On divergence tests for composite hypotheses under composite likelihood

N. Martín1 · L. Pardo2 · K. Zografos3

Abstract It is well-known that in some situations it is not easy to compute the likelihood function as the datasets might be large or the model is too complex. In that contexts composite likelihood, derived by multiplying the likelihoods of subjects of the variables, may be useful. The extension of the classical likelihood ratio test statistics to the framework of composite likelihoods is used as a procedure to solve the problem of testing in the context of composite likelihood. In this paper we introduce and study a new family of test statistics for composite likelihood: Composite $\phi$-divergence test statistics for solving the problem of testing a simple null hypothesis or a composite null hypothesis. To do that we introduce and study the asymptotic distribution of the restricted maximum composite likelihood estimate.

Keywords Composite likelihood · Maximum composite likelihood estimator · Restricted maximum composite likelihood estimator · Composite likelihood $\phi$-divergence test-statistics

Mathematics Subject Classification 62F03 · 62F05 · 62F30 · 62B10

1 Introduction

Hypothesis testing is a cornerstone of mathematical statistics and, subsequently, the theory of log-likelihood ratio tests is a cornerstone in the theory of testing statistical hypotheses. On the other hand, maximum likelihood estimators play a key role in the
development of log-likelihood ratio tests. Albeit maximum likelihood estimators can be easily obtained and they obey nice large sample properties, there are cases, like the case of complicated probabilistic models where the maximum likelihood estimators do not exist or they can not be obtained. In such a case the problem is usually transcended by the use of pseudo-likelihood functions and the respective estimators which result by maximization of such a function. Composite likelihood and the respective composite likelihood estimators are an appealing case of pseudo-likelihood estimators. There is an extensive literature of composite likelihood methods in statistics. The history of the composite likelihood may be traced back to the pseudo-likelihood approach of Besag (1974) for modeling spatial data. The name of composite likelihood was given by Lindsay (1988) to refer a likelihood type object formed by multiplying together individual component likelihoods, each of which corresponds to a marginal or conditional event. Composite likelihood finds applications in a variety of fields, including genetics, spatial statistics, longitudinal analysis, multivariate modeling, to mention a few. The special issue, with guest editors Reid et al. (2011), of the journal *Statistica Sinica* is devoted to composite likelihood methods and applications in several fields and it, moreover, provides with an exhaustive and updated source of knowledge in the subject. The recent papers by Reid et al. (2013) and Cattelan and Sartiri (2016) concentrate on new developments on the composite likelihood inference.

Distance or divergence based on methods of estimation and testing are fundamental tools and constitute a methodological part in the field of statistical inference. The monograph by Kullback (1959) was probably the starting point of applications of divergence measures in testing statistical hypotheses. The next important steps were the monographs by Read and Cressie (1988), Vajda (1989), Pardo (2006) and Basu et al. (2011) where distance, divergence or disparity methods were developed for estimation and testing. Thousands of papers have been also published in this frame and many of them have been exploited and mentioned in the above monographs. For testing a statistical hypothesis in a parametric framework, a test-statistic can be constructed by means of a distance or divergence measure between the empirical model and the model which is specified by the null hypothesis. The empirical model is the parametric model which governs the data with the unknown parameters to be replaced by their maximum likelihood estimators. The asymptotic normality of the maximum likelihood estimators is exploited along with the well known delta method in order to reach the asymptotic distribution of the respective divergence test-statistics.

The divergence test-statistics are based on considering the distance between density functions, chosen in an appropriate way. In statistical situations where we only consider composite densities, it seems completely natural to develop statistical procedures of testing based on divergence measures between the composite densities instead of the densities. This paper is motivated by the necessity to develop divergence based on methods, described above, for testing statistical hypotheses when the maximum composite likelihood estimators are used instead of the classic maximum likelihood estimators. In this setting, we consider divergence measures between composite density functions in order to get an appropriate test-statistic. The paper is organized as follows. The next section introduces the notation which will be used and it reviews composite likelihood estimators. Section 3 is devoted to presentation of the family of $\phi$-divergence test-statistics for testing simple null hypothesis. The formulation of
testing composite null hypotheses by means of \( \phi \)-divergence type test-statistics is the subject of Sect. 5. In order to get the results in relation to the composite null hypothesis it is necessary in Sect. 4 to introduce and study the restricted maximum composite estimator as well as its asymptotic distribution and the relationship between the restricted and the unrestricted maximum composite likelihood estimators. Section 6 presents two numerical examples and a simulation study is finally carried out in Sect. 7. The proofs of the main theoretic results are provided in Appendix.

2 Composite likelihood and divergence tests

We adopt here the notation by Joe et al. (2012) regarding composite likelihood function and the respective maximum composite likelihood estimators. In this regard, let \( \{ f(\cdot; \theta), \theta \in \Theta \subseteq \mathbb{R}^p, p \geq 1 \} \) be a parametric identifiable family of distributions for an observation \( y \), a realization of a random \( m \)-vector \( Y \). In this setting, the composite density based on \( K \) different margins or conditional distributions has the form

\[
CL(\theta, y) = \prod_{k=1}^{K} f_{A_k}(y_j, j \in A_k; \theta)
\]

and the composite log-density based on \( K \) different margins or conditional distributions has the form

\[
c\ell(\theta, y) = \sum_{k=1}^{K} w_k \ell_{A_k}(\theta, y),
\]

with

\[
\ell_{A_k}(\theta, y) = \log f_{A_k}(y_j, j \in A_k; \theta),
\]

where \( \{A_k\}_{k=1}^{K} \) is a family of random variables associated either with marginal or conditional distributions involving some \( y_j, j \in \{1, \ldots, m\} \) and \( w_k, k = 1, \ldots, K \) are non-negative and known weights. If the weights are all equal, then they can be ignored, actually all the statistical procedures produce equivalent results.

Let also \( y_1, \ldots, y_n \) be independent and identically distributed replications of \( y \). We denote by

\[
c\ell(\theta, y_1, \ldots, y_n) = \sum_{i=1}^{n} c\ell(\theta, y_i),
\]

the composite log-likelihood function for the whole sample. In complete accordance with the classic maximum likelihood estimator, the maximum composite likelihood estimator \( \hat{\theta}_c \) is defined by

\[
\hat{\theta}_c = \arg \max_{\theta \in \Theta} \sum_{i=1}^{n} c\ell(\theta, y_i) = \arg \max_{\theta \in \Theta} \sum_{i=1}^{n} \sum_{k=1}^{K} w_k \ell_{A_k}(\theta, y_i).
\]
It can be also obtained by the solution of the equation
\[
\mathbf{u}(\theta, y_1, \ldots, y_n) = \mathbf{0}_p,
\]
where
\[
\mathbf{u}(\theta, y_1, \ldots, y_n) = \frac{\partial c\ell(\theta, y_1, \ldots, y_n)}{\partial \theta} = \sum_{i=1}^{n} \sum_{k=1}^{K} w_k \frac{\partial \ell_k(\theta, y)}{\partial \theta},
\]
is the composite likelihood score function, that is the partial derivative of the composite log-likelihood with respect to the parameter vector.

The maximum composite likelihood estimator \( \hat{\theta}_c \) obeys asymptotic normality and in particular
\[
\sqrt{n}(\hat{\theta}_c - \theta) \xrightarrow{L} N(0, G_*^{-1}(\theta)),
\]
where \( G_*(\theta) \) denotes Godambe information matrix, defined by
\[
G_*(\theta) = H(\theta)J^{-1}(\theta)H(\theta),
\]
with \( H(\theta) \) being the sensitivity or Hessian matrix and \( J(\theta) \) being the variability matrix, defined, respectively, by
\[
H(\theta) = E_\theta[-\frac{\partial}{\partial \theta} \mathbf{u}^T(\theta, Y)],
\]
\[
J(\theta) = \text{Var}_\theta[\mathbf{u}(\theta, Y)] = E_\theta[\mathbf{u}(\theta, Y)\mathbf{u}^T(\theta, Y)],
\]
where the superscript \( T \) denotes the transpose of a vector or a matrix.

The matrices \( H(\theta) \) and \( J(\theta) \) are, by definition, nonnegative definite matrices but throughout this paper both, \( H(\theta) \) and \( J(\theta) \), are assumed to be positive definite matrices. Since the component score functions can be correlated, we have \( H(\theta) \neq J(\theta) \).

If \( c\ell(\theta, y) \) is the full log-likelihood function then \( H(\theta) = J(\theta) = I_F(\theta) \), being \( I_F(\theta) \) the Fisher information matrix of the model. Using multivariate version of the Cauchy-Schwarz inequality we have that the matrix \( G_*(\theta) - I_F(\theta) \) is non-negative definite, i.e., the full likelihood function is more efficient than any other composite likelihood function (cf. Lindsay 1988, Lemma 4A).

For two densities \( p \) and \( q \) associated with two \( m \)-dimensional random variables respectively, Csiszár’s \( \phi \)-divergence between \( p \) and \( q \) is defined by
\[
D_\phi(p, q) = \int_{\mathbb{R}^m} q(y)\phi \left( \frac{p(y)}{q(y)} \right) dy,
\]
where \( \phi, \phi \in \Psi \), is a real valued convex function with
\[
\Psi = \{ \phi : \phi \text{ is strictly convex}, \phi(1) = \phi'(1) = 0, 0\phi \left( \frac{0}{0} \right) = 0, 0\phi \left( \frac{0}{0} \right) = \lim_{v \to \infty} \frac{\phi(v)}{v} \},
\]
On divergence tests for composite hypotheses... 1887

On divergence tests for composite hypotheses... 1887

Csiszár’s \( \phi \)-divergence has been axiomatically characterized and studied extensively by Liese and Vajda (1987, 2006), Vajda (1989), and Stummer and Vajda (2010), among many others. Particular choices of the convex functions \( \phi \), lead to important measures of divergence including Kullback and Leibler (1951) divergence, Rényi (1961) divergence and Cressie and Read (1984) \( \lambda \)-power divergence, to mention a few. Csiszár’s \( \phi \)-divergence can be extended and used in testing hypotheses on more than two distributions (cf. Zografos 1998 and references appeared therein).

In this paper we are going to consider \( \phi \)-divergence measures between the composite densities \( \mathcal{C}L(\theta_1, y) \) and \( \mathcal{C}L(\theta_2, y) \) in order to solve different problems of testing hypotheses. The \( \phi \)-divergence measure between composite densities \( \mathcal{C}L(\theta_1, y) \) and \( \mathcal{C}L(\theta_2, y) \) will be defined by

\[
D_{\phi}(\theta_1, \theta_2) = \int_{\mathbb{R}^m} \mathcal{C}L(\theta_2, y) \phi \left( \frac{\mathcal{C}L(\theta_1, y)}{\mathcal{C}L(\theta_2, y)} \right) dy. \tag{1}
\]

An important particular case is the Kullback–Leibler divergence measure, obtained from (1) with \( \phi(x) = x \log x - x + 1 \), i.e.

\[
D_{\text{Kullback}}(\theta_1, \theta_2) = \int_{\mathbb{R}^m} \mathcal{C}L(\theta_1, y) \log \frac{\mathcal{C}L(\theta_1, y)}{\mathcal{C}L(\theta_2, y)} dy.
\]

Based on (1) we shall present in this paper some new composite test-statistics for testing simple null hypothesis as well as composite null hypothesis. To the best of our knowledge, it is the first time that \( \phi \)-divergences are used for solving testing problems in the context of composite likelihood. However, the Kullback–Leibler divergence has been used, in the context of composite likelihood, by many authors in model selection, see for instance Varin (2008).

3 Hypothesis testing: simple null hypothesis

In this section we are interested in testing

\[
H_0 : \theta = \theta_0 \text{ versus } H_1 : \theta \neq \theta_0. \tag{2}
\]

If we consider the \( \phi \)-divergence between the composite densities \( \mathcal{C}L(\hat{\theta}_c, y) \) and \( \mathcal{C}L(\theta_0, y) \),

\[
D_{\phi}(\hat{\theta}_c, \theta_0) = \int_{\mathbb{R}^m} \mathcal{C}L(\theta_0, y) \phi \left( \frac{\mathcal{C}L(\hat{\theta}_c, y)}{\mathcal{C}L(\theta_0, y)} \right) dy,
\]

it verifies \( D_{\phi}(\hat{\theta}_c, \theta_0) \geq 0 \), and the equality holds if and only if \( \mathcal{C}L(\hat{\theta}_c, y) = \mathcal{C}L(\theta_0, y) \). Small values of \( D_{\phi}(\hat{\theta}_c, \theta_0) \) are in favour of \( H_0 : \theta = \theta_0 \), while large values of \( D_{\phi}(\hat{\theta}_c, \theta_0) \) suggest rejection of \( H_0 \). This is due to the fact that large values of \( D_{\phi}(\hat{\theta}_c, \theta_0) \) suggest that the model \( \mathcal{C}L(\hat{\theta}_c, y) \) is not very close to \( \mathcal{C}L(\theta_0, y) \). Therefore, \( H_0 \) is rejected if \( D_{\phi}(\hat{\theta}_c, \theta_0) > c \), where \( c \) is specified so that the significance level of
the test to be equal to $\alpha$ ($0 < \alpha < 1$). In order to obtain $c$, in the next theorem we shall obtain the asymptotic distribution of

$$T_{\phi,n}(\hat{\theta}_c, \theta_0) = \frac{2n}{\phi''(1)} D_{\phi}(\hat{\theta}_c, \theta_0),$$

which we shall refer to as composite $\phi$-divergence test-statistics for testing a simple null hypothesis. The asymptotic null distribution of $T_{\phi,n}(\hat{\theta}_c, \theta_0)$ is derived in the next theorem and its proof is presented in “Proof of Theorem 1” section in Appendix.

**Theorem 1** Under the null hypothesis $H_0 : \theta = \theta_0$,

$$T_{\phi,n}(\hat{\theta}_c, \theta_0) \xrightarrow{L} \sum_{i=1}^{k} \lambda_i Z_i^2,$$

where $\lambda_i$, $i = 1, \ldots, k$, are the eigenvalues of the matrix $J(\theta_0)G_*^{-1}(\theta_0)$, $k = \text{rank}(J(\theta_0))$ and $Z_1, \ldots, Z_k$ are independent standard normal random variables.

**Remark 1** Based on the previous Theorem we shall reject the null hypothesis $H_0 : \theta = \theta_0$ if $T_{\phi,n}(\hat{\theta}_c, \theta_0) > c_\alpha$, where $c_\alpha$ is the quantile of order $1 - \alpha$ of the asymptotic distribution of $T_{\phi,n}(\hat{\theta}_c, \theta_0)$, given in (3). The value of $k$ is usually $p$, since the components of $\theta$ are assumed to be non-redundant.

In most cases, the power function of this testing procedure can not be calculated explicitly. In the following theorem we present a useful asymptotic result for approximating the power function. The proof is given in “Proof of Theorem 2” section in Appendix.

**Theorem 2** Let $\theta^*$ be the true parameter, with $\theta^* \neq \theta_0$. Then it holds

$$\sqrt{n}(D_{\phi}(\hat{\theta}_c, \theta_0) - D_{\phi}(\theta^*, \theta_0)) \xrightarrow{L} \mathcal{N}(0, \sigma^2_{\phi}(\theta^*)),$$

where

$$\sigma^2_{\phi}(\theta^*) = q^T G_*^{-1}(\theta^*) q,$$

and $q = (q_1, \ldots, q_p)^T$ with $q_j = \frac{\partial D_{\phi}(\theta, \theta_0)}{\partial \theta_j} \bigg|_{\theta = \theta^*}$, $j = 1, \ldots, p$.

From Theorem 2, a first approximation to the power function, at $\theta^* \neq \theta_0$, is given by

$$\beta_{n,\phi}(\theta^*) = 1 - \Phi\left( \frac{\sqrt{n}(D_{\phi}(\hat{\theta}_c, \theta_0) - D_{\phi}(\theta^*, \theta_0))}{\sigma_{\phi}(\theta^*)} \right),$$

where $\Phi$ is the standard normal distribution function. If some $\theta^* \neq \theta_0$ is the true parameter, the probability of rejecting $\theta_0$ with the rejection rule $T_{\phi,n}(\hat{\theta}_c, \theta_0) > c_\alpha$, for fixed significance level $\alpha$, tends to one as $n \to \infty$. Hence, the test is consistent in Fraser's sense.
Remark 2 In the context of composite likelihood, the most common test statistics are the composite likelihood ratio test statistic, the composite Wald type and the composite score-type test statistics. Details about these test statistics can be found, for instance, in Varin et al. (2011). In this paper we pay special attention only to the composite likelihood ratio test statistic because we study the behavior of it in comparison to the family of composite $\phi$-divergence test statistics, considered for the first time in this paper, to the best of our knowledge. The interest is not focused in the presentation of a new family of composite test statistics for solving problems that are not possible to be solved with the composite likelihood ratio test statistic. Our aim is to show that in some situations, the family of composite $\phi$-divergence test statistics, introduced in this paper, has a better behavior in the sense which is presented in the last section which is devoted to the simulation study.

One Referee pointed out the possibility of considering a composite gradient-type test statistic in this context. To the best of our knowledge, there is not any study in the respective literature with the composite gradient-type test statistic. The composite gradient-type test statistic can be defined by,

$$G_n(\theta_0) = u(\theta_0; y_1, \ldots, y_n)^T J^{-1}(\theta_0) H(\theta_0)(\hat{\theta}_c - \theta_0).$$

(4)

It is well known that if $c\ell(\theta, y)$ is the full log-likelihood function then $H(\theta_0) = J(\theta_0) = I_F(\theta_0)$ and the expression of the composite gradient-type test, given in (4), coincides with the classical gradient test, as it has been initially defined by Terrell (2002).

In relation to the asymptotic distribution of $G_n(\theta_0)$ in (4), we have,

$$u(\hat{\theta}_c; y_1, \ldots, y_n) = u(\theta_0; y_1, \ldots, y_n) + \left. \frac{\partial u(\theta; y_1, \ldots, y_n)}{\partial \theta} \right|_{\theta=\theta_0} (\hat{\theta}_c - \theta_0) + o_P(n^{-1/2}).$$

Then, by definition of the maximum composite estimator $\hat{\theta}_c$, $u(\hat{\theta}_c; y_1, \ldots, y_n) = 0_p$ and the above equation leads to

$$\hat{\theta}_c - \theta_0 = - \left( \frac{\partial u(\theta; y_1, \ldots, y_n)}{\partial \theta} \right)^{-1}_{\theta=\theta_0} u(\theta_0; y_1, \ldots, y_n) + o_P(n^{-1/2}),$$

or

$$\sqrt{n} (\hat{\theta}_c - \theta_0) = - \left( \frac{1}{n} \frac{\partial u(\theta; y_1, \ldots, y_n)}{\partial \theta} \right)^{-1}_{\theta=\theta_0} \frac{1}{\sqrt{n}} u(\theta_0; y_1, \ldots, y_n) + o_P(1).$$

Taking into account that

$$- \left( \frac{1}{n} \frac{\partial u(\theta; y_1, \ldots, y_n)}{\partial \theta} \right)^{-1}_{\theta=\theta_0} \xrightarrow{n\to\infty} H^{-1}(\theta_0),$$
it is clear that
\[
\sqrt{n} \left( \hat{\theta}_c - \theta_0 \right) = H^{-1}(\theta_0) \frac{1}{\sqrt{n}} u(\theta_0; y_1, \ldots, y_n) + o_P(1).
\]

Based on this last equation,
\[
G_n(\theta_0) = \frac{1}{n} u(\theta_0; y_1, \ldots, y_n)^T J^{-1}(\theta_0) u(\theta_0; y_1, \ldots, y_n) + o_P(1) \\
= R_n(\theta_0) + o_P(1).
\]

Therefore, the composite gradient-type test statistic has the same asymptotic distribution as that of the composite score-type test statistic \( R_n(\theta_0) \). For more details about the gradient test statistic in the full likelihood setup we mention the initial paper by Terrell (2002) and the recent monograph by Lemonte (2016).

It is very clear that there is enough material in this remark to prepare in the future a new paper, devoted to the composite Wald-type test, composite score-type test, composite gradient-type test in comparison to the composite \( \phi \)-divergence test statistics, introduced in the present paper.

4 Restricted maximum composite likelihood estimator

In some common situations such as the problem of testing composite null hypotheses, it is necessary to get the maximum composite likelihood estimator which is restricted by some constraints of the type
\[
g(\theta) = 0_r,
\]
where \( g \) is a function such that \( g : \Theta \subseteq \mathbb{R}^p \to \mathbb{R}^r \), \( r \) is an integer, with \( r < p \) and \( 0_r \) denotes the null vector of dimension \( r \). The function \( g \) is a vector valued function such that the \( p \times r \) matrix
\[
G(\theta) = \frac{\partial g^T(\theta)}{\partial \theta}
\]
exists and is continuous in \( \theta \) with \( \text{rank}(G(\theta)) = r \). The restricted maximum composite likelihood estimator of \( \theta \) is defined by
\[
\tilde{\theta}_{rc} = \arg \max_{\theta \in \Theta, g(\theta) = 0_r} \sum_{i=1}^n c\ell(\theta, y_i) = \arg \max_{\theta \in \Theta, g(\theta) = 0_r} \sum_{i=1}^n \sum_{k=1}^K w_k \ell_{A_k}(\theta, y_i),
\]
and it is obtained by the solution of the restricted likelihood equations
\[
\sum_{i=1}^n \frac{\partial}{\partial \theta} c\ell(\theta, y_i) + G(\theta)\lambda = 0_p, \\
g(\theta) = 0_r,
\]
where \( \lambda \in \mathbb{R}^r \) is a vector of Lagrange multipliers.
In this section we shall get the asymptotic distribution of the restricted maximum composite likelihood estimator. Consider a random sample \( y_1, \ldots, y_n \) from the parametric model \( f(\cdot; \theta), \theta \in \Theta \subseteq \mathbb{R}^p, p \geq 1 \), and let \( \hat{\theta}_c \) and \( \tilde{\theta}_{rc} \) be the unrestricted and the restricted maximum composite likelihood estimators of \( \theta \). The following result derives the asymptotic distribution of \( \tilde{\theta}_{rc} \).

**Theorem 3** Under the constraints \( g(\theta) = 0_r \) the restricted maximum composite likelihood estimator obeys asymptotic normality in the sense

\[
\sqrt{n}(\tilde{\theta}_{rc} - \theta) \xrightarrow{\mathcal{L}} N(0_p, \Sigma_{rc}),
\]

with

\[
\Sigma_{rc} = P(\theta)J(\theta)P^T(\theta),
\]

\[
P(\theta) = H^{-1}(\theta) + Q(\theta)G^T(\theta)H^{-1}(\theta),
\]

\[
Q(\theta) = -H^{-1}(\theta)G(\theta)\left[G^T(\theta)H^{-1}(\theta)G(\theta)\right]^{-1}.
\]

The proof of the Theorem is outlined in “Proof of Theorem 3” section of Appendix.

The lemma that follows formulates the relationship between the maximum composite and the restricted maximum composite likelihood estimators \( \hat{\theta}_c \) and \( \tilde{\theta}_{rc} \) respectively.

**Lemma 4** The estimators of \( \theta, \hat{\theta}_c \) and \( \tilde{\theta}_{rc} \), satisfy

\[
\sqrt{n}(\tilde{\theta}_{rc} - \theta) = \left(I_p + Q(\theta)G^T(\theta)\right)\sqrt{n}(\hat{\theta}_c - \theta) + o_P(1).
\]

The proof of the lemma is given in “Proof of Lemma 4” section of Appendix.

**5 Hypothesis testing: composite null hypothesis**

Following Basu et al. (2015), consider the null hypothesis

\[ H_0 : \theta \in \Theta_0 \text{ against } H_0 : \theta \notin \Theta_0, \]

which restricts the parameter \( \theta \) to a subset \( \Theta_0 \) of \( \Theta \subseteq \mathbb{R}^p, p \geq 1 \). Based on Sen and Singer (1993, p. 239), we shall assume that the composite null hypothesis \( H_0 : \theta \in \Theta_0 \) can be equivalently formulated in the form

\[ H_0 : g(\theta) = 0_r. \tag{7} \]

For testing the composite null hypothesis (7) on the basis of a random sample \( y_1, \ldots, y_n \) from the parametric model \( f(\cdot; \theta), \theta \in \Theta \subseteq \mathbb{R}^p, p \geq 1 \), there are well-known procedures to be applied. The likelihood ratio test-statistic, the Wald and Rao statistics are used in this direction. Test-statistics based on divergences or disparities,
as they have been described and mentioned above, constitute an appealing procedure for testing this hypothesis. Moreover, there are composite likelihood methods analog to the likelihood ratio test or the Wald test. However, there are not composite likelihood versions of the tests based on divergence measures, to the best of our knowledge. So, our aim in this section is to develop composite test-statistics for testing (7), on the basis of divergence measures and in the composite likelihood framework. The $\phi$-divergence between the composite densities $CL(\theta_c, y)$ and $CL(\theta_{rc}, y)$, is given by

$$D_\phi(\theta_c, \theta_{rc}) = \int_{\mathbb{R}^m} CL(\theta_{rc}, y) \phi \left( \frac{CL(\theta_c, y)}{CL(\theta_{rc}, y)} \right) dy.$$ 

Based on the property $D_\phi(\theta_c, \theta_{rc}) \geq 0$, with equality, if and only if $CL(\theta_{rc}, y) = CL(\theta_c, y)$, small values of $D_\phi(\theta_c, \theta_{rc})$ are in favour of (7), while large values of $D_\phi(\theta_c, \theta_{rc})$ suggest that the composite densities $CL(\theta_{rc}, y)$ and $CL(\theta_c, y)$ are not the same and is expected for the respective theorectic models $f(\cdot; \theta)$ with $\theta \in \Theta$ and $f(\cdot; \theta)$ with $\theta \in \Theta_0$. So, small values of $D_\phi(\theta_c, \theta_{rc})$ are in favor of (7) while large values of $D_\phi(\theta_c, \theta_{rc})$ suggest the rejection of $H_0$. Given the asymptotic normality of the maximum composite likelihood estimator $\hat{\theta}_c$, the asymptotic normality of the respective restricted estimator $\hat{\theta}_{rc}$ has been verified in the previous section. The asymptotic distribution of the test-statistic $D_\phi(\theta_c, \theta_{rc})$ is the subject of this section.

Based on Theorem 3 and Lemma 4, the composite likelihood $\phi$-divergence test-statistic is introduced in the next theorem and its asymptotic distribution is derived under the composite null hypothesis (7). The standard regularity assumptions of asymptotic statistic are assumed to be valid (cf. Serfling 1980, p. 144 and Pardo 2006, p. 58).

**Theorem 5** Under the composite null hypothesis (7),

$$T_{\phi,n}(\hat{\theta}_c, \hat{\theta}_{rc}) = \frac{2n}{\phi''(1)} D_\phi(\hat{\theta}_c, \hat{\theta}_{rc}) \xrightarrow{L} \sum_{i=1}^{k} \beta_i Z_i^2,$$

where $\beta_i, i = 1, \ldots, k$, are the eigenvalues of the matrix

$$J(\theta) Q(\theta) G^T(\theta) G^{-1}_*(\theta) G(\theta) Q^T(\theta),$$

$$k = \text{rank} \left( Q(\theta) G^T(\theta) G^{-1}_*(\theta) G(\theta) Q^T(\theta) \right) J(\theta) Q(\theta) G^T(\theta) G^{-1}_*(\theta) G(\theta) Q^T(\theta),$$

and $Z_1, \ldots, Z_k$ are independent standard normal random variables.

The proof of this theorem is presented in Appendix, “Proof of Theorem 5” section. In the following we refer $T_{\phi,n}(\hat{\theta}_c, \hat{\theta}_{rc})$ by composite $\phi$-divergence test-statistics for testing composite null hypothesis.

**Remark 3** For the testing problem considered in this section it is perhaps well-known the composite likelihood ratio test but it was not possible for us to find it in the statistical literature. This test will be used in Sect. 4 and this is the reason to develop the said test in the present remark.
We shall denote
\[ c \ell (\theta) = \sum_{i=1}^{n} c \ell (\theta, y_i). \]
The composite likelihood ratio test for testing the composite null hypothesis (7), considered in this paper, is defined by
\[ \lambda_n (\hat{\theta}_c, \tilde{\theta}_{rc}) = 2 (c \ell (\hat{\theta}_c) - c \ell (\tilde{\theta}_{rc})). \]
A second order Taylor expansion gives
\[ c \ell (\tilde{\theta}_{rc}, y_i) - c \ell (\hat{\theta}_c, y_i) = \left. \frac{\partial c \ell (\theta, y_i)}{\partial \theta} \right|_{\theta=\hat{\theta}_c} (\hat{\theta}_c - \tilde{\theta}_{rc}) + \frac{1}{2} (\hat{\theta}_c - \tilde{\theta}_{rc})^T \left. \frac{\partial^2 c \ell (\theta, y_i)}{\partial \theta \partial \theta^T} \right|_{\theta=\hat{\theta}_c} (\hat{\theta}_c - \tilde{\theta}_{rc}) + o_P(1). \]
But,
\[ \left. \frac{\partial c \ell (\theta, y_i)}{\partial \theta} \right|_{\theta=\hat{\theta}_c} = 0_p \quad \text{and} \quad 1/n \sum_{i=1}^{n} \left. \frac{\partial^2 c \ell (\theta, y_i)}{\partial \theta \partial \theta^T} \right|_{\theta=\hat{\theta}_c} \xrightarrow{n \to \infty} -H(\theta_0), \]
and therefore,
\[ 2(c \ell (\hat{\theta}_c) - c \ell (\tilde{\theta}_{rc})) = \sqrt{n} (\hat{\theta}_c - \tilde{\theta}_{rc})^T H(\theta) \sqrt{n} (\hat{\theta}_c - \tilde{\theta}_{rc}) + o_P(1). \]
Taking into account that
\[ \sqrt{n} (\hat{\theta}_c - \tilde{\theta}_{rc}) \xrightarrow{\mathcal{L}} \mathcal{N} \left( 0_p, Q(\theta) G^T(\theta) G^{-1}_* (\theta) G(\theta) Q^T(\theta) \right), \]
(cf. Proof of Theorem 5 in Appendix)
\[ \lambda_n (\hat{\theta}_c, \tilde{\theta}_{rc}) = 2(c \ell (\hat{\theta}_c) - c \ell (\tilde{\theta}_{rc})) \xrightarrow{n \to \infty} \sum_{i=1}^{\ell} \gamma_i Z_i^2, \quad (8) \]
where \( \gamma_i, i = 1, \ldots, \ell \) are the non null eigenvalues of the matrix
\[ H(\theta) Q(\theta) G^T(\theta) G^{-1}_* (\theta) G(\theta) Q^T(\theta), \]
with
\[ \ell = \text{rank} \left( Q(\theta) G^T(\theta) G^{-1}_* (\theta) G(\theta) Q^T(\theta) H(\theta) Q(\theta) G^T(\theta) G^{-1}_* (\theta) G(\theta) Q^T(\theta) \right), \]
and \( Z_i, i = 1, \ldots, \ell \) are independent standard normal random variables.
Remark 4 In order to avoid the problem of getting percentiles or probabilities from the distribution of linear combinations of \( \chi^2 \) distributions we are going to present some adjusted composite likelihood \( \phi \)-divergence test-statistics.

Following Corollary 1 of Rao and Scott (1981) one can use the statistic

\[
1 \quad T_{\phi,n}(\theta_c, \tilde{\theta}_{rc}) = \frac{T_{\phi,n}(\theta_c, \tilde{\theta}_{rc})}{\lambda_{\max}} \leq \sum_{i=1}^{k} Z_i^2,
\]

where \( \lambda_{\max} = \max(\beta_1, \ldots, \beta_k) \). As \( \sum_{i=1}^{k} Z_i^2 \sim \chi_k^2 \), a strategy that rejects the null hypothesis \( H_0 : g(\theta) = 0_r \) for \( 1 \quad T_{\phi,n}(\theta_c, \tilde{\theta}_{rc}) > \chi_{k,1-\alpha}^2 \) produces an asymptotically conservative test at a nominal level \( \alpha \), where \( \chi_{k,1-\alpha}^2 \) is the quantile of order \( 1 - \alpha \) for \( \chi_k^2 \).

Another approximation to the asymptotic tail probabilities of \( T_{\phi,n}(\theta_c, \tilde{\theta}_{rc}) \) can be obtained through the modification

\[
2 \quad T_{\phi,n}(\theta_c, \tilde{\theta}_{rc}) = \frac{T_{\phi,n}(\theta_c, \tilde{\theta}_{rc})}{\bar{\lambda}},
\]

where \( \bar{\lambda} = \frac{1}{r} \sum_{i=1}^{k} \beta_i \) (see Satterthwaite 1946). The distribution of this last statistic is approximated by a \( \chi^2 \) distribution with \( k \) degrees of freedom. In this case we can observe that

\[
E \left[ 2 \quad T_{\phi,n}(\theta_c, \tilde{\theta}_{rc}) \right] = k = E \left[ \chi_k^2 \right],
\]

\[
\text{Var} \left[ 2 \quad T_{\phi,n}(\theta_c, \tilde{\theta}_{rc}) \right] = 2 \sum_{i=1}^{k} \beta_i^2 \frac{1}{\bar{\lambda}^2} = 2k + 2 \sum_{i=1}^{r} \frac{(\beta_i - \bar{\lambda})^2}{\bar{\lambda}^2} > 2k = \text{Var} \left[ \chi_k^2 \right].
\]

If we denote by \( \Lambda = \text{diag}(\beta_1, \ldots, \beta_k) \), we get

\[
E \left[ \sum_{i=1}^{k} \beta_i Z_i^2 \right] = \sum_{i=1}^{k} \beta_i = \text{trace}(\Lambda) = \text{trace}(A(\theta) G_*(\theta)).
\]

The test given by the statistic \( 2 \quad T_{\phi,n}(\theta_c, \tilde{\theta}_{rc}) \) is more conservative than the one based on

\[
3 \quad T_{\phi,n}(\theta_c, \tilde{\theta}_{rc}) = \frac{2 \quad T_{\phi,n}(\theta_c, \tilde{\theta}_{rc})}{\nu} = \frac{T_{\phi,n}(\theta_c, \tilde{\theta}_{rc})}{\nu \bar{\lambda}},
\]

and we can find \( \nu \) by imposing the condition \( \text{Var} \left[ 3 \quad T_{\phi,n}(\theta_c, \tilde{\theta}_{rc}) \right] = 2E \left[ 3 \quad T_{\phi,n}(\theta_c, \tilde{\theta}_{rc}) \right] \), as in the \( \chi^2 \) distribution. Since

\[
E \left[ 3 \quad T_{\phi,n}(\theta_c, \tilde{\theta}_{rc}) \right] = \frac{k}{\nu} \quad \text{and} \quad \text{Var} \left[ 3 \quad T_{\phi,n}(\theta_c, \tilde{\theta}_{rc}) \right] = \frac{2k}{\nu},
\]

\[
\nu = 1 + \sum_{i=1}^{k} \frac{(\beta_i - \bar{\lambda})^2}{k \bar{\lambda}^2} = 1 + CV^2(\beta_i)_{i=1}^{k},
\]
where \( CV \) represents the coefficient of variation. Then a \( \chi^2 \) distribution with \( \frac{k}{v} \) degrees of freedom approximates the asymptotic distribution of the statistic \( 3T_{\phi,n}(\hat{\theta}_c, \tilde{\theta}_{rc}) \) for large \( n \).

The degrees of freedom of \( 3T_{\phi,n}(\hat{\theta}_c, \tilde{\theta}_{rc}) \) is \( \frac{k}{v} \), which may not be an integer. To avoid this deficiency one can modify the statistic such that the first two moments match specifically with the \( \chi^2_k \) distribution (rather than with just any other \( \chi^2 \) distribution). Specifically let

\[
X = 2T_{\phi,n}(\hat{\theta}_c, \tilde{\theta}_{rc}).
\]

We have

\[
E[X] = k = E\left[\chi^2_k\right],
\]

\[
Var[X] = \frac{2\sum_{i=1}^{k}\beta_i^2}{\lambda^2} = 2k + 2\sum_{i=1}^{k} \left(\frac{\beta_i - \lambda}{\lambda}\right)^2 = 2k + c,
\]

where \( c \) stands for the last indicated term in the previous expression. We define \( Y = (X - a)/b \), where the constants \( a \) and \( b \) are such that

\[
E(Y) = k, \quad Var(Y) = 2k.
\]

Thus,

\[
\frac{k - a}{b} = k, \quad \frac{2k + c}{b^2} = 2k.
\]

Solving these equations, we get

\[
b = \sqrt{1 + \frac{c}{2k}}, \quad a = k(1 - b).
\]

Thus it makes sense to consider another modification of the statistic given by

\[
4T_{\phi,n}(\hat{\theta}_c, \tilde{\theta}_{rc}) = \frac{2T_{\phi,n}(\hat{\theta}_c, \tilde{\theta}_{rc}) - a}{b},
\]

the large sample distribution of which may be approximated by the \( \chi^2_k \) distribution.

The approximation presented in this remark for the asymptotic distribution of the \( \phi \)-divergence test-statistics, \( T_{\phi,n}(\hat{\theta}_c, \tilde{\theta}_{rc}) \), can be used in the approximation of the \( \phi \)-divergence test-statistics \( T_{\phi,n}(\hat{\theta}_c, \theta_0) \) as well.

By the previous theorem, the null hypothesis should be rejected if \( T_{\phi,n}(\hat{\theta}_c, \tilde{\theta}_{rc}) \geq c_{\alpha} \), where \( c_{\alpha} \) is the quantile of order \( 1 - \alpha \) of the asymptotic distribution of \( T_{\phi,n}(\hat{\theta}_c, \tilde{\theta}_{rc}) \). The following theorem can be used to approximate the power function. Assume that \( \theta \notin \Theta_0 \) is the true value of the parameter so that \( \hat{\theta}_c \overset{a.s.}{\rightarrow} \theta \) and that there exists \( \theta^* \in \Theta_0 \) such that the restricted maximum composite likelihood estimator satisfies \( \tilde{\theta}_{rc} \overset{a.s.}{\rightarrow} \theta^* \), as well as
\[
n^{1/2}(\hat{\theta}_c, \tilde{\theta}_{rc} - (\theta, \theta^*)) \xrightarrow{L} \mathcal{N}\left(\begin{pmatrix} 0_p \\ 0_p \end{pmatrix}, \begin{pmatrix} G_{\theta}(\theta, \theta^*) & A_{12}(\theta, \theta^*) \\ A_{12}^T(\theta, \theta^*) \Sigma(\theta, \theta^*) \end{pmatrix}\right),
\]

where \(A_{12}(\theta, \theta^*)\) and \(\Sigma(\theta, \theta^*)\) are appropriate \(p \times p\) matrices. We have then the following result, the proof of which is outlined in “Proof of Theorem 6” section of Appendix.

**Theorem 6** Under \(H_1\) we have

\[
n^{1/2} \left(D_{\phi}(\hat{\theta}_c, \tilde{\theta}_{rc}) - D_{\phi}(\theta, \theta^*)\right) \xrightarrow{L} \mathcal{N}(0, \sigma^2(\theta, \theta^*)),
\]

where

\[
\sigma^2(\theta, \theta^*) = t^T G_{\theta}^{-1}(\theta)t + 2t^T A_{12}(\theta, \theta^*)s + s^T \Sigma(\theta, \theta^*) s.
\]

Remark 5 On the basis of the previous theorem we can get an approximation of the power function

\[
\pi_n^\phi(\theta) = \Pr_\theta\left(T_{\phi,n}(\hat{\theta}_c, \tilde{\theta}_{rc}) \geq c\right)
\]

at \(\theta\) as

\[
\pi_{n,\alpha}^\beta,\gamma(\theta) = 1 - \Phi\left(\frac{n^{1/2}}{\sigma(\theta, \theta^*)} \left(\frac{c}{2n} - D_{\phi}(\theta, \theta^*)\right)\right),
\]

where \(\Phi(x)\) is the standard normal distribution function and \(\sigma^2(\theta, \theta^*)\) was defined in (9).

If some \(\theta \neq \theta^*\) is the true parameter, then the probability of rejecting \(H_0\) with the rule that it is rejected when \(T_{\phi,n}(\hat{\theta}_c, \tilde{\theta}_{rc}) \geq c\) for a fixed test size \(\alpha\) tends to one as \(n \to \infty\). The test-statistic is consistent in the Fraser’s sense.

Obtaining the approximate sample size \(n\) to guarantee a power of \(\pi\) at a given alternative \(\theta^*\) is an interesting application of formula (10). Let \(n^*\) be the positive root of Eq. (10), i.e.

\[
n^* = \frac{A + B + \sqrt{A(A + 2B)}}{2D_{\phi}^2(\theta, \theta^*)},
\]

where

\[
A = \sigma^2(\theta, \theta^*) \left(\Phi^{-1}(1 - \pi)\right)^2 \quad \text{and} \quad B = cD_{\phi}(\theta, \theta^*).
\]

Then the required sample size is \(n = \left[n^*\right] + 1\), where \([\cdot]\) is used to denote “integer part of”.

Remark 6 The class of \(\phi\)-divergence measures is a wide family of divergence measures but unfortunately there are some classical divergence measures that are not included in this family of \(\phi\)-divergence measures such as the Rényi’s divergence or the Sharma
and Mittal’s divergence. The expression of Rényi’s divergence is given by

\[ D_{\text{Rényi}}^a(\hat{\theta}_c, \tilde{\theta}_{rc}) = \frac{1}{a(a-1)} \log \int_{\mathbb{R}^m} \mathcal{L}(\hat{\theta}_c, y) \mathcal{L}^{1-a}(\tilde{\theta}_{rc}, y) dy, \quad \text{if } a \neq 0, 1, \]

with

\[ D_{\text{Rényi}}^0(\hat{\theta}_c, \tilde{\theta}_{rc}) = \lim_{a \to 0} D_{\text{Rényi}}^a(\hat{\theta}_c, \tilde{\theta}_{rc}) = D_{\text{Kullback}}(\tilde{\theta}_{rc}, \hat{\theta}_c) = \int_{\mathbb{R}^m} \mathcal{L}(\hat{\theta}_c, y) \log \frac{\mathcal{L}(\hat{\theta}_c, y)}{\mathcal{L}(\tilde{\theta}_{rc}, y)} dy \]

and

\[ D_{\text{Rényi}}^1(\hat{\theta}_c, \tilde{\theta}_{rc}) = \lim_{a \to 1} D_{\text{Rényi}}^a(\hat{\theta}_c, \tilde{\theta}_{rc}) = D_{\text{Kullback}}(\hat{\theta}_c, \tilde{\theta}_{rc}). \]

This measure of divergence was introduced in Rényi (1961) for \( a > 0 \) and \( a \neq 1 \) and Liese and Vajda (1987) extended it for all \( a \neq 1, 0 \). An interesting divergence measure related to Rényi divergence measure is the Bhattacharya divergence defined as the Rényi divergence for \( a = 1/2 \) divided by 4. Other interesting example of divergence measure, not included in the family of \( \phi \)-divergence measures, is the divergence measures introduced by Sharma and Mittal (1997).

In order to unify the previous divergence measures as well as another divergence measures Menéndez et al. (1955, 1997) introduced the family of divergences called “\((h, \phi)\)-divergence measures” in the following way

\[ D_{\phi}^h(\hat{\theta}_c, \tilde{\theta}_{rc}) = h \left( D_{\phi}(\hat{\theta}_c, \tilde{\theta}_{rc}) \right), \quad (11) \]

where \( h \) is a differentiable increasing function mapping from \([0, \phi(0) + \lim_{t \to \infty} \frac{\phi(t)}{t}]\) onto \([0, \infty)\), with \( h(0) = 0, h'(0) > 0\), and \( \phi \in \Psi \). In the next table Rényi and Sharma–Mittal divergence measures are presented, along with the corresponding expressions of \( h \) and \( \phi \).

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

Based on the \((h, \phi)\)-divergence measures we can define a new family of \((h, \phi)\)-divergence test-statistics for testing the null hypothesis \( H_0 \) given in (7)

\[ T_{h,\phi,n}(\hat{\theta}_c, \tilde{\theta}_{rc}) = \frac{2n}{\phi''(1)h'(0)} h \left( D_{\phi}(\hat{\theta}_c, \tilde{\theta}_{rc}) \right). \quad (12) \]
Since
\[ h(x) = h(0) + h'(0)x + o(x), \]
the asymptotic distribution of \( T_{h,\phi,n}(\hat{\theta}_c, \tilde{\theta}_{rc}) \) coincides with the asymptotic distribution of \( T_{\phi,n}(\hat{\theta}_c, \tilde{\theta}_{rc}) \). In a similar way we can define the family of \((h, \phi)\)-divergence test-statistics for testing the null hypothesis \( H_0 \) given in (2) by
\[ T_{h,\phi,n}(\hat{\theta}_c, \theta_0) = \frac{2n}{\phi''(1)h'(0)}h \left(D_\phi(\hat{\theta}_c, \theta_0)\right). \]

6 Numerical examples

6.1 Estimation of correlation, Xu and Reid (2011)

In this section we shall consider an example, studied previously by Xu and Reid (2011) on the robustness of maximum composite estimator. The aim of this section is to clarify the different issues which are discussed in the previous sections.

Consider the random vector \( Y = (Y_1, Y_2, Y_3, Y_4)^T \) which follows a four dimensional normal distribution with mean vector \( \mu = (\mu_1, \mu_2, \mu_3, \mu_4)^T \) and variance–covariance matrix
\[
\Sigma = \begin{pmatrix}
1 & \rho & 2\rho & 2\rho \\
\rho & 1 & 2\rho & 2\rho \\
2\rho & 2\rho & 1 & \rho \\
2\rho & 2\rho & \rho & 1
\end{pmatrix}, \tag{13}
\]
i.e., we suppose that the correlation between \( Y_1 \) and \( Y_2 \) is the same as the correlation between \( Y_3 \) and \( Y_4 \). Taking into account that \( \Sigma \) must be semi-definite positive, the following condition is imposed, \(-\frac{1}{5} \leq \rho \leq \frac{1}{4}\). In order to avoid several problems regarding the consistency of the maximum likelihood estimator of the parameter \( \rho \) (cf. Xu and Reid 2011), we shall consider the composite likelihood function
\[ CL(\theta, y) = f_{A_1}(\theta, y) f_{A_2}(\theta, y), \]
where
\[
f_{A_1}(\theta, y) = f_{12}(\mu_1, \mu_2, \rho, y_1, y_2), \quad f_{A_2}(\theta, y) = f_{34}(\mu_3, \mu_4, \rho, y_3, y_4),
\]
where \( f_{12} \) and \( f_{34} \) are the densities of the marginals of \( Y \), i.e. bivariate normal distributions with mean vectors \((\mu_1, \mu_2)^T\) and \((\mu_3, \mu_4)^T\), respectively, and common variance–covariance matrix
\[
\begin{pmatrix}
1 & \rho \\
\rho & 1
\end{pmatrix},
\]
with expressions given by
\[
f_{h,h+1}(\mu_h, \mu_{h+1}, \rho, y_h, y_{h+1}) = \frac{1}{2\pi \sqrt{1 - \rho^2}} \exp \left\{ -\frac{1}{2(1-\rho^2)} Q(y_h, y_{h+1}) \right\}, \quad h \in \{1, 3\},
\]
On divergence tests for composite hypotheses...

being

\[ Q(y_h, y_{h+1}) = (y_h - \mu_h)^2 - 2\rho (y_h - \mu_h)(y_{h+1} - \mu_{h+1}) + (y_{h+1} - \mu_{h+1})^2, \quad h \in \{1, 3\}. \]

In this context, the interest is focused in testing the composite null hypothesis

\[ H_0 : \rho = \rho_0 \text{ against } H_1 : \rho \neq \rho_0, \quad (14) \]

by using the composite \( \phi \)-divergence test-statistics, presented above. In this case, the parameter space is given by

\[ \Theta = \left\{ \theta = (\mu_1, \mu_2, \mu_3, \mu_4, \rho)^T : \mu_i \in \mathbb{R}, \ i = 1, \ldots, 4 \text{ and } -\frac{1}{3} \leq \rho \leq \frac{1}{3} \right\}. \]

If we consider \( g : \Theta \subseteq \mathbb{R}^5 \longrightarrow \mathbb{R} \), with

\[ g(\theta) = g(\mu_1, \mu_2, \mu_3, \mu_4, \rho) = \rho - \rho_0, \quad (15) \]

the parameter space under the null hypothesis is given by

\[ \Theta_0 = \left\{ \theta = (\mu_1, \mu_2, \mu_3, \mu_4, \rho)^T \in \Theta : g(\mu_1, \mu_2, \mu_3, \mu_4, \rho) = 0 \right\}. \]

It is now clear that the dimensions of both parameter spaces are \( \text{dim}(\Theta) = 5 \) and \( \text{dim}(\Theta_0) = 4 \). Consider now a random sample of size \( n \), \( y_i = (y_{i1}, \ldots, y_{i4})^T, \ i = 1, \ldots, n \). The maximum composite likelihood estimators of the parameters \( \mu_i, \ i = 1, 2, 3, 4, \) and \( \rho \) in \( \Theta \) are obtained by standard maximization of the composite log-density function associated to the random sample of size \( n \),

\[
c\ell(\theta, y_1, \ldots, y_n) = \sum_{i=1}^{n} c\ell(\theta, y_i) = \sum_{i=1}^{n} \log C_L(\theta, y_i)
= \sum_{i=1}^{n} \left[ \log f_{A_1}(\theta, y_i) + \log f_{A_2}(\theta, y_i) \right]
= \sum_{i=1}^{n} \ell_{A_1}(\theta, y_i) + \sum_{i=1}^{n} \ell_{A_2}(\theta, y_i)
= -\frac{n}{2} \log \left( 1 - \rho^2 \right) - \frac{1}{2 (1 - \rho^2)} (\varsigma_1^2 + \varsigma_2^2 + \varsigma_3^2 + \varsigma_4^2 - 2\rho (\varsigma_{12} + \varsigma_{34})) + k,
\]

\( \cdot \) Springer
where

\[ \varsigma_j^2 = \sum_{i=1}^n (y_{ij} - \mu_j)^2, \quad j \in \{1, 2, 3, 4\}, \]

\[ \varsigma_{h,h+1} = \sum_{i=1}^n (y_{ih} - \mu_h)(y_{i,h+1} - \mu_{h+1}), \quad h \in \{1, 3\}, \]

and \( k \) is a constant, independent of the unknown parameters, i.e., by solving the following system of five equations

\[ (\bar{y}_1 - \mu_1) - \rho(\bar{y}_2 - \mu_2) = 0, \]
\[ (\bar{y}_2 - \mu_2) - \rho(\bar{y}_1 - \mu_1) = 0, \]
\[ (\bar{y}_3 - \mu_3) - \rho(\bar{y}_4 - \mu_4) = 0, \]
\[ (\bar{y}_4 - \mu_4) - \rho(\bar{y}_3 - \mu_3) = 0, \]
\[ n\rho^3 - \frac{\varsigma_{12} + \varsigma_{34}}{2}\rho^2 + \left(\frac{\varsigma_1^2 + \varsigma_2^2 + \varsigma_3^2 + \varsigma_4^2}{2} - n\right)\rho - \frac{\varsigma_{12} + \varsigma_{34}}{2} = 0, \]

with

\[ \bar{y}_j = \frac{1}{n} \sum_{i=1}^n y_{ij}, \quad j \in \{1, 2, 3, 4\}. \quad (16) \]

From the first two equations we get

\[ \rho(\bar{y}_1 - \mu_1) - \rho^2(\bar{y}_2 - \mu_2) = 0, \]
\[ -\rho(\bar{y}_1 - \mu_1) + \bar{y}_2 - \mu_2 = 0. \]

Therefore

\[ (1 - \rho^2)(\bar{y}_2 - \mu_2) = 0, \]

and since we assume that \( \rho \in (-\frac{1}{5}, \frac{1}{3}) \subset (-1, 1) \), it is obtained that

\[ \hat{\mu}_1 = \bar{y}_1 \quad \text{and} \quad \hat{\mu}_2 = \bar{y}_2. \]

In a similar manner, from the third and fourth equations we can get that

\[ \hat{\mu}_3 = \bar{y}_3 \quad \text{and} \quad \hat{\mu}_4 = \bar{y}_4. \]

The maximum composite likelihood estimator of \( \rho \) under \( \Theta \), \( \hat{\rho} \), is the real solution of the following cubic equation

\[ \rho^3 - \frac{v_{12} + v_{34}}{2}\rho^2 + \left(\frac{v_1^2 + v_2^2 + v_3^2 + v_4^2}{2} - 1\right)\rho - \frac{v_{12} + v_{34}}{2} = 0, \]

\( \odot \) Springer
where

\[ v_j^2 = \frac{1}{n} \bar{\varsigma}_j^2 = \frac{1}{n} \sum_{i=1}^{n} (y_{ij} - \bar{y}_j)^2, \quad j \in \{1, 2, 3, 4\}, \]

\[ v_{h,h+1} = \frac{1}{n} \bar{\varsigma}_{h,h+1} = \frac{1}{n} \sum_{i=1}^{n} (y_{ih} - \bar{y}_h)(y_{i,h+1} - \bar{y}_{h+1}), \quad h \in \{1, 3\}, \]

and \( v_j^2, j \in \{1, 2, 3, 4\} \), are the sample variances while \( v_{h,h+1}, h \in \{1, 3\} \), are the sample covariances.

Under \( \Theta_0 \), the restricted maximum composite likelihood estimators of the parameters \( \mu_j, j \in \{1, 2, 3, 4\} \) are given by,

\[ \tilde{\mu}_j = \bar{y}_j, \quad j \in \{1, 2, 3, 4\}, \]

with \( \bar{y}_j \) given by (16). Therefore, in our model, the maximum composite likelihood estimators are

\[ \tilde{\theta}_c = (\bar{y}_1, \bar{y}_2, \bar{y}_3, \bar{y}_4, \bar{\rho})^T \quad \text{and} \quad \tilde{\theta}_{rc} = (\bar{y}_1, \bar{y}_2, \bar{y}_3, \bar{y}_4, \rho_0)^T, \]

under \( \Theta \) and \( \Theta_0 \) respectively.

After some heavy algebraic manipulations the sensitivity or Hessian matrix \( H(\theta) \) is given by

\[ H(\theta) = \frac{1}{1 - \rho^2} \begin{pmatrix} 1 & -\rho & 0 & 0 & 0 \\ -\rho & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & -\rho & 0 \\ 0 & 0 & -\rho & 1 & 0 \\ 0 & 0 & 0 & 0 & 2\frac{1+\rho^2}{1-\rho^2} \end{pmatrix}. \]

(17)

In a similar manner, the expression of the variability matrix \( J(\theta) \) coincides with that of sensitivity matrix \( H(\theta) \), i.e. \( J(\theta) = H(\theta) \).

In order to get the unique non zero eigenvalue \( \beta_1 \) from Theorem 5, it is necessary to obtain (6), which, in the present setup, is given by

\[ G(\theta) = \frac{\partial g(\theta)}{\partial \theta} = (0, 0, 0, 1)^T, \]

(18)

where \( g(\theta) \) is given by (15) in the context of the present example. In addition, taking into account that \( Q(\theta) = -G(\theta) \) and after some algebra, it is concluded that \( \beta_1 = 1 \) and therefore the asymptotic distribution of the composite \( \phi \)-divergence test-statistics is

\[ T_{\phi,n}(\tilde{\theta}_c, \tilde{\theta}_{rc}) = \frac{2n}{\phi''(1)} D_{\phi}(\tilde{\theta}_c, \tilde{\theta}_{rc}) \xrightarrow{n \to \infty} \chi_1^2, \]

\( \square \) Springer
under the null hypothesis $H_0 : \rho = \rho_0$. In a completely similar manner, the composite likelihood ratio test, presented in Remark 1, for testing $H_0 : \rho = \rho_0$, is

\[
\lambda_n(\hat{\theta}_c, \tilde{\theta}_{rc}) = 2 \left( c \ell(\hat{\theta}_c) - c \ell(\tilde{\theta}_{rc}) \right) \xrightarrow{n \to \infty} \chi^2_1,
\]

because the only non zero eigenvalue of the asymptotic distribution (8) is equal to one.

In a similar way, if we consider the composite $(h, \phi)$-divergence test-statistics, we have

\[
T_{h,\phi,n}(\hat{\theta}_c, \tilde{\theta}_{rc}) = \frac{2n}{\phi''(1)h'(0)} h(D_{\phi}(\hat{\theta}_c, \tilde{\theta}_{rc})) \xrightarrow{n \to \infty} \chi^2_1.
\]

In order to apply the above theoretic issues in practice, it is necessary to consider a particular convex function $\phi$ in order to get a concrete $\phi$-divergence or to consider $\phi$ and $h$ in order to get an $(h, \phi)$-divergence. Using the Rényi’s family of divergences, i.e., a family of $(h, \phi)$-divergences with $\phi$ and $h$ given in Table 1, the family of test-statistics is given by

\[
T^a_n(\hat{\theta}_c, \tilde{\theta}_{rc}) = \frac{2n}{a(a - 1)} \left( \log \int_{\mathbb{R}^2} \frac{f_{12}^a(\hat{\mu}_1, \hat{\mu}_2, \hat{\rho}, y_1, y_2)}{f_{12}^{a-1}(\hat{\mu}_1, \hat{\mu}_2, \rho_0, y_1, y_2)} dy_1 dy_2 
+ \log \int_{\mathbb{R}^2} \frac{f_{34}^a(\hat{\mu}_3, \hat{\mu}_4, \hat{\rho}, y_3, y_4)}{f_{34}^{a-1}(\hat{\mu}_3, \hat{\mu}_4, \rho_0, y_3, y_4)} dy_3 dy_4 \right)
\]
\[
= \frac{4n}{a(a - 1)} \log \int_{\mathbb{R}^2} \frac{f_{12}^a(\hat{\mu}_1, \hat{\mu}_2, \hat{\rho}, y_1, y_2)}{f_{12}^{a-1}(\hat{\mu}_1, \hat{\mu}_2, \rho_0, y_1, y_2)} dy_1 dy_2,
\]

for $a \neq 0, 1$. The last equality follows because the integrals does not depend on $\hat{\mu}_1, \hat{\mu}_2, \hat{\mu}_3$ and $\hat{\mu}_4$. For $a = 1$ we have

\[
T^1_n(\hat{\theta}_c, \tilde{\theta}_{rc}) = 2n \left( \int_{\mathbb{R}^2} f_{34}(\hat{\mu}_3, \hat{\mu}_4, \hat{\rho}, y_3, y_4) dy_3 dy_4 \int_{\mathbb{R}^2} f_{12}(\hat{\mu}_1, \hat{\mu}_2, \hat{\rho}, y_1, y_2) 
\times \log \frac{f_{12}(\hat{\mu}_1, \hat{\mu}_2, \hat{\rho}, y_1, y_2)}{f_{12}(\hat{\mu}_1, \hat{\mu}_2, \rho_0, y_1, y_2)} dy_1 dy_2 
+ \int_{\mathbb{R}^2} f_{12}(\hat{\mu}_1, \hat{\mu}_2, \hat{\rho}, y_1, y_2) dy_1 dy_2 \int_{\mathbb{R}^2} f_{34}(\hat{\mu}_3, \hat{\mu}_4, \hat{\rho}, y_3, y_4) 
\times \log \frac{f_{34}(\hat{\mu}_3, \hat{\mu}_4, \hat{\rho}, y_3, y_4)}{f_{34}(\hat{\mu}_3, \hat{\mu}_4, \rho_0, y_3, y_4)} dy_3 dy_4 \right)
\]
\[
= 4n \int_{\mathbb{R}^2} f_{12}(\hat{\mu}_1, \hat{\mu}_2, \hat{\rho}, y_1, y_2) \log \frac{f_{12}(\hat{\mu}_1, \hat{\mu}_2, \hat{\rho}, y_1, y_2)}{f_{12}(\hat{\mu}_1, \hat{\mu}_2, \rho_0, y_1, y_2)} dy_1 dy_2,
\]

and

\[
T^0_n(\hat{\theta}_c, \tilde{\theta}_{rc}) = T^1_n(\tilde{\theta}_{rc}, \hat{\theta}_c).
\]

Based on the expression for Rényi divergence in normal populations (for more details see Pardo 2006, p. 33) we have
On divergence tests for composite hypotheses...

\[ T_n^a (\tilde{\theta}_c, \tilde{\theta}_{rc}) = \begin{cases} 
\frac{2n}{a(a-1)} \log \frac{(1 - \rho_0^2)^{a}(1 - \hat{\rho}^2)^{(a-1)}}{1 - [a\rho_0 + (1 - a)\hat{\rho}]^2}, & a \notin \{0, 1\}, \hat{\rho} \in \left(\frac{a}{a-1} \rho_0 - \frac{1}{|a-1|}, \frac{a}{a-1} \rho_0 + \frac{1}{|a-1|}\right), \\
+\infty, & a \notin \{0, 1\}, \hat{\rho} \notin \left(\frac{a}{a-1} \rho_0 - \frac{1}{|a-1|}, \frac{a}{a-1} \rho_0 + \frac{1}{|a-1|}\right), \\
2n \left(\log \frac{1-\rho_0^2}{1-\rho^2} + 2 \hat{\rho}(\hat{\rho} - \rho)\right), & a = 1, \\
2n \left(\log \frac{1-\rho_0^2}{1-\rho^2} + 2 \hat{\rho}(\hat{\rho} - \rho)\right), & a = 0.
\end{cases} \tag{19} \]

Similarly, using the Cressie–Read’s family of divergences, the family of test-statistics is given by

\[ T^{\lambda}_n (\tilde{\theta}_c, \tilde{\theta}_{rc}) = \begin{cases} 
\frac{4n}{\lambda(\lambda+1)} \left(\frac{1 - \rho_0^2}{1 - (\lambda + 1)\rho_0} - 1\right)^{1/\lambda}, & \lambda \notin \{0, -1\}, \hat{\rho} \in \left(\frac{\lambda+1}{\lambda} \rho_0 - \frac{1}{|\lambda|}, \frac{\lambda+1}{\lambda} \rho_0 + \frac{1}{|\lambda|}\right), \\
+\infty, & \lambda \notin \{0, -1\}, \hat{\rho} \notin \left(\frac{\lambda+1}{\lambda} \rho_0 - \frac{1}{|\lambda|}, \frac{\lambda+1}{\lambda} \rho_0 + \frac{1}{|\lambda|}\right), \\
2n \left(\log \frac{1-\rho_0^2}{1-\rho^2} + 2 \rho_0 (\rho - \hat{\rho})\right) = T^1_n (\tilde{\theta}_c, \tilde{\theta}_{rc}), & \lambda = 0, \\
2n \left(\log \frac{1-\rho_0^2}{1-\rho^2} + 2 \rho_0 (\rho - \hat{\rho})\right) = T^0_n (\tilde{\theta}_c, \tilde{\theta}_{rc}), & \lambda = -1.
\end{cases} \tag{20} \]

After some algebra we can also obtain the composite likelihood ratio test. This has the following expression

\[ \lambda_n (\tilde{\theta}_c, \tilde{\theta}_{rc}) = 2 \left( c \ell (\tilde{\theta}_c, y_1, \ldots, y_n) - c \ell (\tilde{\theta}_{rc}, y_1, \ldots, y_n) \right) \\
= 2n \left[ \log \frac{1-\rho_0^2}{1-\hat{\rho}^2} + (v_1^2 + v_2^2 + v_3^2 + v_4^2) \left(\frac{1-\rho_0^2}{1-\hat{\rho}^2} - \frac{1}{\hat{\rho}^2}\right) \\
- 2(v_{12} + v_{34}) \left(\frac{\rho_0}{1-\rho_0^2} - \frac{\hat{\rho}}{1-\hat{\rho}^2}\right) \right]. \tag{21} \]

### 6.2 First order autoregression, Pace et al. (2011)

A normal AR(1) process of order one is defined for observations \(Y_{ir} = \mu + \rho (Y_{ir-1} - \mu) + \varepsilon_{ir}, i = 1, \ldots, n, r = 2, \ldots, q\), where \(\varepsilon_{ir}\) are independently normally distributed random variables with \(E(\varepsilon_{ir}) = 0\) and \(Var(\varepsilon_{ir}) = \sigma^2\). This setting implies that we have in hand \(n\) observations \(Y_i\) and each of them follows a \(q\)-dimensional normal distribution with mean vector \(\mu = (\mu, \ldots, \mu)^T\) and variance–covariance matrix \(\Sigma = (Cov(Y_{i1}, Y_{ir})), s, r = 1, \ldots, q\) with \(Y_i = (Y_{i1}, \ldots, Y_{iq})^T\) and \(Cov(Y_{is}, Y_{ir}) = \frac{\sigma^2 \rho^{s-r}}{1-\rho^2}, s, r = 1, \ldots, q\).
The parameters of interest, in this setting, are $\mu, \sigma^2$ and $\rho$ and let the vector $\theta = (\mu, \sigma^2, \rho)^T$. Following Pace et al. (2011), we concentrate only to a single process or serie. Therefore, we drop, in the following, the subscript $i$ from the above notation.

We shall consider as composite likelihood the pairwise likelihood by using only pairs of contiguous components. Hence, for $y = (y_1, \ldots, y_q)^T$, the composite likelihood function $\mathcal{C}L$ is given by

$$\mathcal{C}L(\mu, \sigma, \rho, y) = \prod_{r=2}^{q} f_{Y_{r-1},Y_r}(y_{r-1}, y_r),$$

where $f_{Y_{r-1},Y_r}(y_{r-1}, y_r)$ is the density function of a bivariate normally distributed random vector with mean $\mu = (\mu, \mu)^T$ and variance–covariance matrix

$$ \begin{pmatrix} \sigma^2 & \sigma^2 \\ \sigma^2 & 1-\rho \end{pmatrix}.$$ 

Therefore,

$$\mathcal{C}L(\mu, \rho, \sigma, y) = \frac{(1 - \rho^2)(q-1)/2}{(2\pi)^{(q-1)/2} \sigma^{2(q-1)}} \exp \left\{- \frac{1}{2\sigma^2} \left[ \sum_{r=2}^{q} (y_{r-1} - \mu)^2 + \sum_{r=2}^{q} (y_r - \mu)^2 - 2\rho \sum_{r=2}^{q} (y_{r-1} - \mu)(y_r - \mu) \right] \right\}. \quad (22)$$

We are interested in testing the hypotheses,

$$H_0 : \sigma = \sigma_0, \rho = \rho_0, \text{ against } H_1 : \sigma \neq \sigma_0, \rho \neq \rho_0. \quad (23)$$

by using the composite $\phi$-divergence test statistics, where $\mu$ is considered as a nuisance parameter.

In this context, the parameter space is defined by

$$\Theta = \left\{ \theta = (\mu, \sigma, \rho)^T : \mu \in \mathbb{R}, \sigma \in \mathbb{R}^+ \text{ and } \rho \in (0, 1) \right\}.$$ 

We shall consider the function $g : \Theta \subseteq \mathbb{R}^3 \rightarrow \mathbb{R}^2$, with

$$g(\theta) = g(\mu, \sigma, \rho) = (g_1(\mu, \sigma, \rho), g_2(\mu, \sigma, \rho))^T = (\sigma - \sigma_0, \rho - \rho_0)^T.$$ 

The parameter space under the null hypothesis is given by

$$\Theta_0 = \left\{ \theta = (\mu, \sigma, \rho)^T \in \Theta : g(\mu, \sigma, \rho) = 0 \right\},$$

and it is clear that $\dim(\Theta) = 3$ and $\dim(\Theta_0) = 2$.

In our case, the matrix $G(\theta)$, defined in (6), is given by

$$G(\theta) = \frac{\partial g^T(\theta_0)}{\partial \theta_0} = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$
For testing (23), we shall consider the family of composite \((h, \phi)\)-divergence test-statistics \(T_{h,\phi,n}(\hat{\theta}_c, \hat{\theta}_{rc})\) with \(h\) and \(\phi\) defined properly so as to get the composite Rényi’s test statistics (see, (11) and the subsequent table).

We are now going to get \(\hat{\theta}_c\) and \(\tilde{\theta}_{rc}\). Based on (22), it is obtained that

\[
\log \mathcal{L}(\mu, \rho, \sigma, y) = \frac{q - 1}{2} \log(1 - \rho^2) - \frac{q - 1}{2} \log(2\pi) - 2(q - 1) \log \sigma \\
- \frac{1}{2\sigma^2} \left[ \sum_{r=2}^{q} (y_{r-1} - \mu)^2 + \sum_{r=2}^{q} (y_r - \mu)^2 \\
- 2\rho \sum_{r=2}^{q} (y_{r-1} - \mu)(y_r - \mu) \right],
\]

and

\[
\frac{\partial}{\partial \mu} \log \mathcal{L}(\mu, \rho, \sigma, y) = \frac{1}{2\sigma^2} \left\{ 2 \sum_{r=2}^{q} (y_{r-1} - \mu) + 2 \sum_{r=2}^{q} (y_r - \mu) \\
- 2\rho \left[ \sum_{r=2}^{q} (y_{r-1} - \mu) + \sum_{r=2}^{q} (y_r - \mu) \right] \right\}.
\]

The solution of the equation \(\frac{\partial}{\partial \mu} \log \mathcal{L}(\mu, \rho, \sigma, y) = 0\) in respect to \(\mu\) gives that

\[
(1 - \rho) \left\{ \sum_{r=2}^{q} y_{r-1} + \sum_{r=2}^{q} y_r \right\} = 2(1 - \rho)\mu(q - 1),
\]

and therefore,

\[
\hat{\mu} = \frac{\sum_{r=2}^{q} y_{r-1} + \sum_{r=2}^{q} y_r}{2(q - 1)}. \quad (24)
\]

In a similar manner, \(\frac{\partial}{\partial \sigma} \log \mathcal{L}(\mu, \rho, \sigma, y) = 0\) and \(\frac{\partial}{\partial \rho} \log \mathcal{L}(\mu, \rho, \sigma, y) = 0\), lead, respectively to

\[
\sigma^2 = \frac{1}{2(q - 1)} \left\{ \sum_{r=2}^{q} (y_{r-1} - \mu)^2 + \sum_{r=2}^{q} (y_r - \mu)^2 - 2\rho \sum_{r=2}^{q} (y_{r-1} - \mu)(y_r - \mu) \right\}, \quad (25)
\]

and

\[
\frac{2(\rho^2 - 1)}{(q - 1)\sigma^2} \sum_{r=2}^{q} (y_{r-1} - \mu)(y_r - \mu) - \rho = 0. \quad (26)
\]

Hence, \(\hat{\mu}\) is obtained in an explicit form by (24) while \(\hat{\sigma}^2\) and \(\hat{\rho}\) are obtained by the solution of the system of Eqs. (25) and (26). The respective estimators \(\tilde{\theta}_c = (\hat{\mu}, \hat{\sigma}, \hat{\rho})^T\) and \(\tilde{\theta}_{rc} = (\hat{\mu}, \hat{\sigma}_0, \hat{\rho}_0)^T\) are then easily obtained.
In order to get the composite Rényi’s test statistics in an explicit form, it is necessary to obtain the Rényi’s divergence between $CL(\hat{\mu}, \hat{\sigma}, \hat{\rho}, y)$ and $CL(\tilde{\mu}, \sigma_0, \rho_0, y)$ where $\tilde{\mu}$ coincides with $\hat{\mu}$, given by (24), because the last expression doesn’t depend on $\rho$ and $\sigma^2$.

Based on the definition of Rényi’s divergence by (11) and the subsequent table, it is necessary to obtain in an explicit form the integral

$$I_a = \int_{\mathbb{R}^q} CL(\hat{\mu}, \hat{\sigma}, \hat{\rho}, y)^a CL(\tilde{\mu}, \sigma_0, \rho_0, y)^{1-a} dy, \quad a \neq 0, 1, \quad (27)$$

where

$$CL(\mu, \sigma, \rho, y) = L_{\rho,\sigma}(1) \exp \left\{ -\frac{1}{2\sigma^2} \left[ \sum_{r=2}^{q} (y_r - \mu)^2 \right. \\
+ \sum_{r=2}^{q} (y_r - \mu)^2 - 2\rho \sum_{r=2}^{q} (y_r - \mu)(y_r - \mu) \right] \right\},$$

and

$$L_{\rho,\sigma}(s) = \frac{(1 - \rho^2)^s(q-1)/2}{(2\pi)^s(q-1)/2\sigma^2(q-1)/2}.$$

The algebraic manipulations for the integral $I_a$ are summarized in “Calculation of the integral $I_a$ in (28)” section in Appendix. According to them,

$$I_a = \left( \hat{\sigma}^2 \right)^{(q-1)(a-1)} \left( \sigma_0^2 \right)^{a(q-1)} \left[ \frac{(1 - \hat{\rho}^2)^a(1 - \rho_0^2)^{1-a}}{(a\sigma_0^2 + (1-a)\hat{\sigma}^2)^2 - (a\sigma_0^2\hat{\rho} + (1-a)\rho_0^2\hat{\rho})^2} \right]^{\frac{a-1}{2}}.$$

(28)

Based on $I_a$ in (28), composite Rényi’s test statistics are given by

$$T^{(a)}_n(\hat{\theta}_c, \tilde{\theta}_{rc}) = \frac{2n}{a(a-1)} \log \int_{\mathbb{R}^q} CL(\hat{\mu}, \hat{\sigma}, \hat{\rho}, y)^a CL(\tilde{\mu}, \sigma_0, \rho_0, y)^{1-a} dy,$$

or

$$T^{(a)}_n(\hat{\theta}_c, \tilde{\theta}_{rc}) = \frac{2n}{a(a-1)} \log \left\{ \left( \hat{\sigma}^2 \right)^{(q-1)(a-1)} \left( \sigma_0^2 \right)^{a(q-1)} \right. \\
\times \left[ \frac{(1 - \hat{\rho}^2)^a(1 - \rho_0^2)^{1-a}}{(a\sigma_0^2 + (1-a)\hat{\sigma}^2)^2 - (a\sigma_0^2\hat{\rho} + (1-a)\rho_0^2\hat{\rho})^2} \right]^{\frac{a-1}{2}} \right\}.$$

For $a \to 0$ or $a \to 1$, $T^{(a)}_n(\hat{\theta}_c, \tilde{\theta}_{rc})$ leads to the respective Kullback–Leibler test statistic.
Based on the theoretic results, developed in the previous sections, it is immediately obtained that

\[ T_n^{(a)}(\hat{\theta}_c, \tilde{\theta}_{rc}) \xrightarrow{L} n \to \infty \sum_{i=1}^{2} \lambda_i(\theta) Z_i^2, \]

where \( Z_i, i = 1, 2 \), is a standard normal random variable and \( \lambda_i(\theta), i = 1, 2 \), are the eigenvalues of the matrix

\[
H(\theta)G(\theta)Q^T(\theta)G^{-1}(\theta)Q(\theta)G^T(\theta),
\]

with

\[
Q(\theta) = -H^{-1}(\theta)G(\theta) \left( G^T(\theta)H^{-1}(\theta)G(\theta) \right)^{-1},
\]

and

\[
G_s(\theta) = H(\theta)J^{-1}(\theta)H(\theta).
\]

Therefore, the only we need is to obtain the matrices \( H(\theta) \) and \( J(\theta) \) because the matrix \( G(\theta) \) has been obtained previously and it is given by \( G(\theta) = \left( \begin{array}{cc} 0 & 1 \\ 0 & 0 & 1 \end{array} \right) \). However, the matrices \( H(\theta) \) and \( J(\theta) \) have been obtained in Appendix of the paper by Pace et al. (2011) and the only we have to do is to replace in the expression of \( H(\theta) \) and \( J(\theta) \) the values of \( \sigma \) and \( \rho \) which are specified by the null hypothesis.

### 7 Simulation study

In this section, a simulation study is presented in order to study the behavior of the composite \( \phi \)-divergence test-statistics. The theoretical model, studied in the first example of the previous section, is applied to the composite Cressie–Read test-statistics (20). The composite likelihood ratio test-statistic (CLRT), given in (21), is also considered. Special attention is paid to the hypothesis testing (14) with \( \rho_0 \in \{-0.1, 0.2\} \). The case \( \rho_0 = 0 \) has been considered, but this case is less important since taking into account the way of the theoretical model under consideration and having the case of independent observations, the composite likelihood theory is useless. For finite sample sizes and nominal size \( \alpha = 0.05 \), the estimated significance level for different composite Cressie–Read test-statistics as well as for the CLRT, are given by

\[
\alpha_n^{(\lambda)}(\rho_0) = \Pr(T_n^{(\lambda)}(\hat{\theta}_c, \tilde{\theta}_{rc}) > \chi^2_{1,0.05}|H_0), \quad \lambda \in \mathbb{R}
\]

and

\[
\alpha_n^{(CLRT)}(\rho_0) = \Pr \left( \lambda_n(\hat{\theta}_c, \tilde{\theta}_{rc}) > \chi^2_{1,0.05}|H_0 \right).
\]

More thoroughly, the composite Cressie–Read test-statistics with \( \lambda \in \{-1, -0.5, 0, 2/3, 1, 1.5\} \) have been selected for the study. Following Dale (1986), we consider the inequality

\[
\left| \logit(1 - \alpha_n^{(\lambda)}) - \logit(1 - \alpha) \right| \leq \varepsilon
\]

(29)
α(λ) = −λ valid for the study are limited to those verifying αn(λ) ∈ (0.0325, 0.07625). This criterion has been used in many previous studies, see for instance Cressie et al. (2003), Martin et al. (2014), Martin and Pardo (2012) and references therein.

Through R = 10,000 replications of the simulation experiment, with the model under the null hypothesis, the estimated significance level for different composite Cressie–Read test-statistics are

\[ \hat{α}_n^{(λ)}(ρ_0) = \hat{Pr}(T_n^{I}(\theta_c, \tilde{θ}_{rc}) > \chi^2_{1,0.05}|H_0) = \frac{\sum_{i=1}^{R} I(T_n^{I}(\theta_c, \tilde{θ}_{rc}) > \chi^2_{1,0.05}|H_0)}{R} , \]

with \( I(S) \) being the indicator function (with value 1 if \( S \) is true and 0 otherwise) and the estimated significance level for CLRT

\[ \hat{α}_n^{(CLRT)}(ρ_0) = \hat{Pr}(\alpha_n(\theta_c, \tilde{θ}_{rc}) > \chi^2_{1,0.05}|H_0) = \frac{\sum_{i=1}^{R} I(\alpha_n(\theta_c, \tilde{θ}_{rc}) > \chi^2_{1,0.05}|H_0)}{R} . \]

In Table 1 we present the simulated level for different values of \( λ \in \{-1, -0.5, 0, 2/3, 1, 1.5\} \) as well as for the CLRT, when \( n = 100, n = 200 \) and \( n = 300 \) for \( ρ_0 = -0.1 \) and \( ρ_0 = 0.2 \). In order to investigate the behavior for \( ρ_0 = 0 \) we present in Table 2 the simulated level for \( λ \in \{-1, -0.5, 0, 2/3, 1, 1.5\} \) as well as the simulated level of CLRT for \( n = 50, n = 100, n = 200 \) and \( n = 300 \). Clearly, as expected, the performance of the traditional divergence and likelihood methods is stronger in comparison with the composite divergence and likelihood methods.

For finite sample sizes and nominal size \( α = 0.05 \), the simulated powers are obtained under \( H_1 \) in (14), when \( ρ \in \{-0.2, -0.15, 0.0, 0.1\} \) and \( ρ_0 = -0.1 \) (Table 3) and when \( ρ \in \{0, 0.15, 0.25, 0.3\} \) and \( ρ_0 = 0.2 \) (Table 4). The (simulated) power for different composite Cressie–Read test-statistics is obtained by

| \( n = 100 \) | \( n = 200 \) | \( n = 300 \) |
| --- | --- | --- |
| \( CLRT \) | \( λ = -1 \) | \( ρ_0 = -0.1 \) | \( ρ_0 = 0.2 \) | \( ρ_0 = -0.1 \) | \( ρ_0 = 0.2 \) |
| 0.0688 | 0.0694 | 0.0673 | 0.0687 | 0.0645 | 0.0662 |
| \( λ = -0.5 \) | 0.0756 | 0.0762 | 0.0706 | 0.0740 | 0.0666 | 0.0685 |
| \( λ = 0 \) | 0.0738 | 0.0746 | 0.0697 | 0.0727 | 0.0662 | 0.0670 |
| \( λ = 2/3 \) | 0.0725 | 0.0739 | 0.0691 | 0.0720 | 0.0659 | 0.0672 |
| \( λ = 1 \) | 0.0726 | 0.0739 | 0.0694 | 0.0719 | 0.0662 | 0.0677 |
| \( λ = 1.5 \) | 0.0739 | 0.0747 | 0.0700 | 0.0720 | 0.0662 | 0.0680 |
| 0.0762 | 0.0762 | 0.0726 | 0.0729 | 0.0674 | 0.0677 |

Table 1: Simulated significance level for \( ρ_0 = -0.1 \) and \( ρ_0 = 0.2 \)
Table 2: Simulated significance level for $\rho_0 = 0$

|          | $n = 50$ | $n = 100$ | $n = 200$ | $n = 300$ |
|----------|----------|-----------|-----------|-----------|
| $LRT$    | 0.0543   | 0.0529    | 0.0527    | 0.0526    |
| $\lambda = -1$ | 0.0707 | 0.0605 | 0.0559 | 0.0542 |
| $\lambda = -0.5$ | 0.0677 | 0.0594 | 0.0553 | 0.0540 |
| $\lambda = 0$ | 0.0659 | 0.0577 | 0.0552 | 0.0540 |
| $\lambda = 2/3$ | 0.0670 | 0.0591 | 0.0552 | 0.0540 |
| $\lambda = 1$ | 0.0686 | 0.0597 | 0.0553 | 0.0541 |
| $\lambda = 1.5$ | 0.0726 | 0.0610 | 0.0564 | 0.0544 |

Table 3: Simulated powers for $\rho_0 = -0.1$

|          | $\rho = -0.2$ | $\rho = -0.15$ | $\rho = 0$ | $\rho = 0.1$ |
|----------|----------------|----------------|-----------|-------------|
| $n = 100$ | $CLRT$         | 0.3584         | 0.1604    | 0.2993      | 0.7958      |
|          | $\lambda = -1/2$ | 0.3751         | 0.1750    | 0.3057      | 0.8076      |
| $n = 200$ | $CLRT$         | 0.5455         | 0.2227    | 0.5087      | 0.9705      |
|          | $\lambda = -1/2$ | 0.5512         | 0.2322    | 0.5114      | 0.9737      |
| $n = 300$ | $CLRT$         | 0.7770         | 0.2705    | 0.8087      | 0.9962      |
|          | $\lambda = -1/2$ | 0.7797         | 0.2795    | 0.8112      | 0.9970      |

Table 4: Simulated powers for $\rho_0 = 0.2$

|          | $\rho = 0$ | $\rho = 0.15$ | $\rho = 0.25$ | $\rho = 0.3$ |
|----------|------------|---------------|---------------|-------------|
| $n = 100$ | $CLRT$     | 0.8054        | 0.1227        | 0.1534      | 0.3689      |
|          | $\lambda = -1/2$ | 0.8118         | 0.1305        | 0.1602      | 0.3806      |
| $n = 200$ | $CLRT$     | 0.9813        | 0.1904        | 0.2146      | 0.5818      |
|          | $\lambda = -1/2$ | 0.9825         | 0.1920        | 0.2194      | 0.5957      |
| $n = 300$ | $CLRT$     | 0.9978        | 0.2591        | 0.2870      | 0.7482      |
|          | $\lambda = -1/2$ | 0.9979         | 0.2577        | 0.2935      | 0.7612      |

$$\beta_n^{(\lambda)}(\rho_0, \rho) = \Pr(T_n^{\lambda}(\hat{\theta}_c, \tilde{\theta}_{rc}) > \chi^2_{1.0.05}\mid H_1) \quad \text{and}$$

$$\tilde{\beta}_n^{(\lambda)}(\rho_0, \rho) = \frac{\sum_{i=1}^R I(T_n^{\lambda}(\hat{\theta}_c, \tilde{\theta}_{rc}) > \chi^2_{1.0.05}\mid \rho_0, \rho)}{R},$$

and for the CLRT by

$$\beta_n^{(CLRT)}(\rho_0, \rho) = \Pr(\lambda_n(\hat{\theta}_c, \tilde{\theta}_{rc}) > \chi^2_{1.0.05}\mid H_1) \quad \text{and}$$

$$\tilde{\beta}_n^{(CLRT)}(\rho_0, \rho) = \frac{\sum_{i=1}^R I(\lambda_n(\hat{\theta}_c, \tilde{\theta}_{rc}) > \chi^2_{1.0.05}\mid \rho_0, \rho)}{R}.$$

Among the composite test-statistics with simulated significance levels verifying (29), at first sight the composite test-statistics with higher powers should be selected however
since in general high powers correspond to high significance levels, this choice is not straightforward. For this reason, based on $\beta_n^{LRT} - \alpha_n^{LRT}$ as baseline, the efficiencies relative to the composite likelihood ratio test, given by

$$e_n^{(\lambda)} = \frac{(\beta_n^{(\lambda)} - \alpha_n^{(\lambda)}) - (\beta_n^{LRT} - \alpha_n^{LRT})}{\beta_n^{LRT} - \alpha_n^{LRT}}, \quad \lambda \in \{-1, -0.5, 0, 2/3, 1, 1.5\},$$

were considered for $n = 100, n = 200$ and $n = 300$. Only the values of the power for $\lambda = -1/2$ are included in Tables 3 and 4, in order to show that the corresponding composite test-statistic is a good alternative to the composite likelihood ratio test-statistic. The values of the powers for which the values of $e_n^{(-1/2)}$ are positive, i.e., the case in which the composite test-statistic associated to $\lambda = -1/2$ is better than the composite likelihood ratio test, are shown in bold in Tables 3 and 4. This choice of $\lambda = -1/2$ divergence based test-statistic has been also recommended in Morales et al. (1977) and Martin et al. (2016).

8 Conclusions

This paper presents the theoretical background for the development of statistical tests for testing composite hypotheses when the composite likelihood is used instead of the classic likelihood of the data. The test statistic is based on the notion of phi-divergence and its by products, that is measures of the statistical distance between the theoretical model and the respective empirical one. The notion of divergence or disparity provides with abstract methods of estimation and testing and four monographs, mentioned in the introductory section, developed the state of the art on this subject.

This work is the first, to the best of our knowledge, which try to link the notion of composite likelihood with the notion of divergence between theoretical and empirical models for testing hypotheses. There are several extensions to this framework which can be considered. The theoretical framework, presented here, would be extended to develop statistical tests for testing homogeneity of two or more populations on the basis of composite likelihood. On the other hand, minimum phi-divergence or disparity procedures have been observed to provide strong robustness properties in estimation and testing problems. It would be maybe of interest to proceed in this direction in a composite likelihood setting.

Acknowledgements The authors would like to thank the referees for their helpful comments and suggestions. The third author wants to cordially thank Prof. Alex de Leon, from the University of Calgary, for fruitful discussions on composite likelihood methods, some years ago. This research is partially supported by Grant MTM2015-67057-P from Ministerio de Economía y Competitividad (Spain).
Appendix: Proofs

Proof of Theorem 1

Under the standard regularity assumptions of asymptotic statistics (cf. Serfling 1980, p. 144 and Pardo 2006, p. 58), we have

\[ \frac{\partial D_\phi(\theta, \theta_0)}{\partial \theta} = \int_{\mathbb{R}^m} \frac{\partial CL(\theta, y)}{\partial \theta} \phi' \left( \frac{CL(\theta, y)}{CL(\theta_0, y)} \right) dy, \]

therefore

\[ \frac{\partial D_\phi(\theta, \theta_0)}{\partial \theta} \bigg|_{\theta=\theta_0} = \phi' \left( \int_{\mathbb{R}^m} \frac{\partial CL(\theta, y)}{\partial \theta} \bigg|_{\theta=\theta_0} dy \right) = 0. \]

On the other hand,

\[ \frac{\partial^2 D_\phi(\theta, \theta_0)}{\partial \theta \partial \theta^T} = \int_{\mathbb{R}^m} \frac{\partial^2 CL(\theta, y)}{\partial \theta \partial \theta^T} \phi' \left( \frac{CL(\theta, y)}{CL(\theta_0, y)} \right) dy \]

and

\[ \frac{\partial^2 D_\phi(\theta, \theta_0)}{\partial \theta \partial \theta^T} \bigg|_{\theta=\theta_0} = \phi'' \left( \int_{\mathbb{R}^m} \frac{\partial c\ell(\theta, y)}{\partial \theta} \frac{\partial c\ell(\theta, y)}{\partial \theta^T} \bigg|_{\theta=\theta_0} CL(\theta_0, y)dy \right) = \phi'' \left( \int_{\mathbb{R}^m} \frac{\partial c\ell(\theta, y)}{\partial \theta} \frac{\partial c\ell(\theta, y)}{\partial \theta^T} \bigg|_{\theta=\theta_0} CL(\theta_0, y)dy \right) \]

Then, from

\[ D_\phi(\hat{\theta}_c, \theta_0) = \frac{\phi''}{2} (\hat{\theta}_c - \theta_0)^T J(\theta_0)(\hat{\theta}_c - \theta_0) + o(n^{-1/2}) \]

the desired result is obtained. The value of \( k \) comes from

\[ k = \text{rank} \left( G_\star^{-1}(\theta_0) J^T(\theta_0) G_\star^{-1}(\theta_0) \right) = \text{rank}(J(\theta_0)). \]

Proof of Theorem 2

A first order Taylor expansion gives

\[ D_\phi(\hat{\theta}_c, \theta_0) = D_\phi(\theta^*, \theta_0) + q^T(\hat{\theta}_c - \theta^*) + o(\|\hat{\theta}_c - \theta^*\|). \]

But

\[ \sqrt{n}(\hat{\theta}_c - \theta) \xrightarrow{L} \mathcal{N}(0, G_\star^{-1}(\theta)) \]

and \( \sqrt{n}o(\|\hat{\theta}_c - \theta^*\|) = o_P(1) \). Now the result follows.
Proof of Theorem 3

Following Sen and Singer (1993, pp. 242–243), let \( \theta_n = \theta + n^{-1/2}v \), where \( \|v\| < K^* \), \( 0 < K^* < \infty \). Consider now the following Taylor expansion of the partial derivative of the composite log-density,

\[
\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left. \frac{\partial}{\partial \theta} c \ell(\theta, y_i) \right|_{\theta = \theta_n} = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left. \frac{\partial}{\partial \theta} c \ell(\theta, y_i) \right|_{\theta = \theta_n} + \frac{1}{n} \sum_{i=1}^{n} \left. \frac{\partial^2}{\partial \theta \partial \theta^T} c \ell(\theta, y_i) \right|_{\theta = \theta_n} \sqrt{n} (\theta_n - \theta),
\]

(30)

where \( \theta_n^* \) belongs to the line segment joining \( \theta \) and \( \theta_n \). Then, observing that (cf. Theorem 2.3.6 of Sen and Singer 1993, p. 61)

\[
\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left. \frac{\partial^2}{\partial \theta \partial \theta^T} c \ell(\theta, y_i) \right|_{\theta = \theta_n} \xrightarrow{p} n \rightarrow \infty E_{\theta_n} \left[ \frac{\partial^2}{\partial \theta \partial \theta^T} c \ell(\theta, Y) \right] = E_{\theta_n} \left[ \frac{\partial}{\partial \theta} u^T(\theta, Y) \right] = -H(\theta),
\]

Eq. (30) leads

\[
\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left. \frac{\partial}{\partial \theta} c \ell(\theta, y_i) \right|_{\theta = \theta_n} = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left. \frac{\partial}{\partial \theta} c \ell(\theta, y_i) - H(\theta) \sqrt{n}(\theta_n - \theta) + o_P(1). \right.
\]

(31)

Since \( G(\theta) = \frac{\partial g^T(\theta)}{\partial \theta} \) is continuous in \( \theta \), it is true that,

\[
g(\theta_n) = G^T(\theta) \sqrt{n} (\theta_n - \theta) + o_P(1).
\]

(32)

Since, the restricted maximum composite likelihood estimator \( \tilde{\theta}_{rc} \) should satisfy the likelihood equations

\[
\sum_{i=1}^{n} \left. \frac{\partial}{\partial \theta} c \ell(\theta, y_i) + G(\theta) \lambda = 0_p, \right.
\]

\[
g(\theta) = 0_r,
\]

and in view of (31) and (32) it holds that

\[
\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left. \frac{\partial}{\partial \theta} c \ell(\theta, y_i) - H(\theta) \sqrt{n}(\tilde{\theta}_{rc} - \theta) + G(\theta) \frac{1}{\sqrt{n}} \lambda_n + o_P(1) = 0_p, \right.
\]

\[
G^T(\theta) \sqrt{n}(\tilde{\theta}_{rc} - \theta) + o_P(1) = 0_p.
\]

In matrix notation it may be re-expressed as

\[
\begin{pmatrix}
H(\theta) & -G(\theta) \\
-G^T(\theta) & 0_{r \times r}
\end{pmatrix}
\begin{pmatrix}
\sqrt{n}(\tilde{\theta}_{rc} - \theta) \\
n^{-1/2} \lambda_n
\end{pmatrix}
= \begin{pmatrix}
\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left. \frac{\partial}{\partial \theta} c \ell(\theta, y_i) \\
0_r
\end{pmatrix}
+ o_P(1).
\]
Then
\[
\left( \frac{\sqrt{n}(\tilde{\theta}_{rc} - \theta)}{n^{-1/2} \lambda_n} \right) = \begin{pmatrix} P(\theta) & Q(\theta) \\ Q^T(\theta) & R(\theta) \end{pmatrix} \left( \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{\partial}{\partial \theta} c\ell(\theta, y_i) \right)_{0_r} + o_P(1), \quad (33)
\]

where
\[
\begin{pmatrix} P(\theta) & Q(\theta) \\ Q^T(\theta) & R(\theta) \end{pmatrix} = \begin{pmatrix} H(\theta) & -G(\theta) \\ -G^T(\theta) & 0_{r \times r} \end{pmatrix}^{-1}.
\]

This last equation implies (cf. Sen and Singer 1993, p. 243, Eq. (5.6.24)),
\[
P(\theta) = H^{-1}(\theta) \left( I_p - G(\theta) \left( G^T(\theta) H^{-1}(\theta) G(\theta) \right)^{-1} G^T(\theta) H^{-1}(\theta) \right),
\]
\[
Q(\theta) = -H^{-1}(\theta) G(\theta) \left( G^T(\theta) H^{-1}(\theta) G(\theta) \right)^{-1},
\]
\[
R(\theta) = -\left( G^T(\theta) H^{-1}(\theta) G(\theta) \right)^{-1}.
\]

Based on the central limit theorem (Theorem 3.3.1 of Sen and Singer 1993, p. 107) and the Cramér–Wald theorem (Theorem 3.2.4 of Sen and Singer 1993, p. 106) it is obtained
\[
\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{\partial}{\partial \theta} c\ell(\theta, y_i) \xrightarrow{\mathcal{L}} \mathcal{N}(0_p, \text{Var}_\theta[u(\theta, Y)]).
\]

with \(\text{Var}_\theta[u(\theta, Y)] = J(\theta)\). Then, it follows from (33) that
\[
\left( \frac{\sqrt{n}(\tilde{\theta}_{rc} - \theta)}{n^{-1/2} \lambda_n} \right) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \Sigma),
\]

with
\[
\Sigma = \begin{pmatrix} P(\theta) & Q(\theta) \\ Q^T(\theta) & R(\theta) \end{pmatrix} \begin{pmatrix} J(\theta) & 0_{p \times r} \\ 0_{r \times p} & 0_{r \times r} \end{pmatrix} \begin{pmatrix} P^T(\theta) & Q(\theta) \\ Q^T(\theta) & R^T(\theta) \end{pmatrix}.
\]

or
\[
\Sigma = \begin{pmatrix} P(\theta) J(\theta) P^T(\theta) & P(\theta) J(\theta) Q(\theta) \\ Q^T(\theta) J(\theta) P^T(\theta) & Q^T(\theta) J(\theta) Q(\theta) \end{pmatrix}.
\]

Therefore,
\[
\sqrt{n}(\tilde{\theta}_{rc} - \theta) \xrightarrow{\mathcal{L}} \mathcal{N}(0_p, P(\theta) J(\theta) P^T(\theta)),
\]

with
\[
P(\theta) = H^{-1}(\theta) \left( I_p - G(\theta) \left( G^T(\theta) H^{-1}(\theta) G(\theta) \right)^{-1} G^T(\theta) H^{-1}(\theta) \right)
\]
\[
= H^{-1}(\theta) - H^{-1}(\theta) G(\theta) \left( G^T(\theta) H^{-1}(\theta) G(\theta) \right)^{-1} G^T(\theta) H^{-1}(\theta)
\]
\[
= H^{-1}(\theta) + Q(\theta) G^T(\theta) H^{-1}(\theta),
\]
and the proof of the lemma is now completed.

**Proof of Lemma 4**

Based on Eq. (33), above,

$$\sqrt{n}(\hat{\theta}_rc - \theta) = P(\theta) \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{\partial}{\partial \theta} c\ell(\theta, y_i) + o_P(1).$$  \hspace{1cm} (34)

The Taylor series expansion (30) gives that

$$0 = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{\partial}{\partial \theta} c\ell(\theta, y_i) \bigg|_{\theta=\hat{\theta}_c} = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{\partial}{\partial \theta} c\ell(\theta, y_i) + \frac{1}{n} \sum_{i=1}^{n} \frac{\partial^2}{\partial \theta \partial \theta^T} c\ell(\theta, y_i) \bigg|_{\theta=\theta^*_n} \sqrt{n}(\hat{\theta}_c - \theta),$$

or

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{\partial}{\partial \theta} c\ell(\theta, y_i) = -\frac{1}{n} \sum_{i=1}^{n} \frac{\partial^2}{\partial \theta \partial \theta^T} c\ell(\theta, y_i) \bigg|_{\theta=\theta^*_n} \sqrt{n}(\hat{\theta}_c - \theta),$$

where $\theta^*_n$ belongs to the line segment joining $\theta$ and $\hat{\theta}_c$. Taking into account Theorem 2.3.6 of Sen and Singer (1993, p. 61),

$$\frac{1}{n} \sum_{i=1}^{n} \frac{\partial^2}{\partial \theta \partial \theta^T} c\ell(\theta, y_i) \bigg|_{\theta=\theta^*_n} \xrightarrow{n \to \infty} -H(\theta),$$

and the above two equations lead

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{\partial}{\partial \theta} c\ell(\theta, y_i) = H(\theta) \sqrt{n}(\hat{\theta}_c - \theta) + o_P(1).$$  \hspace{1cm} (35)

Equations (34), (35) and the fact that $P(\theta) = H^{-1}(\theta) + Q(\theta)G^T(\theta)H^{-1}(\theta)$ give that

$$\sqrt{n}(\hat{\theta}_rc - \theta) = P(\theta) \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{\partial}{\partial \theta} c\ell(\theta, y_i) + o_P(1)$$

$$= P(\theta) H(\theta) \sqrt{n}(\hat{\theta}_c - \theta) + o_P(1)$$

$$= \left( H^{-1}(\theta) + Q(\theta)G^T(\theta)H^{-1}(\theta) \right) H(\theta) \sqrt{n}(\hat{\theta}_c - \theta) + o_P(1),$$

which completes the proof of the lemma.
Proof of Theorem 5

A second order Taylor expansion of $D_\phi(\hat{\theta}_{rc}, \tilde{\theta}_{rc})$, considered as a function of $\hat{\theta}_c$, around $\tilde{\theta}_{rc}$, gives

$$D_\phi(\hat{\theta}_c, \tilde{\theta}_{rc}) = D_\phi(\hat{\theta}_{rc}, \tilde{\theta}_{rc}) + \frac{\partial}{\partial \theta} D_\phi(\theta, \tilde{\theta}_{rc}) \bigg|_{\theta = \hat{\theta}_{rc}} (\hat{\theta}_c - \tilde{\theta}_{rc}) + \frac{1}{2} (\hat{\theta}_c - \tilde{\theta}_{rc})^T \frac{\partial^2}{\partial \theta \partial T} D_\phi(\theta, \tilde{\theta}_{rc}) \bigg|_{\theta = \hat{\theta}_{rc}} (\hat{\theta}_c - \tilde{\theta}_{rc}) + o(\|\hat{\theta}_c - \tilde{\theta}_{rc}\|^2).$$

Based on Pardo (2006, pp. 411–412), we obtain $D_\phi(\hat{\theta}_{rc}, \tilde{\theta}_{rc}) = 0$ and $\frac{\partial^2}{\partial \theta \partial T} D_\phi(\theta, \tilde{\theta}_{rc}) \bigg|_{\theta = \hat{\theta}_{rc}} = \phi''(1) \mathbf{J}(\tilde{\theta}_{rc})$. Then, the above equation leads

$$\frac{2n}{\phi''(1)} D_\phi(\hat{\theta}_c, \tilde{\theta}_{rc}) = n(\hat{\theta}_c - \tilde{\theta}_{rc})^T \mathbf{J}(\tilde{\theta}_{rc}) \sqrt{n}(\hat{\theta}_c - \tilde{\theta}_{rc}) + n o(\|\hat{\theta}_c - \tilde{\theta}_{rc}\|^2),$$

or

$$T_{\phi,n}(\hat{\theta}_c, \tilde{\theta}_{rc}) = \sqrt{n}(\hat{\theta}_c - \tilde{\theta}_{rc})^T \mathbf{J}(\tilde{\theta}_{rc}) \sqrt{n}(\hat{\theta}_c - \tilde{\theta}_{rc}) + n o(\|\hat{\theta}_c - \tilde{\theta}_{rc}\|^2). \quad (36)$$

On the other hand (cf., Pardo 2006, p. 63),

$$n o(\|\hat{\theta}_c - \tilde{\theta}_{rc}\|^2) \leq n o(\|\hat{\theta}_c - \theta\|^2) + n o(\|\tilde{\theta}_{rc} - \theta\|^2),$$

and $n o(\|\hat{\theta}_c - \theta\|^2) = o_P(1)$, $n o(\|\tilde{\theta}_{rc} - \theta\|^2) = o_P(1)$, $o(\|\hat{\theta}_c - \tilde{\theta}_{rc}\|^2) = o_P(1)$. To apply the Slutsky’s theorem, it remains to obtain the asymptotic distribution of the quantity

$$\sqrt{n}(\hat{\theta}_c - \tilde{\theta}_{rc})^T \mathbf{J}(\tilde{\theta}_{rc}) \sqrt{n}(\hat{\theta}_c - \tilde{\theta}_{rc}).$$

From Lemma 4 it is immediately obtained that

$$\sqrt{n}(\hat{\theta}_c - \tilde{\theta}_{rc}) = \mathbf{Q}(\theta) \mathbf{G}^T(\theta) \sqrt{n}(\hat{\theta}_c - \theta) + o_P(1).$$

On the other hand, we know that

$$\sqrt{n}(\hat{\theta}_c - \theta) \xrightarrow{L} \mathcal{N}(0, \mathbf{G}^{-1}(\theta)).$$

Therefore,

$$\sqrt{n}(\hat{\theta}_c - \tilde{\theta}_{rc}) \xrightarrow{L} \mathcal{N}(0, \mathbf{Q}(\theta) \mathbf{G}^T(\theta) \mathbf{G}^{-1}(\theta) \mathbf{G}(\theta) \mathbf{Q}^T(\theta)).$$

and taking into account (36) and Corollary 2.1 of Dic and Gunst (1985), $T_{\phi,n}(\hat{\theta}_c, \tilde{\theta}_{rc})$ converge in law to the random variable $\sum_{i=1}^k \beta_i Z_i^2$, where $\beta_i, i = 1, \ldots, k$, are the
eigenvalues of the matrix \( J(\theta)Q(\theta)G^T(\theta)G^{-1}(\theta)G(\theta)Q^T(\theta) \) and
\[
k = \text{rank} \left( Q(\theta)G^T(\theta)G^{-1}(\theta)G(\theta)Q^T(\theta)J(\theta)Q(\theta)G^T(\theta)G^{-1}(\theta)G(\theta)Q^T(\theta) \right).
\]

**Proof of Theorem 6**

The result follows in a straightforward manner by considering a first order Taylor expansion of \( D_\phi(\hat{\theta}_c, \tilde{\theta}_{rc}) \), which yields
\[
D_\phi(\hat{\theta}_c, \tilde{\theta}_{rc}) = D_\phi(\theta, \theta^*) + t^T(\hat{\theta}_c - \theta) + s^T(\tilde{\theta}_{rc} - \theta^*) + o(\|\hat{\theta}_c - \theta\| + \|\tilde{\theta}_{rc} - \theta^*\|).
\]

**Calculation of the integral \( I_a \)** in (28)

The integral \( I_a \) is given by
\[
I_a = \int_{\mathbb{R}^q} \mathcal{C}\mathcal{L}(\hat{\mu}, \hat{\sigma}, \hat{\rho}, y)^a \mathcal{C}\mathcal{L}(\tilde{\mu}, \sigma_0, \rho_0, y)^{1-a} dy, \quad a \neq 0, 1,
\]
where
\[
\mathcal{C}\mathcal{L}(\mu, \sigma, \rho, y) = L_{\rho,\sigma}(1) \exp\left\{ -\frac{1}{2\sigma^2} \left[ \sum_{r=2}^{q} (y_{r-1} - \mu)^2 + \sum_{r=2}^{q} (y_r - \mu)^2 \right] -2\rho \sum_{r=2}^{q} (y_{r-1} - \mu)(y_r - \mu) \right\},
\]
and
\[
L_{\rho,\sigma}(s) = \frac{(1 - \rho^2)^{s(q-1)/2}}{(2\pi)^{s(q-1)/2}\sigma^2(q-1)_s}. \quad (37)
\]

Then,
\[
I_a = L_{\rho,\sigma}(a)L_{\rho_0,\sigma_0}(1-a) \times \int_{\mathbb{R}^q} \exp\left\{ -\frac{a}{2\sigma^2} \left[ \sum_{r=2}^{q} (y_{r-1} - \hat{\mu})^2 + \sum_{r=2}^{q} (y_r - \hat{\mu})^2 - 2\hat{\rho} \sum_{r=2}^{q} (y_{r-1} - \hat{\mu})(y_r - \hat{\mu}) \right] \right\} \\
\times \exp\left\{ -\frac{(1-a)}{2\sigma_0^2} \left[ \sum_{r=2}^{q} (y_{r-1} - \tilde{\mu})^2 + \sum_{r=2}^{q} (y_r - \tilde{\mu})^2 - 2\rho_0 \sum_{r=2}^{q} (y_{r-1} - \tilde{\mu})(y_r - \tilde{\mu}) \right] \right\} dy_1 \ldots dy_q. \quad (38)
\]
But, if

$$E_1 = \exp \left\{ -\frac{a}{2\hat{\sigma}^2} \left[ \sum_{r=2}^{q} (y_{r-1} - \hat{\mu})^2 + \sum_{r=2}^{q} (y_r - \hat{\mu})^2 - 2\hat{\rho} \sum_{r=2}^{q} (y_{r-1} - \hat{\mu})(y_r - \hat{\mu}) \right] \right\},$$

$$E_2 = \exp \left\{ -\frac{(1-a)}{2\sigma_0^2} \left[ \sum_{r=2}^{q} (y_{r-1} - \hat{\mu})^2 + \sum_{r=2}^{q} (y_r - \hat{\mu})^2 - 2\rho_0 \sum_{r=2}^{q} (y_{r-1} - \hat{\mu})(y_r - \hat{\mu}) \right] \right\},$$

then

$$E_1 \times E_2 = \exp \left\{ -\frac{1}{2} \left( \frac{a}{\hat{\sigma}^2} - \frac{1-a}{\sigma_0^2} \right) \left[ \sum_{r=2}^{q} (y_{r-1} - \hat{\mu})^2 + \sum_{r=2}^{q} (y_r - \hat{\mu})^2 \right] + \left( \frac{\hat{\rho}}{\hat{\sigma}^2} + \frac{(1-a)\rho_0}{\sigma_0^2} \right) \sum_{r=2}^{q} (y_{r-1} - \hat{\mu})(y_r - \hat{\mu}) \right\}$$

$$= \exp \left\{ -\frac{1}{2} \left( \frac{a}{\hat{\sigma}^2} - \frac{1-a}{\sigma_0^2} \right) \left[ \sum_{r=2}^{q} (y_{r-1} - \hat{\mu})^2 + \sum_{r=2}^{q} (y_r - \hat{\mu})^2 \right] -2 \frac{a\hat{\rho}}{\hat{\sigma}^2} - \frac{1-a}{\sigma_0^2} \sum_{r=2}^{q} (y_{r-1} - \hat{\mu})(y_r - \hat{\mu}) \right\}$$

$$= \exp \left\{ -\frac{1}{2\sigma_*^2} \left[ \sum_{r=2}^{q} (y_{r-1} - \hat{\mu})^2 + \sum_{r=2}^{q} (y_r - \hat{\mu})^2 \right] -2\rho^* \sum_{r=2}^{q} (y_{r-1} - \hat{\mu})(y_r - \hat{\mu}) \right\},$$

(39)

with

$$\frac{1}{\sigma_*^2} = \frac{a}{\hat{\sigma}^2} - \frac{1-a}{\sigma_0^2} = \frac{a\sigma_0^2}{\hat{\sigma}^2\sigma_0^2} - \frac{(1-a)\hat{\sigma}^2}{\hat{\sigma}^2\sigma_0^2},$$

or

$$\sigma_*^2 = \frac{\hat{\sigma}^2\sigma_0^2}{a\sigma_0^2 - (1-a)\hat{\sigma}^2},$$

(40)

and

$$\rho^* = \frac{a\hat{\rho} + (1-a)\rho_0}{\frac{a}{\hat{\sigma}^2} - \frac{1-a}{\sigma_0^2}} = \frac{a\hat{\rho}\sigma_0^2 + (1-a)\rho_0\hat{\sigma}^2}{\frac{a\hat{\sigma}^2\sigma_0^2}{\hat{\sigma}^2\sigma_0^2} - (1-a)\hat{\sigma}^2} = \frac{a\hat{\rho}\sigma_0^2 + (1-a)\rho_0\sigma_0^2}{a\sigma_0^2 - (1-a)\hat{\sigma}^2}.$$

(41)
Based on (38)–(41),

\[
I_a = \frac{L_{\hat{\rho}, \hat{\sigma}}(a)L_{\rho_0, \sigma_0}(1-a)}{L_{\rho^*, \sigma^*}(1)} \times \int_{\mathbb{R}^q} L_{\rho^*, \sigma^*}(1) \exp \left\{ -\frac{1}{2\sigma^*_{\rho}} \left[ \sum_{r=2}^q (y_{r-1} - \hat{\mu})^2 + \sum_{r=2}^q (y_r - \hat{\mu})^2 \right] -2\rho^* \sum_{r=2}^q (y_{r-1} - \hat{\mu})(y_r - \hat{\mu}) \right\} \, dy,
\]

and taking into account that the last integral is equal to one,

\[
I_a = \frac{L_{\hat{\rho}, \hat{\sigma}}(a)L_{\rho_0, \sigma_0}(1-a)}{L_{\rho^*, \sigma^*}(1)},
\]  

(42)

with \(L_{\rho, \sigma}(s)\), defined by (37). After some algebraic manipulations, (42) leads to the explicit expression of the integral \(I_a\), given by (28).

References

Ali S, Silvey S (1966) A general class of coefficients of divergence of one distribution from another. J R Stat Soc Ser B 28:131–142
Basu A, Shioya H, Park C (2011) Statistical inference. The minimum distance approach. Chapman & Hall/CRC, Boca Raton
Basu A, Mandal A, Martin N, Pardo L (2015) Density power divergence tests for composite null hypotheses. arXiv: 1403.0330v3
Besag J (1974) Spatial interaction and the statistical analysis of lattice systems. J R Stat Soc Ser B 36:192–236
Cattelan M, Sartiri N (2016) Empirical and simulated adjustments of composite likelihood ratio statistics. J Stat Comput Simul 86:1056–1067
Cressie N, Read TRC (1984) Multinomial goodness-of-fit tests. J R Stat Soc Ser B 46:440–464
Cressie N, Pardo L, Pardo MC (2003) Size and power considerations for testing loglinear models using \(\phi\)-divergence test statistics. Stat Sin 13:550–557
Csiszár I (1963) Eine informationstheoretische Ungleichung und ihre Anwendung auf den Beweis der Ergodizität von Markffschen Ketten. Magyar Tud. Akad Mat Kutato Int Kozl 8:85–108
Csiszár I (1967) Information-type measures of difference of probability distributions and indirect observations. Stud Sci Math Hung 2:299–318
Dale JR (1986) Asymptotic normality of goodness-of-fit statistics for sparse product multinomials. J R Stat Soc Ser B 48:48–59
Dic JJ, Gunst MCM (1985) The distribution of general quadratic forms in normal variables. Stat Neerl 39:14–26
Joe H, Reid N, Song PX, Firth D, Varin C (2012) Composite likelihood methods. In: Report on the workshop on composite likelihood. http://www.birs.ca/events/2012/5-day-workshops/12w5046
Kullback S (1959) Information theory and statistics. Wiley, New York
Kullback S, Leibler RA (1951) On information and sufficiency. Ann Math Stat 22:79–86
Leomnte AJ (2016) The gradient test: another likelihood-based test. Elsevier, San Diego
Liese F, Vajda I (1987) Convex statistical distances. Teubner texts in mathematics. Teubner, Leipzig
Liese F, Vajda I (2006) On divergences and informations in statistics and information theory. IEEE Trans Inf Theory 52:4394–4412
Lindsay G (1988) Composite likelihood methods. Contemp Math 80:221–239

\(\copyright\) Springer
Martin N, Pardo L (2012) Poisson-loglinear modeling with linear constraints on the expected cell frequencies. Sankhya B 74:238–267
Martin N, Mata R, Pardo L (2014) Phi-divergence statistics for the likelihood ratio order: an approach based on log-linear models. J Multivar Anal 130:387–408
Martin N, Mata R, Pardo L (2016) Wald type and phi-divergence based test-statistics for isotonic binomial proportions. Math Comput Simul 120:31–49
Menéndez ML, Morales D, Pardo L, Salicrú M (1995) Asymptotic behavior and statistical applications of divergence measures in multinomial populations: a unified study. Stat Pap 36:1–29
Menéndez ML, Pardo JA, Pardo L, Pardo MC (1997) Asymptotic approximations for the distributions of the \((h, \phi)\)-divergence goodness-of-fit statistics: applications to Rényi’s statistic. Kybernetes 26:442–452
Morales D, Pardo L, Vajda I (1997) Some new statistics for testing hypotheses in parametric models. J Multivar Anal 62:137–168
Pace L, Salvan A, Sartori N (2011) Adjusting composite likelihood ratio statistics. Stat Sin 21:129–148
Pardo L (2006) Statistical inference based on divergence measures. Chapman & Hall/CRC, Boca Raton
Rao JNK, Scott AJ (1981) The analysis of categorical data from complex sample surveys: chi-squared tests for goodness-of-fit and independence in two-way tables. J Am Stat Assoc 76:221–230
Read TRC, Cressie N (1988) Goodness-of-fit statistics for discrete multivariate data. Springer, New York
Reid N (2013) Aspects of likelihood inference. Bernoulli 19:1404–1418
Reid N, Lindsay B, Liang K-Y (2011) Introduction to special issue. Stat Sin 21:1–3
Rényi A (1961) On measures of entropy and information. In: Proceedings of the 4th Berkeley symposium on mathematical statistics and probability, vol I, Berkeley, pp 547–561
Satterthwaite FE (1946) An approximate distribution of estimates of variance components. Biom Bull 2:110–114
Sen PK, Singer JM (1993) Large sample methods in statistics. Chapman & Hall/CRC, New York
Serfling RJ (1980) Aproximation theorems of mathematical statistics. Wiley, New York
Sharma BD, Mittal DP (1997) New non-additive measures of relative information. J Combin Inf Syst Sci 2:122–133
Stummer W, Vajda I (2010) On divergences of finite measures and their applicability in statistics and information theory. Statistics 44:169–187
Terrell GR (2002) The gradient statistic. Comput Sci Stat 34:206–215
Vajda I (1989) Theory of statistical inferences and information. Kluwer, Dordrecht
Varin C (2008) On composite marginal likelihoods. Adv Stat Anal 92:1–28
Varin C, Reid N, Firth D (2011) An overview of composite likelihood methods. Stat Sin 21:5–42
Xu X, Reid N (2011) On the robustness of maximum composite estimate. J Stat Plan Inference 141:3047–3054
Zografos K (1998) \( f\)-dissimilarity of several distributions in testing statistical hypotheses. Ann Inst Stat Math 50:295–310