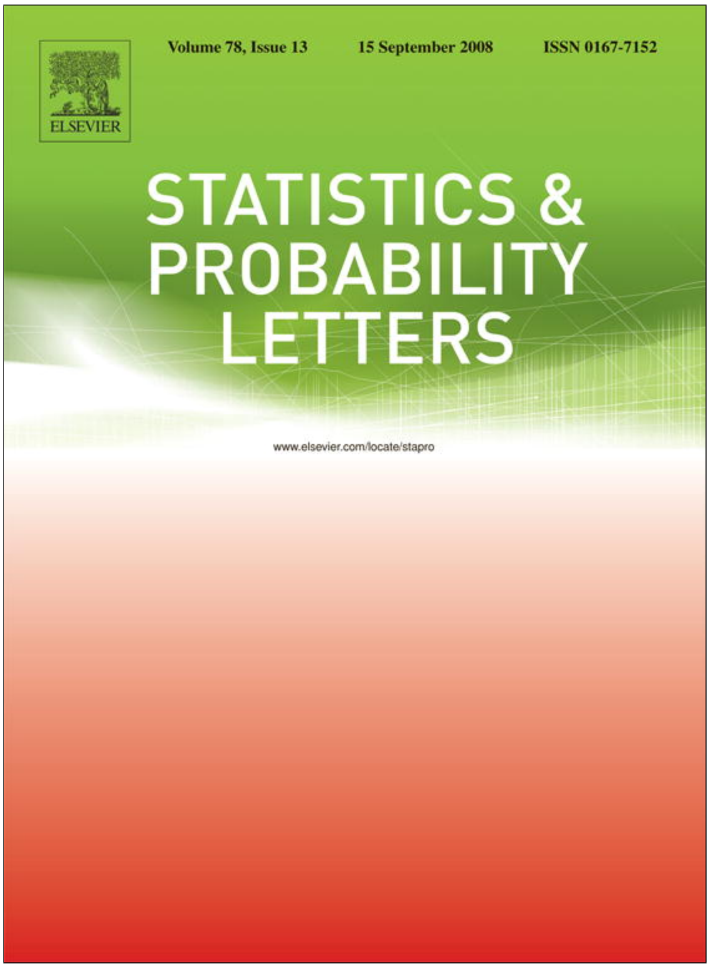
Volume 78, Issue 13
15 September 2008
ISSN 0167-7152
ELSEVIER
STATISTICS & PROBABILITY LETTERS
www.elsevier.com/locate/stapro

Available online at www.sciencedirect.com
ScienceDirect
ELSEVIER
Statistics and Probability Letters 78 (2008) 1702–1710
STATISTICS & PROBABILITY LETTERS
www.elsevier.com/locate/stapro

Divergence-based tests for model diagnostic[☆]

M.D. Esteban[a], T. Hobza[b], D. Morales[a], Y. Marhuenda[a,*]

[a] *Operations Research Center, Miguel Hernández University of Elche, Elche, Spain*
[b] *Department of Mathematics, Czech Technical University, Prague, Czech Republic*

Received 2 December 2005; received in revised form 14 September 2007; accepted 3 January 2008
Available online 15 January 2008

Abstract

Pearson's χ^2 test, and more generally, divergence-based tests of goodness-of-fit are asymptotically χ^2-distributed with $m-1$ degrees of freedom if the numbers of cells m is fixed, the observations are i.i.d and the cell probabilities and model parameters are completely specified. Jiang [Jiang, J., 2001. A nonstandard χ^2-test with application to generalized linear model diagnostics. Statistics and Probability Letters 53, 101–109] proposed a nonstandard χ^2 test to check distributional assumptions for the case of observations not identically distributed. Under the same setup, in this paper a family of divergence-based tests are introduced and their asymptotic distributions are derived. In addition bootstrap tests based on the given divergence test statistics are considered. Applications to generalized linear models diagnostic are proposed. A simulation study is carried out to investigate performance of several power-divergence tests.

1. Introduction

The problem of goodness-of-fit to a distribution in the real line, $H_0 : F = F_0$, is frequently treated by partitioning the range of data in disjoint intervals and by testing the hypothesis $H_0 : \boldsymbol{p} = \boldsymbol{p}_0$ of a multinomial distribution.

Let $Y_1, \dots, Y_n$ be i.i.d. random variables with c.d.f. F. Let $E_1, \dots, E_m$ be a partition of $R = (-\infty, \infty)$ in m intervals. Let $\boldsymbol{p} = (p_1, \dots, p_m)$ and $\boldsymbol{p}_0 = (p_{01}, \dots, p_{0m})$ be the true and hypothetical probabilities of the intervals E_k, i.e.

$$p_{0k} = \int_{E_k} \mathrm{d}F_0, \qquad p_k = \int_{E_k} \mathrm{d}F, \quad k = 1, \dots, m.$$

Define the observed cell counts

$$N_k = \sum_{j=1}^{n} 1_{(Y_j \in E_k)} = \#(1 \le j \le n : Y_j \in E_k), \quad k = 1, \dots, m,$$

[☆] Supported by the grants MTM2006-05693 and MSMT 1M0572.

* Corresponding author.
E-mail address: y.marhuenda@umh.es (Y. Marhuenda).

0167-7152/$ - see front matter © 2008 Elsevier B.V. All rights reserved.
doi:10.1016/j.spl.2008.01.007

and the estimated cell probabilities $\widehat{\boldsymbol{p}} = (\widehat{p}_1, \ldots, \widehat{p}_m)$ with $\widehat{p}_k = N_k/n$, $k = 1, \ldots, m$. To test $H_0 : \boldsymbol{p} = \boldsymbol{p}_0$ the most commonly used test statistic is Pearson's χ^2 statistic

$$\chi^2_{nP}(\widehat{\boldsymbol{p}}, \boldsymbol{p}_0) = n \sum_{k=1}^{m} \frac{(\widehat{p}_k - p_{0k})^2}{p_{0k}}, \tag{1}$$

which is a particular case of the family of power-divergence statistics introduced by Cressie and Read (1984) and given by

$$T_n^r(\widehat{\boldsymbol{p}}, \boldsymbol{p}_0) = \frac{2n}{r(r+1)} \sum_{k=1}^{m} \widehat{p}_k \left[\left(\frac{\widehat{p}_k}{p_{0k}} \right)^r - 1 \right], \quad -\infty < r < \infty. \tag{2}$$

The test statistics $T_n^0(\widehat{\boldsymbol{p}}, \boldsymbol{p}_0)$ and $T_n^{-1}(\widehat{\boldsymbol{p}}, \boldsymbol{p}_0)$ are defined by continuity. Well-known test statistics are obtained from particular values of r in (2). Some examples are $r = 1$ for Pearson's test statistic, $r = 0$ for the log-likelihood-ratio statistic, $r = -1/2$ for the Freeman–Tukey test statistic, $r = -2$ for the Neyman-modified test statistic and $r = 2/3$ for the Cressie–Read statistic.

More generally, $T_n^r(\widehat{\boldsymbol{p}}, \boldsymbol{p}_0)$ is a particular case of the ϕ-divergence test statistic

$$T_n^{\phi}(\widehat{\boldsymbol{p}}, \boldsymbol{p}_0) = \frac{2n}{\phi''(1)} D_{\phi}(\widehat{\boldsymbol{p}}, \boldsymbol{p}_0) = \frac{2n}{\phi''(1)} \sum_{k=1}^{m} p_{0k} \phi \left(\frac{\widehat{p}_k}{p_{0k}} \right), \tag{3}$$

where $D_{\phi}(\cdot, \cdot)$ denotes the ϕ-divergence of two probability distributions introduced by Csiszár (1963) and Ali and Silvey (1966) for every ϕ in the set Φ of real convex functions defined on $[0, \infty)$, continuously differentiable in the neighborhood of 1 and satisfying $\phi(1) = \phi'(1) = 0$, $\phi''(1) > 0$. In formula (3) if either p_{0k} or p_{0k} and $\widehat{p}_k$ are zero, expressions $0\phi(x/0)$ and $0\phi(0/0)$ are defined as $x \cdot \lim_{u\to\infty} \phi(u)/u$ and 0 respectively. Properties of ϕ-divergences have been extensively studied by Liese and Vajda (1987) and Vajda (1989). Zografos et al. (1990) proved that $T_n^{\phi}(\widehat{\boldsymbol{p}}, \boldsymbol{p}_0) \xrightarrow{L} \chi^2_{m-1}$ as $n \to \infty$ under $H_0 : \boldsymbol{p} = \boldsymbol{p}_0$, where $\xrightarrow{L}$ stands for convergence in law.

It is common to deal with the problem of testing the composite hypothesis that the c.d.f. F is a member of a parametric family $\{F_\theta\}_{\theta \in \Theta}$ for a given open subset $\Theta \subset R^d$. In such cases cell probabilities depend on the unknown parameter θ, i.e.

$$p_k(\theta) = \int_{E_k} \mathrm{d}F_\theta, \quad k = 1, \ldots, m,$$

so they may be estimated with minimum ϕ-divergence estimators satisfying

$$\widehat{\theta}_{\phi} = \arg \inf_{\theta \in \Theta} D_{\phi}(\widehat{\boldsymbol{p}}, \boldsymbol{p}(\theta)),$$

which contains as a particular case the maximum likelihood estimator (MLE) based on the quantized data. Morales et al. (1996) proved that if regularity conditions given by Birch (1964) hold, then

$$T_n^{\phi_1}(\widehat{\boldsymbol{p}}, \boldsymbol{p}(\widehat{\theta}_{\phi_2})) \xrightarrow[n\to\infty]{L} \chi^2_{m-d-1}$$

under $H_0 : F = F_\theta$ for any $\phi_1, \phi_2 \in \Phi$. However, if MLE estimator $\widehat{\theta}$ is based on the original data, then the asymptotic distribution of $T_n^{\phi}(\widehat{\boldsymbol{p}}, \boldsymbol{p}(\widehat{\theta}))$ under $H_0 : F = F_\theta$ is a linear combination of independent χ^2_1 variables. This result was originally proved by Chernoff and Lehmann (1954) and extended to any $\phi \in \Phi$ by Morales et al. (1996).

If original variables are independent with c.d.f.s $F_1, \ldots, F_n$, depending on an unknown parameter $\theta \in \Theta \subset R^d$ open, the hypothesis of interest is

$$H_0 : Y_1 \sim F_1, \ldots, Y_n \sim F_n. \tag{4}$$

Let us define $p_k(\theta) = \mathbf{E}_\theta[N_k]/n$, with $\mathbf{E}_\theta[N_k] = \sum_{j=1}^{n} P_\theta(Y_j \in E_k)$. Jiang (2001) proposed to test H_0 with

$$\chi^2_{nJ}(\widehat{\boldsymbol{p}}, \boldsymbol{p}(\widehat{\theta})) = n \sum_{k=1}^{m} (\widehat{p}_k - p_k(\widehat{\theta}))^2, \tag{5}$$

where $\widehat{\theta}$ is a consistent estimator of θ, and gave regularity conditions under which asymptotic distribution of $\chi^2_{nJ}(\widehat{\boldsymbol{p}}, \boldsymbol{p}(\widehat{\theta}))$ is a linear combination of independent χ^2_1 variables.

The targets of this paper are to extend Jiang's result to the class of test statistics $T_n^{\phi}(\widehat{\boldsymbol{p}}, \boldsymbol{p}(\widehat{\theta}))$, to introduce their bootstrap versions and finally to give some recommendations on the choice of ϕ based on the results obtained from Monte Carlo simulation experiments. The rest of the paper is organized as follows: In Section 2 the asymptotic distribution of $T_n^{\phi}(\widehat{\boldsymbol{p}}, \boldsymbol{p}(\widehat{\theta}))$ is derived. In Section 3 the corresponding bootstrap tests are introduced. In Section 4 applications to GLM diagnostics are suggested, a simulation experiment is carried out to investigate the finite sample performance of the introduced test statistics and some conclusions are given.

2. Asymptotic distribution of T_n^{ϕ} statistics

In this section we derive the asymptotic distribution of the T_n^{ϕ} statistics

$$T_n^{\phi} = T_n^{\phi}(\widehat{\boldsymbol{p}}, \boldsymbol{p}(\widehat{\theta})) = \frac{2n}{\phi''(1)} \sum_{k=1}^{m} p_k(\widehat{\theta}) \phi\left(\frac{\widehat{p}_k}{p_k(\widehat{\theta})}\right) \tag{6}$$

for the class of functions $\phi \in \Phi$ under the null hypotheses (4). This leads to a goodness-of-fit test, which can be used to check the distributional assumptions in the model involving independent but not identically distributed random variables. Essential for us will be the result of Jiang (2001) where asymptotic distribution of the statistics $\chi^2_{nJ}(\widehat{\boldsymbol{p}}, \boldsymbol{p}(\widehat{\theta}))$ was given. Let us start with introducing some notation and regularity conditions used in Jiang (2001).

It is known that the choice of $\widehat{\theta}$ has an impact on the asymptotic distribution of T_n^{ϕ}. Throughout this paper it is assumed that $\widehat{\theta}$ is a consistent estimator of θ and has an asymptotic expansion

$$\sqrt{n}(\widehat{\theta} - \theta) = A_n \left(\frac{1}{\sqrt{n}} \sum_{j=1}^{n} \psi_j(Y_j, \theta) \right) + o_P(1). \tag{7}$$

For example, under some regularity conditions, the MLE of θ has the expansion (7), where ψ_j is the score function corresponding to the jth observation and A_n is equal to n times the inverse of the Fisher information matrix (based on all data).

Let further $\xi_n = (\xi_{nk})_{1 \le k \le m}$, where $\xi_{nk} = N_k - \mathbf{E}_{\widehat{\theta}} N_k$; $\boldsymbol{p}_j(\theta) = (p_{jk}(\theta))_{1 \le k \le m}$ and $p_{jk}(\theta) = P_{\theta}(Y_j \in E_k)$. Define

$$h_{nj} = (1_{(Y_j \in E_k)} - p_{jk}(\theta))_{1 \le k \le m} - \left(\frac{1}{n} \sum_{j=1}^{n} \frac{\partial}{\partial \theta} \boldsymbol{p}_j(\theta) \right) A_n \psi_j(Y_j, \theta)$$

and $\Sigma_n = \Sigma_n(\theta) = n^{-1} \sum_{j=1}^{n} \mathbf{Var}(h_{nj})$. Let Q_n be an orthogonal matrix such that

$$Q_n^t \Sigma_n Q_n = D_n = \operatorname{diag}(\lambda_{n1}, \ldots, \lambda_{nm}),$$

where $\lambda_{n1} \ge \cdots \ge \lambda_{nm}$ are the eigenvalues of Σ_n.

The following set of assumptions is supposed: (i) $Y_1, \ldots, Y_n$ are independent, (ii) $\Sigma_n \longrightarrow \Sigma$ as $n \to \infty$, (iii) (7) holds with $\mathbf{E}\psi_j(Y_j, \theta) = 0$, $1 \le j \le n$, and (iv) it holds

$$\frac{1}{n} \max_{1 \le j \le n} \mathbf{E}|A_n \psi_j(Y_j, \theta)|^4 \longrightarrow 0, \qquad \max_{1 \le j \le n} \left| \frac{\partial}{\partial \theta} \boldsymbol{p}_j(\theta) \right| = O(1),$$

and there exists $\delta > 0$ such that

$$\frac{1}{n} \sum_{j=1}^{n} \sup_{|\widetilde{\theta} - \theta| \le \delta} \left\| \frac{\partial^2}{\partial \theta^2} p_{jk}(\widetilde{\theta}) \right\| = O(1), \quad 1 \le k \le m.$$

Under the assumptions (i)–(iv) Jiang proved that the asymptotic distribution of χ^2_{nJ} is the same as that of $\sum_{k=1}^{m} \lambda_k Z_k^2$ where $Z_1, \ldots, Z_m$ are i.i.d. $N(0, 1)$ random variables and $\lambda_1, \ldots, \lambda_m$ are the eigenvalues of Σ.

First we extend Jiang's result to the Pearson statistics $\chi^2_{nP}(\widehat{\boldsymbol{p}}, \boldsymbol{p}(\widehat{\theta}))$ defined in (1). To achieve this aim we need to put an additional assumption about the partition and the probability model:

$$\boldsymbol{p}(\theta) = \frac{1}{n}\sum_{j=1}^{n} \boldsymbol{p}_j(\theta) \underset{n\to\infty}{\longrightarrow} \boldsymbol{q}, \quad \text{where } q_k > 0 \text{ for all } k \in \{1, \dots, m\}. \tag{8}$$

The above-mentioned extension is stated in the following lemma.

Lemma 1. *If the assumptions* (i)–(iv) *and* (8) *are fulfilled then the statistic*

$$T_n^1 = \chi^2_{nP}(\widehat{\boldsymbol{p}}, \boldsymbol{p}(\widehat{\theta}))$$

has, under the null hypothesis (4)*, the same asymptotic distribution as* $\sum_{k=1}^m (\lambda_k/q_k) Z_k^2$*, where* $Z_1, \dots, Z_m$ *are i.i.d.* $N(0, 1)$ *random variables, and* $\lambda_1 \geq \cdots \geq \lambda_m$ *are the eigenvalues of* Σ.

Proof. In the proof of his Theorem 1, Jiang (2001) showed that under the assumptions (i)–(iv) it holds

$$X_n \triangleq n^{-1/2} Q_n^t \xi_n \xrightarrow{L} X \sim N_m(0, D), \tag{9}$$

where $D = \operatorname{diag}(\lambda_1, \dots, \lambda_m)$. Let us define the random vector

$$\widetilde{X}_n \triangleq \operatorname{diag}\,(\boldsymbol{p}(\theta))^{-1/2} \cdot X_n = B_n X_n,$$

where matrix B_n has diagonal elements $(B_n)_{kk} = p_k(\theta)^{-1/2} = (\mathbf{E}_\theta N_k/n)^{-1/2}$, $k = 1, \dots, m$. Then, (8) imply

$$B_n \longrightarrow B = \operatorname{diag}(q_1^{-1/2}, \dots, q_m^{-1/2})$$

for $n \to \infty$ and using the Slutsky theorem we get

$$\widetilde{X}_n \xrightarrow{L} BX \sim N_m(0, BDB^t).$$

From this it already follows that the asymptotic distribution of $\widetilde{X}_n^t \widetilde{X}_n$ is the same as that of $\sum_{k=1}^m (\lambda_k/q_k) Z_k^2$. To finish the proof we will show that $T_n^1 = \widetilde{X}_n^t \widetilde{X}_n + o_P(1)$. Let us start with a partial problem. For $k = 1, \dots, m$ we can write

$$p_k(\widehat{\theta}) = \frac{\mathbf{E}_{\widehat{\theta}} N_k}{n} = \frac{1}{n}\sum_{j=1}^n p_{jk}(\widehat{\theta}) = \frac{1}{n}\sum_{j=1}^n p_{jk}(\theta) + \frac{1}{n}\sum_{j=1}^n (p_{jk}(\widehat{\theta}) - p_{jk}(\theta)). \tag{10}$$

Using the Taylor expansion

$$p_{jk}(\widehat{\theta}) = p_{jk}(\theta) + \left(\frac{\partial}{\partial\theta} p_{jk}(\theta)\right)(\widehat{\theta} - \theta) + \frac{1}{2}(\widehat{\theta} - \theta)^t \left(\frac{\partial^2}{\partial\theta^2} p_{jk}(\theta^{(j,k)})\right)(\widehat{\theta} - \theta),$$

where $\theta^{(j,k)}$ lies in the line between θ and $\widehat{\theta}$, we get

$$\begin{aligned}
&\frac{1}{n}\sum_{j=1}^n (p_{jk}(\widehat{\theta}) - p_{jk}(\theta)) \\
&\quad = n^{-1/2}\left[\left(\frac{1}{n}\sum_{i=1}^n \frac{\partial}{\partial\theta} p_{ik}(\theta)\right)\sqrt{n}(\widehat{\theta} - \theta) + \frac{1}{2}\sqrt{n}(\widehat{\theta} - \theta)^t \left(\frac{1}{n}\sum_{i=1}^n \frac{\partial^2}{\partial\theta^2} p_{ik}(\theta^{(i,k)})\right)(\widehat{\theta} - \theta)\right] = o_P(1),
\end{aligned}$$

as follows from the assumptions of the lemma. Substituting this results into (10) for all k we finally get the asymptotic relation

$$\boldsymbol{p}(\widehat{\theta}) = \boldsymbol{p}(\theta) + o_P(1). \tag{11}$$

For the statistic of interest we have

$$\begin{aligned}
T_n^1 &= n\sum_{k=1}^m \frac{(\widehat{p}_k - p_k(\widehat{\theta}))^2}{p_k(\widehat{\theta})} = n^{-1}\sum_{k=1}^m \frac{(N_k - \mathbf{E}_{\widehat{\theta}} N_k)^2}{\mathbf{E}_{\widehat{\theta}} N_k/n} = n^{-1}\xi_n^t \operatorname{diag}\left(\boldsymbol{p}(\widehat{\theta})\right)^{-1} \xi_n \\
&= n^{-1}\xi_n^t Q_n \operatorname{diag}\left(\boldsymbol{p}(\widehat{\theta})\right)^{-1} Q_n^t \xi_n = X_n^t \operatorname{diag}\left(\boldsymbol{p}(\widehat{\theta})\right)^{-1} X_n
\end{aligned}$$

and thus

$$T_n^1 = \widetilde{X}_n^t \widetilde{X}_n + X_n^t \left(\operatorname{diag}\left(\boldsymbol{p}(\widehat{\theta})\right)^{-1} - \operatorname{diag}\left(\boldsymbol{p}(\theta)\right)^{-1} \right) X_n = \widetilde{X}_n^t \widetilde{X}_n + o_P(1)$$

as can be seen from (8), (9) and (11). □

The main result of this section stating the asymptotic distribution of T_n^ϕ is presented in the following theorem.

Theorem 1. *If the assumptions* (i)–(iv) *and* (8) *are fulfilled then for all* $\phi \in \Phi$ *the statistics*

$$T_n^\phi = T_n^\phi(\widehat{\boldsymbol{p}}, \boldsymbol{p}(\widehat{\theta}))$$

defined in (6) *has, under the null hypothesis* (4)*, the same asymptotic distribution as* $\sum_{k=1}^m (\lambda_k/q_k) Z_k^2$*, where* $Z_1, \ldots, Z_m$ *are i.i.d.* $N(0, 1)$ *random variables, and* $\lambda_1 \geq \cdots \geq \lambda_m$ *are the eigenvalues of* Σ*.*

Proof. The proof is based on the Lemma 4.1 of Menéndez et al. (1998) which states that for any random stochastic m-vectors $\boldsymbol{s}_n, \boldsymbol{t}_n$ and all functions $\phi \in \Phi$ it holds

$$T_n^\phi(\boldsymbol{s}_n, \boldsymbol{t}_n) = \chi_{nP}^2(\boldsymbol{s}_n, \boldsymbol{t}_n) + o_P(1)$$

provided that the conditions

$$\|\boldsymbol{s}_n - \boldsymbol{t}_n\| = O_P(n^{-1/2})$$

and

$$\Pi(\boldsymbol{t}_{n_\ell}) = o_P(1) \quad \text{for no subsequence } \boldsymbol{t}_{n_\ell} \text{ of } \boldsymbol{t}_n, \tag{12}$$

where $\Pi(\boldsymbol{t}_n) = \prod_{k=1}^m t_{nk}$, are satisfied.

Since the validity of (12) for $\boldsymbol{t}_n = \boldsymbol{p}(\widehat{\theta})$ follows directly from (8) and (11), to prove the assertion we need to check the condition

$$\|\widehat{\boldsymbol{p}} - \boldsymbol{p}(\widehat{\theta})\| = O_P(n^{-1/2})$$

and apply Lemma 4.1 of Menéndez et al. (1998) and Lemma 1 of the present paper. From the definition of $\xi_n, \widehat{\boldsymbol{p}}$ and $\boldsymbol{p}(\widehat{\theta})$ it follows that $\widehat{p}_k - p_k(\widehat{\theta}) = \xi_{nk}/n$, $k = 1, \ldots, m$, and thus

$$\sqrt{n}\|\widehat{\boldsymbol{p}} - \boldsymbol{p}(\widehat{\theta})\| = \left(n \sum_{k=1}^m (\widehat{p}_k - p_k(\widehat{\theta}))^2 \right)^{\frac{1}{2}} = \left(\frac{1}{n} \sum_{k=1}^m \xi_{nk}^2 \right)^{\frac{1}{2}} = \left(\frac{1}{n} \xi_n^t \xi_n \right)^{\frac{1}{2}}.$$

As $(1/n)\xi_n^t \xi_n = \chi_{nJ}^2(\widehat{\boldsymbol{p}}, \boldsymbol{p}(\widehat{\theta}))$ is the Jiang statistics which has under the assumed conditions the asymptotic distribution stated in Theorem 1 of Jiang (2001) and is thus $O_P(1)$, the proof is finished. □

Let us note that to use the class of statistics T_n^ϕ for testing, the eigenvalues $\lambda_1, \ldots, \lambda_m$ as well as the stochastic vector $\boldsymbol{q} = (q_1, \ldots, q_m)$ have to be replaced by their estimators. From (11) it follows that $\boldsymbol{p}(\widehat{\theta}) = (1/n) \sum_{j=1}^n \boldsymbol{p}_j(\widehat{\theta})$ is a consistent estimator of the vector $\boldsymbol{q}$. If we denote $\widehat{\lambda}_{n1}, \ldots, \widehat{\lambda}_{nm}$ the eigenvalues of $\widehat{\Sigma}_n = \Sigma_n(\widehat{\theta})$ then, by Weyl's eigenvalue perturbation theorem (e.g. Bhatia (1997)), $|\widehat{\lambda}_{nk} - \lambda_{nk}| \leq \|\Sigma_n(\widehat{\theta}) - \Sigma_n(\theta)\|$ which can be expected to go to 0 since $\widehat{\theta}$ is consistent. By the same theorem it can be seen also that $\lambda_{nk} \to \lambda_k$ and so $\widehat{\lambda}_{nk}$ is a consistent estimator of λ_k, $k = 1, \ldots, m$. The following testing procedure can be thus proposed: Reject H_0 if T_n^ϕ exceeds the critical value of $\sum_{k=1}^m (\widehat{\lambda}_{nk}/p_k(\widehat{\theta}))\, Z_k^2$.

3. Bootstrap goodness-of-fit tests

The application of the Jiang statistic (5) and the T_n^ϕ statistics (6) to test the hypothesis (4) requires the use of their asymptotic distribution given in Theorem 1 of Jiang (2001) and Theorem 1 of the present paper respectively. Practitioners will find the following difficulties in applying this approach: (1) in most cases, derivation of Σ_n is not straightforward and numerical computations may be needed, and (2) Σ_n is estimated with $\widehat{\Sigma}_n = \Sigma_n(\widehat{\theta})$ and $\Sigma_n(\widehat{\theta})$ is assumed to be close to $\Sigma_n(\theta)$. Therefore sample size should be large enough to fulfil the desired test size. Bootstrap tests avoid the mentioned difficulties because they only require the calculation of the test statistics in independent bootstrap samples and they approximate the required distribution under H_0.

Let $Y_1, \ldots, Y_n$ be random variables and let $F_{1\theta}, \ldots, F_{n\theta}$ be c.d.f. depending on a common parameter $\theta \in \Theta \subset R^d$ open. The hypothesis (4) under consideration is of the form

$$H_0 : Y_1 \sim F_{1\theta}, \ldots, Y_n \sim F_{n\theta} \quad \text{independent}, \theta \in \Theta.$$

Let $T_n = T_n(Y_1, \ldots, Y_n)$ be a given test statistic for this problem and assume that H_0 is rejected if $T_n > c_n$ for a given critical value $c_n > 0$. Let $F_{T_n\theta}(x) = P_\theta^n(T_n \le x)$ be the distribution of T_n under H_0, where P_θ^n is the probability corresponding to the joint distribution $\prod_{j=1}^n F_{j\theta}$. Suppose that we have an estimator $\widehat{\theta}$ of θ such that $\widehat{\theta}$ is consistent under H_0 in the sense that

$$P_\theta^n \left(\|\widehat{\theta} - \theta\| > \varepsilon \right) \underset{n\to\infty}{\longrightarrow} 0, \quad \text{for any } \varepsilon > 0.$$

Assuming that $F_{T_n\theta}$ is continuous a bootstrap estimator of c_n is

$$\widehat{c}_n = F^{-1}_{T_n\widehat{\theta}}(1-\alpha),$$

where $\alpha \in (0, 1)$ is the size of the test. The computation of $\widehat{c}_n$ can be done by Monte Carlo simulation in the following way. Generate B independent bootstrap samples $\{Y^*_{1b}, \ldots, Y^*_{nb}\}$ from the joint distribution $\prod_{j=1}^n F_{j\widehat{\theta}}$. Then $\widehat{c}_n$ is approximated by the $\{[(1-\alpha)B]+1\}$th order statistic of $T_n(Y^*_{1b}, \ldots, Y^*_{nb})$, $b = 1, \ldots, B$.

Alternatively bootstrap estimated p-value can be used to decide if H_0 is rejected or not. Let $Y_1 = y_1, \ldots, Y_n = y_n$ be the observed values. For the test of the form $T_n > c$, its p-value is defined by

$$p_n = P_\theta^n(T_n(Y_1, \ldots, Y_n) > T_n(y_1, \ldots, y_n)),$$

and hypothesis is rejected if $p_n < \alpha$. A bootstrap estimator of p_n is

$$\widehat{p}_n = P_*(T_n(Y_1^*, \ldots, Y_n^*) > T_n(y_1, \ldots, y_n))$$

where $Y_1^* \sim F_{1\widehat{\theta}}, \ldots, Y_n^* \sim F_{n\widehat{\theta}}$ are the bootstrap independent data. The computation of $\widehat{p}_n$ can be done by Monte Carlo simulation in the following way. Generate B independent bootstrap samples $\{Y^*_{1b}, \ldots, Y^*_{nb}\}$ from the joint distribution $\prod_{j=1}^n F_{j\widehat{\theta}}$. Then $\widehat{p}_n$ is approximated by

$$\widehat{p}_n = \frac{\#\left(T_n(y_1^*, \ldots, y_n^*) > T_n(y_1, \ldots, y_n)\right)}{B}.$$

This approach is also used in Section 4 to calculate the p-value

$$p_n = P\left(\sum_{k=1}^m (\widehat{\lambda}_{nk}/p_k(\widehat{\theta}))Z_k^2 > t\right)$$

when $T_n^\phi = t$ has been observed.

4. Example and simulation

This section contains an example that illustrates results and proposals of Sections 2 and 3, as well as a simulation study designed to investigate the performance of several test statistics. Let us consider the linear model

$$H_0 : y_j = \beta x_j + e_j, \quad j = 1, \ldots, n, \tag{13}$$

with e_j i.i.d. $N(0, \sigma^2)$. Let $\theta = (\beta, \sigma^2)$ be the unknown parameter and let

$$\widehat{\beta} = \frac{\sum_{j=1}^n y_j x_j}{\sum_{j=1}^n x_j^2}, \quad \widehat{\sigma}^2 = \frac{1}{n-1}\sum_{j=1}^n (y_j - x_j\widehat{\beta})^2.$$

be the corresponding maximum likelihood estimators. Consider the interval partition defined by the cut points

$$c_1 = 1 + F^{-1}_{N(0,1)}(1/m), \dots, c_{m-1} = 1 + F^{-1}_{N(0,1)}((m-1)/m),$$

i.e. $E_1 = (-\infty, c_1]$, $E_m = (c_{m-1}, \infty]$ and $E_k = (c_{k-1}, c_k]$, $k = 2, \dots, m-1$.

Let $E_{\theta k} = E_\theta[N_k] = \sum_{j=1}^n p_{jk}$ and define

$$\Sigma_{kk} = \frac{1}{n}\sum_{j=1}^n \text{var}(h_{njk}), \qquad \Sigma_{k_1k_2} = \frac{1}{n}\sum_{j=1}^n \text{cov}(h_{njk_1}, h_{njk_2}), \quad k_1 \neq k_2, \Sigma = \left(\Sigma_{k_1k_2}\right)_{k_1,k_2=1,\dots,m},$$

for the h_{njk}'s introduced in Section 2. Let $\lambda_1, \dots, \lambda_m$ be the eigenvalues of $A = n\,\text{diag}(E^{-1}_{\widehat{\theta}1}, \dots, E^{-1}_{\widehat{\theta}m})\widehat{\Sigma}$, then

$$T_n^r(\widehat{\boldsymbol{p}}, \boldsymbol{p}(\widehat{\theta})) \sim \sum_{k=1}^m \lambda_k Z_k^2,$$

where $Z_1, \dots, Z_m$ are i.i.d. $N(0, 1)$.

Regarding the introduced example a simulation experiment has been implemented to analyze the performance of Jiang and Cressie–Read statistics

$$\chi^2_{nJ} = n\sum_{k=1}^m (\widehat{p}_k - p_k(\widehat{\theta}))^2, \qquad T_n^r = \frac{2n}{r(r+1)}\sum_{k=1}^m \widehat{p}_k\left[\left(\frac{\widehat{p}_k}{p_k(\widehat{\theta})}\right)^r - 1\right], \quad r = -1/2, 0, 2/3, 1.$$

For every considered test statistics, T_n, the simulation follows the next steps.

1. Repeat $I = 10000$ times ($i = 1, \dots, I$)
   - 1.1. Generate a sample $(y_j^{(i)}, x_j^{(i)})$, $j = 1, \dots, n$, from model (13) with $\beta = 1$, $\sigma^2 = 1$ and $x_j^{(i)}$ i.i.d. Unif(0, 2). Calculate $\widehat{\beta}^{(i)}, \widehat{\sigma}^{2(i)}, \widehat{\lambda}_1, \dots, \widehat{\lambda}_m$ and $T_n^{(i)}$.
   - 1.2. Simulate $v_1, \dots, v_A$ from $\sum_{i=1}^m \widehat{\lambda}_i Z_i^2$, with $Z_1, \dots, Z_m$ i.i.d $N(0, 1)$ and $A = 5000$. Calculate
   $$p_n^{(i)} = \frac{\#\{v_\ell : v_\ell \geq T_n^{(i)}\}}{A} \quad \text{and} \quad \xi_n^{(i)} = \begin{cases} 1 & \text{if } p_n^{(i)} < 0.05 \\ 0 & \text{otherwise.} \end{cases}$$
   - 1.3. Repeat $B = 1000$ times ($b = 1, \dots, B$)
     - 1.3.1. Generate $e_j^{*(ib)} \sim N(0, \widehat{\sigma}^{2(i)})$, $j = 1, \dots, n$. Generate a bootstrap sample $(y_j^{*(ib)}, x_j^{(i)})$, $j = 1, \dots, n$, from model $y_j^{*(ib)} = \widehat{\beta}^{(i)} x_j^{(i)} + e_j^{*(ib)}$.
     - 1.3.2. Calculate $\widehat{\beta}^{*(ib)}, \widehat{\sigma}^{2*(ib)}$ and $T_n^{*(ib)}$.
   - 1.4. Calculate
   $$\alpha_n^{*(i)} = \frac{\#(T_{n,}^{*(ib)} \geq T_{n,}^{(i)})}{B} \quad \text{and} \quad \xi_n^{*(i)} = \begin{cases} 1 & \text{if } \alpha_{n,}^{*(i)} < 0.05 \\ 0 & \text{otherwise.} \end{cases}$$
2. Output:

$$\xi_n = \frac{1}{I}\sum_{i=1}^I \xi_n^{(i)}, \qquad \xi_n^* = \frac{1}{I}\sum_{i=1}^I \xi_n^{*(i)}.$$

It should occur that both ξ_n and ξ_n^* are close to 0.05. In Table 1 test sizes of bootstrap and asymptotic tests are given. We observe that bootstrap tests attain the desired size even for small sample sizes ($n = 40$), where some asymptotic tests fails. To be sure that asymptotic distribution works properly under the null hypothesis, sample size should not be much lower than 100. At this point it is worthwhile to emphasize that the asymptotic distribution is in fact also approximated in some sense because eigenvalues are calculated from the estimated matrix $\widehat{\Sigma}_n$ and not from Σ.

Powers are calculated, and presented in Tables 2–4, for the following alternatives to (13):

1. $y_j = g_a(\beta x_j) + e_j$, with $g_a(x) = x^a$ and a varying from 0 to 2.5,
2. $e_j \sim (1-p)N(0, \sigma_1^2) + p\,Gumbel(0, \sigma_2^2)$, with $\sigma_1^2 = \sigma_2^2 = 1$ and $p = 0, 0.2, 0.5, 0.8, 1$,
3. $y_j = \sum_{i=1}^\kappa \beta_i x_j^i + e_j$, with $\beta_1 = \dots, \beta_5 = 1$ and $\kappa = 1, \dots, 5$.

Table 1
Test sizes for $\alpha = 0.05$ (asymptotic | bootstrap)

n	$-1/2$	0	2/3	1	Jiang	$-1/2$	0	2/3	1	Jiang
40	.1000	.0683	.0457	.0462	.0440	.0512	.0537	.0548	.0547	.0547
100	.0571	.0500	.0444	.0445	.0458	.0539	.0534	.0524	.0527	.0533
200	.0584	.0545	.0511	.0521	.0518	.0498	.0498	.0493	.0495	.0507
500	.0563	.0543	.0530	.0531	.0533	.0525	.0525	.0523	.0524	.0528
1000	.0537	.0517	.0515	.0516	.0509	.0512	.0506	.0503	.0502	.0510

Table 2
Powers for case 1, $\alpha = 0.05$ and $n = 200$ (asymptotic | bootstrap)

a	$-1/2$	0	2/3	1	Jiang	$-1/2$	0	2/3	1	Jiang
.0	.9408	.9349	.9278	.9242	.9822	.9397	.9397	.9359	.9326	.9844
.2	.6188	.6063	.5919	.5895	.7629	.6079	.6097	.6068	.6029	.7714
.4	.2765	.2672	.2657	.2664	.3824	.2646	.2724	.2802	.2818	.3937
.6	.1234	.1212	.1224	.1266	.1588	.1100	.1152	.1213	.1243	.1587
.8	.0705	.0674	.0685	.0694	.0749	.0640	.0680	.0702	.0722	.0759
1.0	.0584	.0545	.0511	.0521	.0518	.0498	.0498	.0493	.0495	.0507
1.2	.0725	.0654	.0603	.0592	.0602	.0624	.0617	.0608	.0602	.0635
1.4	.1075	.1007	.0944	.0937	.1035	.1052	.1053	.1011	.1006	.1130
1.6	.1985	.1912	.1881	.1887	.2182	.1889	.1945	.1970	.1998	.2314
1.8	.3652	.3634	.3712	.3796	.4470	.3375	.3562	.3768	.3855	.4521
2.0	.5822	.5918	.6119	.6282	.7557	.5413	.5773	.6120	.6253	.7627
2.5	.9848	.9873	.9904	.9921	.9998	.9779	.9852	.9896	.9910	.9999

Table 3
Powers for case 2, $\alpha = 0.05$ and $n = 100$ (asymptotic | bootstrap)

p	$-1/2$	0	2/3	1	Jiang	$-1/2$	0	2/3	1	Jiang
.0	.0571	.0500	.0444	.0445	.0458	.0539	.0534	.0524	.0527	.0533
.2	.0777	.0650	.0579	.0559	.0618	.0647	.0636	.0612	.0598	.0633
.5	.1622	.1313	.1160	.1134	.1630	.1255	.1337	.1358	.1336	.1656
.8	.3265	.1967	.1200	.1018	.1709	.2860	.2110	.1470	.1304	.1737
1	.4518	.3140	.1852	.1543	.1667	.3936	.3057	.2058	.1718	.1701

Table 4
Powers for case 3, $\alpha = 0.05$ and $n = 40$ (asymptotic | bootstrap)

κ	$-1/2$	0	2/3	1	Jiang	$-1/2$	0	2/3	1	Jiang
1	.1000	.0683	.0457	.0462	.0440	.0512	.0537	.0548	.0547	.0547
2	.2610	.1249	.1239	.1479	.2516	.0425	.0800	.1371	.1522	.2588
3	.4659	.4799	.6483	.7290	.9529	.0989	.4996	.6941	.7142	.9554
4	.7746	.8600	.9525	.9740	.9868	.7030	.9312	.9577	.9582	.9835
5	.9004	.9622	.9928	.9967	.9764	.9713	.9916	.9931	.9915	.9746

One can conclude that Jiang's test statistic has an excellent performance in relation with the more classical power-divergence statistics. Comparing the Cressie–Read statistics no dramatic differences were observed. Just in the case 2 the Freeman–Tukey statistic ($r = -1/2$) seems to have the best behavior in the sense of powers, in this case even better than Jiang's statistic.

References

Ali, S.M., Silvey, S.D., 1966. A general class of coefficient of divergence of one distribution from another. Journal of the Royal Statistical Society, Series B 286, 131–142.

Bhatia, R., 1997. Matrix Analysis. Springer, New York.

Birch, M.W., 1964. A new proof of the Pearson–Fisher Theorem. Annals of Mathematical Statistics 35, 817–824.

Chernoff, H., Lehmann, E.L., 1954. The use of maximum likelihood estimates in χ^2 tests of goodness of fit. Annals of Mathematical Statistics 25, 579–586.

Csiszár, I., 1963. Eine Informationstheoretische Ungleichung und ihre Anwendung auf den Beweis der Ergodizität von Markoffschen Ketten. Publications of the Mathematical Institute of Hungarian Academy of Sciences, Series A 8, 85–108.

Cressie, N.A.C., Read, T.R.C., 1984. Multinomial goodness of fit tests. Journal of the Royal Statistical Society, Series B 46, 440–464.

Jiang, J., 2001. A nonstandard χ^2-test with application to generalized linear model diagnostics. Statistics and Probability Letters 53, 101–109.

Liese, F., Vajda, I., 1987. Convex Statistical Distances. Teubner, Leipzig.

Morales, D., Pardo, L., Vajda, I., 1996. Divergence between various estimates of quantized information sources. Kybernetika 32 (4), 395–407.

Menéndez, M.L., Morales, D., Pardo, L., Vajda, I., 1998. Two approaches to grouping of data and related disparity statistics. Communications in Statistics—Theory and Methods 27 (3), 609–633.

Vajda, I., 1989. Theory of Statistical Inference and Information. Kluwer, Dordrecht.

Zografos, K., Ferentinos, K., Papaioannou, T., 1990. ϕ-divergence statistics: Sampling properties, multinomial goodness of fit and divergence tests. Communications in Statistics — Theory and Methods 19, 1785–1802.