend on ϕ or h ,

$$T_n^{\phi,h}(\hat{\theta}_{E,n}, \theta_0) = \mathbf{U}_n(\hat{\theta}_{E,n}, \theta_0)^T \mathbf{U}_n(\hat{\theta}_{E,n}, \theta_0) + o_p(1), \quad \text{and} \quad S_n^{\phi,h}(\hat{\theta}_{E,n}, \theta_0) = \mathbf{U}_n(\hat{\theta}_{E,n}, \theta_0)^T \mathbf{U}_n(\hat{\theta}_{E,n}, \theta_0) + o_p(1),$$

with $\mathbf{U}_n(\hat{\theta}_{E,n}, \theta_0) = \sqrt{n} \mathbf{V}^{-1/2}(\theta_0) (\hat{\theta}_{E,n} - \theta_0) \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \mathcal{N}(\mathbf{0}_r, \mathbf{I}_p)$. With regard to their influence function, a second order influence function have to be considered since the first order one vanishes (see van der Vaart (2000, Section 20.1.1)), and its expression is given by

$$\begin{aligned} \mathcal{IF}_2 \left(\mathbf{x}, T_n^{\phi,h}(\hat{\theta}_{E,n}, \theta_0), F_{n,\theta_0} \right) &= \mathcal{IF}_2 \left(\mathbf{x}, S_n^{\phi,h}(\hat{\theta}_{E,n}, \theta_0), F_{n,\theta_0} \right) = \mathcal{IF}_2 \left(\mathbf{x}, \left\| \mathbf{U}_n(\hat{\theta}_{E,n}, \theta_0) \right\|^2, F_{n,\theta_0} \right) \quad (23) \\ &= 2 \left\| \mathbf{V}^{-1/2}(\theta_0) \mathcal{IF}(\mathbf{x}, \hat{\theta}_{E,n}, F_{n,\theta_0}) \right\|^2 = 2 \mathbf{g}(\mathbf{x}, \theta_0)^T \mathbf{S}_{11}^{-1}(\theta_0) \mathbf{S}_{12}(\theta_0) \mathbf{V}(\theta_0) \mathbf{S}_{12}(\theta_0)^T \mathbf{S}_{11}^{-1}(\theta_0) \mathbf{g}(\mathbf{x}, \theta_0). \end{aligned}$$

The first and second equality come from the previous second order Taylor expansion, the third one from Heritier and Ronchetti (1994) and the last one from Lemma 1. It suggests that under the simple null hypothesis, all members of both family of test-statistics have the same infinitesimal robustness. This is not a strange result, since in Lemma 1 of Toma (2009) a similar conclusion has been reached based on power divergence measures for parametric models.

4 Two approximations of power

Based on the null distribution presented in Theorem 3, we reject the null hypothesis in (8) in favour of the alternative hypothesis, if $S_n^\phi(\widehat{\boldsymbol{\theta}}_{E,n}, \boldsymbol{\theta}_0) > \chi_{p,\alpha}^2$, where $\chi_{p,\alpha}^2$ is the $(1 - \alpha)$ -th quantile of the chi-squared distribution with p degrees of freedom.

In most cases, the power function of this test procedure can not be derived explicitly. In the following theorem, we present an asymptotic result, which provides an approximation of the power of the test.

Theorem 7 *Let $\boldsymbol{\theta}^*$ be the true parameter value with $\boldsymbol{\theta}^* \neq \boldsymbol{\theta}_0$. Under the assumptions of Theorem 3 and $E_F \left[\sup_{\boldsymbol{\theta} \in \Theta} \|\mathbf{g}(\mathbf{X}, \boldsymbol{\theta})\|^2 \right] < \infty$, we have*

$$\sqrt{n} \left(D_\phi(F_{n,\widehat{\boldsymbol{\theta}}_{E,n}}, F_{n,\boldsymbol{\theta}_0}) - D_\phi(F_{n,\boldsymbol{\theta}^*}, F_{n,\boldsymbol{\theta}_0}) \right) \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \mathcal{N}(0, \sigma_\phi^2(\boldsymbol{\theta}^*, \boldsymbol{\theta}_0)),$$

where

$$\sigma_\phi^2(\boldsymbol{\theta}^*, \boldsymbol{\theta}_0) = \boldsymbol{\tau}_\phi(\boldsymbol{\theta}^*, \boldsymbol{\theta}_0)^T \mathbf{V}(\boldsymbol{\theta}_0) \boldsymbol{\tau}_\phi(\boldsymbol{\theta}^*, \boldsymbol{\theta}_0)$$

with

$$\boldsymbol{\tau}_\phi(\boldsymbol{\theta}^*, \boldsymbol{\theta}_0) = \left(\tau_1^\phi(\boldsymbol{\theta}^*, \boldsymbol{\theta}_0), \dots, \tau_p^\phi(\boldsymbol{\theta}^*, \boldsymbol{\theta}_0) \right)^T = \left. \frac{\partial}{\partial \boldsymbol{\theta}} D_\phi(F_{n,\boldsymbol{\theta}}, F_{n,\boldsymbol{\theta}_0}) \right|_{\boldsymbol{\theta}=\boldsymbol{\theta}^*}, \quad (24)$$

and the matrix $\mathbf{V}(\boldsymbol{\theta}_0)$ as defined in (15).

Proof. A first-order Taylor expansion gives

$$D_\phi(F_{n,\widehat{\boldsymbol{\theta}}_{E,n}}, F_{n,\boldsymbol{\theta}_0}) = D_\phi(F_{n,\boldsymbol{\theta}^*}, F_{n,\boldsymbol{\theta}_0}) + \boldsymbol{\tau}_\phi(\boldsymbol{\theta}^*, \boldsymbol{\theta}_0)^T (\widehat{\boldsymbol{\theta}}_{E,n} - \boldsymbol{\theta}^*) + o(\|\widehat{\boldsymbol{\theta}}_{E,n} - \boldsymbol{\theta}^*\|).$$

But,

$$\sqrt{n}(\widehat{\boldsymbol{\theta}}_{E,n} - \boldsymbol{\theta}^*) \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \mathcal{N}(\mathbf{0}_p, \mathbf{V}(\boldsymbol{\theta}_0)),$$

and so $\sqrt{n} o(\|\widehat{\boldsymbol{\theta}}_{E,n} - \boldsymbol{\theta}^*\|) = o_p(\mathbf{1}_r)$. Thus, the random variables

$$\sqrt{n} \left(D_\phi(F_{n,\widehat{\boldsymbol{\theta}}_{E,n}}, F_{n,\boldsymbol{\theta}_0}) - D_\phi(F_{n,\boldsymbol{\theta}^*}, F_{n,\boldsymbol{\theta}_0}) \right) \text{ and } \boldsymbol{\tau}_\phi(\boldsymbol{\theta}^*, \boldsymbol{\theta}_0) \sqrt{n}(\widehat{\boldsymbol{\theta}}_{E,n} - \boldsymbol{\theta}^*)$$

have the same asymptotic distribution, and hence the desired result. ■

Remark 8 *It is straightforward to extend Theorem 7 for (h, ϕ) -divergence measures $D_\phi^h(F_{n,\boldsymbol{\theta}}, F_{n,\boldsymbol{\theta}_0}) = h(D_\phi(F_{n,\boldsymbol{\theta}}, F_{n,\boldsymbol{\theta}_0}))$, as follows. Since $h(x) - h(x_0) = h'(x_0)(x - x_0) + o(x - x_0)$, we have*

$$\begin{aligned} & \sqrt{n} \left(D_\phi^h(F_{n,\widehat{\boldsymbol{\theta}}_{E,n}}, F_{n,\boldsymbol{\theta}_0}) - D_\phi^h(F_{n,\boldsymbol{\theta}^*}, F_{n,\boldsymbol{\theta}_0}) \right) \\ &= \frac{\partial}{\partial x} h(x) \Big|_{x=D_\phi^h(F_{n,\boldsymbol{\theta}^*}, F_{n,\boldsymbol{\theta}_0})} \left(D_\phi(F_{n,\widehat{\boldsymbol{\theta}}_{E,n}}, F_{n,\boldsymbol{\theta}_0}) - D_\phi(F_{n,\boldsymbol{\theta}^*}, F_{n,\boldsymbol{\theta}_0}) \right) + o_p(1), \end{aligned}$$

and so

$$\sqrt{n} \left(D_\phi^h(F_{n,\widehat{\boldsymbol{\theta}}_{E,n}}, F_{n,\boldsymbol{\theta}_0}) - D_\phi^h(F_{n,\boldsymbol{\theta}^*}, F_{n,\boldsymbol{\theta}_0}) \right) \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \mathcal{N}(0, \sigma_{h,\phi}^2(\boldsymbol{\theta}^*, \boldsymbol{\theta}_0)),$$

where

$$\sigma_{h,\phi}^2(\boldsymbol{\theta}^*, \boldsymbol{\theta}_0) = \kappa_{h,\phi}^2(\boldsymbol{\theta}^*, \boldsymbol{\theta}_0) \sigma_\phi^2(\boldsymbol{\theta}^*, \boldsymbol{\theta}_0),$$

$\kappa_{h,\phi}(\boldsymbol{\theta}^*, \boldsymbol{\theta}_0) = \frac{\partial}{\partial x} h(x) \Big|_{x=D_\phi^h(F_{n,\boldsymbol{\theta}^*}, F_{n,\boldsymbol{\theta}_0})}$ and $\sigma_\phi^2(\boldsymbol{\theta}^*, \boldsymbol{\theta}_0)$ as defined in Theorem 7.

Remark 9 From Theorem 7, we can present a first approximation of the power function, at $\boldsymbol{\theta}^* \neq \boldsymbol{\theta}_0$, of the test based on ϕ -divergence measure, as

$$\beta_{n,\phi}^1(\boldsymbol{\theta}^*, \boldsymbol{\theta}_0) = 1 - \Phi \left(\frac{\sqrt{n}}{\sigma_\phi(\boldsymbol{\theta}^*, \boldsymbol{\theta}_0)} \left(\frac{\phi''(1)\chi_{p,\alpha}^2}{2n} - D_\phi(F_{n,\boldsymbol{\theta}^*}, F_{n,\boldsymbol{\theta}_0}) \right) \right), \quad (25)$$

with $\Phi(x)$ being the standard normal distribution function.

If some alternative $\boldsymbol{\theta}^* \neq \boldsymbol{\theta}_0$ is the true parameter, then the probability of rejecting $\boldsymbol{\theta}_0$ with the rejection rule $S_n^\phi(\widehat{\boldsymbol{\theta}}_{E,n}, \boldsymbol{\theta}_0) > \chi_{p,\alpha}^2$, for fixed significance level α , tends to one as $n \rightarrow \infty$. Thus, the test is consistent in the sense of Fraser (1957). Similarly, from Theorem 7, we obtain a first approximation of the power function, at $\boldsymbol{\theta}^* \neq \boldsymbol{\theta}_0$, of the test based on (h, ϕ) -divergence measure as

$$\beta_{n,h,\phi}^1(\boldsymbol{\theta}^*, \boldsymbol{\theta}_0) = 1 - \Phi \left(\frac{\sqrt{n}}{\sigma_{h,\phi}(\boldsymbol{\theta}^*, \boldsymbol{\theta}_0)} \left(\frac{\phi''(1)\chi_{p,\alpha}^2}{2n} - D_\phi^h(F_{n,\boldsymbol{\theta}^*}, F_{n,\boldsymbol{\theta}_0}) \right) \right). \quad (26)$$

To produce some less trivial asymptotic powers that are not all equal to 1, we can use a Pitman-type local analysis, as developed by Le Cam (1960), by confining attention to $n^{1/2}$ -neighborhoods of the true parameter values. A key tool to get the asymptotic distribution of the statistic $S_n^\phi(\widehat{\boldsymbol{\theta}}_{E,n}, \boldsymbol{\theta}_0)$ under such a contiguous hypothesis is Le Cam's third lemma, as presented in Hájek and Sidák (1967). Instead of relying on these results, we present in the following theorem a proof which is easy and direct to follow. This proof is based on the results of Morales and Pardo (2001). Specifically, we consider the power at contiguous alternatives of the form

$$H_{1,n} : \boldsymbol{\theta}_n = \boldsymbol{\theta}_0 + n^{-1/2} \mathbf{f}, \quad (27)$$

where \mathbf{f} is a fixed vector in \mathbb{R}^p such that $\boldsymbol{\theta}_n \in \Theta \subset \mathbb{R}^p$.

Theorem 10 Under the assumptions of Theorem 3 and $H_{1,n}$ in (27), we have the asymptotic distribution of the empirical ϕ -divergence test statistic $S_n^\phi(\widehat{\boldsymbol{\theta}}_{E,n}, \boldsymbol{\theta}_0)$ to be non-central chi-squared with p degrees of freedom and non-centrality parameter

$$S_n^\phi(\widehat{\boldsymbol{\theta}}_{E,n}, \boldsymbol{\theta}_0) \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \chi_p^2(\delta(\boldsymbol{\theta}_0)),$$

where

$$\delta(\boldsymbol{\theta}_0) = \mathbf{f}^T \mathbf{V}^{-1}(\boldsymbol{\theta}_0) \mathbf{f}. \quad (28)$$

Proof. We can write

$$\sqrt{n}(\widehat{\boldsymbol{\theta}}_{E,n} - \boldsymbol{\theta}_0) = \sqrt{n}(\widehat{\boldsymbol{\theta}}_{E,n} - \boldsymbol{\theta}_n) + \sqrt{n}(\boldsymbol{\theta}_n - \boldsymbol{\theta}_0) = \sqrt{n}(\widehat{\boldsymbol{\theta}}_{E,n} - \boldsymbol{\theta}_n) + \mathbf{f}.$$

Under $H_{1,n}$, we have

$$\sqrt{n}(\widehat{\boldsymbol{\theta}}_{E,n} - \boldsymbol{\theta}_n) \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \mathcal{N}(\mathbf{0}, \mathbf{V}(\boldsymbol{\theta}_0))$$

and

$$\sqrt{n}(\widehat{\boldsymbol{\theta}}_{E,n} - \boldsymbol{\theta}_0) \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \mathcal{N}(\mathbf{f}, \mathbf{V}(\boldsymbol{\theta}_0)).$$

In Theorem 3, it has been shown that

$$S_n^\phi(\widehat{\boldsymbol{\theta}}_{E,n}, \boldsymbol{\theta}_0) = \left(\mathbf{V}(\boldsymbol{\theta}_0)^{-1/2} \sqrt{n}(\widehat{\boldsymbol{\theta}}_{E,n} - \boldsymbol{\theta}_0) \right)^T \mathbf{V}(\boldsymbol{\theta}_0)^{-1/2} \sqrt{n}(\widehat{\boldsymbol{\theta}}_{E,n} - \boldsymbol{\theta}_0) + o_p(1).$$

On the other hand, we have

$$\mathbf{V}(\boldsymbol{\theta}_0)^{-1/2} \sqrt{n}(\widehat{\boldsymbol{\theta}}_{E,n} - \boldsymbol{\theta}_0) \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \mathcal{N}(\mathbf{V}(\boldsymbol{\theta}_0)^{-1/2} \mathbf{f}, \mathbf{I}_p).$$

We thus obtain

$$S(F_{n,\widehat{\theta}_{E,n}}, F_{n,\theta_0}) \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \chi_p^2(\delta(\theta_0)),$$

with $\delta(\theta_0)$ as in (28). ■

Remark 11 Notice that $D_\phi^h(F_{n,\widehat{\theta}_{E,n}}, F_{n,\theta_0})$ and $h'(0)D_\phi(F_{n,\widehat{\theta}_{E,n}}, F_{n,\theta_0})$ have the same asymptotic distribution under the contiguous alternative hypothesis in (27). Hence, it is straightforward to extend Theorem 10 for $T_n^{\phi,h}(\widehat{\theta}_{E,n}, \theta_0)$ and to conclude that both test statistics, $T_n^{\phi,h}(\widehat{\theta}_{E,n}, \theta_0)$ and $S_n^\phi(\widehat{\theta}_{E,n}, \theta_0)$, have the same asymptotic distribution under the contiguous alternative hypothesis.

Remark 12 Using Theorem 10 and the preceding remark, we obtain a second approximation to the power function, at $\theta_n = \theta_0 + n^{-1/2}\mathbf{f}$, as

$$\begin{aligned} \beta_n^2(\theta_n) &= \Pr(S_n^\phi(\widehat{\theta}_{E,n}, \theta_0) > \chi_{p,\alpha}^2 | \theta = \theta_n) \\ &= \Pr(T_n^{\phi,h}(\widehat{\theta}_{E,n}, \theta_0) > \chi_{p,\alpha}^2 | \theta = \theta_n) = 1 - F_{\chi_p^2(\delta(\theta_0))}(\chi_{p,\alpha}^2), \end{aligned} \quad (29)$$

where $F_{\chi_p^2(\delta(\theta_0))}$ is the distribution function of $\chi_p^2(\delta(\theta_0))$.

If we want to approximate the power at the alternative $\theta \neq \theta_0$, then we can take $\mathbf{f} = \mathbf{f}(n, \theta, \theta_0) = n^{1/2}(\theta - \theta_0)$. It should be noted that this approximation is independent of the function ϕ in Φ , as well as of the h function.

5 Analysis of Newcomb's measurements on the passage time of light

In this section, we make use of a well-known data to illustrate the application of the empirical phi-divergence test statistics for testing a simple null hypothesis. In 1882, the astronomer and mathematician Simon Newcomb, measured the time required for a light signal to pass from his laboratory on the Potomac River to a mirror at the base of the Washington Monument and back, a distance of 7443.73 meters. Table 2 contains these measurements from three samples, as deviations from 24, 800 nanoseconds. For example, for the first observation, 28 represents the 24, 828 nanoseconds that were spent for the light to travel the required 7443.73 meters. The data contain three samples, of sizes 20, 20 and 26, respectively, corresponding to three different days. They have been analyzed previously by a number of authors including Stigler (1973) and Voinov et al. (2013).

Our interest here is in obtaining confidence intervals for the mean, μ , of the passage time of light, X . This means that a unique estimating function, $g(X, \mu) = X - \mu$, is required for a univariate random variable ($k = 1$). Note that $r = p = 1$, and so $t(\widehat{\mu}_{E,n}) = 0$. Based on the results of Theorem 5, we shall use the empirical Rényi test statistics

$$S_{n,a}^{\text{Rényi}}(\mu_0) = S_{n,a}^{\text{Rényi}}(F_{n,\widehat{\mu}_{E,n}}, F_{n,\mu_0}) = \begin{cases} \frac{2n}{a(a-1)} \log \sum_{i=1}^n \frac{1}{n} (1 + t(\mu_0)(X_i - \mu_0))^{a-1}, & a \neq 0, 1, \\ -2 \sum_{i=1}^n \frac{\log(1 + t(\mu_0)(X_i - \mu_0))}{1 + t(\mu_0)(X_i - \mu_0)}, & a = 0, \\ 2 \sum_{i=1}^n \log(1 + t(\mu_0)(X_i - \mu_0)) & a = 1, \end{cases} \quad (30)$$

where $t(\mu_0)$ is the solution of the equation $\frac{1}{n} \sum_{i=1}^n \frac{X_i - \mu_0}{1 + t(\mu_0)(X_i - \mu_0)} = 0$ subject to $\min\{t(\mu_0)(X_i - \mu_0)\}_{i=1,\dots,n} > \frac{1-n}{n}$ and $\frac{1}{n} \sum_{i=1}^n \frac{1}{1 + t(\mu_0)(X_i - \mu_0)} = 1$.

A confidence interval, with confidence level $1 - \alpha$, based on (30) is given by

$$CI_{1-\alpha}^a(\mu) = \{\mu \in \mathbb{R} : S_{n,a}^{\text{Rényi}}(\mu) \leq \chi_{1,\alpha}^2\},$$

with $\chi_{1,\alpha}^2$ being the (right hand) quantile of order α for the χ_1^2 distribution. By taking different choices for the parameter a as $\{-1, -0.5, 0, 0.5, 1, 1.5, 2.5\}$, we obtain different confidence intervals, and this is done for

day 1	28	26	33	24	34	-44	27	16	40	-2	29	22	24	21	25	30	23	29	31	19						
day 2	24	20	36	32	36	28	25	21	28	29	37	25	28	26	30	32	36	26	30	22						
day 3	36	23	27	27	28	27	31	27	26	33	26	32	32	24	39	28	24	25	32	25	29	27	28	29	16	23

Table 2: Newcomb's data on the passage time of light.

each of the three days. Table 3 summarizes these results obtained at 95% confidence level. If the confidence interval based on the Likelihood Ratio test statistic, i.e., $CI_{0.95}^1(\mu)$ are compared with $CI_{0.95}^a(\mu)$, we observe that narrower confidence intervals are obtained for $a < 1$, and wider confidence intervals are obtained for $a > 1$. This behavior is more evident for the first day, due to the presence of some outliers in the sample; see for example, Bennett and Lewis (1993). Next, we approximate the power function, assuming that μ_0 is a value very close to $\hat{\mu}_{E,n}$. Despite the fact that the exact significance may not be approximated well (since the true value μ_0 is unknown), the power values can provide meaningful information on the quality of performance of the different Rényi test statistics.

	$CI_{0.95}^{-1}(\mu)$	$CI_{0.95}^{-0.5}(\mu)$	$CI_{0.95}^0(\mu)$	$CI_{0.95}^{0.5}(\mu)$	$CI_{0.95}^1(\mu)$	$CI_{0.95}^{1.5}(\mu)$	$CI_{0.95}^{2.5}(\mu)$
day 1	(14.22, 27.24)	(13.65, 27.25)	(10.85, 27.11)	(12.00, 27.24)	(12.92, 27.24)	(9.46, 26.97)	(6.90, 26.60)
day 2	(26.48, 30.70)	(26.47, 30.70)	(26.40, 30.78)	(26.43, 30.74)	(26.45, 30.71)	(26.35, 30.83)	(26.24, 30.97)
day 3	(26.24, 29.42)	(26.27, 29.48)	(26.00, 29.70)	(26.10, 29.61)	(26.20, 29.54)	(25.88, 29.79)	(25.62, 30.01)

Table 3: Confidence intervals, based on the empirical Rényi test statistics, for Newcomb's data for the three days.

In order to calculate the power function by using the first approximation

$$\beta_{n,a}^{1,\text{Rényi}}(\mu^*) = 1 - \Phi \left(\frac{\sqrt{n}}{\sigma_{\text{Rényi}}^a(\mu_0, \mu^*)} \left(\frac{\chi_{1,\alpha}^2}{2n} - D_{\text{Rényi}}^a(F_{n,\mu^*}, F_{n,\mu_0}) \right) \right),$$

we need to obtain

$$\sigma_{\text{Rényi}}^a(\mu_0, \mu^*) = \kappa_{\text{Rényi}}^a(\mu_0, \mu^*) |\tau_{\text{Rényi}}^a(\mu_0, \mu^*)| V^{\frac{1}{2}}(\mu_0)$$

in terms of

$$\begin{aligned} \kappa_{\text{Rényi}}^a(\mu_0, \mu^*) &= \frac{\partial}{\partial x} h(x)|_{x=D_{\phi_a}(F_{n,\mu^*}, F_{n,\mu_0})} = \begin{cases} \frac{1}{a(a-1)D_{\phi_a}(F_{n,\mu^*}, F_{n,\mu_0})+1}, & a(a-1) \neq 0, \\ 1, & a(a-1) = 0, \end{cases} \\ &= \left(\sum_{i=1}^n p_i^a(\mu^*) p_i^{1-a}(\mu_0) \right)^{-1}, \end{aligned}$$

$$p_i(\mu) = \frac{1}{n} \frac{1}{1 + t(\mu)(X_i - \mu)}, \quad \mu \in \{\mu_0, \mu^*\},$$

$$\tau_{\text{Rényi}}^a(\mu_0, \mu^*) = \begin{cases} \frac{n}{1-a} \sum_{i=1}^n [t'(\mu^*)(X_i - \mu^*) - t(\mu^*)] p_i^{a+1}(\mu^*) p_i^{1-a}(\mu_0), & a \neq 1, \\ n \sum_{i=1}^n [t'(\mu^*)(X_i - \mu^*) - t(\mu^*)] p_i^2(\mu^*) \left(\log \frac{p_i(\mu_0)}{p_i(\mu^*)} - 1 \right), & a = 1, \end{cases}$$

$$t'(\mu^*) = - \left(\sum_{j=1}^n p_j^2(\mu^*) \right) \left(\sum_{j=1}^n (X_j - \mu^*)^2 p_j^2(\mu^*) \right)^{-1},$$

and

$$V(\mu_0) = (S_{12}^T(\mu_0) S_{11}^{-1}(\mu_0) S_{12}(\mu_0))^{-1} = E[(X - \mu_0)^2].$$

Since $V(\mu_0)$ is unknown, it can be replaced by a consistent estimator

$$\tilde{V}(\mu_0) = \sum_{j=1}^n (X_j - \mu_0)^2 p_j(\mu_0),$$

so that

$$\tilde{\beta}_{n,a}^1(\mu^*) = 1 - \Phi\left(\frac{\sqrt{n}}{\tilde{\sigma}_{\text{Rényi}}^a(\mu_0, \mu^*)} \left(\frac{\chi_{1,\alpha}^2}{2n} - D_{\text{Rényi}}^a(F_{n,\mu^*}, F_{n,\mu_0})\right)\right),$$

where

$$\begin{aligned} \tilde{\sigma}_{\text{Rényi}}^a(\mu_0, \mu^*) &= \kappa_{\text{Rényi}}^a(\mu_0, \mu^*) |\tau_{\text{Rényi}}^a(\mu_0, \mu^*)| \tilde{V}^{\frac{1}{2}}(\mu_0), \\ D_{\text{Rényi}}^a(F_{n,\mu^*}, F_{n,\mu_0}) &= \begin{cases} -\frac{1}{a(a-1)} \log \kappa_{\text{Rényi}}^a(\mu_0, \mu^*), & a(a-1) \neq 0, \\ \sum_{i=1}^n p_i(\mu_0) \log \frac{p_i(\mu_0)}{p_i(\mu^*)}, & a = 0, \\ \sum_{i=1}^n p_i(\mu^*) \log \frac{p_i(\mu^*)}{p_i(\mu_0)}, & a = 1. \end{cases} \end{aligned}$$

In Figure 1, the power functions are plotted for the three days, for three different values of the parameter $a \in \{-1, 1, 2.5\}$. We do observe from these plots that the most powerful test statistics are for the case when $a > 1$, which is in fact the case when the confidence intervals are wider.

Notice that there are two outliers on Day 1 (-44 and -2) and one possible outlier on Day 3 (16); see Stigler (1973) and Barnett and Lewis (1994). By dropping these outliers, we reproduced the plots of approximated power functions for Days 1 and 3, and we found that in the new plots the approximated power functions are more similar to that of Day 2. In the presence of outliers (Days 1 and 3), the comparison of shape of the power function for different values provides an interesting information: the power function is flatter for $S_{n,a}^{\text{Rényi}}(\mu)$, with $a = -1$ in comparison with $a = 1, 2$. This tends to be associated with higher capacity of the empirical likelihood ratio test ($S_{n,a}^{\text{Rényi}}(\mu)$, with $a = 1$) for detecting samples not fulfilling the null hypothesis, but a lack of robustness in the presence of outliers. In Section 6.2, a simulation study is carried out to examine this robustness aspect.

6 Simulation study

6.1 Evaluation of procedures under normality

In this section, we will pay special attention to a subfamily of ϕ -divergence measures in (11), the so-called power divergence measures (see Cressie and Read (1984)) for which $\phi(x) \equiv \phi_\lambda(x) \in \Phi$ is given by

$$\begin{aligned} \phi_\lambda(x) &= \frac{1}{\lambda(1+\lambda)} [x^{\lambda+1} - x - \lambda(x-1)], \quad \text{if } \lambda(1+\lambda) \neq 0, \\ \phi_0(x) &= \lim_{\lambda \rightarrow 0} \phi_\lambda(x) = x \log x - x + 1, \\ \phi_{-1}(x) &= \lim_{\lambda \rightarrow -1} \phi_\lambda(x) = -\log x + x - 1. \end{aligned} \tag{31}$$

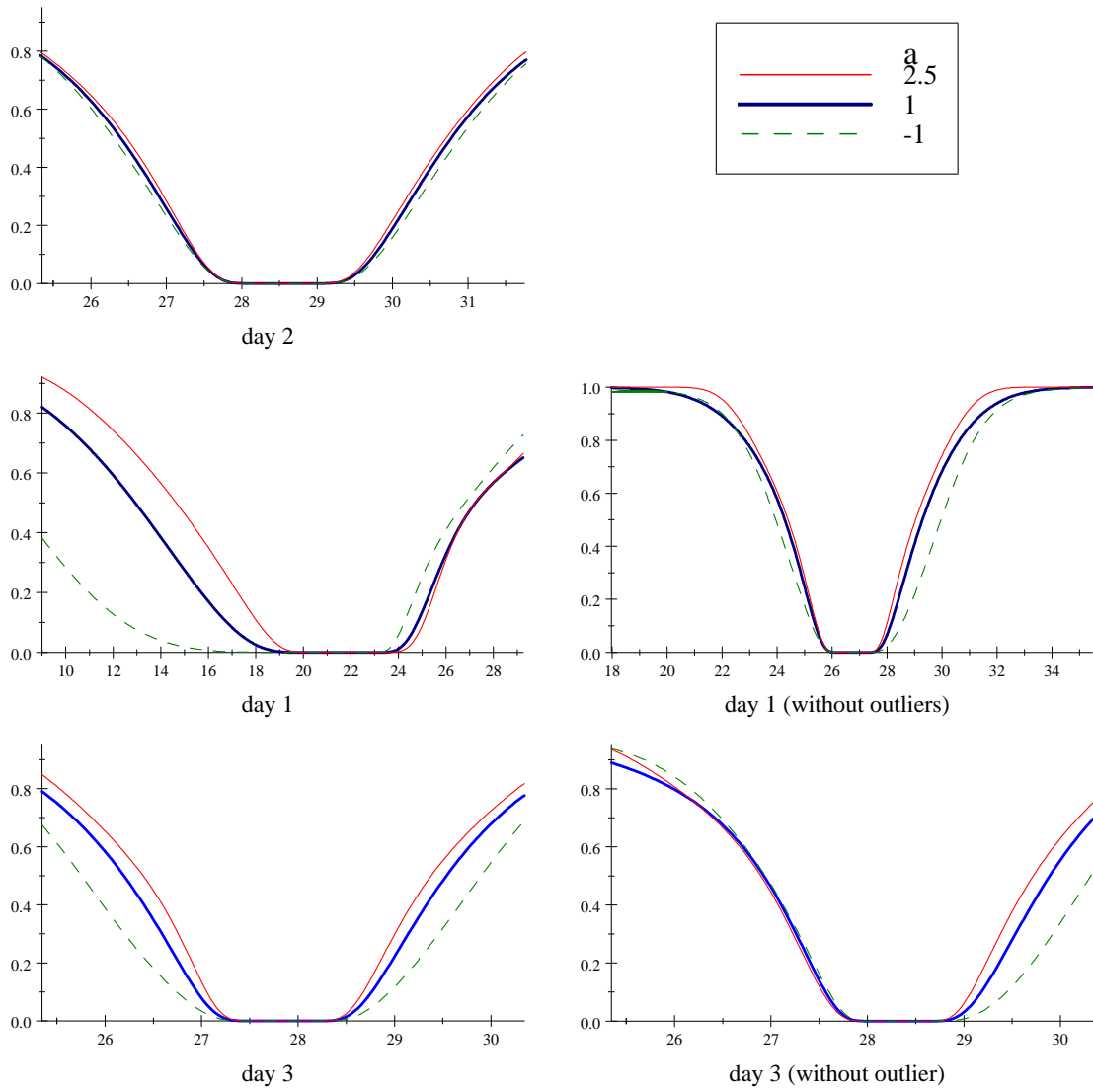


Figure 1: Approximate power functions $\tilde{\beta}_{n,a}^1(\mu^*)$ for Newcomb's measurements for three days.

In this case, (12) can be expressed as

$$\begin{aligned}
T_n^\lambda(\widehat{\boldsymbol{\theta}}_{E,n}, \boldsymbol{\theta}_0) &= \frac{2}{\lambda(1+\lambda)} \left(\sum_{i=1}^n (1 + \mathbf{t}(\boldsymbol{\theta}_0)^T \mathbf{g}(\mathbf{X}_i, \boldsymbol{\theta}_0))^\lambda - \sum_{i=1}^n (1 + \mathbf{t}(\widehat{\boldsymbol{\theta}}_{E,n})^T \mathbf{g}(\mathbf{X}_i, \widehat{\boldsymbol{\theta}}_{E,n}))^\lambda \right), \quad \text{if } \lambda(1+\lambda) \neq 0, \\
&= 2 \sum_{i=1}^n \log \left(\frac{1 + \mathbf{t}(\boldsymbol{\theta}_0)^T \mathbf{g}(\mathbf{X}_i, \boldsymbol{\theta}_0)}{1 + \mathbf{t}(\widehat{\boldsymbol{\theta}}_{E,n})^T \mathbf{g}(\mathbf{X}_i, \widehat{\boldsymbol{\theta}}_{E,n})} \right) = L_{E,n}(\widehat{\boldsymbol{\theta}}_E, \boldsymbol{\theta}_0), \quad \text{if } \lambda = 0, \\
&= 2 \left(\sum_{i=1}^n \frac{\log(1 + \mathbf{t}(\widehat{\boldsymbol{\theta}}_{E,n})^T \mathbf{g}(\mathbf{X}_i, \widehat{\boldsymbol{\theta}}_{E,n}))}{1 + \mathbf{t}(\widehat{\boldsymbol{\theta}}_{E,n})^T \mathbf{g}(\mathbf{X}_i, \widehat{\boldsymbol{\theta}}_{E,n})} - \sum_{i=1}^n \frac{\log(1 + \mathbf{t}(\boldsymbol{\theta}_0)^T \mathbf{g}(\mathbf{X}_i, \boldsymbol{\theta}_0))}{1 + \mathbf{t}(\boldsymbol{\theta}_0)^T \mathbf{g}(\mathbf{X}_i, \boldsymbol{\theta}_0)} \right), \quad \text{if } \lambda = -1.
\end{aligned} \tag{32}$$

This family of test statistics include several commonly used tests such as the empirical modified Kullback-Leibler statistic ($T_n^\lambda(\widehat{\boldsymbol{\theta}}_{E,n}, \boldsymbol{\theta}_0)$, with $\lambda = -1$), the empirical Freeman-Tukey statistic ($T_n^\lambda(\widehat{\boldsymbol{\theta}}_{E,n}, \boldsymbol{\theta}_0)$, with $\lambda = -\frac{1}{2}$), the empirical likelihood ratio test statistic ($T_n^\lambda(\widehat{\boldsymbol{\theta}}_{E,n}, \boldsymbol{\theta}_0)$, with $\lambda = 0$), the empirical Cressie-Read statistic ($T_n^\lambda(\widehat{\boldsymbol{\theta}}_{E,n}, \boldsymbol{\theta}_0)$, with $\frac{2}{3}$), and the empirical Pearson's chi-square statistic ($T_n^\lambda(\widehat{\boldsymbol{\theta}}_{E,n}, \boldsymbol{\theta}_0)$, with $\lambda = 1$). Similarly, (13) can be expressed as

$$\begin{aligned}
S_n^\lambda(\widehat{\boldsymbol{\theta}}_{E,n}, \boldsymbol{\theta}_0) &= \frac{2}{\lambda(1+\lambda)} \left(\sum_{i=1}^n \frac{(1 + \mathbf{t}(\boldsymbol{\theta}_0)^T \mathbf{g}(\mathbf{X}_i, \boldsymbol{\theta}_0))^\lambda}{(1 + \mathbf{t}(\widehat{\boldsymbol{\theta}}_{E,n})^T \mathbf{g}(\mathbf{X}_i, \widehat{\boldsymbol{\theta}}_{E,n}))^{\lambda+1}} - n \right), \quad \text{if } \lambda(1+\lambda) \neq 0, \\
&= 2 \sum_{i=1}^n \frac{1}{1 + \mathbf{t}(\widehat{\boldsymbol{\theta}}_{E,n})^T \mathbf{g}(\mathbf{X}_i, \widehat{\boldsymbol{\theta}}_{E,n})} \log \left(\frac{1 + \mathbf{t}(\boldsymbol{\theta}_0)^T \mathbf{g}(\mathbf{X}_i, \boldsymbol{\theta}_0)}{1 + \mathbf{t}(\widehat{\boldsymbol{\theta}}_{E,n})^T \mathbf{g}(\mathbf{X}_i, \widehat{\boldsymbol{\theta}}_{E,n})} \right), \quad \text{if } \lambda = 0, \\
&= 2 \sum_{i=1}^n \frac{1}{1 + \mathbf{t}(\boldsymbol{\theta}_0)^T \mathbf{g}(\mathbf{X}_i, \boldsymbol{\theta}_0)} \log \left(\frac{1 + \mathbf{t}(\widehat{\boldsymbol{\theta}}_{E,n})^T \mathbf{g}(\mathbf{X}_i, \widehat{\boldsymbol{\theta}}_{E,n})}{1 + \mathbf{t}(\boldsymbol{\theta}_0)^T \mathbf{g}(\mathbf{X}_i, \boldsymbol{\theta}_0)} \right), \quad \text{if } \lambda = -1.
\end{aligned} \tag{33}$$

The r -dimensional vectors $\mathbf{t}(\boldsymbol{\theta}_0)$ and $\mathbf{t}(\widehat{\boldsymbol{\theta}}_{E,n})$ are obtained by solving (6) with $\boldsymbol{\theta}_0$ and $\widehat{\boldsymbol{\theta}}_{E,n}$ being specified.

We now examine, through Monte Carlo simulations, the performance of the confidence intervals (CIs) obtained through the empirical power divergence test statistics in (32) and (33) for the choices $\lambda \in \{-1, -\frac{1}{2}, 0, \frac{2}{3}, 1\}$, when the sample sizes are small, $n = 30, 45, 60$, and the nominal confidence levels are 90% and 95%. Even though the asymptotic distribution is equivalent for the different empirical divergence based statistics, in practice we need to work with finite sample sizes and it is therefore important to compare the CIs based on $T_n^\lambda(\widehat{\boldsymbol{\theta}}_{E,n}, \boldsymbol{\theta}_0)$ and $S_n^\lambda(\widehat{\boldsymbol{\theta}}_{E,n}, \boldsymbol{\theta}_0)$, in different scenarios. We focus on the coverage probability and average width of the CIs of a mean based on 5000 simulated samples under the same setting as Example 1 of Qin and Lawless (1994). In this case, a sample of univariate ($k = 1$) i.i.d. random variables, X_1, \dots, X_n , of size n was considered, with mean θ and variance $\theta^2 + 1$, i.e., $r = 2 > p = 1$ and

$$\mathbf{g}(X, \theta) = (g_1(X, \theta), g_2(X, \theta))^T = (X - \theta, X^2 - 2\theta^2 - 1)^T.$$

In the present study, we have considered an additional sample of size $n = 45$, compared to those of Qin and Lawless (1994). For constructing the test statistic, an unknown distribution of X_i is assumed, but for this simulation study, normally distributed random variables were taken with $\theta_0 \in \{0, 1\}$. It is important to mention here that Baggerly (1998) studied the coverage probabilities for the normal distribution, but only for the family of statistics in (32). The results of the simulation study are summarized in Table 4 and Figure 2 for the empirical power divergence test statistics in (32) and (33) for two different samples when the nominal confidence level is 95%. In Figure 2, it can be seen that the confidence intervals based on both statistics, $T_n^\lambda(\widehat{\boldsymbol{\theta}}_{E,n}, \boldsymbol{\theta}_0)$ and $S_n^\lambda(\widehat{\boldsymbol{\theta}}_{E,n}, \boldsymbol{\theta}_0)$, with small values of $\lambda \in \{-1, -\frac{1}{2}, 0, \frac{2}{3}, 1\}$ tend to be narrower in width, and this is also seen in the simulation results of Table 4. In fact, the simulation study of Baggerly showed that $T_n^{-1}(\widehat{\boldsymbol{\theta}}_{E,n}, \boldsymbol{\theta}_0)$ had the worst

$\mathcal{N}(\theta, \theta^2 + 1), \theta_0 = 0$						$\mathcal{N}(\theta, \theta^2 + 1), \theta_0 = 1$					
		$T_n^\lambda(\hat{\theta}_{E,n}, \theta_0)$		$S_n^\lambda(\hat{\theta}_{E,n}, \theta_0)$				$T_n^\lambda(\hat{\theta}_{E,n}, \theta_0)$		$S_n^\lambda(\hat{\theta}_{E,n}, \theta_0)$	
		Cov (%)	Avw	Cov (%)	Avw			Cov (%)	Avw	Cov (%)	Avw
$n = 30$						$n = 30$					
90%	λ					90%	λ				
	-1	86.36	0.524	88.35	0.540		-1	83.59	0.627	79.37	0.542
	$-\frac{1}{2}$	87.11	0.542	87.98	0.551		$-\frac{1}{2}$	84.64	0.635	79.73	0.547
	0	87.77	0.557	87.64	0.560		0	85.16	0.642	80.11	0.551
	$\frac{2}{3}$	87.46	0.571	86.88	0.570		$\frac{2}{3}$	83.01	0.608	80.51	0.555
	1	87.19	0.575	86.55	0.573		1	82.81	0.605	80.39	0.556
95%	-1	91.97	0.617	93.85	0.637	95%	-1	88.70	0.728	84.36	0.645
	$-\frac{1}{2}$	92.67	0.642	93.77	0.654		$-\frac{1}{2}$	89.83	0.740	85.08	0.652
	0	93.05	0.663	93.48	0.669		0	90.39	0.749	85.52	0.657
	$\frac{2}{3}$	93.05	0.684	92.76	0.684		$\frac{2}{3}$	89.08	0.721	85.88	0.663
	1	92.88	0.690	92.28	0.689		1	88.62	0.714	85.82	0.664
$n = 45$						$n = 45$					
90%	λ					90%	λ				
	-1	87.62	0.445	88.95	0.455		-1	85.13	0.476	84.44	0.460
	$-\frac{1}{2}$	88.45	0.457	88.85	0.462		$-\frac{1}{2}$	86.03	0.482	85.19	0.464
	0	88.83	0.469	88.66	0.468		0	86.46	0.488	85.73	0.468
	$\frac{2}{3}$	88.70	0.480	88.11	0.474		$\frac{2}{3}$	86.20	0.485	86.03	0.473
	1	88.51	0.484	87.81	0.476		1	86.20	0.487	86.01	0.475
95%	-1	93.09	0.526	94.42	0.538	95%	-1	90.37	0.563	89.73	0.546
	$-\frac{1}{2}$	93.43	0.544	94.34	0.550		$-\frac{1}{2}$	91.12	0.571	90.18	0.553
	0	94.02	0.559	94.13	0.559		0	91.45	0.579	90.94	0.559
	$\frac{2}{3}$	94.23	0.575	93.51	0.569		$\frac{2}{3}$	91.53	0.580	91.02	0.567
	1	93.81	0.581	93.22	0.573		1	91.62	0.582	91.19	0.570
$n = 60$						$n = 60$					
90%	λ					90%	λ				
	-1	87.19	0.394	88.47	0.401		-1	85.46	0.413	84.91	0.403
	$-\frac{1}{2}$	87.86	0.404	88.45	0.405		$-\frac{1}{2}$	85.99	0.417	85.46	0.406
	0	88.36	0.413	88.13	0.409		0	86.28	0.421	85.80	0.410
	$\frac{2}{3}$	88.28	0.421	87.63	0.413		$\frac{2}{3}$	86.37	0.423	85.90	0.415
	1	88.03	0.425	87.29	0.415		1	86.39	0.425	86.09	0.417
95%	-1	93.22	0.467	93.85	0.475	95%	-1	91.02	0.489	90.83	0.479
	$-\frac{1}{2}$	93.45	0.481	93.72	0.483		$-\frac{1}{2}$	92.08	0.496	91.80	0.485
	0	93.95	0.493	93.55	0.489		0	92.58	0.502	92.25	0.490
	$\frac{2}{3}$	93.78	0.505	93.13	0.496		$\frac{2}{3}$	92.92	0.506	92.56	0.498
	1	93.70	0.509	92.90	0.498		1	92.86	0.509	92.54	0.501

Table 4: Coverage and average width of 90% and 95% confidence intervals.

coverage probability, and this is also seen in the present simulation study. This is in accordance with the results obtained for two specific samples analyzed in Figure 2. The smallest average width of confidence intervals is obtained for $S_n^{-1}(\hat{\theta}_{E,n}, \theta_0)$, which is the so-called empirical modified likelihood ratio test statistic or empirical minimum discrimination information statistic (see Gokhale and Kullback, 1978), and this might explain, when $\theta_0 = 0$, the low coverage probability as well. Often, the coverage probability closest to the nominal level is also obtained for the same test statistics, $S_n^{-1}(\hat{\theta}_{E,n}, \theta_0)$ when $\theta_0 = 1$ but the empirical likelihood ratio test has a closer coverage probability when $\theta_0 = 0$. When both characteristics of a CI, viz., average width and coverage probability, are taken into account, the empirical modified likelihood ratio test statistic, $S_n^{-1}(\hat{\theta}_{E,n}, \theta_0)$, is a good compromise and it turns out to be the best for $\theta_0 = 0$ in all the cases considered. Finally, another important feature seen in Figure 2 is that even though the support of the asymptotic distribution of the test statistics is strictly positive, it is possible to find samples for which $T_n^\lambda(\hat{\theta}_{E,n}, \theta_0)$ is negative, while $S_n^\lambda(\hat{\theta}_{E,n}, \theta_0)$ is always strictly positive. This usually happens when $\hat{\theta}_{E,n}$ is very close to θ_0 .

6.2 Robustness of procedures to contamination in the data

We conducted a simulation study using the same design as in the preceding subsection, but by considering $\frac{15}{100}\%$ of shifted observations in each sample: 2 shifted observations out of 30 in the sample, 3 shifted observations out of 45, 4 shifted observations out of 60. The underlying distribution for shifted observations is taken to be $\mathcal{N}(\theta + \sqrt{\theta^2 + 1}, \theta^2 + 1)$, given that $\mathcal{N}(\theta, \theta^2 + 1)$ is the true distribution. Therefore, the shifted observations follow $\mathcal{N}(1, 1)$ distribution when the true distribution is $\mathcal{N}(0, 1)$, while they follow $\mathcal{N}(2.4142, 2)$ distribution when the true distribution is $\mathcal{N}(1, 2)$. In Table 5, the results of introducing $\frac{15}{100}\%$ of shifted observations is shown. Our interest is focused on identifying the statistics that are less sensitive (robust statistics with respect to shifted observations) as well as those that are quite sensitive (non-robust statistics with respect to shifted observations). As pointed out in Remark 6, all the test-statistics have the same infinitesimal robustness, so the differences in robustness could be attributable to the magnitude of the shift. When $\theta_0 = 0$, the best statistics with shifted observations are exactly the same as ones without shifted observations, but when the dispersion of the data is higher, i.e. $\theta_0 = 1$, the closest coverage probability to the nominal level is not for the empirical likelihood ratio test; we further observe that $T_n^\lambda(\hat{\theta}_{E,n}, \theta_0)$ and $S_n^\lambda(\hat{\theta}_{E,n}, \theta_0)$ not only have the narrowest CIs but also the highest coverage probabilities. This means that the empirical modified likelihood ratio test statistic, $S_n^{-1}(\hat{\theta}_{E,n}, \theta_0)$, is a robust statistic with respect to shifted observations in comparison to the empirical likelihood ratio test, with respect to the coverage probability and width of CIs. This also agrees with the conclusion obtained from Table 3.

7 Testing for a composite null hypothesis

Now, let us consider the problem of testing the composite null hypothesis

$$H_0 : \mathbf{c}(\boldsymbol{\theta}) = \mathbf{0}_q . \quad (34)$$

The rest of this section, proceeds as follows. In Section 7.1, we introduce the empirical phi-divergence test statistic for the composite null hypothesis in (34). Section 7.2 is devoted to the asymptotic results. In Section 7.3, we present some results regarding the power function of the family of empirical test statistics proposed here. In Section 7.4, a simulation study is carried out to evaluate the performance of the proposed test procedure in comparison to some other tests.

7.1 Empirical phi-divergence test statistics

In the following, we shall assume as in Qin and Lawless (1995), that the unknown parameter vector $\boldsymbol{\theta}$ is defined through $r = p$ estimating functions, given in (1). With regard to (34), we shall assume that $\mathbf{c} : \mathbb{R}^p \rightarrow \mathbb{R}^q$

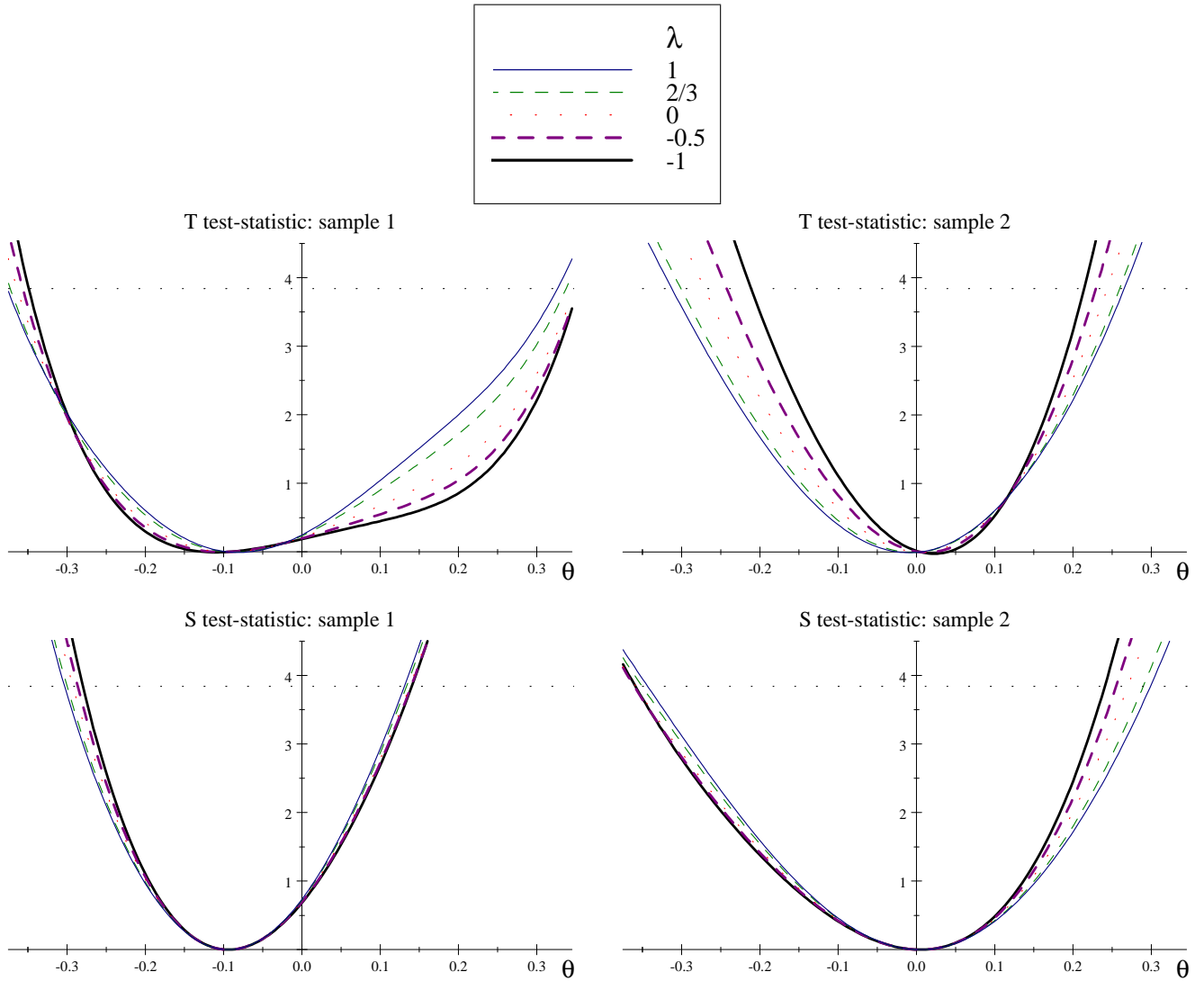


Figure 2: Test statistics for samples 1 and 2 of size 30 from the standard normal population.

$\mathcal{N}(\theta, \theta^2 + 1), \theta_0 = 0$						$\mathcal{N}(\theta, \theta^2 + 1), \theta_0 = 1$					
		$T_n^\lambda(\hat{\theta}_{E,n}, \theta_0)$		$S_n^\lambda(\hat{\theta}_{E,n}, \theta_0)$				$T_n^\lambda(\hat{\theta}_{E,n}, \theta_0)$		$S_n^\lambda(\hat{\theta}_{E,n}, \theta_0)$	
		Cov (%)	Avw	Cov (%)	Avw			Cov (%)	Avw	Cov (%)	Avw
$n = 30$						$n = 30$					
90%	λ					90%	λ				
	-1	83.54	0.553	84.47	0.542		-1	85.09	0.564	85.58	0.575
	$-\frac{1}{2}$	84.39	0.568	84.15	0.552		$-\frac{1}{2}$	85.13	0.571	85.37	0.580
	0	84.51	0.580	83.84	0.560		0	84.43	0.577	84.72	0.584
	$\frac{2}{3}$	84.17	0.590	83.15	0.569		$\frac{2}{3}$	83.11	0.582	83.62	0.589
	1	83.80	0.592	82.61	0.572		1	82.31	0.583	83.19	0.590
95%	-1	90.27	0.652	90.70	0.639	95%	-1	91.13	0.669	91.50	0.683
	$-\frac{1}{2}$	91.00	0.673	90.65	0.655		$-\frac{1}{2}$	91.28	0.680	91.42	0.691
	0	91.40	0.691	90.05	0.670		0	90.97	0.688	90.89	0.697
	$\frac{2}{3}$	90.90	0.707	89.07	0.684		$\frac{2}{3}$	89.87	0.695	90.13	0.702
	1	90.33	0.711	88.65	0.689		1	89.13	0.696	89.56	0.704
$n = 45$						$n = 45$					
90%	λ					90%	λ				
	-1	83.59	0.471	83.08	0.457		-1	86.25	0.473	86.50	0.479
	$-\frac{1}{2}$	84.10	0.479	83.08	0.463		$-\frac{1}{2}$	85.96	0.479	85.90	0.483
	0	84.24	0.486	82.75	0.469		0	85.14	0.484	85.57	0.488
	$\frac{2}{3}$	83.63	0.492	82.03	0.474		$\frac{2}{3}$	84.05	0.490	84.30	0.493
	1	83.32	0.493	81.62	0.476		1	83.15	0.493	83.56	0.496
95%	-1	90.30	0.557	90.59	0.541	95%	-1	92.48	0.562	92.25	0.569
	$-\frac{1}{2}$	91.10	0.570	90.18	0.551		$-\frac{1}{2}$	92.09	0.570	92.25	0.576
	0	91.41	0.581	89.81	0.560		0	91.66	0.578	91.72	0.583
	$\frac{2}{3}$	91.06	0.590	89.22	0.570		$\frac{2}{3}$	90.44	0.588	90.63	0.592
	1	90.55	0.593	88.56	0.573		1	89.46	0.592	89.85	0.595
$n = 60$						$n = 60$					
90%	λ					90%	λ				
	-1	83.15	0.418	82.33	0.402		-1	85.22	0.416	85.24	0.419
	$-\frac{1}{2}$	83.56	0.424	82.14	0.406		$-\frac{1}{2}$	84.61	0.420	84.67	0.423
	0	83.65	0.428	81.73	0.410		0	83.95	0.425	83.91	0.427
	$\frac{2}{3}$	83.34	0.431	81.28	0.414		$\frac{2}{3}$	82.49	0.431	82.66	0.432
	1	82.86	0.431	80.93	0.415		1	81.70	0.434	82.03	0.435
95%	-1	90.10	0.495	89.30	0.477	95%	-1	91.93	0.494	91.62	0.498
	$-\frac{1}{2}$	90.39	0.505	89.12	0.484		$-\frac{1}{2}$	91.41	0.501	91.27	0.504
	0	90.17	0.512	88.89	0.490		0	90.29	0.508	90.35	0.510
	$\frac{2}{3}$	89.55	0.517	88.34	0.497		$\frac{2}{3}$	89.23	0.517	89.35	0.519
	1	89.43	0.518	87.92	0.499		1	88.42	0.522	88.96	0.523

Table 5: Coverage probability and average width of 90% and 95% confidence intervals with data contaminated by shifted observations.

is a vector-valued function such that the $q \times p$ matrix $\mathbf{C}(\boldsymbol{\theta}) = \frac{\partial \mathbf{c}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^T}$ exists and is continuous at $\boldsymbol{\theta}$ and that $\text{rank}(\mathbf{C}(\boldsymbol{\theta})) = q$ ($q \leq p$). We denote by $\tilde{\boldsymbol{\theta}}_{E,n}$ the empirical restricted maximum likelihood estimator of $\boldsymbol{\theta}$, obtained by minimizing $\ell_{E,n}(\boldsymbol{\theta}) = -\sum_{i=1}^n \log [1 + \mathbf{t}(\boldsymbol{\theta})^T \mathbf{g}(\mathbf{X}_i, \boldsymbol{\theta})]$, subject to $\sum_{i=1}^n p_i(\boldsymbol{\theta}) \mathbf{g}(\mathbf{X}_i, \boldsymbol{\theta}) = \mathbf{0}_r$ and $\mathbf{c}(\boldsymbol{\theta}) = \mathbf{0}_q$, i.e., (6) or

$$\boldsymbol{\xi}_{1n}(\boldsymbol{\theta}, \mathbf{t}(\boldsymbol{\theta}), \boldsymbol{\nu}) = \mathbf{0}_p, \quad (35)$$

where

$$\boldsymbol{\xi}_{1n}(\boldsymbol{\theta}, \mathbf{t}(\boldsymbol{\theta}), \boldsymbol{\nu}) = \boldsymbol{\xi}_{1n}(\boldsymbol{\theta}, \mathbf{t}(\boldsymbol{\theta})) = \frac{1}{n} \sum_{i=1}^n \frac{1}{1 + \mathbf{t}(\boldsymbol{\theta})^T \mathbf{g}(\mathbf{X}_i, \boldsymbol{\theta})} \mathbf{g}(\mathbf{X}_i, \boldsymbol{\theta})$$

provides $\mathbf{t}(\boldsymbol{\theta})$ in terms of $\boldsymbol{\theta}$. Upon using the Lagrange multiplier method, once again, we have

$$\zeta_{E,n}(\boldsymbol{\theta}, \mathbf{t}(\boldsymbol{\theta}), \boldsymbol{\nu}) = \frac{1}{n} \ell_{E,n}(\boldsymbol{\theta}) + \boldsymbol{\nu}^T \mathbf{c}(\boldsymbol{\theta}),$$

where $\boldsymbol{\nu}$ is a q -dimensional vector of Lagrange multipliers, and differentiating $\zeta_{E,n}(\boldsymbol{\theta}, \mathbf{t}(\boldsymbol{\theta}), \boldsymbol{\nu})$ with respect to $\boldsymbol{\theta}$ and $\boldsymbol{\nu}$, we obtain

$$\begin{aligned} \boldsymbol{\xi}_{2n}(\boldsymbol{\theta}, \mathbf{t}(\boldsymbol{\theta}), \boldsymbol{\nu}) &= \frac{1}{n} \sum_{i=1}^n \frac{1}{1 + \mathbf{t}(\boldsymbol{\theta})^T \mathbf{g}(\mathbf{X}_i, \boldsymbol{\theta})} \mathbf{G}_{\mathbf{X}_i}(\boldsymbol{\theta})^T \mathbf{t}(\boldsymbol{\theta}) + \mathbf{C}(\boldsymbol{\theta})^T \boldsymbol{\nu}, \\ \boldsymbol{\xi}_{3n}(\boldsymbol{\theta}, \mathbf{t}(\boldsymbol{\theta}), \boldsymbol{\nu}) &= \boldsymbol{\xi}_{3n}(\boldsymbol{\theta}) = \mathbf{c}(\boldsymbol{\theta}). \end{aligned}$$

Therefore, $(\tilde{\boldsymbol{\theta}}_{E,n}^T, \tilde{\boldsymbol{\nu}}_{E,n}^T)^T$ is obtained as the solution of (35) and

$$\boldsymbol{\xi}_{2n}(\boldsymbol{\theta}, \mathbf{t}(\boldsymbol{\theta}), \boldsymbol{\nu}) = \mathbf{0}_p, \quad \boldsymbol{\xi}_{3n}(\boldsymbol{\theta}, \mathbf{t}(\boldsymbol{\theta}), \boldsymbol{\nu}) = \mathbf{0}_q, \quad (36)$$

and conditions (3) and (5) must be also satisfied.

The empirical likelihood ratio test for testing (34) has the expression

$$L_{E,n}(\hat{\boldsymbol{\theta}}_{E,n}, \tilde{\boldsymbol{\theta}}_{E,n}) = 2\ell_{E,n}(\hat{\boldsymbol{\theta}}_{E,n}) - 2\ell_{E,n}(\tilde{\boldsymbol{\theta}}_{E,n}), \quad (37)$$

where $\hat{\boldsymbol{\theta}}_{E,n}$ is the empirical maximum likelihood estimator of the parameter $\boldsymbol{\theta}$, defined in Section 1 (case $r = p$), for which $\ell_{E,n}(\hat{\boldsymbol{\theta}}_{E,n}) = 0$. Taking into account the results stated in Section 1 for $r = p$, we have $F_n = F_{n, \hat{\boldsymbol{\theta}}_{E,n}}$, and so it is easy to show that the expression in (37) can be written as

$$L_{E,n}(\hat{\boldsymbol{\theta}}_{E,n}, \tilde{\boldsymbol{\theta}}_{E,n}) = 2n \left(D_{Kullback}(F_n, F_{n, \tilde{\boldsymbol{\theta}}_{E,n}}) - D_{Kullback}(F_n, F_{n, \hat{\boldsymbol{\theta}}_{E,n}}) \right) = 2n D_{Kullback}(F_{n, \hat{\boldsymbol{\theta}}_{E,n}}, F_{n, \tilde{\boldsymbol{\theta}}_{E,n}}).$$

Defining a family of empirical phi-divergence test statistics for testing the hypothesis in (34), we have

$$\begin{aligned} T_n^\phi(\hat{\boldsymbol{\theta}}_{E,n}, \tilde{\boldsymbol{\theta}}_{E,n}) &= \frac{2n}{\phi''(1)} \left(D_\phi(F_n, F_{n, \tilde{\boldsymbol{\theta}}_{E,n}}) - D_\phi(F_n, F_{n, \hat{\boldsymbol{\theta}}_{E,n}}) \right), \\ S_n^\phi(\hat{\boldsymbol{\theta}}_{E,n}, \tilde{\boldsymbol{\theta}}_{E,n}) &= \frac{2n}{\phi''(1)} D_\phi(F_{n, \hat{\boldsymbol{\theta}}_{E,n}}, F_{n, \tilde{\boldsymbol{\theta}}_{E,n}}) \end{aligned}$$

where the ϕ -divergence measures are as defined in (11). Since $F_n = F_{n, \hat{\boldsymbol{\theta}}_{E,n}}$, both families of empirical phi-divergence test statistics are equivalent and taking into account (13) and $\mathbf{t}(\hat{\boldsymbol{\theta}}_{E,n}) = \mathbf{0}_p$, we have

$$\begin{aligned} T_n^\phi(\hat{\boldsymbol{\theta}}_{E,n}, \tilde{\boldsymbol{\theta}}_{E,n}) &= S_n^\phi(\hat{\boldsymbol{\theta}}_{E,n}, \tilde{\boldsymbol{\theta}}_{E,n}) \\ &= \frac{2}{\phi''(1)} \sum_{i=1}^n \frac{\phi \left(1 + \mathbf{t}(\tilde{\boldsymbol{\theta}}_{E,n})^T \mathbf{g}(\mathbf{x}_i, \tilde{\boldsymbol{\theta}}_{E,n}) \right)}{1 + \mathbf{t}(\tilde{\boldsymbol{\theta}}_{E,n})^T \mathbf{g}(\mathbf{x}_i, \tilde{\boldsymbol{\theta}}_{E,n})}. \end{aligned} \quad (38)$$

7.2 Asymptotic null distributions

Under the same conditions of Theorem 3 and the conditions given in Section 7.1 for $\mathbf{c}(\boldsymbol{\theta})$, it is assumed that there exists a neighbourhood of $\boldsymbol{\theta}_0$ in which $\left\| \frac{\partial \mathbf{C}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right\|$ is bounded by some constant, we have

$$\sqrt{n} \begin{pmatrix} \tilde{\boldsymbol{\theta}}_{E,n} - \boldsymbol{\theta}_0 \\ \tilde{\boldsymbol{\nu}}_{E,n} \end{pmatrix} \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \mathcal{N} \left(\mathbf{0}_{p+q}, \begin{pmatrix} \mathbf{P}(\boldsymbol{\theta}_0) & \mathbf{0}_{p \times q} \\ \mathbf{0}_{q \times p} & \mathbf{Q}(\boldsymbol{\theta}_0) \end{pmatrix} \right), \quad (39)$$

where $\boldsymbol{\theta}_0$ is the true value of the parameter $\boldsymbol{\theta}$ and

$$\mathbf{P}(\boldsymbol{\theta}_0) = \mathbf{V}(\boldsymbol{\theta}_0) - \mathbf{V}(\boldsymbol{\theta}_0) \mathbf{C}(\boldsymbol{\theta}_0)^T \mathbf{Q}(\boldsymbol{\theta}_0) \mathbf{C}(\boldsymbol{\theta}_0) \mathbf{V}(\boldsymbol{\theta}_0), \quad (40)$$

$$\mathbf{Q}(\boldsymbol{\theta}_0) = (\mathbf{C}(\boldsymbol{\theta}_0) \mathbf{V}(\boldsymbol{\theta}_0) \mathbf{C}(\boldsymbol{\theta}_0)^T)^{-1}. \quad (41)$$

This result is derived by taking into account, the facts that

$$\sqrt{n}(\tilde{\boldsymbol{\theta}}_{E,n} - \boldsymbol{\theta}_0) = -\mathbf{S}_{12}^{-1}(\boldsymbol{\theta}_0) \sqrt{n} \bar{\mathbf{g}}_n(\boldsymbol{\theta}_0) - \mathbf{V}(\boldsymbol{\theta}_0) \mathbf{C}(\boldsymbol{\theta}_0)^T \sqrt{n} \tilde{\boldsymbol{\nu}}_{E,n} + o_p(1), \quad (42)$$

$$\sqrt{n} \tilde{\boldsymbol{\nu}}_{E,n} = -\mathbf{Q}(\boldsymbol{\theta}_0) \mathbf{C}(\boldsymbol{\theta}_0) \mathbf{S}_{12}^{-1}(\boldsymbol{\theta}_0) \sqrt{n} \bar{\mathbf{g}}_n(\boldsymbol{\theta}_0) + o_p(1). \quad (43)$$

For a complete proof, one may refer to Qin and Lawless (1995).

Theorem 13 Under H_0 in (34) and the assumptions above, we have

$$T_n^\phi(\hat{\boldsymbol{\theta}}_{E,n}, \tilde{\boldsymbol{\theta}}_{E,n}) \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \chi_q^2. \quad (44)$$

Proof. Let us consider the function

$$\ell_\phi(\mathbf{t}) = \frac{1}{n} \sum_{i=1}^n \frac{\phi(1 + \mathbf{t}^T \mathbf{g}(\mathbf{x}_i, \tilde{\boldsymbol{\theta}}_{E,n}))}{1 + \mathbf{t}^T \mathbf{g}(\mathbf{x}_i, \tilde{\boldsymbol{\theta}}_{E,n})}.$$

A second-order Taylor expansion of $\ell_\phi(\tilde{\mathbf{t}})$ around $\mathbf{0}_p$, with $\tilde{\mathbf{t}} = \mathbf{t}(\tilde{\boldsymbol{\theta}}_{E,n})$, gives

$$T_n^\phi(\hat{\boldsymbol{\theta}}_{E,n}, \tilde{\boldsymbol{\theta}}_{E,n}) = \frac{2n\ell_\phi(\tilde{\mathbf{t}})}{\phi''(1)} = n\mathbf{t}(\tilde{\boldsymbol{\theta}}_{E,n})^T \mathbf{S}_{11}(\tilde{\boldsymbol{\theta}}_{E,n}) \mathbf{t}(\tilde{\boldsymbol{\theta}}_{E,n}) + o\left(\|\sqrt{n}\mathbf{t}(\tilde{\boldsymbol{\theta}}_{E,n})\|^2\right).$$

But,

$$\sqrt{n}\mathbf{t}(\tilde{\boldsymbol{\theta}}_{E,n}) = \sqrt{n}\mathbf{S}_{11}^{-1}(\boldsymbol{\theta}_0) \bar{\mathbf{g}}_n(\tilde{\boldsymbol{\theta}}_{E,n}) + o_p(\mathbf{1}_r),$$

and

$$\begin{aligned} \bar{\mathbf{g}}_n(\tilde{\boldsymbol{\theta}}_{E,n}) &= \bar{\mathbf{g}}_n(\boldsymbol{\theta}_0) + \frac{\partial}{\partial \boldsymbol{\theta}^T} \bar{\mathbf{g}}_n(\boldsymbol{\theta})|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} (\tilde{\boldsymbol{\theta}}_{E,n} - \boldsymbol{\theta}_0) + o_p(n^{-1/2} \mathbf{1}_r) \\ &= \bar{\mathbf{g}}_n(\boldsymbol{\theta}_0) + \mathbf{S}_{12}(\boldsymbol{\theta}_0) (\tilde{\boldsymbol{\theta}}_{E,n} - \boldsymbol{\theta}_0) + o_p(n^{-1/2} \mathbf{1}_r) \\ &= \mathbf{S}_{12}(\boldsymbol{\theta}_0) \mathbf{V}(\boldsymbol{\theta}_0) \mathbf{C}(\boldsymbol{\theta}_0)^T \tilde{\boldsymbol{\nu}}_{E,n} + o_p(n^{-1/2} \mathbf{1}_r). \end{aligned}$$

By the strong law of large numbers, $\frac{\partial}{\partial \boldsymbol{\theta}^T} \bar{\mathbf{g}}_n(\boldsymbol{\theta})|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} \xrightarrow[n \rightarrow \infty]{a.s.} \mathbf{S}_{12}(\boldsymbol{\theta}_0)$, and so (42) can be rewritten as

$$\mathbf{S}_{12}(\boldsymbol{\theta}_0) (\tilde{\boldsymbol{\theta}}_{E,n} - \boldsymbol{\theta}_0) = -\bar{\mathbf{g}}_n(\boldsymbol{\theta}_0) - \mathbf{S}_{12}(\boldsymbol{\theta}_0) \mathbf{V}(\boldsymbol{\theta}_0) \mathbf{C}(\boldsymbol{\theta}_0)^T \tilde{\boldsymbol{\nu}}_{E,n} + o_p(n^{-1/2} \mathbf{1}_r).$$

Consequently,

$$T_n^\phi(\hat{\boldsymbol{\theta}}_{E,n}, \tilde{\boldsymbol{\theta}}_{E,n}) = (\sqrt{n} \tilde{\boldsymbol{\nu}}_{E,n})^T \mathbf{Q}^{-1}(\boldsymbol{\theta}_0) \sqrt{n} \tilde{\boldsymbol{\nu}}_{E,n} + o_p(1), \quad (45)$$

with $\sqrt{n} \tilde{\boldsymbol{\nu}}_{E,n} \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \mathcal{N}(\mathbf{0}_q, \mathbf{Q}(\boldsymbol{\theta}_0))$, which means that (44) holds. ■

7.3 Approximations of the power function

Assume that $\boldsymbol{\theta}^* \notin \Theta_0$ is the true value of the unknown parameter so that $\widehat{\boldsymbol{\theta}}_{E,n} \xrightarrow[n \rightarrow \infty]{a.s.} \boldsymbol{\theta}^*$, and that there exists a $\boldsymbol{\theta}_0 \in \Theta_0$ such that the restricted empirical maximum likelihood estimator satisfies $\widetilde{\boldsymbol{\theta}}_{E,n} \xrightarrow[n \rightarrow \infty]{a.s.} \boldsymbol{\theta}_0$. Then, we have

$$\sqrt{n} \left(\begin{pmatrix} \widehat{\boldsymbol{\theta}}_{E,n} \\ \widetilde{\boldsymbol{\theta}}_{E,n} \end{pmatrix} - \begin{pmatrix} \boldsymbol{\theta}^* \\ \boldsymbol{\theta}_0 \end{pmatrix} \right) \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \mathcal{N} \left(\begin{pmatrix} \mathbf{0}_p \\ \mathbf{0}_p \end{pmatrix}, \begin{pmatrix} \mathbf{V}(\boldsymbol{\theta}_0) & \mathbf{A}_{12}(\boldsymbol{\theta}^*, \boldsymbol{\theta}_0) \\ \mathbf{A}_{12}(\boldsymbol{\theta}^*, \boldsymbol{\theta}_0)^T & \mathbf{P}(\boldsymbol{\theta}_0) \end{pmatrix} \right), \quad (46)$$

where $\mathbf{A}_{12}(\boldsymbol{\theta}^*, \boldsymbol{\theta}_0)$ is a $q \times q$ matrix.

Theorem 14 *Under the conditions stated above, we have*

$$\sqrt{n} \left(D_\phi(\widehat{\boldsymbol{\theta}}_{E,n}, \widetilde{\boldsymbol{\theta}}_{E,n}) - D_\phi(\boldsymbol{\theta}^*, \boldsymbol{\theta}_0) \right) \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \mathcal{N} \left(0, \varsigma_\phi^2(\boldsymbol{\theta}^*, \boldsymbol{\theta}_0) \right),$$

where

$$\varsigma_\phi^2(\boldsymbol{\theta}^*, \boldsymbol{\theta}_0) = \boldsymbol{\tau}_\phi^T(\boldsymbol{\theta}^*, \boldsymbol{\theta}_0) \mathbf{V}(\boldsymbol{\theta}_0) \boldsymbol{\tau}_\phi(\boldsymbol{\theta}^*, \boldsymbol{\theta}_0) + 2\boldsymbol{\tau}_\phi(\boldsymbol{\theta}^*, \boldsymbol{\theta}_0)^T \mathbf{A}_{12}(\boldsymbol{\theta}^*, \boldsymbol{\theta}_0) \boldsymbol{\nu}_\phi(\boldsymbol{\theta}^*, \boldsymbol{\theta}_0) + \boldsymbol{\nu}_\phi(\boldsymbol{\theta}^*, \boldsymbol{\theta}_0)^T \mathbf{P}(\boldsymbol{\theta}_0) \boldsymbol{\nu}_\phi(\boldsymbol{\theta}^*, \boldsymbol{\theta}_0),$$

with $\boldsymbol{\tau}_\phi(\boldsymbol{\theta}^*, \boldsymbol{\theta}_0)$ given by (24) and

$$\boldsymbol{\nu}_\phi(\boldsymbol{\theta}^*, \boldsymbol{\theta}_0) = (\nu_1^\phi(\boldsymbol{\theta}^*, \boldsymbol{\theta}_0), \dots, \nu_p^\phi(\boldsymbol{\theta}^*, \boldsymbol{\theta}_0))^T = \left. \frac{\partial D_\phi(\boldsymbol{\theta}^*, \boldsymbol{\vartheta})}{\partial \boldsymbol{\vartheta}} \right|_{\boldsymbol{\vartheta}=\boldsymbol{\theta}_0}.$$

Remark 15 *Based on Theorem 14, we get an approximation of the power function at $\boldsymbol{\theta}^*$ as*

$$\begin{aligned} \beta_{n,\phi}^1(\boldsymbol{\theta}^*, \boldsymbol{\theta}_0) &= P_{\boldsymbol{\theta}^*}(T_n^\phi(\widehat{\boldsymbol{\theta}}_{E,n}, \widetilde{\boldsymbol{\theta}}_{E,n}) > \chi_{q,\alpha}^2) \\ &= 1 - \Phi \left(\frac{\sqrt{n}}{\varsigma_\phi(\boldsymbol{\theta}^*, \boldsymbol{\theta}_0)} \left(\frac{\phi''(1)}{2n} \chi_{q,\alpha}^2 - D_\phi(F_{n,\boldsymbol{\theta}^*}, F_{n,\boldsymbol{\theta}_0}) \right) \right). \end{aligned}$$

If some alternative $\boldsymbol{\theta}^* \neq \boldsymbol{\theta}_0$ is the true parameter, then the probability of rejecting $\boldsymbol{\theta}_0$ with the rejection rule $T_n^\phi(\widehat{\boldsymbol{\theta}}_{E,n}, \widetilde{\boldsymbol{\theta}}_{E,n}) > \chi_{q,\alpha}^2$, for a fixed significance level α , tends to one as $n \rightarrow \infty$. Thus, the test is consistent in the sense of Fraser (1957).

We may also find an approximation to the power of $T_n^\phi(\widehat{\boldsymbol{\theta}}_{E,n}, \widetilde{\boldsymbol{\theta}}_{E,n})$ at an alternative hypothesis close to the null hypothesis. Let $\boldsymbol{\theta}_n \in \Theta - \Theta_0$ be a given alternative, and let $\boldsymbol{\theta}_0$ be the element in Θ_0 closest to $\boldsymbol{\theta}_n$ in terms of Euclidean distance. In order to introduce contiguous alternative hypotheses, we may consider a fixed $\mathbf{f} \in \mathbb{R}^p$ and allow $\boldsymbol{\theta}_n$ to tend to $\boldsymbol{\theta}_0$ as n increases in the following manner:

$$H_{1,n} : \boldsymbol{\theta}_n = \boldsymbol{\theta}_0 + n^{-1/2} \mathbf{f}. \quad (47)$$

Theorem 16 *Under the assumptions of Theorem 13 and $H_{1,n}$ given in (47), we have*

$$T_n^\phi(\widehat{\boldsymbol{\theta}}_{E,n}, \widetilde{\boldsymbol{\theta}}_{E,n}) \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \chi_q^2(\varrho(\boldsymbol{\theta}_0)), \quad (48)$$

where $\varrho(\boldsymbol{\theta}_0) = \mathbf{f}^T \mathbf{C}(\boldsymbol{\theta}_0)^T \mathbf{Q}(\boldsymbol{\theta}_0) \mathbf{C}(\boldsymbol{\theta}_0) \mathbf{f}$.

Proof. If we substitute (43) into (45), we obtain

$$T_n^\phi(\widehat{\boldsymbol{\theta}}_{E,n}, \widetilde{\boldsymbol{\theta}}_{E,n}) = \sqrt{n} \bar{\mathbf{g}}_n^T(\boldsymbol{\theta}_0) \mathbf{S}_{12}^{-1}(\boldsymbol{\theta}_0) \mathbf{C}(\boldsymbol{\theta}_0)^T \mathbf{Q}(\boldsymbol{\theta}_0) \mathbf{C}(\boldsymbol{\theta}_0) \mathbf{S}_{12}^{-1}(\boldsymbol{\theta}_0) \sqrt{n} \bar{\mathbf{g}}_n(\boldsymbol{\theta}_0) + o_p(1).$$

Since $\bar{\mathbf{g}}_n(\hat{\boldsymbol{\theta}}_{E,n}) = \mathbf{0}_p$, the Taylor expansion $\bar{\mathbf{g}}_n(\hat{\boldsymbol{\theta}}_{E,n}) = \bar{\mathbf{g}}_n(\boldsymbol{\theta}_0) + \frac{\partial}{\partial \boldsymbol{\theta}^T} \bar{\mathbf{g}}_n(\boldsymbol{\theta})|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} (\hat{\boldsymbol{\theta}}_{E,n} - \boldsymbol{\theta}_0) + o_p(n^{-1/2} \mathbf{1}_r)$ yields

$$-\mathbf{S}_{12}^{-1}(\boldsymbol{\theta}_0) \sqrt{n} \bar{\mathbf{g}}_n(\boldsymbol{\theta}_0) = \hat{\boldsymbol{\theta}}_{E,n} - \boldsymbol{\theta}_0 + o_p(\mathbf{1}_r).$$

By following the same steps as in the proof of Theorem 10, we have that under $H_{1,n}$, $\sqrt{n}(\hat{\boldsymbol{\theta}}_{E,n} - \boldsymbol{\theta}_0) \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \mathcal{N}(\mathbf{f}, \mathbf{V}(\boldsymbol{\theta}_0))$. Hence, we obtain

$$\begin{aligned} T_n^\phi(\hat{\boldsymbol{\theta}}_{E,n}, \tilde{\boldsymbol{\theta}}_{E,n}) &= (\hat{\boldsymbol{\theta}}_{E,n} - \boldsymbol{\theta}_0)^T \mathbf{C}(\boldsymbol{\theta}_0)^T \mathbf{Q}(\boldsymbol{\theta}_0) \mathbf{C}(\boldsymbol{\theta}_0) (\hat{\boldsymbol{\theta}}_{E,n} - \boldsymbol{\theta}_0) + o_p(1) \\ &= \mathbf{Z}^T \mathbf{Z} + o_p(1), \end{aligned}$$

where $\mathbf{Z} = \mathbf{Q}(\boldsymbol{\theta}_0)^{1/2} \mathbf{C}(\boldsymbol{\theta}_0) \sqrt{n}(\hat{\boldsymbol{\theta}}_{E,n} - \boldsymbol{\theta}_0) \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \mathcal{N}(\mathbf{Q}(\boldsymbol{\theta}_0)^{1/2} \mathbf{C}(\boldsymbol{\theta}_0) \mathbf{f}, \mathbf{I}_q)$. We thus obtain

$$T_n^\phi(\hat{\boldsymbol{\theta}}_{E,n}, \tilde{\boldsymbol{\theta}}_{E,n}) \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \chi_q^2(\mathbf{f}^T \mathbf{C}(\boldsymbol{\theta}_0)^T \mathbf{Q}(\boldsymbol{\theta}_0) \mathbf{C}(\boldsymbol{\theta}_0) \mathbf{f}),$$

which completes the proof. ■

A second way to consider contiguous alternative hypotheses is to relax the condition $\mathbf{c}(\boldsymbol{\theta}) = \mathbf{0}_q$ defining the null hypothesis Θ_0 . Let $\mathbf{f} \in \mathbb{R}^d$ be such that $\mathbf{f} \neq \mathbf{0}_p$. Consider the following sequence $\boldsymbol{\theta}_n$ of parameters approaching Θ_0 :

$$H_{1,n}^* : \mathbf{c}(\boldsymbol{\theta}_n) = n^{-1/2} \bar{\mathbf{f}}. \quad (49)$$

A Taylor expansion of $\mathbf{c}(\boldsymbol{\theta}_n)$ around $\boldsymbol{\theta}_0$ yields

$$\mathbf{c}(\boldsymbol{\theta}_n) = \mathbf{c}(\boldsymbol{\theta}_0) + \mathbf{C}(\boldsymbol{\theta}_0) (\boldsymbol{\theta}_n - \boldsymbol{\theta}_0) + o(\|\boldsymbol{\theta}_n - \boldsymbol{\theta}_0\|).$$

Upon substituting $\boldsymbol{\theta}_n = \boldsymbol{\theta}_0 + n^{-1/2} \mathbf{f}$ in the previous formula and taking into account that $\mathbf{c}(\boldsymbol{\theta}_0) = \mathbf{0}_r$, we have

$$\mathbf{c}(\boldsymbol{\theta}_n) = n^{-1/2} \mathbf{C}(\boldsymbol{\theta}_0) \mathbf{f}.$$

Then, the equivalence between $H_{1,n}^*$ and $H_{1,n}$ is obtained for $\bar{\mathbf{f}} = \mathbf{C}(\boldsymbol{\theta}_0) \mathbf{f}$. We thus have the following result.

Theorem 17 *Under the contiguous alternative hypothesis in (49), we have*

$$T_n^\phi(\tilde{\boldsymbol{\theta}}_{E,n}) \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \chi_q^2(\varrho(\boldsymbol{\theta}_0)),$$

with $\varrho(\boldsymbol{\theta}_0) = \bar{\mathbf{f}}^T \mathbf{Q}(\boldsymbol{\theta}_0) \bar{\mathbf{f}}$ and $\mathbf{Q}(\boldsymbol{\theta}_0)$ as defined in (41).

7.4 Simulation results

From a sample X_1, \dots, X_n of i.i.d. random variables with $E[X_i] = \mu$ and $Var[X_i] = \sigma^2$, we wish to test that the coefficient of variation is 1, i.e., $H_0 : \mu = \sigma$ against $H_0 : \mu \neq \sigma$. In order to make a decision, we need to obtain the maximum likelihood estimator under the restriction $c(\boldsymbol{\vartheta}) = \sigma - \mu = 0$, with $\boldsymbol{\vartheta} = (\mu, \sigma)^T$. The estimating equations in this case are $g_1(x_i, \boldsymbol{\vartheta}) = x_i - \mu$, $g_2(x_i, \boldsymbol{\vartheta}) = x_i^2 - (\sigma^2 + \mu^2)$. If we establish a bijective transformation between $\boldsymbol{\vartheta} = (\mu, \sigma)^T$ and $\boldsymbol{\theta} = (u, v)^T$, and are able to obtain the empirical restricted maximum likelihood estimator of $\boldsymbol{\theta}$, $\tilde{\boldsymbol{\theta}}_{E,n}$, due to the invariance property, we can obtain the empirical restricted maximum likelihood estimator of $\boldsymbol{\vartheta}$, $\tilde{\boldsymbol{\vartheta}}_R$, by taking the inverse of the transformation. Let $\mu = u$ and $\sigma^2 = v - u^2$. Then

$$H_0 : c(\boldsymbol{\theta}) = 0 \quad \text{vs.} \quad H_0 : c(\boldsymbol{\theta}) \neq 0$$

with $c(\boldsymbol{\theta}) = c(u, v) = v - 2u^2 = 0$. The estimating equations under the new parameterization are

$$\begin{aligned} g_1(X_i, \boldsymbol{\theta}) &= g_1(X_i, u, v) = X_i - u, \\ g_2(X_i, \boldsymbol{\theta}) &= g_2(X_i, u, v) = X_i^2 - v. \end{aligned}$$

In general, we have the system of $2p + r = 5$ equations as follows:

$$\begin{aligned}\xi_{1,j,n}(u, v, t_1, t_2, \gamma) &= \frac{1}{n} \sum_{i=1}^n \frac{g_j(X_i, u, v)}{1 + t_1 g_1(X_i, u, v) + t_2 g_2(X_i, u, v)} = 0, \quad j = 1, 2, \\ \xi_{2,1,n}(u, v, t_1, t_2, \gamma) &= \frac{1}{n} \sum_{i=1}^n \frac{t_1 \frac{\partial}{\partial u} g_1(X_i, u, v) + t_2 \frac{\partial}{\partial u} g_2(X_i, u, v)}{1 + t_1 g_1(X_i, u, v) + t_2 g_2(X_i, u, v)} + \frac{\partial}{\partial u} c(u, v) \gamma = 0, \\ \xi_{3,1,n}(u, v, t_1, t_2, \gamma) &= \frac{1}{n} \sum_{i=1}^n \frac{t_1 \frac{\partial}{\partial v} g_1(X_i, u, v) + t_2 \frac{\partial}{\partial v} g_2(X_i, u, v)}{1 + t_1 g_1(X_i, u, v) + t_2 g_2(X_i, u, v)} + \frac{\partial}{\partial v} c(u, v) \gamma = 0, \\ c(u, v) &= v - 2u^2 = 0,\end{aligned}$$

which are equivalent to

$$\begin{aligned}\frac{1}{n} \sum_{i=1}^n \frac{X_i - u}{1 + t_1(X_i - u) + t_2(X_i^2 - v)} &= 0, \\ \frac{1}{n} \sum_{i=1}^n \frac{X_i^2 - v}{1 + t_1(X_i - u) + t_2(X_i^2 - v)} &= 0, \\ -\frac{1}{n} \sum_{i=1}^n \frac{t_1}{1 + t_1 g_1(X_i, u, v) + t_2 g_2(X_i, u, v)} - 4u\gamma &= 0, \\ -\frac{1}{n} \sum_{i=1}^n \frac{t_2}{1 + t_1 g_1(X_i, u, v) + t_2 g_2(X_i, u, v)} + \gamma &= 0, \\ v - 2u^2 &= 0,\end{aligned}$$

for obtaining $\tilde{\boldsymbol{\theta}}_{E,n} = (\tilde{u}, \tilde{v})^T$. Observe that

$$\frac{1}{n} \sum_{i=1}^n \frac{1}{1 + t_1 g_1(X_i, u, v) + t_2 g_2(X_i, u, v)} = 1,$$

since the sum of the probabilities is 1, and consequently the third and fourth equations become $t_1 = -4u\gamma$, $t_2 = \gamma$, and we also know that $v = 2u^2$. So, the optimization problem is reduced to simply to

$$\begin{aligned}f_1(u, \gamma) &= \sum_{i=1}^n \frac{X_i - u}{1 + \gamma(X_i^2 - 4X_i u + 2u^2)} = 0, \\ f_2(u, \gamma) &= \sum_{i=1}^n \frac{X_i^2 - 2u^2}{1 + \gamma(X_i^2 - 4X_i u + 2u^2)} = 0.\end{aligned}$$

The solution $\tilde{\boldsymbol{\theta}}_{E,n}$ must satisfy that the n probabilities are not less than zero, i.e.,

$$\begin{aligned}\frac{1}{n(1 + t_1 g_1(X_i, u, v) + t_2 g_2(X_i, u, v))} &= \frac{1}{n(1 + \gamma(X_i^2 - 4x_i u + 2u^2))} \geq 0, \\ \gamma(X_i^2 - 4X_i u + 2u^2) &\geq \frac{1-n}{n}, \quad i = 1, \dots, n,\end{aligned}$$

and if such a solution exists, it is unique; see Qin and Lawless (1994). In our empirical study, we solved the system of these two equations by using the NAG subroutine in Fortran, C05PBF.

To compare the exact coverage probabilities of the confidence intervals for the coefficient of variation based on some empirical phi-divergence test statistics, a simulation study was conducted separately for continuous and discrete distributions since we found that the rate of convergence to the asymptotic distribution is much faster for discrete distributions (Poisson) than for continuous distributions (normal and Student t). Based on

100,000 samples of sizes 30, 45, 60, 75, 90, 105 from $\mathcal{N}(1,1)$ and $1 + \sqrt{0.6}t_5$ distributions and sample sizes 15, 20, 25, 30 from Poisson, $\mathcal{P}(1)$, the so-called power-divergence measures (see Cressie and Read (1984)) were considered to construct the phi-divergence test statistics,

$$\begin{aligned} T_n^\lambda(\widehat{\boldsymbol{\theta}}_{E,n}, \widetilde{\boldsymbol{\theta}}_{E,n}) &= \frac{2}{\lambda(1+\lambda)} \left(\sum_{i=1}^n (1 + \widetilde{\gamma}(X_i^2 - 4X_i\widetilde{u} + 2\widetilde{u}^2))^\lambda - n \right), \quad \text{if } \lambda(1+\lambda) \neq 0, \\ &= 2 \sum_{i=1}^n \log(1 + \widetilde{\gamma}(X_i^2 - 4X_i\widetilde{u} + 2\widetilde{u}^2)), \quad \text{if } \lambda = 0, \\ &= -2 \sum_{i=1}^n \frac{\log(1 + \widetilde{\gamma}(X_i^2 - 4X_i\widetilde{u} + 2\widetilde{u}^2))}{1 + \widetilde{\gamma}(X_i^2 - 4X_i\widetilde{u} + 2\widetilde{u}^2)}, \quad \text{if } \lambda = -1, \end{aligned}$$

that is, for each $\lambda \in \mathbb{R}$, we have a different divergence measure by taking $\phi = \phi_\lambda$, where $\phi_\lambda(x)$ is as defined earlier in Section 6. In addition, the empirical generalized Wald test statistic, the empirical generalized score test statistic, and the empirical Lagrange multiplier test statistic were also obtained. For this purpose, the following auxiliary matrices are necessary:

$$\begin{aligned} \mathbf{V}(\boldsymbol{\theta}) &= \left(\mathbf{S}_{12}(\boldsymbol{\theta})^T \mathbf{S}_{11}^{-1}(\boldsymbol{\theta}) \mathbf{S}_{12}(\boldsymbol{\theta}) \right)^{-1}, \\ \mathbf{C}(\boldsymbol{\theta}) &= \frac{\partial}{\partial \boldsymbol{\theta}^T} \mathbf{c}(\boldsymbol{\theta}) = \left(\frac{\partial}{\partial u}(v - 2u^2), \frac{\partial}{\partial v}(v - 2u^2) \right) = (-4u, \quad 1), \\ \mathbf{Q}(\boldsymbol{\theta}) &= (\mathbf{C}(\boldsymbol{\theta})\mathbf{V}(\boldsymbol{\theta})\mathbf{C}(\boldsymbol{\theta})^T)^{-1}, \end{aligned}$$

where $\mathbf{S}_{11}(\boldsymbol{\theta}) = E[\mathbf{g}(\mathbf{X}, \boldsymbol{\theta})\mathbf{g}(\mathbf{X}, \boldsymbol{\theta})^T]$ and $\mathbf{S}_{12}(\boldsymbol{\theta}) = E[\mathbf{G}_{\mathbf{X}}(\boldsymbol{\theta})]$ are unknown. The matrices $\mathbf{S}_{11}(\boldsymbol{\theta})$ and $\mathbf{S}_{12}(\boldsymbol{\theta})$ can be replaced by any consistent estimator

$$\begin{aligned} \widehat{\mathbf{S}}_{12}(\boldsymbol{\theta}) &= \frac{1}{n} \sum_{i=1}^n \mathbf{g}(X_i, \boldsymbol{\theta})\mathbf{g}(X_i, \boldsymbol{\theta})^T = \begin{pmatrix} \frac{1}{n} \sum_{i=1}^n (X_i - u)^2 & \frac{1}{n} \sum_{i=1}^n (X_i - u)(X_i^2 - v) \\ \frac{1}{n} \sum_{i=1}^n (X_i - u)(X_i^2 - v) & \frac{1}{n} \sum_{i=1}^n (X_i^2 - v)^2 \end{pmatrix}, \\ \widehat{\mathbf{S}}_{11}(\boldsymbol{\theta}) &= \frac{\partial}{\partial \boldsymbol{\theta}^T} \widehat{\mathbf{g}}_n(\boldsymbol{\theta}) = -\mathbf{I}_2, \end{aligned}$$

and also $\mathbf{V}(\boldsymbol{\theta})$ and $\mathbf{Q}(\boldsymbol{\theta})$ by

$$\begin{aligned} \widehat{\mathbf{V}}(\boldsymbol{\theta}) &= \widehat{\mathbf{S}}_{11}^{-1}(\boldsymbol{\theta}) \widehat{\mathbf{S}}_{12}(\boldsymbol{\theta}) \widehat{\mathbf{S}}_{11}^{-1}(\boldsymbol{\theta}) = \widehat{\mathbf{S}}_{12}(\boldsymbol{\theta}), \\ \widehat{\mathbf{Q}}(\boldsymbol{\theta}) &= \left(\mathbf{C}(\boldsymbol{\theta})\widehat{\mathbf{V}}(\boldsymbol{\theta})\mathbf{C}(\boldsymbol{\theta})^T \right)^{-1} \\ &= \frac{n}{\sum_{i=1}^n (X_i^4 - 2vX_i^2 + v^2) - 8u \sum_{i=1}^n (-vX_i + X_i^3 - uX_i^2 + uv) + 16u^2 \sum_{i=1}^n (-2uX_i + X_i^2 + u^2)}. \end{aligned}$$

The unconstrained maximum likelihood estimators were calculated as the solution of the system of $p = 2$ equations:

$$\begin{aligned} \frac{1}{n} \sum_{i=1}^n g_1(X_i, u, v) &= \frac{1}{n} \sum_{i=1}^n (X_i - u) = 0, \\ \frac{1}{n} \sum_{i=1}^n g_2(X_i, u, v) &= \frac{1}{n} \sum_{i=1}^n (X_i^2 - v) = 0, \end{aligned}$$

that is,

$$\hat{u} = \frac{1}{n} \sum_{i=1}^n X_i, \quad \hat{v} = \frac{1}{n} \sum_{i=1}^n X_i^2.$$

First, the expression of the empirical generalized Wald test statistic is obtained as

$$W_n(\hat{\theta}_{E,n}) = nc(\hat{\theta}_{E,n})\hat{Q}(\hat{\theta}_{E,n})c(\hat{\theta}_{E,n}) = \frac{n(2\hat{u}^2 - \hat{v})^2}{\frac{1}{n} \sum_{i=1}^n X_i^4 - 8\hat{u}\frac{1}{n} \sum_{i=1}^n X_i^3 - \hat{v}^2 + 24\hat{u}^2\hat{v} - 16\hat{u}^4};$$

next, the empirical generalized score test statistic is obtained as

$$\begin{aligned} S_n(\tilde{\theta}_{E,n}) &= n\bar{g}_n(\tilde{\theta}_{E,n})^T \hat{V}(\tilde{\theta}_{E,n}) C(\tilde{\theta}_{E,n})^T \hat{Q}(\tilde{\theta}_{E,n}) C(\tilde{\theta}_{E,n}) \hat{V}(\tilde{\theta}_{E,n}) \bar{g}_n(\tilde{\theta}_{E,n}) \\ &= \frac{n[4\tilde{u}(\tilde{u} - \hat{u}) + (\tilde{v} - \hat{v})]^2}{\frac{1}{n} \sum_{i=1}^n X_i^4 - 8\tilde{u}\frac{1}{n} \sum_{i=1}^n X_i^3 - 2\tilde{v}\hat{v} + \tilde{v}^2 + 8\tilde{u}\tilde{u}\hat{v} + 24\tilde{u}^2\hat{v} - 8\tilde{u}^2\tilde{v} - 32\tilde{u}\tilde{u}^3 + 16\tilde{u}^4} \\ &= \frac{n[\tilde{v} + \tilde{v} - 4\tilde{u}\hat{u}]^2}{\frac{1}{n} \sum_{i=1}^n X_i^4 - 8\tilde{u}\frac{1}{n} \sum_{i=1}^n X_i^3 - 8\tilde{u}\tilde{u}\hat{v} + 10\tilde{v}\hat{v} + \tilde{v}^2}, \end{aligned}$$

where the second equality is obtained by taking into account that $\tilde{v} = 2\tilde{u}^2$; finally, the empirical Lagrange multiplier test statistic is obtained as

$$\lambda_n(\tilde{\theta}_{E,n}) = n\tilde{\gamma}\hat{Q}^{-1}(\tilde{\theta}_{E,n})\tilde{\gamma} = n\tilde{\gamma}^2 \left(\frac{1}{n} \sum_{i=1}^n X_i^4 - 8\tilde{u}\frac{1}{n} \sum_{i=1}^n X_i^3 - 8\tilde{u}\tilde{u}\hat{v} + 10\tilde{v}\hat{v} + \tilde{v}^2 \right).$$

n	Nom. level	T_n^{-1}	$T_n^{-1/2}$	T_n^0	$T_n^{2/3}$	T_n^1	T_n^2	W_n	S_n	λ_n
30	0.90	0.8537	0.8615	0.8672	0.8717	0.8723	0.8683	0.8676	0.8671	0.8689
45	0.90	0.8682	0.8745	0.8789	0.8828	0.8836	0.8811	0.8778	0.8780	0.8842
60	0.90	0.8775	0.8824	0.8860	0.8895	0.8908	0.8899	0.8852	0.8854	0.8930
75	0.90	0.8815	0.8856	0.8892	0.8923	0.8932	0.8922	0.8870	0.8880	0.8962
90	0.90	0.8851	0.8892	0.8919	0.8947	0.8955	0.8951	0.8900	0.8900	0.8996
105	0.90	0.8870	0.8905	0.8938	0.8965	0.8976	0.8976	0.8916	0.8917	0.9009
30	0.95	0.9062	0.9143	0.9210	0.9251	0.9258	0.9211	0.9228	0.9169	0.9149
45	0.95	0.9199	0.9265	0.9312	0.9352	0.9358	0.9321	0.9308	0.9275	0.9285
60	0.95	0.9286	0.9342	0.9392	0.9425	0.9426	0.9401	0.9374	0.9336	0.9360
75	0.95	0.9322	0.9379	0.9417	0.9446	0.9446	0.9427	0.9399	0.9367	0.9398
90	0.95	0.9350	0.9398	0.9432	0.9457	0.9460	0.9443	0.9415	0.9387	0.9419
105	0.95	0.9372	0.9417	0.9449	0.9477	0.9480	0.9464	0.9428	0.9403	0.9447

Table 6: Simulated values of interval coverage probabilities with underlying $\mathcal{N}(1, 1)$ distribution.

n	Nom. level	T_n^{-1}	$T_n^{-1/2}$	T_n^0	$T_n^{2/3}$	T_n^1	T_n^2	W_n	S_n	λ_n
30	0.90	0.8003	0.8086	0.8145	0.8175	0.8174	0.8130	0.8302	0.8148	0.8208
45	0.90	0.8232	0.8306	0.8356	0.8387	0.8384	0.8343	0.8465	0.8333	0.8417
60	0.90	0.8354	0.8424	0.8473	0.8489	0.8488	0.8446	0.8561	0.8446	0.8526
75	0.90	0.8440	0.8500	0.8541	0.8568	0.8573	0.8535	0.8621	0.8531	0.8597
90	0.90	0.8517	0.8571	0.8610	0.8631	0.8629	0.8586	0.8683	0.8603	0.8645
105	0.90	0.8577	0.8627	0.8657	0.8668	0.8665	0.8621	0.8737	0.8654	0.8665
30	0.95	0.8584	0.8696	0.8776	0.8825	0.8822	0.8740	0.8897	0.8665	0.8709
45	0.95	0.8807	0.8906	0.8976	0.9012	0.9008	0.8938	0.9037	0.8847	0.8923
60	0.95	0.8925	0.9007	0.9074	0.9103	0.9098	0.9026	0.9107	0.8958	0.9004
75	0.95	0.9013	0.9088	0.9140	0.9165	0.9159	0.9087	0.9169	0.9040	0.9072
90	0.95	0.9080	0.9150	0.9199	0.9218	0.9209	0.9146	0.9220	0.9108	0.9112
105	0.95	0.9135	0.9201	0.9234	0.9244	0.9235	0.9160	0.9265	0.9163	0.9125

Table 7: Simulated values of interval coverage probabilities with underlying $1 + \sqrt{0.6}t_5$ distribution.

n	Nom. level	T_n^{-1}	$T_n^{-1/2}$	T_n^0	$T_n^{2/3}$	T_n^1	T_n^2	W_n	S_n	λ_n
15	0.90	0.8856	0.8878	0.9010	0.9089	0.9089	0.9071	0.8962	0.9075	0.9099
20	0.90	0.8814	0.8849	0.8875	0.8904	0.8925	0.8923	0.9074	0.8962	0.9013
25	0.90	0.8758	0.8793	0.8874	0.8934	0.8930	0.8923	0.9048	0.8882	0.8974
30	0.90	0.8790	0.8848	0.8861	0.8916	0.8923	0.8901	0.9027	0.8933	0.8928
15	0.95	0.9329	0.9392	0.9529	0.9606	0.9613	0.9639	0.9546	0.9440	0.9670
20	0.95	0.9359	0.9452	0.9486	0.9522	0.9539	0.9544	0.9501	0.9505	0.9519
25	0.95	0.9336	0.9388	0.9409	0.9434	0.9438	0.9442	0.9502	0.9468	0.9472
30	0.95	0.9318	0.9336	0.9406	0.9442	0.9449	0.9446	0.9519	0.9408	0.9466

Table 8: Simulated values of interval coverage probabilities with underlying Poisson $\mathcal{P}(1)$ distribution.

Focusing on the coefficients of variation of the two continuous distributions, the exact coverage probabilities of the confidence intervals based on the empirical power divergence test statistics with $\lambda \in \{-1, \frac{1}{2}, 0, \frac{2}{3}, 1, 2\}$ are presented in Tables 6 and 7 when the nominal coverage probability based on the asymptotic distribution is either 90% or 95%. In addition, the empirical generalized Wald test statistic, the empirical generalized score test statistic and the empirical Lagrange multiplier test statistic are also considered for comparative purposes. From these results, we note that the empirical likelihood ratio test is not satisfactory and that among the empirical power divergence test statistics there is a good choice, for the underlying normal distribution, in the empirical chi-squared test statistic ($\lambda = 1$), and for the non-normal underlying distribution in the empirical Cressie-Read test statistic ($\lambda = \frac{2}{3}$), but there is not much difference between their performance. If we consider other test statistics, for the normally distributed observations, with theoretical asymptotic coverage as 95%, the empirical chi-squared test statistic ($\lambda = 1$) while for 90% coverage probability, the empirical Lagrange multiplier test statistic are seen to be the best ones but there is very little difference with respect to the empirical chi-squared test statistic ($\lambda = 1$). For the non-normal distribution, the empirical Wald test statistic is slightly superior than the empirical Cressie-Read test statistic, while the opposite seems to be the case for the normal distribution. Since in practice we do not know the form of the underlying distribution, based on this simulation study, we would recommend the use of either the empirical Cressie-Read test statistic or the empirical Wald test statistic. In Table 8, the same study has been carried out but for the case of Poisson distribution. The empirical likelihood ratio test is once again unsatisfactory and that the empirical Cressie-Read test statistic and the empirical chi-squared test statistic have a better coverage probability close to the nominal level. Further, the empirical Wald test statistic seems to be slightly superior than the empirical Cressie Read test statistic since it has a greater coverage probability in 5 of the 8 cases.

8 Concluding remarks

In a non-parametric setting, we have proposed here a broad family of empirical test statistics based on ϕ -divergence measures, first for a simple null hypothesis and then for a composite null hypothesis. Through numerical examples and simulations, it has been shown that confidence intervals constructed through the empirical ϕ -divergence based statistics improve the coverage probability of the empirical likelihood ratio test slightly. However, the most promising advantage of this new family of test statistics is that some members outperform the empirical likelihood ratio test in the presence of some shifted observations in the data. The approximation of the power function based on a specific sample provides an insight about the most appropriate robust ϕ -divergence statistic. These robust statistics tend to yield narrower confidence intervals in comparison to the empirical likelihood ratio test.

The development of these empirical ϕ -divergence tests in the two-sample and multi-sample situations will be of great interest. We are currently working in this direction and hope to report these findings in a future paper.

References

- [1] Baggerly, K. A. (1998). Empirical likelihood as a goodness-of-fit measure. *Biometrika*, **85**, 535–547.
- [2] Barnett, V. and Lewis, T. (1994). *Outliers in Statistical Data*. Third Edition. John Wiley & Sons, Chichester, England.
- [3] Basu, A., Shioya, H. and Park, C. (2011). *Statistical Inference: The Minimum Distance Approach*. Chapman & Hall/CRC Press, Boca Raton, Florida.
- [4] Bhattacharyya, A. (1943). On a measure of divergence between two statistical populations defined by their probability distributions. *Bulletin of the Calcutta Mathematical Society*, **35**, 99–109.
- [5] Broniatowski, M. and Keziou, A. (2012). Divergences and duality for estimating and test under moment condition models. *Journal of Statistical Planning and Inference*, **142**, 2554-2573.
- [6] Cressie, N. and Read, T. R. C. (1984). Multinomial goodness-of-fit tests. *Journal of the Royal Statistical Society, Series B*, **46**, 440–464.
- [7] Ferguson, T. S. (1996). *A Course in Large Sample Theory*. Chapman and Hall, New York.
- [8] Fraser, D. A. S. (1957). *Nonparametric Methods in Statistics*. John Wiley & Sons, New York.
- [9] Gokhale, D. V. and Kullback, S. (1978). *The Information in Contingency Tables*. Marcel Dekker, New York.
- [10] Hájek, J. and Sidák, Z. (1967). *Theory of Rank Tests*. Academic Press, New York.
- [11] Heritier, S. and Ronchetti, E. (1994). Robust Bounded-Influence Tests in General Parametric Models. *Journal of the American Statistical Association*, **89**, 897–904.
- [12] Le Cam, L. (1960). *Locally Asymptotic Normal Families of Distributions*. Universality of California Press, Berkeley, California.
- [13] Menéndez, M. L., Morales, D., Pardo, L. and Salicrú, M. (1995). Asymptotic behavior and statistical applications of divergence measures in multinomial populations: A unified study. *Statistical Papers*, **36**, 1-29.
- [14] Menéndez, M. L., Pardo, J. A., Pardo, L. and Pardo, M. C. (1997). Asymptotic approximations for the distributions of the (h, ϕ) -divergence goodness-of-fit statistics: Applications to Rényi's statistic. *Kybernetes*, **26**, 442-452.
- [15] Morales, D. and Pardo, L. (2001). Some approximations to power functions of ϕ -divergence tests in parametric models. *TEST*, **10**, 249-269.
- [16] Owen, A. B. (1988). Empirical likelihood ratio confidence interval for a single functional. *Biometrika*, **75**, 308-313.
- [17] Owen, A. B. (1990). Empirical likelihood confidence regions. *The Annals of Statistics*, **18**, 90-120.
- [18] Pardo, L. (2006). *Statistical Inference Based on Divergence Measures*. Chapman & Hall/ CRC Press, Boca Raton, Florida.
- [19] Qin, J. and Lawless, J. (1994). Empirical likelihood and general estimating equations. *The Annals of Statistics*, **22**, 300-325.

- [20] Ragusa, G. (2011). Minimum Divergence, Generalized Empirical likelihoods, and Higher Order Expansions. *Econometric Reviews*, **30**, 4, 406-456.
- [21] Rényi, A. (1961). On measures of entropy and information. *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability*, **1**, 547-561.
- [22] Schennach, S. M. (2007). Point Estimation with Exponentially Tilted Empirical Likelihood. *The Annals of Statistics*, **35**, 634-672.
- [23] Stigler, S. M. (1973). Simon Newcomb, Percy Daniell, and the history of robust estimation, 1885-1920. *Journal of the American Statistical Association*, **68**, 872-879.
- [24] Sharma, B. D. and Mittal, D. P. (1997). New non-additive measures of relative information. *Journal of Combinatorics, Information & Systems Science*, **2**, 122-133.
- [25] Toma, A. (2009) Optimal robust M-estimators using divergences. *Statistics & Probability Letters*, **79**, 1-5.
- [26] Toma, A. (2013) Robustness of dual divergence estimators for models satisfying linear constraints. *C. R. Acad. Sci. Paris, Ser. I*, **351**, 311-316.
- [27] van der Vaart, A. W. (2000). *Asymptotic Statistics*. Cambridge University Press, Cambridge.
- [28] Voinov, V., Nikulin, M. S. and Balakrishnan, N. (2013). *Chi-Squared Goodness of Fit Tests with Applications*. Academic Press, Boston.