Ann Inst Stat Math (2013) 65:319–348
DOI 10.1007/s10463-012-0372-y

Testing statistical hypotheses based on the density power divergence

A. Basu · A. Mandal · N. Martin · L. Pardo

Received: 22 August 2011 / Revised: 21 March 2012 / Published online: 20 July 2012

Abstract The family of density power divergences is an useful class which generates robust parameter estimates with high efficiency. None of these divergences require any non-parametric density estimate to carry out the inference procedure. However, these divergences have so far not been used effectively in robust testing of hypotheses. In this paper, we develop tests of hypotheses based on this family of divergences. The asymptotic variances of the estimators are generally different from the inverse of the Fisher information matrix, so that the usual drop-in-divergence type statistics do not lead to standard Chi-square limits. It is shown that the alternative test statistics proposed herein have asymptotic limits which are described by linear combinations of Chi-square statistics. Extensive simulation results are presented to substantiate the theory developed.

Keywords Density power divergence · Linear combination of Chi-squares · Robustness · Tests of hypotheses

1 Introduction

Let $\mathcal{F} = \{F_{\boldsymbol{\theta}} : \boldsymbol{\theta} \in \Theta \subset \mathbb{R}^p\}$ be a parametric family which models the distribution of the random variable X of interest.

A. Basu · A. Mandal
Indian Statistical Institute, Kolkata 700108, India

N. Martin (✉)
Department of Statistics, Carlos III University of Madrid, 28903 Getafe (Madrid), Spain
e-mail: nmapaola@est-econ.uc3m.es

L. Pardo
Department of Statistics and O.R., Complutense University of Madrid, 28040 Madrid, Spain

We assume that the densities $f_{\boldsymbol{\theta}}(x)$ of the probability measures $P_{\boldsymbol{\theta}}$ exist with respect to a dominating measure μ. Let $X_1, \ldots, X_n$ be a random sample from the distribution of X, and suppose that we are interested in testing parametric hypothesis about $\boldsymbol{\theta}$. It has been shown (see Simpson 1989; Lindsay 1994; Pardo 2006, Chapter 9; Basu et al. 2011, Chapter 5) that test statistics based on ϕ-divergence measures or disparities provide useful robust alternatives to the classical tests. In continuous models, however, these procedures suffer from the drawback that it is necessary to use some non-parametric density estimation technique—such as kernel density estimation—to generate a continuous density estimate of the unknown population density; thus, the procedure inherits all the associated complications and difficulties of the kernel density estimation method such as bandwidth selection. Handling continuous models with bounded support (or at least bounded at one end) becomes even more difficult in such cases because the kernel requires further modification to preserve the support. In fact the associated complications become so overwhelming that one rarely sees, if ever, the application of the minimum ϕ-divergence method for estimation and hypotheses testing for any continuous model other than the normal.

Basu et al. (1998) introduced a new family of density-based divergence measures called the density power divergence family; a single parameter β controls the trade-off between robustness and asymptotic efficiency of the parameter estimates which are the minimizers of this family of divergences. When $\beta = 0$, the density power divergence is the Kullback–Leibler divergence and the corresponding minimizer is the maximum likelihood estimator of $\boldsymbol{\theta}$. For positive values of β, the corresponding minimum distance procedure is substantially more robust, with the degree of robustness increasing with β. All minimum distance estimators within this class have bounded influence functions except for the case when $\beta = 0$. The remarkable thing about this family is that none of its members requires any non-parametric density estimation procedure for the minimization routine. Although minimum density power divergence estimation provides excellent robust estimators in standard parametric models, this family of divergences has not been, so far, used successfully in parametric hypothesis testing problems. This is partly because the asymptotic variance of ($\sqrt{n}$ times) the minimum density power divergence estimator is different from the Fisher information matrix (except in the case $\beta = 0$), and the usual analogues of the disparity difference tests (e.g. Lindsay 1994) do not have standard Chi-squared limits in this case. Here, we consider alternative tests of hypothesis based on the density power divergence. The distribution of our statistics can asymptotically be described as linear combinations of independent Chi-square variables. These tests provide excellent robust alternatives to the likelihood ratio test and the tests based on ϕ-divergences. The construction of these test statistics do not require any non-parametric smoothing in any parametric model.

Methods based on maximum likelihood form the backbone of the theory of statistical inference. However, it is well known that under this method even small deviations from the assumed conditions can have a substantial undesirable impact on the inference procedure and the final conclusions. We allow the possibility of the true (data generating) distribution G being in a small neighborhood of the assumed parametric model $\mathcal{F}$, rather than being strictly inside it. The best fitting parameter in this case corresponds to the model element nearest to the true distribution, where the measure of closeness is in terms of the divergence in question. “Robust” divergences, such as those corresponding to moderate and large values of the tuning parameter β within the density power divergence family, can effectively downweight small deviations from the model so that the parametric modeling often provides a good fit to the large majority of the data while leaving out a small proportion of outlying observations. This is in contrast with the methods based on maximum likelihood, which try to fit all the observations in such cases, and end up providing a poor fit for all. In similar spirit, the tests of hypotheses based on robust divergences can provide meaningful tests for the parameter of the larger, dominant, component represented by the majority of the data, while essentially eliminating the small deviations; these tests largely preserve their level and power in comparison to the quantities which one would obtain under pure data. However, the likelihood ratio test can experience significant magnification in the level or loss in power, depending on the situation, for such problems.

It is worth mentioning that although the test statistics proposed in this paper are similar in spirit to the disparity difference (or generally, drop-in-divergence) type tests, they are different from the latter in one significant way. The test statistics depend on the data only through the parameter estimates; in this development, these estimates are the minimum density power divergence estimates of the parameters. Thus, the robustness of the test statistics are directly linked to the robustness of the estimators, which has already been discussed in the literature. Also, in our numerical studies, the model misspecifications are envisaged to be in the form of mixture contaminations; this is the most common set up to study the robustness of proposed methods. Model violations of other natures could also possibly be studied, but we believe that could be the topic of a different article.

The rest of the paper is organized as follows. In Sect. 2, we discuss the density power divergence family and the resulting parametric estimation based on this divergence. Tests of hypothesis based on the density power divergence are introduced in Sect. 3. The different cases under consideration and relevant theory are developed in this section. Numerical illustrations of the performance of these test statistics are presented in Sect. 4 and 5. Section 6 has some concluding remarks.

Throughout the paper, we will make the standard assumptions about asymptotic inference as given by Assumptions A, B, C and D of Lehmann (1983, p. 429). We will refer to them as the Lehmann conditions.

2 The density power divergence and parametric estimation

2.1 The divergence

The density power divergence family (Basu et al. 1998) represents a rich class of density-based divergences. Let $\mathcal{G}$ denote the set of all distributions having densities with respect to the dominating measure. Given densities $g, f \in \mathcal{G}$ the density power divergence between them is defined, as a function of a non-negative parameter β, as

$$d_\beta(g, f) = \begin{cases} \int \left\{ f^{1+\beta}(x) - \left(1 + \frac{1}{\beta}\right) f^\beta(x) g(x) + \frac{1}{\beta} g^{1+\beta}(x) \right\} \mathrm{d}x, & \text{for } \beta > 0, \\ \int g(x) \log\left(\frac{g(x)}{f(x)}\right) \mathrm{d}x, & \text{for } \beta = 0. \end{cases} \quad (1)$$

The case corresponding to $\beta = 0$ may be derived from the general case by taking the continuous limit as $\beta \to 0$. The quantities defined in Eq. (1) are genuine divergences in the sense $d_\beta(g, f) \geq 0$ for all $g, f \in \mathcal{G}$ and all $\beta \geq 0$, and $d_\beta(g, f)$ is equal to zero if and only if the densities g and f are identically equal.

We consider the parametric model of densities $\{f_{\boldsymbol{\theta}} : \boldsymbol{\theta} \in \Theta \subset \mathbb{R}^p\}$; suppose we are interested in the estimation of $\boldsymbol{\theta}$. Let G represent the distribution function corresponding to the density g. The minimum density power divergence functional $T_\beta(G)$ at G is defined by the requirement $d_\beta(g, f_{T_\beta(G)}) = \min_{\boldsymbol{\theta}\in\Theta} d_\beta(g, f_{\boldsymbol{\theta}})$. Clearly the term $\int g^{1+\beta}(x)\mathrm{d}x$ has no role in the minimization of $d_\beta(g, f_{\boldsymbol{\theta}})$ over $\boldsymbol{\theta} \in \Theta$. Thus, the essential objective function to be minimized in the computation of the minimum density power divergence functional $T_\beta(G)$ reduces to

$$\int \left\{ f_{\boldsymbol{\theta}}^{1+\beta}(x) - \left(1 + \frac{1}{\beta}\right) f_{\boldsymbol{\theta}}^{\beta}(x) g(x) \right\} \mathrm{d}x = \int f_{\boldsymbol{\theta}}^{1+\beta}(x)\mathrm{d}x - \left(1 + \frac{1}{\beta}\right) \int f_{\boldsymbol{\theta}}^{\beta}(x) dG(x).$$

Notice that in the above objective function, the density g appears only as a linear term (unlike, say, the computation of the minimum Hellinger distance functional where the square root of the density g is the relevant quantity). Thus, given a random sample $X_1, \ldots, X_n$ from the distribution G, we can approximate the above objective function by replacing G with its empirical estimate G_n. For a given tuning parameter β, therefore, the minimum density power divergence estimator $\widehat{\boldsymbol{\theta}}_\beta$ of $\boldsymbol{\theta}$ can be obtained by minimizing

$$\begin{aligned} \int f_{\boldsymbol{\theta}}^{1+\beta}(x)\mathrm{d}x - \left(1 + \frac{1}{\beta}\right) \int f_{\boldsymbol{\theta}}^{\beta}(x) dG_n(x) &= \int f_{\boldsymbol{\theta}}^{1+\beta}(x)\mathrm{d}x - \left(1 + \frac{1}{\beta}\right) \frac{1}{n} \sum_{i=1}^{n} f_{\boldsymbol{\theta}}^{\beta}(X_i) \\ &= \frac{1}{n} \sum_{i=1}^{n} V_{\boldsymbol{\theta}}(X_i) \end{aligned} \tag{2}$$

over $\boldsymbol{\theta} \in \Theta$, where $V_{\boldsymbol{\theta}}(x) = \int f_{\boldsymbol{\theta}}^{1+\beta}(y)\mathrm{d}y - \left(1 + \frac{1}{\beta}\right) f_{\boldsymbol{\theta}}^{\beta}(x)$. The minimization of the above expression over $\boldsymbol{\theta}$ does not require the use of a non-parametric density estimate. Existing theory (e.g. De Angelis and Young 1992) shows that in general there is little or no advantage in introducing smoothing for such functionals which may be empirically estimated using the empirical distribution function alone, except in very special cases. Using G_n to substitute G, if possible, is, therefore, a natural step.

Let $u_{\boldsymbol{\theta}}(x) = \frac{\partial}{\partial \boldsymbol{\theta}} \log f_{\boldsymbol{\theta}}$ be the score function of the model. Under differentiability of the model, the maximization of the objective function in Eq. (2) leads to an estimating equation of the form

$$\frac{1}{n} \sum_{i=1}^{n} u_{\boldsymbol{\theta}}(X_i) f_{\boldsymbol{\theta}}^{\beta}(X_i) - \int u_{\boldsymbol{\theta}}(x) f_{\boldsymbol{\theta}}^{1+\beta}(x)\mathrm{d}x = 0, \tag{3}$$

which is an unbiased estimating equation under the model. Since the corresponding estimating equation weights the score $u_{\boldsymbol{\theta}}(X_i)$ with the power of the density $f_{\boldsymbol{\theta}}^{\beta}(X_i)$, the outlier resistant behavior of the estimator is intuitively apparent. See Basu et al. (1998) and Jones et al. (2001) for more details.

The functional $T_\beta(G)$ is Fisher consistent; it takes the value $\boldsymbol{\theta}$ when the true density $g = f_{\boldsymbol{\theta}}$ is in the model. When it does not, $\boldsymbol{\theta}_\beta^g = T_\beta(G)$ represents the best fitting parameter. For brevity we will suppress the β subscript in the notation for $\boldsymbol{\theta}_\beta^g$; then $f_{\boldsymbol{\theta}^g}$ is the model element closest to g in the density power divergence sense.

2.2 The asymptotic distribution of the minimum density power divergence estimator

Let g be the true data generating density and $\boldsymbol{\theta}^g = T_\beta(G)$ be the best fitting parameter. We define

$$\begin{aligned} \boldsymbol{J}_\beta(\boldsymbol{\theta}) = \int u_{\boldsymbol{\theta}}(x) u_{\boldsymbol{\theta}}^T(x) f_{\boldsymbol{\theta}}^{1+\beta}(x)\mathrm{d}x + \int \{ \boldsymbol{i}_{\boldsymbol{\theta}}(x) - \beta u_{\boldsymbol{\theta}}(x) u_{\boldsymbol{\theta}}^T(x)\} \\ \times \{g(x) - f_{\boldsymbol{\theta}}(x)\} f_{\boldsymbol{\theta}}^\beta(x)\mathrm{d}x \end{aligned} \tag{4}$$

and

$$\boldsymbol{K}_\beta(\boldsymbol{\theta}) = \int u_{\boldsymbol{\theta}}(x) u_{\boldsymbol{\theta}}^T(x) f_{\boldsymbol{\theta}}^{2\beta}(x) g(x)\mathrm{d}x - \xi_\beta(\boldsymbol{\theta})\xi_\beta^T(\boldsymbol{\theta}), \tag{5}$$

where $\xi_\beta(\boldsymbol{\theta}) = \int u_{\boldsymbol{\theta}}(x) f_{\boldsymbol{\theta}}^\beta(x) g(x)\mathrm{d}x$, and $\boldsymbol{i}_{\boldsymbol{\theta}}(x) = -\frac{\partial}{\partial \boldsymbol{\theta}} u_{\boldsymbol{\theta}}(x)$, the so-called information function of the model.

For the rest of the paper, we will assume the conditions D1–D5 of Basu et al. (2011, p. 304) which we will refer to as the Basu et al. conditions. The following results are then available about the asymptotic distribution of the minimum density power divergence estimators:

(a) The minimum density power divergence estimating Eq. (3) has a consistent sequence of roots $\widehat{\boldsymbol{\theta}}_\beta = \widehat{\boldsymbol{\theta}}_n$.

(b) $n^{1/2}(\hat{\boldsymbol{\theta}}_\beta - \boldsymbol{\theta}^g)$ has an asymptotic multivariate normal distribution with (vector) mean zero and covariance matrix $\boldsymbol{J}^{-1}\boldsymbol{K}\boldsymbol{J}^{-1}$, where $\boldsymbol{J} = \boldsymbol{J}_\beta(\boldsymbol{\theta}^g)$, $\boldsymbol{K} = \boldsymbol{K}_\beta(\boldsymbol{\theta}^g)$, and $\boldsymbol{J}_\beta$ and $\boldsymbol{K}_\beta$ are as in (4) and (5) respectively; see Basu et al. (1998) and Basu et al. (2011).

When the true distribution G belongs to the model so that $G = F_{\boldsymbol{\theta}}$ for some $\boldsymbol{\theta} \in \Theta$, the formula for $\boldsymbol{J} = \boldsymbol{J}_\beta(\boldsymbol{\theta})$, $\boldsymbol{K} = \boldsymbol{K}_\beta(\boldsymbol{\theta})$ and $\xi = \xi_\beta(\boldsymbol{\theta})$ simplify to

$$\begin{aligned} \boldsymbol{J} = \int u_{\boldsymbol{\theta}}(x) u_{\boldsymbol{\theta}}^T(x) f_{\boldsymbol{\theta}}^{1+\beta}(x)\mathrm{d}x, \quad \boldsymbol{K} = \int u_{\boldsymbol{\theta}}(x) u_{\boldsymbol{\theta}}^T(x) f_{\boldsymbol{\theta}}^{1+2\beta}(x)\mathrm{d}x - \xi\xi^T, \\ \xi = \int u_{\boldsymbol{\theta}}(x) f_{\boldsymbol{\theta}}^{1+\beta}(x)\mathrm{d}x. \end{aligned} \tag{6}$$

3 Testing parametric hypotheses using the density power divergence

In this section, we will develop tests of parametric hypothesis based on the density power divergence family. As the usual drop-in-divergence type statistics constructed with the density power divergence do not lead to standard Chi-square limits, we consider an alternative test statistic based on the density power divergence. To keep a clear focus in our presentations, we consider the following specific cases:

(a) The one sample problem described by the hypotheses

$$H_0 : \boldsymbol{\theta} = \boldsymbol{\theta}_0 \text{ against } H_1 : \boldsymbol{\theta} \neq \boldsymbol{\theta}_0. \tag{7}$$

when a random sample of size n is available from the population of interest.

(b) The two-sample problem described by the hypotheses

$$H_0 : \boldsymbol{\theta}_1 = \boldsymbol{\theta}_2 \text{ against } H_1 : \boldsymbol{\theta}_1 \neq \boldsymbol{\theta}_2,$$

where random samples of size m and n are available from two different populations, and $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$ are the parameters of the model density describing the two different populations.

3.1 The one sample problem

We consider a parametric family of densities $\{f_{\boldsymbol{\theta}} : \boldsymbol{\theta} \in \Theta \subseteq \mathbb{R}^p\}$ and a random sample $X_1, \ldots, X_n$ of size n from the population. We denote by $\boldsymbol{\theta}_0$ a known value of the parameter and our interest is in testing $H_0 : \boldsymbol{\theta} = \boldsymbol{\theta}_0$ against $H_1 : \boldsymbol{\theta} \neq \boldsymbol{\theta}_0$. There is no uniformly most powerful test for the problem under consideration for most parametric models. When the model is correctly specified and the null hypothesis is correct, $f_{\boldsymbol{\theta}_0}$ is the data generating density. In the next theorem, we shall obtain the asymptotic distribution of our proposed test statistic

$$\begin{aligned} T_\gamma(\widehat{\boldsymbol{\theta}}_\beta, \boldsymbol{\theta}_0) &= 2n d_\gamma(f_{\widehat{\boldsymbol{\theta}}_\beta}, f_{\boldsymbol{\theta}_0}), \\ d_\gamma(f_{\widehat{\boldsymbol{\theta}}_\beta}, f_{\boldsymbol{\theta}_0}) &= \begin{cases} \int \left(f_{\boldsymbol{\theta}_0}^{1+\gamma}(x) - \left(1 + \frac{1}{\gamma}\right) f_{\boldsymbol{\theta}_0}^{\gamma}(x) f_{\widehat{\boldsymbol{\theta}}_\beta}(x) + \frac{1}{\gamma} f_{\widehat{\boldsymbol{\theta}}_\beta}^{1+\gamma}(x) \right) \mathrm{d}x, & \text{for } \gamma > 0 \\ \int f_{\widehat{\boldsymbol{\theta}}_\beta}(x) \log \left(\frac{f_{\widehat{\boldsymbol{\theta}}_\beta}(x)}{f_{\boldsymbol{\theta}_0}(x)} \right) \mathrm{d}x, & \text{for } \gamma = 0, \end{cases} \end{aligned} \tag{8}$$

under the null hypothesis H_0, where $\widehat{\boldsymbol{\theta}}_\beta$ is the minimum density power divergence estimator of $\boldsymbol{\theta}$. Observe that the test statistic defined in (8) depends on two different tuning parameters β and γ. The density power divergence associated with the parameter β is used for estimating the unknown parameters and the divergence associated with the parameter γ is used for obtaining the test statistic to test the hypotheses of interest.

Theorem 1 *Suppose that the model satisfies the Lehmann and the Basu et al. conditions. Under the null hypothesis $H_0: \boldsymbol{\theta} = \boldsymbol{\theta}_0$, the asymptotic distribution of $T_\gamma(\widehat{\boldsymbol{\theta}}_\beta, \boldsymbol{\theta}_0)$ coincides with the distribution of*

$$\sum_{i=1}^{r} \lambda_i^{\gamma,\beta}(\boldsymbol{\theta}_0) Z_i^2,$$

where $Z_1, \ldots, Z_r$ are independent standard normal variables, $\lambda_1^{\gamma,\beta}, \ldots, \lambda_r^{\gamma,\beta}$ are the non-zero eigenvalues of $\boldsymbol{A}_\gamma(\boldsymbol{\theta}_0) \boldsymbol{J}_\beta^{-1}(\boldsymbol{\theta}_0) \boldsymbol{K}_\beta(\boldsymbol{\theta}_0) \boldsymbol{J}_\beta^{-1}(\boldsymbol{\theta}_0)$, and the matrix $\boldsymbol{A}_\gamma(\boldsymbol{\theta}_0)$ is as defined later in (9), and

$$r = \operatorname{rank}\left(\boldsymbol{K}_\beta(\boldsymbol{\theta}_0) \boldsymbol{J}_\beta^{-1}(\boldsymbol{\theta}_0) \boldsymbol{A}_\gamma(\boldsymbol{\theta}_0) \boldsymbol{J}_\beta^{-1}(\boldsymbol{\theta}_0) \boldsymbol{K}_\beta(\boldsymbol{\theta}_0)\right).$$

Proof A second-order Taylor expansion of $d_\gamma(f_{\boldsymbol{\theta}}, f_{\boldsymbol{\theta}_0})$ around $\boldsymbol{\theta} = \boldsymbol{\theta}_0$ at $\boldsymbol{\theta} = \widehat{\boldsymbol{\theta}}_\beta$ gives,

$$\begin{aligned} d_\gamma(f_{\widehat{\boldsymbol{\theta}}_\beta}, f_{\boldsymbol{\theta}_0}) = d_\gamma(f_{\boldsymbol{\theta}_0}, f_{\boldsymbol{\theta}_0}) &+ \sum_{i=1}^{p} \left(\frac{\partial d_\gamma(f_{\boldsymbol{\theta}}, f_{\boldsymbol{\theta}_0})}{\partial \theta_i}\right)_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} \left(\widehat{\theta}_\beta^i - \theta_{0i}\right) \\ &+ \frac{1}{2} \sum_{i=1}^{p} \sum_{j=1}^{p} \left(\frac{\partial^2 d_\gamma(f_{\boldsymbol{\theta}}, f_{\boldsymbol{\theta}_0})}{\partial \theta_i \partial \theta_j}\right)_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} \left(\widehat{\theta}_\beta^i - \theta_{0i}\right) \left(\widehat{\theta}_\beta^j - \theta_{0j}\right) \\ &+ o\left(\left\|\widehat{\boldsymbol{\theta}}_\beta - \boldsymbol{\theta}_0\right\|^2\right), \end{aligned}$$

where the scripts denote the indicated components. Clearly $d_\gamma(f_{\boldsymbol{\theta}_0}, f_{\boldsymbol{\theta}_0}) = 0$ and $(\frac{\partial d_\gamma(f_{\boldsymbol{\theta}}, f_{\boldsymbol{\theta}_0})}{\partial \theta_i})_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} = 0$. Also

$$a_{ij}^\gamma(\boldsymbol{\theta}_0) = \left(\frac{\partial^2 d_\gamma(f_{\boldsymbol{\theta}}, f_{\boldsymbol{\theta}_0})}{\partial \theta_i \partial \theta_j}\right)_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} = (1+\gamma) \int_{\mathcal{X}} f_{\boldsymbol{\theta}_0}^{\gamma-1}(x) \left(\frac{\partial f_{\boldsymbol{\theta}}(x)}{\partial \theta_j} \frac{\partial f_{\boldsymbol{\theta}}(x)}{\partial \theta_i}\right)_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} \mathrm{d}x,$$

which proves that the asymptotic distributions of

$$T_\gamma(\widehat{\theta}_\beta, \theta_0) = 2n d_\gamma(f_{\widehat{\boldsymbol{\theta}}_\beta}, f_{\boldsymbol{\theta}_0}) \text{ and } n^{1/2}(\widehat{\boldsymbol{\theta}}_\beta - \boldsymbol{\theta}_0)^T \boldsymbol{A}_\gamma(\boldsymbol{\theta}_0) n^{1/2}(\widehat{\boldsymbol{\theta}}_\beta - \boldsymbol{\theta}_0)$$

are the same because

$$n \times o\left(\left\|\widehat{\boldsymbol{\theta}}_\beta - \boldsymbol{\theta}_0\right\|^2\right) = o_p(1).$$

The matrix $\boldsymbol{A}_\gamma(\boldsymbol{\theta}_0)$ is defined by

$$\boldsymbol{A}_\gamma(\boldsymbol{\theta}_0) = \left(a_{ij}^\gamma(\boldsymbol{\theta}_0)\right)_{i,j=1,\ldots,p}. \tag{9}$$

For $\boldsymbol{X} \sim N_q(\mathbf{0}, \boldsymbol{\Sigma})$, and a q dimensional real symmetric matrix A, the distribution of the quadratic form $\boldsymbol{X}^T \boldsymbol{A} \boldsymbol{X}$ is the same as that of $\sum_{i=1}^{r} \lambda_i Z_i^2$, where $Z_1, \ldots, Z_r$ are independent standard normal variables, $r = \text{rank}(\boldsymbol{\Sigma} \boldsymbol{A} \boldsymbol{\Sigma})$, $r \geq 1$ and $\lambda_1, \ldots, \lambda_r$ are the non-zero eigenvalues of $\boldsymbol{A}\boldsymbol{\Sigma}$ (Dik and de Ghunst 1985, Corollary 2.1). An application of the same establishes that the asymptotic distribution of $T_\gamma(\widehat{\boldsymbol{\theta}}_\beta, \boldsymbol{\theta}_0)$ is described by the random variable $\sum_{i=1}^{r} \lambda_i^{\gamma,\beta} Z_i^2$, where $\lambda_1^{\gamma,\beta}, \ldots, \lambda_r^{\gamma,\beta}$, are the non-zero eigenvalues of $\boldsymbol{A}_\gamma(\boldsymbol{\theta}_0) \boldsymbol{J}_\beta^{-1}(\boldsymbol{\theta}_0) \boldsymbol{K}_\beta(\boldsymbol{\theta}_0) \boldsymbol{J}_\beta^{-1}(\boldsymbol{\theta}_0)$, and

$$\begin{aligned} r &= \text{rank}\left(\boldsymbol{J}_\beta^{-1}(\boldsymbol{\theta}_0) \boldsymbol{K}_\beta(\boldsymbol{\theta}_0) \boldsymbol{J}_\beta^{-1}(\boldsymbol{\theta}_0) \boldsymbol{A}_\gamma(\boldsymbol{\theta}_0) \boldsymbol{J}_\beta^{-1}(\boldsymbol{\theta}_0) \boldsymbol{K}_\beta(\boldsymbol{\theta}_0) \boldsymbol{J}_\beta^{-1}(\boldsymbol{\theta}_0)\right) \\ &= \text{rank}\left(\boldsymbol{K}_\beta(\boldsymbol{\theta}_0) \boldsymbol{J}_\beta^{-1}(\boldsymbol{\theta}_0) \boldsymbol{A}_\gamma(\boldsymbol{\theta}_0) \boldsymbol{J}_\beta^{-1}(\boldsymbol{\theta}_0) \boldsymbol{K}_\beta(\boldsymbol{\theta}_0)\right), \end{aligned} \tag{10}$$

where the last equality follows from Corollary 8.3.3 of Harville (2008). □

Remark 2 The most common situation with usual density functions is r = rank $(\boldsymbol{K}_\beta(\boldsymbol{\theta}_0)) = \text{rank}(A_\gamma(\boldsymbol{\theta}_0)) = p$. This is observed, for example, in the test involving the two parameter normal model in Sect. 5.1.

Remark 3 The works of Rao and Scott (1981) and Modarres and Jernigan (1992) facilitate the calculation of the tail probabilities of linear combinations of Chi-square variables. A variety of problems in statistical inference and applied probability require percentiles or probabilities from the distribution of linear combinations of Chi-squares (see Jensen and Solomon 1972). Following Corollary 1 of Rao and Scott (1981), one can use the statistic

$${}^1T_\gamma(\widehat{\boldsymbol{\theta}}_\beta, \boldsymbol{\theta}_0) = \frac{T_\gamma(\widehat{\boldsymbol{\theta}}_\beta, \boldsymbol{\theta}_0)}{\lambda_{\max}^{\gamma,\beta}} \leq \sum_{i=1}^{r} Z_i^2,$$

where $\lambda_{\max}^{\gamma,\beta} = \max(\lambda_1^{\gamma,\beta}, \ldots, \lambda_r^{\gamma,\beta})$. As $\sum_{i=1}^{r} Z_i^2 \sim \chi_r^2$, a strategy that rejects the null in (7) for ${}^1T_\gamma(\widehat{\boldsymbol{\theta}}_\beta, \boldsymbol{\theta}_0) > \chi_{r,\alpha}^2$ produces an asymptotically conservative test at nominal level α, where $\chi_{r,\alpha}^2$ is the quantile of order $1-\alpha$ for χ_r^2. Another approximation to the asymptotic tail probabilities of $T_\gamma(\widehat{\boldsymbol{\theta}}_\beta, \boldsymbol{\theta}_0)$ can be obtained through the modification

$${}^2T_\gamma(\widehat{\boldsymbol{\theta}}_\beta, \boldsymbol{\theta}_0) = \frac{T_\gamma(\widehat{\boldsymbol{\theta}}_\beta, \boldsymbol{\theta}_0)}{\overline{\lambda}^{\gamma,\beta}},$$

where $\overline{\lambda}^{\gamma,\beta} = \frac{1}{r}\sum_{i=1}^{r} \lambda_i^{\gamma,\beta}$ (see Satterthwaite 1946), considered approximated by a Chi-squared distribution with r degrees of freedom. In this case, we can observe that

$$E\left[{}^2T_\gamma(\widehat{\boldsymbol{\theta}}_\beta, \boldsymbol{\theta}_0)\right] = r = E\left[\chi_r^2\right],$$

$$\text{Var}\left[{}^2T_\gamma(\widehat{\boldsymbol{\theta}}_\beta, \boldsymbol{\theta}_0)\right] = \frac{2\sum_{i=1}^{r} (\lambda_i^{\gamma,\beta})^2}{(\overline{\lambda}^{\gamma,\beta})^2} = 2r + 2\sum_{i=1}^{r} \frac{\left(\lambda_i^{\gamma,\beta} - \overline{\lambda}^{\gamma,\beta}\right)^2}{(\overline{\lambda}^{\gamma,\beta})^2} > 2r = \text{Var}\left[\chi_r^2\right].$$

If we denote by $\Lambda^{\gamma,\beta} = \operatorname{diag}(\lambda_1^{\gamma,\beta}, \ldots, \lambda_r^{\gamma,\beta})$, we get

$$\begin{aligned} E\left[\sum_{i=1}^{r} \lambda_i^{\gamma,\beta} Z_i^2\right] &= \sum_{i=1}^{r} \lambda_i^{\gamma,\beta} = \operatorname{trace}\left(\Lambda^{\gamma,\beta}\right) \\ &= \operatorname{trace}\left(\boldsymbol{A}_\gamma\left(\boldsymbol{\theta}_0\right) \boldsymbol{J}_\beta^{-1}\left(\boldsymbol{\theta}_0\right) \boldsymbol{K}_\beta\left(\boldsymbol{\theta}_0\right) \boldsymbol{J}_\beta^{-1}\left(\boldsymbol{\theta}_0\right)\right). \end{aligned}$$

The test given by the statistic ${}^2T_\gamma(\widehat{\boldsymbol{\theta}}_\beta, \boldsymbol{\theta}_0)$ is more conservative than the one based on $T_\gamma(\widehat{\boldsymbol{\theta}}_\beta, \boldsymbol{\theta}_0)$. However, if we consider the test-statistic

$${}^3T_\gamma(\widehat{\boldsymbol{\theta}}_\beta, \boldsymbol{\theta}_0) = \frac{{}^2T_\gamma(\widehat{\boldsymbol{\theta}}_\beta, \boldsymbol{\theta}_0)}{\nu^{\gamma,\beta}} = \frac{T_\gamma(\widehat{\boldsymbol{\theta}}_\beta, \boldsymbol{\theta}_0)}{\nu^{\gamma,\beta}\overline{\lambda}^{\gamma,\beta}},$$

we can find $\nu^{\gamma,\beta}$ imposing $\operatorname{Var}[{}^3T_\gamma(\widehat{\boldsymbol{\theta}}_\beta, \boldsymbol{\theta}_0)] = 2E[{}^3T_\gamma(\widehat{\boldsymbol{\theta}}_\beta, \boldsymbol{\theta}_0)]$ as in the Chi-squared distribution. Because

$$E\left[{}^3T_\gamma(\widehat{\boldsymbol{\theta}}_\beta, \boldsymbol{\theta}_0)\right] = \frac{r}{\nu^{\gamma,\beta}} \text{ and } \operatorname{Var}\left[{}^3T_\gamma(\widehat{\boldsymbol{\theta}}_\beta, \boldsymbol{\theta}_0)\right] = \frac{2r}{\nu^{\gamma,\beta}},$$

$$\nu^{\gamma,\beta} = 1 + \sum_{i=1}^{r} \frac{\left(\lambda_i^{\gamma,\beta} - \overline{\lambda}^{\gamma,\beta}\right)^2}{r(\overline{\lambda}^{\gamma,\beta})^2} = 1 + \mathrm{CV}^2(\{\lambda_i^{\gamma,\beta}\}_{i=1}^r),$$

where CV represents the coefficient of variation. Then a Chi-square distribution with $\frac{r}{\nu^{\gamma,\beta}}$ degrees of freedom approximates the asymptotic distribution of the statistic ${}^3T_\gamma(\widehat{\boldsymbol{\theta}}_\beta, \boldsymbol{\theta}_0)$ for large n.

The degrees of freedom of ${}^3T_\gamma(\widehat{\boldsymbol{\theta}}_\beta, \boldsymbol{\theta}_0)$ is $\frac{r}{\nu^{\gamma,\beta}}$, which may not be an integer. To avoid this difficulty one can modify the statistic such that the first two moments match specifically with the $\chi^2(r)$ distribution (rather than with just any other χ^2 distribution). Specifically let

$$X = {}^2T_\gamma(\widehat{\boldsymbol{\theta}}_\beta, \boldsymbol{\theta}_0).$$

We have

$$E[X] = r = E\left[\chi_r^2\right],$$

$$\operatorname{Var}[X] = \frac{2\sum_{i=1}^r (\lambda_i^{\gamma,\beta})^2}{(\overline{\lambda}^{\gamma,\beta})^2} = 2r + 2\sum_{i=1}^{r} \frac{\left(\lambda_i^{\gamma,\beta} - \overline{\lambda}^{\gamma,\beta}\right)^2}{(\overline{\lambda}^{\gamma,\beta})^2} = 2r + c,$$

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

$$E(Y) = r, \quad \operatorname{Var}(Y) = 2r.$$

Thus,

$$\frac{r-a}{b}=r, \quad \frac{2r+c}{b^2}=2r.$$

Solving these set of equations, we get

$$b=\sqrt{1+\frac{c}{2r}}, \quad a=r(1-b).$$

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

$$^{4}T_{\gamma}(\widehat{\boldsymbol{\theta}}_{\beta},\boldsymbol{\theta}_0)=\frac{^{2}T_{\gamma}(\widehat{\boldsymbol{\theta}}_{\beta},\boldsymbol{\theta}_0)-a}{b},$$

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

Apart from the above approximations, it is possible to consider tables of the cumulative distribution $\sum_{i=1}^{r} a_i Z_i^2$ in the case of small r (see Solomon 1960; Johnson and Kotz 1968; Eckler 1969; Gupta 1963).

Now we consider the power of the density power divergence test at contiguous alternative hypotheses described by

$$H_{1,n} : \boldsymbol{\theta}_n = \boldsymbol{\theta}_0 + n^{-1/2}\boldsymbol{d}, \tag{11}$$

where $\boldsymbol{d}$ is a fixed vector in $\mathbb{R}^p$ such that $\boldsymbol{\theta}_n \in \Theta \subset \mathbb{R}^p$. In the following theorem, we present the asymptotic distribution of $T_{\gamma}(\widehat{\boldsymbol{\theta}}_{\beta},\boldsymbol{\theta}_0)$ under (11). The proof is based on Corollary 2.2 in Dik and de Gunst (1985). This Corollary establishes the following: Let $\boldsymbol{X} \sim \mathcal{N}_q(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, a q-variate normal distribution. Let $\boldsymbol{Q}$ be a real symmetric non-negative definite matrix of order q. Let $r = \operatorname{rank}(\boldsymbol{\Sigma}\boldsymbol{Q}\boldsymbol{\Sigma})$, $r \geq 1$, and let $\lambda_1, \ldots, \lambda_r$ be the positive eigenvalues of $\boldsymbol{Q}\boldsymbol{\Sigma}$. Then, the quadratic form $\boldsymbol{X}^T\boldsymbol{Q}\boldsymbol{X}$ has the same distribution as the random variable

$$\sum_{i=1}^{r} \lambda_i\,(U_i + w_i)^2 + \boldsymbol{\xi},$$

where $U_1, \ldots, U_r$ are independent, each having a standard normal distribution. Values of $\boldsymbol{w}$ and $\boldsymbol{\xi}$ are given by

$$\boldsymbol{w} = \boldsymbol{\Lambda}_p^{-1}\,\boldsymbol{P}^T\boldsymbol{S}^T\boldsymbol{Q}\boldsymbol{\mu}, \quad \boldsymbol{\xi} = \boldsymbol{\mu}^T\boldsymbol{Q}\boldsymbol{\mu} - \boldsymbol{w}^T\boldsymbol{\Lambda}_p\boldsymbol{w},$$

where $\boldsymbol{S}$ is any $q \times s$ square root of $\boldsymbol{\Sigma}$, $\boldsymbol{\Lambda}_p = \operatorname{diag}(\lambda_1, \ldots, \lambda_r)$ and $\boldsymbol{P}$ is the matrix of corresponding orthonormal eigenvectors. This result leads to the following theorem.

Theorem 4 *Suppose that the model satisfies the Lehmann and the Basu et al. conditions. Under the contiguous alternative hypotheses $H_{1,n}$ given in (11), the asymptotic distribution of $T_\gamma(\widehat{\boldsymbol{\theta}}_\beta, \boldsymbol{\theta}_0)$, coincides with the distribution of*

$$\sum_{i=1}^{r} \lambda_i^{\gamma,\beta}(\boldsymbol{\theta}_0)\,(Z_i + w_i)^2 + \boldsymbol{\xi},$$

where $Z_1, \ldots, Z_r$ are independent standard normal variables, $\lambda_1^{\gamma,\beta}(\boldsymbol{\theta}_0), \ldots, \lambda_r^{\gamma,\beta}(\boldsymbol{\theta}_0)$ are the positive eigenvalues of $\boldsymbol{A}_\gamma(\boldsymbol{\theta}_0)\boldsymbol{J}_\beta^{-1}(\boldsymbol{\theta}_0)\boldsymbol{K}_\beta(\boldsymbol{\theta}_0)\boldsymbol{J}_\beta^{-1}(\boldsymbol{\theta}_0)$, the values $\boldsymbol{w} = (w_1, \ldots, w_r)$ and $\boldsymbol{\xi}$ are given by

$$\boldsymbol{w} = \boldsymbol{\Lambda}_r^{-1}\boldsymbol{P}^T\boldsymbol{S}^T\boldsymbol{A}_\gamma(\boldsymbol{\theta}_0)\,\boldsymbol{d}, \quad \boldsymbol{\xi} = \boldsymbol{d}^T\boldsymbol{A}_\gamma(\boldsymbol{\theta}_0)\,\boldsymbol{d} - \boldsymbol{w}^T\boldsymbol{\Lambda}_r\boldsymbol{w},$$

$\boldsymbol{S}$ is any square root of $\boldsymbol{J}_\beta^{-1}(\boldsymbol{\theta}_0)\boldsymbol{K}_\beta(\boldsymbol{\theta}_0)\boldsymbol{J}_\beta^{-1}(\boldsymbol{\theta}_0)$, $\boldsymbol{\Lambda}_r = diag(\lambda_1^{\gamma,\beta}(\boldsymbol{\theta}_0), \ldots, \lambda_r^{\gamma,\beta}(\boldsymbol{\theta}_0))$ and $\boldsymbol{P}$ is the matrix of corresponding orthonormal eigenvectors.

Proof We can write

$$\sqrt{n}(\widehat{\boldsymbol{\theta}}_\beta - \boldsymbol{\theta}_0) = \sqrt{n}(\widehat{\boldsymbol{\theta}}_\beta - \boldsymbol{\theta}_n) + \sqrt{n}\,(\boldsymbol{\theta}_n - \boldsymbol{\theta}_0) = \sqrt{n}(\widehat{\boldsymbol{\theta}}_\beta - \boldsymbol{\theta}_n) + \boldsymbol{d}.$$

Under $H_{1,n}$ one has

$$\sqrt{n}(\widehat{\boldsymbol{\theta}}_\beta - \boldsymbol{\theta}_n) \xrightarrow[n\to\infty]{L} N\left(\boldsymbol{0}, \boldsymbol{J}_\beta^{-1}(\boldsymbol{\theta}_0)\boldsymbol{K}_\beta(\boldsymbol{\theta}_0)\boldsymbol{J}_\beta^{-1}(\boldsymbol{\theta}_0)\right)$$

and

$$\sqrt{n}(\widehat{\boldsymbol{\theta}}_\beta - \boldsymbol{\theta}_0) \xrightarrow[n\to\infty]{L} N\left(\boldsymbol{d}, \boldsymbol{J}_\beta^{-1}(\boldsymbol{\theta}_0)\boldsymbol{K}_\beta(\boldsymbol{\theta}_0)\boldsymbol{J}_\beta^{-1}(\boldsymbol{\theta}_0)\right)$$

We know that

$$T_\gamma(\widehat{\boldsymbol{\theta}}_\beta, \boldsymbol{\theta}_0) = n^{1/2}(\widehat{\boldsymbol{\theta}}_\beta - \boldsymbol{\theta}_0)^T\boldsymbol{A}_\gamma(\boldsymbol{\theta}_0)\,n^{1/2}(\widehat{\boldsymbol{\theta}}_\beta - \boldsymbol{\theta}_0) + n \quad o\left(\left\|\widehat{\boldsymbol{\theta}}_\beta - \boldsymbol{\theta}_0\right\|^2\right).$$

Then, $T_\gamma(\widehat{\boldsymbol{\theta}}_\beta, \boldsymbol{\theta}_0)$ has the same asymptotic distribution as the quadratic form $n^{1/2}(\widehat{\boldsymbol{\theta}}_\beta - \boldsymbol{\theta}_0)^T\boldsymbol{A}_\gamma(\boldsymbol{\theta}_0)\,n^{1/2}(\widehat{\boldsymbol{\theta}}_\beta - \boldsymbol{\theta}_0)$. Now the result follows from Corollary 2.2 of Dik and de Gunst (1985). □

While the above theorem is in many ways instructive, it is not necessarily helpful in determining a quick approximation to the power function of our proposed tests. In the next theorem we will get an approximation of the power function for $T_\gamma(\widehat{\boldsymbol{\theta}}_\beta, \boldsymbol{\theta}_0)$, the test statistic given in (8).

Theorem 5 *Suppose that the model satisfies the Lehmann and the Basu et al. conditions. An approximation to the power function of the test statistic $T_\gamma(\widehat{\boldsymbol{\theta}}_\beta, \boldsymbol{\theta}_0)$ for* *testing* $H_0: \boldsymbol{\theta} = \boldsymbol{\theta}_0$ *against* $H_1: \boldsymbol{\theta} \neq \boldsymbol{\theta}_0$ *is given by*

$$\pi_{n,\alpha}^{\beta,\gamma}\left(\boldsymbol{\theta}^*\right) = 1 - \Phi\left(\frac{\sqrt{n}}{\sigma_{\beta,\gamma}\left(\boldsymbol{\theta}^*\right)}\left(\frac{t_\alpha^{\gamma,\beta}}{2n} - d_\gamma(f_{\boldsymbol{\theta}^*}, f_{\boldsymbol{\theta}_0})\right)\right)$$

for $\boldsymbol{\theta}^* \neq \boldsymbol{\theta}_0$, *where* $t_\alpha^{\gamma,\beta}$ *is the quantile of order* $(1-\alpha)$ *of the asymptotic distribution of* $T_\gamma(\widehat{\boldsymbol{\theta}}_\beta, \boldsymbol{\theta}_0)$, *and* $\sigma_{\beta,\gamma}\left(\boldsymbol{\theta}^*\right)$ *is as in* (12) *defined later.*

Proof A first-order Taylor expansion of $d_\gamma(f_{\widehat{\boldsymbol{\theta}}_\beta}, f_{\boldsymbol{\theta}_0})$ under $f_{\boldsymbol{\theta}^*}$, where $\boldsymbol{\theta}^* \neq \boldsymbol{\theta}_0$, gives

$$d_\gamma(f_{\widehat{\boldsymbol{\theta}}_\beta}, f_{\boldsymbol{\theta}_0}) = d_\gamma(f_{\boldsymbol{\theta}^*}, f_{\boldsymbol{\theta}_0}) + \boldsymbol{B}_\gamma^T(\widehat{\boldsymbol{\theta}}_\beta - \boldsymbol{\theta}^*) + o\left(\left\|\widehat{\boldsymbol{\theta}}_\beta - \boldsymbol{\theta}^*\right\|\right).$$

where $\boldsymbol{B}_\gamma = (B_1^\gamma, \ldots, B_k^\gamma)^T$ and $B_j^\gamma = (\frac{\partial d_\gamma(f_{\boldsymbol{\theta}}, f_{\boldsymbol{\theta}_0})}{\partial \theta_j})_{\boldsymbol{\theta}=\boldsymbol{\theta}^*}$. We know

$$\sqrt{n}(\widehat{\boldsymbol{\theta}}_\beta - \boldsymbol{\theta}^*) \xrightarrow[n\to\infty]{L} N\left(\mathbf{0}, \boldsymbol{J}_\beta^{-1}\left(\boldsymbol{\theta}^*\right)\boldsymbol{K}_\beta\left(\boldsymbol{\theta}^*\right)\boldsymbol{J}_\beta^{-1}\left(\boldsymbol{\theta}^*\right)\right)$$

and

$$\sqrt{n} \times o\left(\left\|\widehat{\boldsymbol{\theta}}_\beta - \boldsymbol{\theta}^*\right\|\right) = o_P(1).$$

Then, it is clear that the random variables,

$$\sqrt{n}\left(d_\gamma(f_{\widehat{\boldsymbol{\theta}}_\beta}, f_{\boldsymbol{\theta}_0}) - d_\gamma(f_{\boldsymbol{\theta}^*}, f_{\boldsymbol{\theta}_0})\right) \quad \text{and} \quad \boldsymbol{B}_\gamma^T\sqrt{n}(\widehat{\boldsymbol{\theta}}_\beta - \boldsymbol{\theta}^*)$$

have the same asymptotic distribution. Therefore,

$$\sqrt{n}\left(d_\gamma(f_{\widehat{\boldsymbol{\theta}}_\beta}, f_{\boldsymbol{\theta}_0}) - d_\gamma(f_{\boldsymbol{\theta}^*}, f_{\boldsymbol{\theta}_0})\right) \xrightarrow[n\to\infty]{L} N(\mathbf{0}, \sigma_{\beta,\gamma}^2\left(\boldsymbol{\theta}^*\right),$$

where

$$\sigma_{\beta,\gamma}^2\left(\boldsymbol{\theta}^*\right) = \boldsymbol{B}_\gamma^T\boldsymbol{J}_\beta^{-1}\left(\boldsymbol{\theta}^*\right)\boldsymbol{K}_\beta\left(\boldsymbol{\theta}^*\right)\boldsymbol{J}_\beta^{-1}\left(\boldsymbol{\theta}^*\right)\boldsymbol{B}_\gamma. \tag{12}$$

Based on this result a first approximation to the power function, at $\boldsymbol{\theta}^* \neq \boldsymbol{\theta}_0$, is

$$\pi_{n,\alpha}^{\beta,\gamma}\left(\boldsymbol{\theta}^*\right) = 1 - \Phi\left(\frac{\sqrt{n}}{\sigma_{\beta,\gamma}\left(\boldsymbol{\theta}^*\right)}\left(\frac{t_\alpha^{\gamma,\beta}}{2n} - d_\gamma(f_{\boldsymbol{\theta}^*}, f_{\boldsymbol{\theta}_0})\right)\right),$$

where $\Phi(x)$ is the standard normal distribution function. □

If $\boldsymbol{\theta}^*$ is different from $\boldsymbol{\theta}_0$, the probability of rejecting the null with the rejection rule $T_\gamma(\widehat{\boldsymbol{\theta}}_\beta, \boldsymbol{\theta})_0 > t_\alpha^{\gamma,\beta}$ tends to 1 for any fixed significance level $\alpha > 0$ as $n \to \infty$. The test statistic is consistent in the Fraser's sense.

Theorems 1 and 4 demonstrate that the test statistic defined in (8) has asymptotic properties similar to the likelihood ratio type tests of Heritier and Ronchetti (1994, Proposition 1 and 3). Although the properties of the robust bounded influence type tests are more difficult to determine for the likelihood ratio type tests compared to the Wald and score statistics, the robustness of the density power divergence tests are apparent since these statistics involve the data only through the parameter estimates. The robustness and bounded influence properties of the density power divergence measures have been described in Basu et al. (1998). The link between the density power divergence test statistic and the minimum density power divergence estimator is clearly observed in the example of Sect. 5.1 later.

3.2 The normal model

The normal model is perhaps the most widely used model in statistics. While the technique we have developed is quite general, it is useful in this connection to present the specific values in the context of a normal distribution. Here, we derive the expressions in case of the two parameter normal model.

Let $X_1, \ldots, X_n$ be a random sample from a normal population with mean μ and variance σ^2. Let $\Theta = \mathbb{R} \times \mathbb{R}^+$ be the parameter space, where $\mathbb{R}$ is the set of real numbers and $\mathbb{R}^+$ is the set of positive real numbers; also let $\boldsymbol{\theta} = (\mu, \sigma) \in \Theta$. We are interested in testing the null hypothesis

$$H_0 : \boldsymbol{\theta} = \boldsymbol{\theta}_0 = (\mu_0, \sigma_0) \text{ versus } H_1 : \boldsymbol{\theta} \neq \boldsymbol{\theta}_0.$$

The minimum density power divergence estimator $\widehat{\boldsymbol{\theta}}_\beta = \left(\widehat{\mu}_\beta, \widehat{\sigma}_\beta\right)$ is obtained by maximizing in σ and μ, for each β,

$$\frac{1}{n\beta}\sum_{i=1}^{n}\left(\frac{1}{\sigma\sqrt{2\pi}}\right)^{\beta}\exp\left(-\frac{1}{2}\left(\frac{X_i-\mu}{\sigma}\right)^2\beta\right) \\ -\frac{1}{1+\beta}\int_{\mathbb{R}}\left(\frac{1}{\sigma\sqrt{2\pi}}\right)^{1+\beta}\exp\left(-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2(1+\beta)\right)\mathrm{d}x.$$

Since

$$\int_{\mathbb{R}}\left(\frac{1}{\sigma\sqrt{2\pi}}\right)^{1+\beta}\exp\left(-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2(1+\beta)\right)\mathrm{d}x = \frac{1}{\sigma^{\beta}\sqrt{1+\beta}\,(2\pi)^{\frac{\beta}{2}}},$$

it is necessary to maximize in μ and σ, for each β, the expression

$$\frac{1}{\sigma^{\beta}\,(2\pi)^{\frac{\beta}{2}}}\left\{\frac{1}{n\beta}\sum_{i=1}^{n}\exp\left(-\frac{1}{2}\left(\frac{X_i-\mu}{\sigma}\right)^2\beta\right)-\frac{1}{(1+\beta)^{3/2}}\right\}$$

to get the estimates $\widehat{\mu}_\beta$ and $\widehat{\sigma}_\beta$. Subsequently, we need to derive the expression of the density power divergence test statistic for our problem. After some algebra, we obtain the density power divergence measure $d_\gamma\left(N\left(\mu_1,\sigma_1\right),N\left(\mu_2,\sigma_2\right)\right)$ between the indicated normal densities as

$$
\begin{aligned}
d_\gamma\left(N\left(\mu_1,\sigma_1\right),N\left(\mu_2,\sigma_2\right)\right) = & \frac{1}{\sigma_2^\gamma\sqrt{1+\gamma}\,(2\pi)^{\gamma/2}} \\
& -\left(1+\frac{1}{\gamma}\right)\frac{1}{\sigma_2^{\gamma-1}(\gamma\sigma_1^2+\sigma_2^2)^{1/2}\,(2\pi)^{\gamma/2}} \\
& \times\exp\left\{-\frac{1}{2}\left(\frac{\mu_2^2}{\left(\frac{\sigma_2}{\sqrt{\gamma}}\right)^2}+\frac{\mu_1^2}{\sigma_1^2}\right)\right\} \\
& \times\exp\left\{\frac{1}{2}\frac{\left(\sigma_1^2\mu_2+\mu_1\left(\frac{\sigma_2}{\sqrt{\gamma}}\right)^2\right)^2}{\left(\sigma_1^2+\left(\frac{\sigma_2}{\sqrt{\gamma}}\right)^2\right)\left(\frac{\sigma_2}{\sqrt{\gamma}}\right)^2\sigma_1^2}\right\} \\
& +\frac{1}{\gamma\sigma_1^\gamma\sqrt{1+\gamma}\,(2\pi)^{\gamma/2}}.
\end{aligned}
$$

Therefore,

$$
\begin{aligned}
T_\gamma\left(\widehat{\boldsymbol{\theta}}_\beta,\boldsymbol{\theta}_0\right) &= 2nd_\gamma\left(N\left(\widehat{\mu}_\beta,\widehat{\sigma}_\beta\right),N\left(\mu_0,\sigma_0\right)\right) \\
&= \frac{2n}{(2\pi)^{\gamma/2}\sqrt{1+\gamma}}\left(\frac{1}{\sigma_0^\gamma}-\left(1+\frac{1}{\gamma}\right)\frac{\sqrt{1+\gamma}}{\sigma_0^{\gamma-1}(\gamma\widehat{\sigma}_\beta^2+\sigma_0^2)^{1/2}}\right. \\
&\quad\times\exp\left\{-\frac{1}{2}\left(\frac{\mu_0^2}{\left(\frac{\sigma_0}{\sqrt{\gamma}}\right)^2}+\frac{\widehat{\mu}_\beta^2}{\widehat{\sigma}_\beta^2}\right)\right\} \\
&\quad\left.\times\exp\left\{\frac{1}{2}\frac{\left(\widehat{\sigma}_\beta^2\mu_0+\widehat{\mu}_\beta\left(\frac{\sigma_0}{\sqrt{\gamma}}\right)^2\right)^2}{\left(\widehat{\sigma}_\beta^2+\left(\frac{\sigma_0}{\sqrt{\gamma}}\right)^2\right)\left(\frac{\sigma_0}{\sqrt{\gamma}}\right)^2\widehat{\sigma}_\beta^2}\right\}+\frac{1}{\gamma\widehat{\sigma}_\beta^\gamma}\right).
\end{aligned}
$$

For $\boldsymbol{\theta}=(\mu,\sigma)$, the score function for the normal model is given by

$$
u_{\boldsymbol{\theta}}(x)=\begin{pmatrix}\dfrac{x-\mu}{\sigma^2}\\[2mm] \dfrac{1}{\sigma}\left(\dfrac{x-\mu}{\sigma}\right)^2-\dfrac{1}{\sigma}\end{pmatrix}.
$$

The expressions of $\boldsymbol{J}_{\boldsymbol{\beta}}(\boldsymbol{\theta})$, $\boldsymbol{K}_{\boldsymbol{\beta}}(\boldsymbol{\theta})$ and $\boldsymbol{A}_{\boldsymbol{\gamma}}(\boldsymbol{\theta})$ can be obtained after some algebra and using expressions for the moments of a normal distribution as

$$\boldsymbol{J}_{\boldsymbol{\beta}}(\boldsymbol{\theta}) = \frac{1}{\sqrt{1+\beta}\,(2\pi)^{\beta/2}\,\sigma^{2+\beta}} \begin{pmatrix} \frac{1}{1+\beta} & 0 \\ 0 & \frac{\beta^2+2}{(1+\beta)^2} \end{pmatrix}$$

$$= \frac{1}{(1+\beta)^{3/2}\,(2\pi)^{\beta/2}\,\sigma^{2+\beta}} \begin{pmatrix} 1 & 0 \\ 0 & \frac{\beta^2+2}{(1+\beta)} \end{pmatrix}$$

$$\boldsymbol{K}_{\boldsymbol{\beta}}(\boldsymbol{\theta}) = \frac{1}{\sigma^{2+2\beta}\,(2\pi)^{\beta}} \left\{ \frac{1}{(1+2\beta)^{3/2}} \begin{pmatrix} 1 & 0 \\ 0 & \frac{4\beta^2+2}{1+2\beta} \end{pmatrix} - \begin{pmatrix} 0 & 0 \\ 0 & \frac{\beta^2}{(1+\beta)^3} \end{pmatrix} \right\},$$

$$\boldsymbol{A}_{\gamma}(\boldsymbol{\theta}) = \frac{1}{(2\pi)^{\gamma/2}\,\sigma^{2+\gamma}\,(1+\gamma)^{1/2}} \begin{pmatrix} 1 & 0 \\ 0 & \frac{\gamma^2+2}{(1+\gamma)} \end{pmatrix}.$$

Tests involving only one of the two parameters μ and σ (with the other parameter known) are simpler since in such cases the parameter is a scalar (and hence so are the matrices $\boldsymbol{J}_{\boldsymbol{\beta}}, \boldsymbol{K}_{\boldsymbol{\beta}}$ and $\boldsymbol{A}_{\boldsymbol{\gamma}}$). One simply needs to pick out the correct components from the 2×2 matrices needed to perform the test for $\boldsymbol{\theta} = (\mu, \sigma)$ to determine the single eigenvalue that is now involved in the distribution of our test statistic. An example will be given in Sect. 5.1.

3.3 The two-sample problem

We consider a parametric model $f_{\boldsymbol{\theta}}(x), \boldsymbol{\theta} \in \Theta \subseteq \mathbb{R}^k$, random samples of sizes n and m from two populations with parameters $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$ respectively, and the corresponding estimators, ${}^{(1)}\widehat{\boldsymbol{\theta}}_{\beta} = (\widehat{\theta}_{\beta}^{11}, \ldots, \widehat{\theta}_{\beta}^{1p})^T$ and ${}^{(2)}\widehat{\boldsymbol{\theta}}_{\beta} = (\widehat{\theta}_{\beta}^{21}, \ldots, \widehat{\theta}_{\beta}^{2p})^T$, associated with them. Here, we will derive the asymptotic distribution of the test statistic

$$S_{\gamma}\left({}^{(1)}\widehat{\boldsymbol{\theta}}_{\beta}, {}^{(2)}\widehat{\boldsymbol{\theta}}_{\beta}\right) = \frac{2nm}{(m+n)} d_{\gamma}(f_{{}^{(1)}\widehat{\boldsymbol{\theta}}_{\beta}}, f_{{}^{(2)}\widehat{\boldsymbol{\theta}}_{\beta}}) \tag{13}$$

in order to test

$$H_0 : \boldsymbol{\theta}_1 = \boldsymbol{\theta}_2 \text{ against } H_1 : \boldsymbol{\theta}_1 \neq \boldsymbol{\theta}_2.$$

Theorem 6 *Suppose that the model satisfies the Lehmann and the Basu et al. conditions. Under the null hypothesis $H_0 : \boldsymbol{\theta}_1 = \boldsymbol{\theta}_2$, the asymptotic distribution of $S_{\gamma}\left({}^{(1)}\widehat{\boldsymbol{\theta}}_{\beta}, {}^{(2)}\widehat{\boldsymbol{\theta}}_{\beta}\right)$ coincides with the distribution of*

$$\sum_{i=1}^{r} \lambda_i^{\gamma,\beta} Z_i^2,$$

where $\lambda_1^{\gamma,\beta}, \ldots, \lambda_r^{\gamma,\beta}$ are the non-zero eigenvalues of $\boldsymbol{A}_{\gamma}(\boldsymbol{\theta}_1) \boldsymbol{J}_{\beta}^{-1}(\boldsymbol{\theta}_1) \boldsymbol{K}_{\beta}(\boldsymbol{\theta}_1) \boldsymbol{J}_{\beta}^{-1}(\boldsymbol{\theta}_1)$ and

$$r = \operatorname{rank}\left(\boldsymbol{K}_{\beta}\left(\boldsymbol{\theta}_{1}\right)\boldsymbol{J}_{\beta}^{-1}\left(\boldsymbol{\theta}_{1}\right)\boldsymbol{A}_{\gamma}\left(\boldsymbol{\theta}_{1}\right)\boldsymbol{J}_{\beta}^{-1}\left(\boldsymbol{\theta}_{1}\right)\boldsymbol{K}_{\beta}\left(\boldsymbol{\theta}_{1}\right)\right).$$

Proof We know

$$\begin{aligned}
\sqrt{n}\left({}^{(1)}\widehat{\boldsymbol{\theta}}_{\beta}-\boldsymbol{\theta}_{1}\right) &\underset{n\to\infty}{\xrightarrow{L}} N\left(\boldsymbol{0},\boldsymbol{J}_{\beta}^{-1}\left(\boldsymbol{\theta}_{1}\right)\boldsymbol{K}_{\beta}\left(\boldsymbol{\theta}_{1}\right)\boldsymbol{J}_{\beta}^{-1}\left(\boldsymbol{\theta}_{1}\right)\right),\\
\sqrt{m}\left({}^{(2)}\widehat{\boldsymbol{\theta}}_{\beta}-\boldsymbol{\theta}_{2}\right) &\underset{m\to\infty}{\xrightarrow{L}} N\left(\boldsymbol{0},\boldsymbol{J}_{\beta}^{-1}\left(\boldsymbol{\theta}_{2}\right)\boldsymbol{K}_{\beta}\left(\boldsymbol{\theta}_{2}\right)\boldsymbol{J}_{\beta}^{-1}\left(\boldsymbol{\theta}_{2}\right)\right),
\end{aligned}$$

then

$$\begin{aligned}
\sqrt{\frac{mn}{m+n}}\left({}^{(1)}\widehat{\boldsymbol{\theta}}_{\beta}-\boldsymbol{\theta}_{1}\right) &\underset{n,m\to\infty}{\xrightarrow{L}} N\left(\boldsymbol{0},\omega\boldsymbol{J}_{\beta}^{-1}\left(\boldsymbol{\theta}_{1}\right)\boldsymbol{K}_{\beta}\left(\boldsymbol{\theta}_{1}\right)\boldsymbol{J}_{\beta}^{-1}\left(\boldsymbol{\theta}_{1}\right)\right),\\
\sqrt{\frac{mn}{m+n}}\left({}^{(2)}\widehat{\boldsymbol{\theta}}_{\beta}-\boldsymbol{\theta}_{2}\right) &\underset{n,m\to\infty}{\xrightarrow{L}} N\left(\boldsymbol{0},(1-\omega)\boldsymbol{J}_{\beta}^{-1}\left(\boldsymbol{\theta}_{2}\right)\boldsymbol{K}_{\beta}\left(\boldsymbol{\theta}_{2}\right)\boldsymbol{J}_{\beta}^{-1}\left(\boldsymbol{\theta}_{2}\right)\right),
\end{aligned}$$

where

$$\omega = \lim_{n,m\to\infty}\frac{m}{m+n}.$$

Under the hypothesis $\boldsymbol{\theta}_1 = \boldsymbol{\theta}_2$, we have

$$\sqrt{\frac{mn}{m+n}}\left({}^{(1)}\widehat{\boldsymbol{\theta}}_{\beta}-{}^{(2)}\widehat{\boldsymbol{\theta}}_{\beta}\right) \underset{n,m\to\infty}{\xrightarrow{L}} N\left(\boldsymbol{0},\boldsymbol{J}_{\beta}^{-1}\left(\boldsymbol{\theta}_{1}\right)\boldsymbol{K}_{\beta}\left(\boldsymbol{\theta}_{1}\right)\boldsymbol{J}_{\beta}^{-1}\left(\boldsymbol{\theta}_{1}\right)\right).$$

A second Taylor expansion of $d_{\gamma}(f_{\boldsymbol{\theta}_1}, f_{\boldsymbol{\theta}_2})$ around $\boldsymbol{\theta}_1 = \boldsymbol{\theta}_2$ at $\left({}^{(1)}\widehat{\boldsymbol{\theta}}_{\beta}, {}^{(2)}\widehat{\boldsymbol{\theta}}_{\beta}\right)$ gives,

$$\begin{aligned}
d_{\gamma}(f_{{}^{(1)}\widehat{\boldsymbol{\theta}}_{\beta}}, f_{{}^{(2)}\widehat{\boldsymbol{\theta}}_{\beta}}) &= \frac{1}{2}\sum_{i,j=1}^{p}\left(\frac{\partial^2 d_{\gamma}(f_{\boldsymbol{\theta}_1}, f_{\boldsymbol{\theta}_2})}{\partial\theta_{1i}\partial\theta_{1j}}\right)_{\boldsymbol{\theta}_1=\boldsymbol{\theta}_2}(\widehat{\theta}_{\beta}^{1i}-\theta_{1i})(\widehat{\theta}_{\beta}^{1j}-\theta_{1j})\\
&+\sum_{i,j=1}^{p}\left(\frac{\partial^2 d_{\gamma}(f_{\boldsymbol{\theta}_1}, f_{\boldsymbol{\theta}_2})}{\partial\theta_{1i}\partial\theta_{2j}}\right)_{\boldsymbol{\theta}_1=\boldsymbol{\theta}_2}(\widehat{\theta}_{\beta}^{1i}-\theta_{1i})(\widehat{\theta}_{\beta}^{2j}-\theta_{2j})\\
&+\frac{1}{2}\sum_{i,j=1}^{p}\left(\frac{\partial^2 d_{\gamma}(f_{\boldsymbol{\theta}_1}, f_{\boldsymbol{\theta}_2})}{\partial\theta_{2i}\partial\theta_{2j}}\right)_{\boldsymbol{\theta}_1=\boldsymbol{\theta}_2}(\widehat{\theta}_{\beta}^{2i}-\theta_{2i})(\widehat{\theta}_{\beta}^{2j}-\theta_{2j})\\
&+o\left(\left\|{}^{(1)}\widehat{\boldsymbol{\theta}}_{\beta}-\boldsymbol{\theta}_{1}\right\|^2\right)+o\left(\left\|{}^{(2)}\widehat{\boldsymbol{\theta}}_{\beta}-\boldsymbol{\theta}_{2}\right\|^2\right).
\end{aligned}$$

But

$$\frac{\partial d_{\gamma}(f_{\boldsymbol{\theta}_1}, f_{\boldsymbol{\theta}_2})}{\partial\theta_{1i}} = \left(1+\frac{1}{\gamma}\right)\int_{\mathcal{X}}\left\{-f_{\boldsymbol{\theta}_2}(x)^{\gamma}+f_{\boldsymbol{\theta}_1}(x)^{\gamma}\right\}\frac{\partial f_{\boldsymbol{\theta}_1}(x)}{\partial\theta_{1i}}\mathrm{d}\mu(x) \quad i=1,\ldots,k$$

then

$$\begin{aligned}
\left(\frac{\partial^2 d_\gamma(f_{\boldsymbol{\theta}_1}, f_{\boldsymbol{\theta}_2})}{\partial\theta_{1i}\partial\theta_{1j}}\right)_{\boldsymbol{\theta}_1=\boldsymbol{\theta}_2} &= (1+\gamma)\left(\int_{\mathcal{X}} f_{\boldsymbol{\theta}_2}(x)^{\gamma-1}\frac{\partial f_{\theta_1}(x)}{\partial\theta_{1i}}\frac{\partial f_{\boldsymbol{\theta}_1}(x)}{\partial\theta_{1j}}d\mu(x)\right)_{\boldsymbol{\theta}_1=\boldsymbol{\theta}_2}\\
&= a_{ij}^{\gamma}(\boldsymbol{\theta}_1),\\
\left(\frac{\partial^2 d_\gamma(f_{\boldsymbol{\theta}_1}, f_{\boldsymbol{\theta}_2})}{\partial\theta_{1i}\partial\theta_{2j}}\right)_{\boldsymbol{\theta}_1=\boldsymbol{\theta}_2} &= -\left(\frac{\partial^2 d_\gamma(f_{\boldsymbol{\theta}_1}, f_{\boldsymbol{\theta}_2})}{\partial\boldsymbol{\theta}_{1i}\partial\boldsymbol{\theta}_{1j}}\right)_{\boldsymbol{\theta}_1=\boldsymbol{\theta}_2},\\
\left(\frac{\partial^2 d_\gamma(f_{\boldsymbol{\theta}_1}, f_{\boldsymbol{\theta}_2})}{\partial\theta_{2i}\partial\theta_{2j}}\right)_{\boldsymbol{\theta}_1=\boldsymbol{\theta}_2} &= \left(\frac{\partial^2 d_\gamma(f_{\boldsymbol{\theta}_1}, f_{\boldsymbol{\theta}_2})}{\partial\theta_{1i}\partial\theta_{1j}}\right)_{\boldsymbol{\theta}_1=\boldsymbol{\theta}_2}.
\end{aligned}$$

Therefore,

$$\begin{aligned}
2d_\gamma(f_{\widehat{\boldsymbol{\theta}}_1}, f_{\widehat{\boldsymbol{\theta}}_2}) &= ({}^{(1)}\widehat{\boldsymbol{\theta}}_\beta - \boldsymbol{\theta}_1)^T \boldsymbol{A}_\gamma(\boldsymbol{\theta}_1)\,({}^{(1)}\widehat{\boldsymbol{\theta}}_\beta - \boldsymbol{\theta}_1)\\
&\quad - 2({}^{(1)}\widehat{\boldsymbol{\theta}}_\beta - \boldsymbol{\theta}_1)^T \boldsymbol{A}_\gamma(\boldsymbol{\theta}_1)\,({}^{(2)}\widehat{\boldsymbol{\theta}}_\beta - \boldsymbol{\theta}_2)\\
&\quad + ({}^{(2)}\widehat{\boldsymbol{\theta}}_\beta - \boldsymbol{\theta}_1)^T \boldsymbol{A}_\gamma(\boldsymbol{\theta}_1)\,({}^{(2)}\widehat{\boldsymbol{\theta}}_\beta - \boldsymbol{\theta}_1) + o\left(\left\|{}^{(1)}\widehat{\boldsymbol{\theta}}_\beta - \boldsymbol{\theta}_1\right\|^2\right)\\
&\quad + o\left(\left\|{}^{(2)}\widehat{\boldsymbol{\theta}}_\beta - \boldsymbol{\theta}_2\right\|^2\right)\\
&= ({}^{(1)}\widehat{\boldsymbol{\theta}}_\beta - {}^{(2)}\widehat{\boldsymbol{\theta}}_\beta)^T \boldsymbol{A}_\gamma(\boldsymbol{\theta}_1)\,({}^{(1)}\widehat{\boldsymbol{\theta}}_\beta - {}^{(2)}\widehat{\boldsymbol{\theta}}_\beta) + o\left(\left\|{}^{(1)}\widehat{\boldsymbol{\theta}}_\beta - \boldsymbol{\theta}_1\right\|^2\right)\\
&\quad + o\left(\left\|{}^{(2)}\widehat{\boldsymbol{\theta}}_\beta - \boldsymbol{\theta}_2\right\|^2\right),
\end{aligned}$$

and the asymptotic distribution of

$$S_\gamma\left({}^{(1)}\widehat{\boldsymbol{\theta}}_\beta, {}^{(2)}\widehat{\boldsymbol{\theta}}_\beta\right) = \frac{2mn}{m+n} d_\gamma(f_{\widehat{\boldsymbol{\theta}}_1}, f_{\widehat{\boldsymbol{\theta}}_2})$$

coincides with the distribution of the random variable $\sum_{i=1}^{r}\lambda_i^{\gamma,\beta}Z_i^2$ because $o(\|{}^{(1)}\widehat{\boldsymbol{\theta}}_\beta - \boldsymbol{\theta}_1\|^2) = o_P(n^{-1})$ and $o(\|{}^{(2)}\widehat{\boldsymbol{\theta}}_\beta - \boldsymbol{\theta}_2\|^2) = o_P(m^{-1})$. □

Remark 7 To approximate the power function of the statistic in (13) consider the Taylor expansion

$$\begin{aligned}
d_\gamma(f_{{}^{(1)}\widehat{\boldsymbol{\theta}}_\beta}, f_{{}^{(2)}\widehat{\boldsymbol{\theta}}_\beta}) &= d_\gamma(f_{\boldsymbol{\theta}_1}, f_{\boldsymbol{\theta}_2}) + \sum_{i=1}^{p}\frac{\partial d_\gamma(f_{\boldsymbol{\theta}_1}, f_{\boldsymbol{\theta}_2})}{\partial\theta_{1i}}(\widehat{\theta}_\beta^{1i} - \theta_{1i})\\
&\quad + \sum_{i=1}^{p}\frac{\partial d_\gamma(f_{\boldsymbol{\theta}_1}, f_{\boldsymbol{\theta}_2})}{\partial\theta_{2i}}(\widehat{\theta}_\beta^{2i} - \theta_{2i})\\
&\quad + o\left(\left\|{}^{(1)}\widehat{\boldsymbol{\theta}}_\beta - \boldsymbol{\theta}_1\right\|\right) + o\left(\left\|{}^{(2)}\widehat{\boldsymbol{\theta}}_\beta - \boldsymbol{\theta}_2\right\|\right),
\end{aligned}$$

then

$$\sqrt{\frac{nm}{n+m}}\left(d_\gamma(f_{^{(1)}\widehat{\boldsymbol{\theta}}_\beta}, f_{^{(2)}\widehat{\boldsymbol{\theta}}_\beta}) - d_\gamma(f_{\boldsymbol{\theta}_1}, f_{\boldsymbol{\theta}_2})\right)$$
$$= \boldsymbol{G}_\gamma^T(^{(1)}\widehat{\boldsymbol{\theta}}_\beta - \boldsymbol{\theta}_1) + \boldsymbol{H}_\gamma^T(^{(2)}\widehat{\boldsymbol{\theta}}_\beta - \boldsymbol{\theta}_2) + o\left(\left\| ^{(1)}\widehat{\boldsymbol{\theta}}_\beta - \boldsymbol{\theta}_1\right\|\right) + o\left(\left\| ^{(2)}\widehat{\boldsymbol{\theta}}_\beta - \boldsymbol{\theta}_2\right\|\right)$$

where

$$\boldsymbol{G}_\gamma = \left(g_1^\gamma, \ldots, g_p^\gamma\right)^T \text{ and } \boldsymbol{H}_\gamma = \left(h_1^\gamma, \ldots, h_p^\gamma\right)^T$$

with $g_i^\gamma = \dfrac{\partial d_\gamma(f_{\boldsymbol{\theta}_1}, f_{\boldsymbol{\theta}_2})}{\partial \theta_{1i}}$, $i = 1, \ldots, p$ and $h_i^\gamma = \dfrac{\partial d_\gamma(f_{\boldsymbol{\theta}_1}, f_{\boldsymbol{\theta}_2})}{\partial \theta_{2i}}$, $i = 1, \ldots, p$. On the other hand

$$\sqrt{n}\boldsymbol{G}_\gamma^T(^{(1)}\widehat{\boldsymbol{\theta}}_\beta - \boldsymbol{\theta}_1) \xrightarrow[n\to\infty]{L} N\left(\boldsymbol{0}, \boldsymbol{J}_\beta^{-1}(\boldsymbol{\theta}_1)\boldsymbol{K}_\beta(\boldsymbol{\theta}_1)\boldsymbol{J}_\beta^{-1}(\boldsymbol{\theta}_1)\right),$$
$$\sqrt{m}\boldsymbol{H}_\gamma^T(^{(2)}\widehat{\boldsymbol{\theta}}_\beta - \boldsymbol{\theta}_2) \xrightarrow[n\to\infty]{L} N\left(\boldsymbol{0}, \boldsymbol{J}_\beta^{-1}(\boldsymbol{\theta}_2)\boldsymbol{K}_\beta(\boldsymbol{\theta}_2)\boldsymbol{J}_\beta^{-1}(\boldsymbol{\theta}_2)\right).$$

Therefore, the random variable

$$\sqrt{\frac{nm}{n+m}}\left(d_\gamma(f_{^{(1)}\widehat{\boldsymbol{\theta}}_\beta}, f_{^{(2)}\widehat{\boldsymbol{\theta}}_\beta}) - d_\gamma(f_{\boldsymbol{\theta}_1}, f_{\boldsymbol{\theta}_2})\right)$$

is asymptotically distributed as a normal distribution with mean zero and variance

$$\sigma_{\gamma,\beta}^2(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2) = \omega\boldsymbol{G}_\gamma^T\boldsymbol{J}_\beta^{-1}(\boldsymbol{\theta}_1)\boldsymbol{K}_\beta(\boldsymbol{\theta}_1)\boldsymbol{J}_\beta^{-1}(\boldsymbol{\theta}_1)\boldsymbol{G}_\gamma + (1-\omega)\boldsymbol{H}_\gamma^T\boldsymbol{J}_\beta^{-1}(\boldsymbol{\theta}_2)\boldsymbol{K}_\beta(\boldsymbol{\theta}_2)\boldsymbol{J}_\beta^{-1}(\boldsymbol{\theta}_2)\boldsymbol{H}_\gamma, \tag{14}$$

because $o(\|^{(1)}\widehat{\boldsymbol{\theta}}_\beta - \boldsymbol{\theta}_1\|) = o_P(n^{-1/2})$ and $o(\|^{(2)}\widehat{\boldsymbol{\theta}}_\beta - \boldsymbol{\theta}_2\|) = o_P(m^{-1/2})$.

As in the one sample case, simple calculations show that the power of the test statistic in (13) is approximately

$$\pi_{m,n,\alpha}^{\beta,\gamma}(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2) = 1 - \Phi\left(\frac{\sqrt{\frac{nm}{n+m}}}{\sigma_{\gamma,\beta}^2(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)}\left(\frac{s_\alpha^{\gamma,\beta}}{2}\frac{n+m}{nm} - d_\gamma(f_{\boldsymbol{\theta}_1}, f_{\boldsymbol{\theta}_2})\right)\right),$$

where $\Phi(x)$ is the standard normal distribution function. If some alternative $\boldsymbol{\theta}_1 \neq \boldsymbol{\theta}_2$ is the true parameter, then the probability of rejecting $\boldsymbol{\theta}_1 = \boldsymbol{\theta}_2$ with the rejection rule $S_\gamma\left(^{(1)}\widehat{\boldsymbol{\theta}}_\beta, ^{(2)}\widehat{\boldsymbol{\theta}}_\beta\right) > s_\alpha^{\beta,\gamma}$, for fixed significance level α, tends to one as $n, m \to \infty$. The test statistic is consistent in Fraser's sense.

4 Simulation study

To focus on a case where the support is bounded on one side and estimation or testing using the minimum ϕ-divergence technique is problematic, we consider the exponential distribution with mean θ and explore the performance of our proposed statistics for testing hypothesis about θ. Routine calculations show that the minimum density power divergence estimator $\widehat{\theta}_\beta$ of θ corresponding to tuning parameter β can be obtained by iteratively solving the equation

$$\widehat{\theta}_\beta = \frac{\sum_{i=1}^n X_i \exp\left\{-\frac{\beta X_i}{\widehat{\theta}_\beta}\right\}}{\sum_{i=1}^n \exp\left\{-\frac{\beta X_i}{\widehat{\theta}_\beta}\right\} - \frac{n\beta}{(1+\beta)^2}}, \tag{15}$$

where $X_1, \ldots, X_n$ represents a random sample of size n from $\mathcal{E}xp(\theta)$, the exponential distribution with mean θ. The MLE of θ (minimum density power divergence estimator with $\beta = 0$) has an explicit expression given by $\widehat{\theta}_0 = n \big/ \sum_{i=1}^n X_i$. For general β, the components in Eq. (6) simplify to provide the expression

$$\text{Var}[n^{\frac{1}{2}}\widehat{\theta}_\beta] = h(\beta)\theta^2, \tag{16}$$

where

$$h(\beta) = \frac{(1+\beta)^2 P(\beta)}{(1+\beta^2)^2(1+2\beta)^3}, \quad P(\beta) = 1+4\beta+9\beta^2+14\beta^3+13\beta^4+8\beta^5+4\beta^6. \tag{17}$$

The function $h(\beta)$ is increasing with β; the relative efficiencies of $\widehat{\theta}_\beta$, therefore, are decreasing with β.

In testing the hypothesis $H_0 : \theta = \theta_0$ against $H_1 : \theta \neq \theta_0$ about a scalar parameter θ, it is readily seen that

$$\frac{T_\gamma(\widehat{\theta}_\beta, \theta_0)}{\lambda^{\gamma,\beta}(\theta_0)} \underset{n\to\infty}{\longrightarrow} \chi_1^2, \tag{18}$$

where $\lambda^{\gamma,\beta}(\theta_0)$ is the unique eigenvalue in question and the tests given by T_γ, ${}^1T_\gamma$ and ${}^2T_\gamma$ are equivalent. The expression of $\lambda^{\gamma,\beta}(\theta_0)$ for $\beta > 0$ is easily obtained through standard calculations as (see Theorem 1)

$$\lambda^{\gamma,\beta}(\theta_0) = a_\gamma(\theta_0)\text{Var}[n^{\frac{1}{2}}\widehat{\theta}_\beta] = \frac{P(\beta)Q(\gamma,\beta)\theta_0^{-\gamma}}{\left(1+\beta^2\right)(1+2\beta)^3},$$

with

$$a_\gamma(\theta_0) = \left.\frac{\partial^2 d_\gamma(f_\theta, f_{\theta_0})}{\partial\theta^2}\right]_{\theta=\theta_0} = \frac{\theta_0^{-\gamma-2}}{(1+\gamma)^2}\left(1+\gamma^2\right),$$

$$d_\gamma(f_\theta, f_{\theta_0}) = \frac{\theta_0^{-\gamma}}{1+\gamma} + \frac{\theta^{-\gamma}}{\gamma(1+\gamma)} - \frac{\gamma+1}{\gamma}\frac{\theta_0^{-\gamma}}{\gamma\frac{\theta}{\theta_0}+1}, \quad Q(\gamma,\beta) = \frac{(1+\gamma^2)(1+\beta)^2}{(1+\beta^2)(1+\gamma)^2}.$$

Note that $Q(\beta, \beta) = 1$ and when $\beta = 0$, $\lambda^{\beta,\beta}(\theta_0) = 1$, as it happens with the likelihood ratio statistic. On the other hand, Eq. (8) gives

$$T_\gamma(\widehat{\theta}_\beta, \theta_0) = \begin{cases} 2n\left(\frac{\theta_0^{-\gamma}}{1+\gamma} + \frac{\widehat{\theta}_\beta^{-\gamma}}{\gamma(1+\gamma)} - \frac{\gamma+1}{\gamma}\frac{\theta_0^{-\gamma}}{\gamma\frac{\widehat{\theta}_\beta}{\theta_0}+1}\right), & \text{if } \gamma > 0 \\ 2n\left(\log\frac{\theta_0}{\widehat{\theta}_\beta} + \left(1 - \frac{\theta_0}{\widehat{\theta}_\beta}\right)\right), & \text{if } \gamma = 0, \end{cases}$$

and hence

$$\frac{T_\gamma(\widehat{\theta}_\beta, \theta_0)}{\lambda^{\gamma,\beta}(\theta_0)} = \begin{cases} 2n\frac{(1+\beta^2)(1+2\beta)^3}{P(\beta)Q(\gamma,\beta)}\frac{1}{\gamma(1+\gamma)}\left(\gamma + \left(\frac{\widehat{\theta}_\beta}{\theta_0}\right)^{-\gamma} - \frac{(1+\gamma)^2}{\gamma\frac{\widehat{\theta}_\beta}{\theta_0}+1}\right), & \text{if } \gamma > 0 \\ 2n\left(\log\frac{\theta_0}{\widehat{\theta}_\beta} + \left(1 - \frac{\theta_0}{\widehat{\theta}_\beta}\right)\right), & \text{if } \gamma = 0. \end{cases} \quad (19)$$

To analyze the performance of our family of test statistics (19), we will compare the observed (empirical) sizes of $\frac{T_\gamma(\widehat{\theta}_\beta)}{\lambda^{\gamma,\beta}(\theta_0)}$ for $\beta = 0, 0.1, 0.2, \ldots, 1$. Without any loss of generality, we choose $\theta_0 = 2$ for our simulation since an observation from $X \sim \mathcal{E}xp(\theta_0)$ can be generated through an observation from $X' \sim \mathcal{E}xp(1)$ as $X = \theta_0 X'$, which means that the ratio $\theta_0/\widehat{\theta}_\beta$ will remain constant and, therefore, so will $T_\gamma(\widehat{\theta}_\beta, \theta_0)/\lambda^{\gamma,\beta}(\theta_0)$; the scale equivariance of the minimum density power divergence estimator is obvious from Eq. (15). Our Monte-Carlo study is performed with 100,000 replications for each sample size.

To study the stability of the level, we generate samples from the exponential mixture

$$f_{\theta_0}^{X(\epsilon)} = (1-\epsilon)f_{\theta_0}(x) + \epsilon f_{2\theta_0}(x), \quad (20)$$

where $f_\theta(x)$ represents the density function of the $\mathcal{E}xp(\theta)$ distribution. The pure model corresponds to $\epsilon = 0$, and in this case we expect the nominal level—here chosen as 0.05—will be closely approximated by the empirical level; the latter is determined, for each ϵ, by the empirical proportion of rejections given by

$$\widehat{\alpha}(\epsilon) = \#\left\{\frac{T_\gamma^{(i)}(\widehat{\theta}_\beta(\epsilon), \theta_0)}{\lambda^{\gamma,\beta}(\theta_0)} > \chi^2_{1,0.05} = 3.84146\right\}/100{,}000, \quad i = 1, \ldots, 100{,}000. \quad (21)$$

Good level stability is indicated if the procedure guards against large displacements in $\hat{\alpha}(\epsilon)$, compared to $\hat{\alpha}(0)$. In our study, we will present the values of $\hat{\alpha}(0.05)$ together with $\hat{\alpha}(0)$.

The performance of a robust test of hypothesis cannot be described by its level stability alone; one must also investigate the stability of the power of the statistic when true distributions outside the null are contaminated leading to possible loss in power. To investigate the power scenario, we generate data from the mixture

$$f_{\theta_0'}^{W(\epsilon)}(w) = (1-\epsilon)f_{\theta_0'}(w) + \epsilon f_{\theta_0'/3}(w) \quad (22)$$

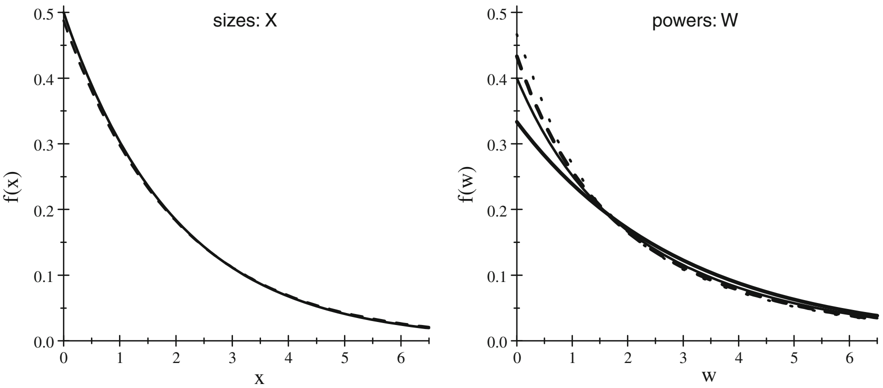

Fig. 1 Probability density functions for the pure and the contaminated distributions

where the real value of the parameter is $\theta_0' = 3\frac{\theta_0}{2}$. For $\epsilon = 0$, this represents the test of $H_0 : \theta = 2$ when $\theta = 3$ is the true parameter, so that one should expect reasonable power. For $\epsilon > 0$ the contaminating proportion brings the mean closer to that of the null, and some loss in power is likely. As in Eq. (21), let the empirical power $\hat{\pi}(\epsilon)$ be the empirical proportion of rejections, and good power stability will be indicated by a slow decline in $\hat{\pi}(\epsilon)$ over increasing ϵ. To compare with $\hat{\pi}(0)$, here we will present values of $\hat{\pi}(\epsilon)$ for $\epsilon = 0.1, 0.15$ and 0.2.

In the first panel of Fig. 1, the density of the pure $\mathcal{E}xp(2)$ distribution is provided by the solid line, while the 5 % contaminated density ($\epsilon = 0.05$) is presented by the dashed line. The shift between the densities is small, almost invisible to the naked eye, but is sufficient to produce an large inflation in the level of the likelihood ratio test in our simulations. In the second panel, the density of the $\mathcal{E}xp(3)$ distribution is presented by the thick solid line, and the three contaminated densities used in our power calculations by the thin solid line ($\epsilon = 0.1$), dashed line ($\epsilon = 0.15$) and the dotted line ($\epsilon = 0.2$). Clearly, the contaminations do force the density somewhat closer to the null, and a major loss in power for the likelihood ratio test is not unexpected.

Tables 1, 2, 3, 4 and 5 provide extensive representations of the simulated sizes and powers, $\widehat{\alpha}(\epsilon)$ and $\widehat{\pi}(\epsilon)$, for several samples sizes, $n \in \{10, 25, 50, 75, 300\}$. We restrict ourselves to the $\beta = \gamma$ case here. The likelihood based method ($\beta = \gamma = 0$) performs well under the model, although values of γ close to zero generate procedures which are competitive in this respect. For contaminated data, the situation changes drastically, and the likelihood based method turns out to be a poor choice. Values of γ around 0.3 or 0.4 appear to provide good compromise solutions, exhibiting small losses at the model and greatly improved performances under contamination. At a sample size of $n = 300$, e.g. a contamination of $\epsilon = 0.2$ pulls down the power of the likelihood ratio test to 28.6 from 100 %, while those for $\gamma = 0.3$ and 0.4 hold their own close to 85 %. Similar stabilities are observed in case of the attained size of these tests. Overall, the density power divergence statistics provide great alternatives to the likelihood ratio test for testing problems about the exponential mean. The boldfaced entries in the tables indicate the optimal results.

Table 1 Simulated sizes and powers with and without contamination when $n = 10$

$\beta = \gamma$	$\widehat{\alpha}(0)$	$\widehat{\alpha}(0.05)$	$\widehat{\pi}(0)$	$\widehat{\pi}(0.1)$	$\widehat{\pi}(0.15)$	$\widehat{\pi}(0.2)$
0.0	**0.05186**	0.06235	0.22627	0.15268	0.13266	0.12187
0.1	0.05205	0.05816	0.23347	0.15528	0.13084	0.11421
0.2	0.05257	0.05497	0.23516	0.16600	0.14089	0.12119
0.3	0.05420	**0.05458**	0.23444	0.17574	0.15182	0.13182
0.4	0.05783	0.05681	0.23412	0.18224	0.16001	0.14048
0.5	0.06135	0.05877	0.23296	0.18568	0.16471	0.14607
0.6	0.06464	0.06141	0.23320	0.18886	0.16833	0.15069
0.7	0.06731	0.06344	0.23352	0.19074	0.17102	0.15389
0.8	0.06967	0.06556	0.23397	0.19295	0.17356	0.15639
0.9	0.07194	0.06740	0.23518	0.19402	0.17499	0.15825
1.0	0.07398	0.06911	**0.23660**	**0.19594**	**0.17678**	**0.15987**

Table 2 Simulated sizes and powers with and without contamination when $n = 25$

$\beta = \gamma$	$\widehat{\alpha}(0)$	$\widehat{\alpha}(0.05)$	$\widehat{\pi}(0)$	$\widehat{\pi}(0.1)$	$\widehat{\pi}(0.15)$	$\widehat{\pi}(0.2)$
0.0	0.05008	0.06590	0.50177	0.24823	0.17799	0.13669
0.1	**0.05004**	0.05913	**0.50161**	0.27259	0.19594	0.14131
0.2	0.04987	0.05465	0.48533	0.29479	0.22168	0.16486
0.3	0.05105	0.05314	0.46323	0.30450	0.23906	0.18488
0.4	0.05232	0.05233	0.44237	**0.30573**	0.24805	0.19661
0.5	0.05330	0.05199	0.42395	0.30251	**0.25010**	0.20299
0.6	0.05494	**0.05188**	0.41001	0.29929	0.25007	0.20526
0.7	0.05601	0.05245	0.39829	0.29561	0.24872	0.20644
0.8	0.05745	0.05279	0.39047	0.29206	0.24697	**0.20673**
0.9	0.05866	0.05322	0.38415	0.28948	0.24547	0.20637
1.0	0.05987	0.05383	0.37952	0.28743	0.24395	0.20603

5 Real data examples

5.1 Telephone-fault data: normal example

We consider the data on telephone line faults presented and analyzed by Welch (1987); the data were also analyzed by Simpson (1989). The data are given in Table 6 and consist of the ordered differences between the inverse test rates and the inverse control rates in 14 matched pairs of areas. A simple parametric approach would be to model these data as a random sample from a normal distribution with mean μ and standard deviation σ. One fact immediately noticeable is that the first observation of this dataset is a huge outlier with respect to the normal model, while the remaining 13 observations appear to be reasonable with respect to the same. In Fig. 2, we present a kernel density estimate for these data, the normal model fit based on the maximum

Table 3 Simulated sizes and powers with and without contamination when $n = 50$

$\beta = \gamma$	$\widehat{\alpha}(0)$	$\widehat{\alpha}(0.05)$	$\widehat{\pi}(0)$	$\widehat{\pi}(0.1)$	$\widehat{\pi}(0.15)$	$\widehat{\pi}(0.2)$
0.0	0.04898	0.07258	**0.81121**	0.38287	0.23805	0.15453
0.1	0.04915	0.06258	0.80430	0.44973	0.29753	0.18599
0.2	0.04910	0.05716	0.78038	0.49277	0.35238	0.23477
0.3	0.04949	0.05430	0.75005	**0.50492**	0.38170	0.27141
0.4	**0.05014**	0.05261	0.71781	0.50164	**0.39232**	0.29166
0.5	0.05060	0.05186	0.68791	0.49089	0.39220	0.29991
0.6	0.05128	0.05100	0.66197	0.47797	0.38663	**0.30164**
0.7	0.05194	**0.05024**	0.63966	0.46579	0.38020	0.29908
0.8	0.05267	0.04978	0.62166	0.45447	0.37279	0.29647
0.9	0.05298	0.04961	0.60760	0.44483	0.36586	0.29284
1.0	0.05359	0.04946	0.59591	0.43697	0.36046	0.28970

Table 4 Simulated sizes and powers with and without contamination when $n = 75$

$\beta = \gamma$	$\widehat{\alpha}(0)$	$\widehat{\alpha}(0.05)$	$\widehat{\pi}(0)$	$\widehat{\pi}(0.1)$	$\widehat{\pi}(0.15)$	$\widehat{\pi}(0.2)$
0.0	0.05014	0.08150	**0.93975**	0.49559	0.29245	0.16693
0.1	0.05020	0.07037	0.93475	0.59902	0.39282	0.22999
0.2	0.04972	0.06273	0.91925	0.65181	0.47018	0.30457
0.3	0.04988	0.05787	0.89802	**0.66476**	0.50925	0.35600
0.4	**0.05008**	0.05501	0.87429	0.65893	0.52161	0.38406
0.5	0.05029	0.05317	0.84879	0.64511	**0.51979**	0.39510
0.6	0.05081	0.05181	0.82460	0.62877	0.51209	**0.39577**
0.7	0.05098	0.05063	0.80295	0.61258	0.50145	0.39197
0.8	0.05157	**0.04964**	0.78486	0.59735	0.49058	0.38568
0.9	0.05169	0.04906	0.76840	0.58388	0.48043	0.37904
1.0	0.05201	0.04881	0.75507	0.57196	0.47246	0.37332

likelihood estimates of μ and σ, a normal model fit based on the minimum density power divergence estimates (with tuning parameter 0.15) of the normal parameters, and a normal model fit based on the minimum Hellinger distance estimates of the normal parameters (Simpson 1989). The figure shows that if a small hump to the extreme left could be ignored, these data would have a nice unimodal structure which could be well modeled by an appropriate normal density. Apart from the minimum Hellinger distance estimates, such a normal density is provided in this figure by the minimum density power divergence estimates which correspond to $\mu = 121.3$ and $\sigma = 134.2$. The maximum likelihood estimates, on the other hand, try to be inclusive and generate a result which neither models the outlier deleted data, nor provides a fit to the outlier component.

For the full data, the t test for the null hypothesis $H_0 : \mu = 0$ against $H_1 : \mu > 0$ fails to reject the null due to the presence of the large outlier; however, the

Table 5 Simulated sizes and powers with and without contamination when $n = 300$

$\beta = \gamma$	$\widehat{\alpha}(0)$	$\widehat{\alpha}(0.05)$	$\widehat{\pi}(0)$	$\widehat{\pi}(0.1)$	$\widehat{\pi}(0.15)$	$\widehat{\pi}(0.2)$
0.0	0.04945	0.14692	**1.00000**	0.942270	0.66559	0.28573
0.1	0.04949	0.11434	**1.00000**	0.98903	0.87863	0.56135
0.2	0.04962	0.09495	**1.00000**	0.99579	0.94733	0.75300
0.3	0.04962	0.08261	**1.00000**	**0.99685**	0.96668	0.83884
0.4	0.04979	0.07404	**1.00000**	0.99647	**0.97114**	0.87163
0.5	0.04982	0.06889	0.99998	0.99552	0.97068	0.88287
0.6	0.05014	0.06507	0.99989	0.99408	0.96769	**0.88387**
0.7	0.04990	0.06241	0.99988	0.99228	0.96345	0.87870
0.8	0.05002	0.06038	0.99983	0.99006	0.95829	0.87117
0.9	**0.04999**	0.05847	0.99970	0.98768	0.95242	0.86276
1.0	0.05028	**0.05743**	0.99959	0.985310	0.94677	0.85388

Table 6 Telephone-line faults data

Pair	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Difference	−988	−135	−78	3	59	83	93	110	189	197	204	229	289	310

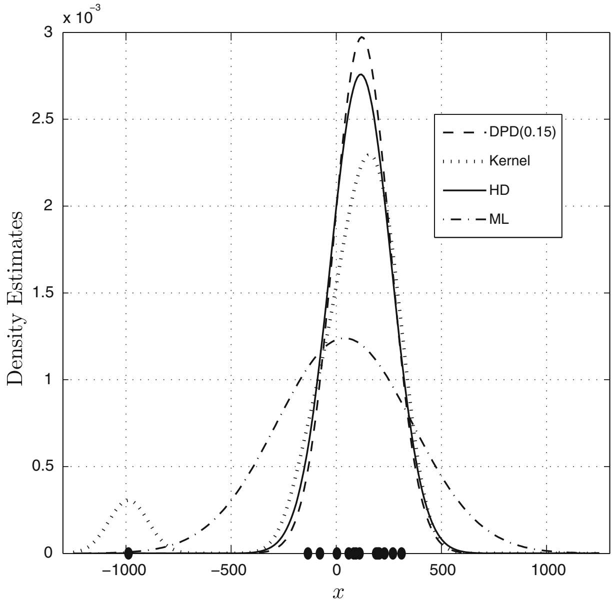

Fig. 2 Kernel density estimate and different normal fits for the drosophila data

robust Hellinger deviance test (Simpson 1989) comfortably rejects the null, as does the t-test based on the cleaned data after the removal of the large outlier. The minimum Hellinger distance estimate of the mean parameter as reported by Simpson (116.8) is

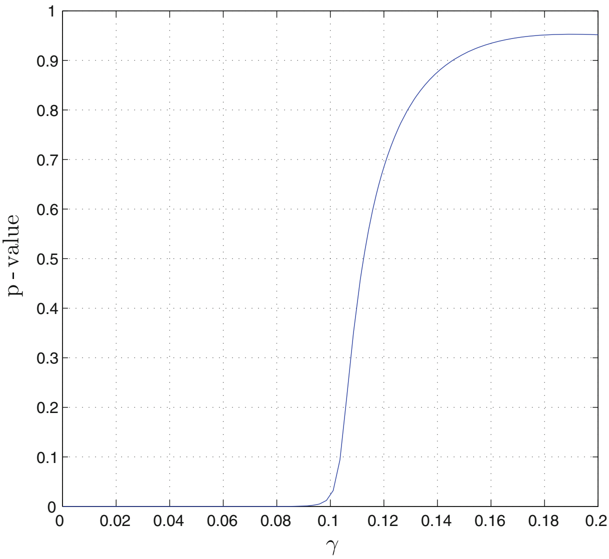

Fig. 3 The p values of the density power divergence tests for $H_0 : \sigma = 132$ against $H_1 : \sigma \neq 132$, under normal model with $\mu = 115$, for different values of γ

remarkably close to the maximum likelihood estimate of the mean parameter based on the cleaned data (117.9).

Under the normal model, the usual estimates of scale are highly inflated due to the presence of the large outlier, and as a result the likelihood ratio test under the normal model is likely to reject null hypotheses about the scale parameter where the null value is chosen to be close to the standard deviation of the normal model fitting the last 13 observations (132.82), but far off from the standard deviation based on the full data (321.94). From the robustness perspective, this is precisely what we will like to avoid, and here we demonstrate that proper choices of the tuning parameter within the class of tests developed in this paper achieve this goal. We consider a $N(115, \sigma^2)$ model for the data, and test the null hypotheses $H_0 : \sigma = \sigma_0$ against $H_1 : \sigma \neq \sigma_0$ for the null value of $\sigma_0 = 132$. For the density power divergence test with $\beta = \gamma$, the expression for the sole eigenvalue is given by

$$\frac{(1+\gamma)^{7/2}}{\sigma_0^{2\gamma}(2\pi)^{\gamma/2}(\gamma^2+2)}\left[\frac{4\gamma^2+2}{(1+2\gamma)^{5/2}} - \frac{\gamma^2}{(1+\gamma)^3}\right],$$

which reduces to 1 when $\gamma = 0$. Figure 3 represents the p values of the tests for different values of γ in a region of interest. While it is clearly seen that the tests fail to recognize 132 as a likely scale value for these data at very small values of γ, the decision turns in favor of $\sigma_0 = 132$, sharply, as γ crosses and goes beyond 0.1. This stable behavior of the test statistic based on the density power divergence approximately coincides with the stability of the density power divergence estimate of σ itself, which is presented in Fig. 4. The density power divergence estimate of σ (for $\mu = 115$) drops from the neighborhood of 319.41 at $\gamma = 0$ to approximately 153.79 at $\gamma = 0.11$,

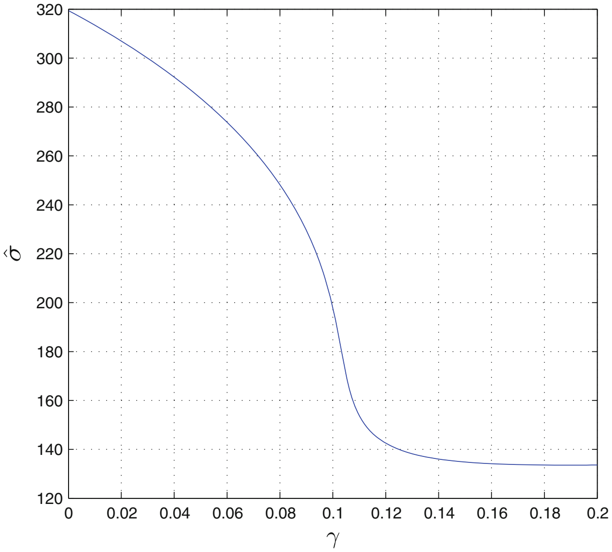

Fig. 4 Estimates of σ for the telephone line fault data under the normal model

about 142.56 at $\gamma = 0.12$, and about 134.81 at $\gamma = 0.15$. At least in this example, the robustness of the test statistic is clearly linked to the robustness of the estimator.

We also consider testing the hypotheses $H_0 : \mu = 115$, $\sigma = 132$ against the not equal to alternative. For this problem, where the dimension of the parameter vector is greater than 1, we have discussed four approximations to the null distribution in Sect. 3.1. The hypothesis testing results based on these four approximations are quite similar, and to keep a clear focus in our presentations we have presented only the results for the statistic ${}^2T_\gamma(\widehat{\boldsymbol{\theta}}_\beta, \boldsymbol{\theta}_0)$ with $\gamma = \beta$. Figure 5 presents the p values of this test for values of $\gamma \in [0, 0.2]$, and the observations are quite similar to those of Fig. 3.

5.2 Two-sample drosophila data: Poisson example

Here, we present a two-sample real data example. This experiment, considered in Woodruff et al. (1984) and Simpson (1989), is known to produce occasional spurious counts. Male flies were exposed either to a certain degree of chemical to be screened or to control conditions and the responses are the numbers of recessive lethal mutations among the daughters of such flies. The responses are assumed to be Poisson with means θ_1 (control group) and θ_2 (treated group) respectively. The data are given in Table 7. We want to test the hypothesis

$$H_0 : \theta_1 \geq \theta_2 \text{ against } H_1 : \theta_1 < \theta_2. \tag{23}$$

The apprehension is that the presence of the two large counts for the treated group may make the mean of the second group appear larger, although this conclusion may not be supported by the rest of the data.

Suppose $P_{\text{oisson}}(\theta)$ denotes the Poisson probability mass function with parameter θ, n and m are the sample size from the two populations, ${}^{(1)}\widehat{\theta}_\beta$ and ${}^{(2)}\widehat{\theta}_\beta$ are the

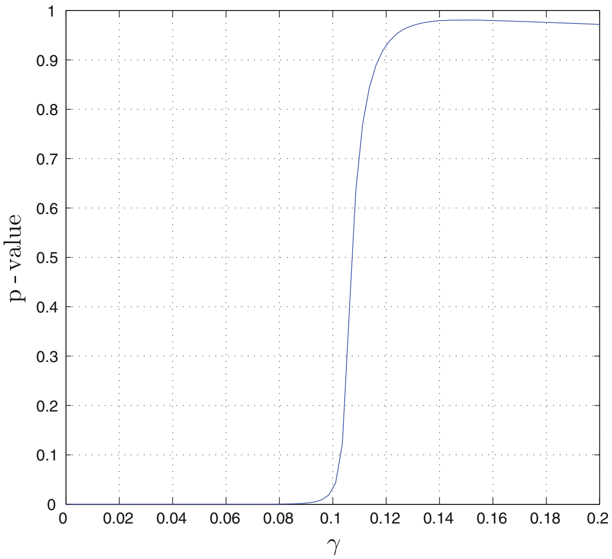

Fig. 5 The p values of the density power divergence tests for $H_0 : \sigma = 132$ and $\mu = 115$ against $H_0 : \sigma \neq 132$ or $\mu \neq 115$, under normal model, for different values of γ

Table 7 Frequencies of the number of recessive lethal daughters for Drosophila data

	0	1	2	3	4	5	6	7
Observed (control)	159	15	3	0	0	0	0	0
Observed (treated)	110	11	5	0	0	0	1	1

minimum density power divergence estimators (MDPDEs) of the two populations and ${}^{(\text{Null})}\widehat{\theta}_\beta$ is the common MDPDE under the null. We consider the hypothesis

$$H_0 : \theta_1 = \theta_2 \text{ against } H_1 : \theta_1 \neq \theta_2. \tag{24}$$

It follows from Theorem 6 that the test statistic for testing the last set of hypotheses is given by

$${}^*S_\gamma\left({}^{(1)}\widehat{\theta}_\beta, {}^{(2)}\widehat{\theta}_\beta\right) = \frac{1}{\lambda\left({}^{(\text{Null})}\widehat{\theta}_\beta\right)} \frac{2nm}{(m+n)} d_\gamma\left(P_{oisson}\left({}^{(1)}\widehat{\theta}_\beta\right), P_{oisson}\left({}^{(2)}\widehat{\theta}_\beta\right)\right),$$

where

$$\lambda\left({}^{(\text{Null})}\widehat{\theta}_\beta\right) = A_\gamma\left({}^{(\text{Null})}\widehat{\theta}_\beta\right) K_\beta\left({}^{(\text{Null})}\widehat{\theta}_\beta\right) \Big/ J_\beta^2\left({}^{(\text{Null})}\widehat{\theta}_\beta\right),$$

$$a_\gamma(\theta) = (1+\gamma) \sum_{x=0}^{\infty} \left(\frac{e^{-\theta}\theta^x}{x!}\right)^{\gamma+1} \left(\frac{x}{\theta} - 1\right)^2,$$

$$J_\beta(\theta) = \sum_{x=0}^{\infty} \left(\frac{e^{-\theta}\theta^x}{x!}\right)^{1+\beta} \left(\frac{x}{\theta} - 1\right)^2$$

Table 8 Estimated Poisson parameters for the two-sample example

	$\beta = 0$	$\beta = 0.25$	$\beta = 0.5$	$\beta = 0.75$	$\beta = 1$
${}^{(1)}\widehat{\theta}_\beta$	0.1186	0.1075	0.1021	0.1008	0.1016
${}^{(2)}\widehat{\theta}_\beta$	0.2656	0.1418	0.1270	0.1249	0.1287
${}^{(\mathrm{Null})}\widehat{\theta}_\beta$	0.1803	0.1211	0.1120	0.1105	0.1127

Table 9 The signed test statistics and the associated p values for the two-sample data

$\beta = \gamma$	Full data	Outlier deleted data
LRT	2.9586 (0.0015)	1.0986 (0.1359)
0.00	2.7622 (0.0029)	1.0646 (0.1435)
0.25	0.8043 (0.2106)	0.7400 (0.2296)
0.50	0.5991 (0.2746)	0.5151 (0.3032)
0.75	0.5835 (0.2798)	0.4475 (0.3273)
1.00	0.6543 (0.2565)	0.4638 (0.3311)

and

$$K_\beta(\theta) = \sum_{x=0}^{\infty} \left(\frac{e^{-\theta}\theta^x}{x!} \right)^{1+2\beta} \left(\frac{x}{\theta} - 1 \right)^2 - \left[\sum_{x=0}^{\infty} \left(\frac{e^{-\theta}\theta^x}{x!} \right)^{1+\beta} \left(\frac{x}{\theta} - 1 \right) \right]^2 .$$

The asymptotic distribution of the statistic ${}^*S_\gamma({}^{(1)}\widehat{\theta}_\beta, {}^{(2)}\widehat{\theta}_\beta)$ is Chi-square with one degree of freedom. An approximate test statistic for testing (23) is then given by the signed square root

$${}^*T^{\beta,\gamma} = \sqrt{{}^*S_\gamma\left({}^{(1)}\widehat{\theta}_\beta, {}^{(2)}\widehat{\theta}_\beta\right)}\, sign\left({}^{(2)}\widehat{\theta}_\beta - {}^{(1)}\widehat{\theta}_\beta\right)$$

whose asymptotic distribution may be approximated by the standard normal distribution. Table 8 presents the estimators for different values of β for the null and the unrestricted situations for the full data.

The value of the signed square root of the likelihood ratio test (LRT) statistic is 2.9586, which gives a p value of 0.0015 using normal approximation. It is clear that this apparently significant result is due to the two large observations in the treated group; the exclusion of the outliers produces a signed statistic of 1.0986 for the likelihood ratio test with an associated p value of 0.1359. Unlike the one sample problem, the statistic for $\gamma = \beta = 0$ case no longer equals the likelihood ratio statistic; however, the density power divergence test for this case exhibits the same behavior. But all the other signed square root statistics generate p values which are insignificant by any usual standard, whether with or without the outliers. Thus, the two outliers do not affect the decision for our robust tests, although the likelihood ratio test and the $\gamma = \beta = 0$ statistic fail to hold up against the outliers. The values of ${}^*T^{\beta,\gamma}$ for the full and the outlier deleted (with the two outliers removed) data are given in Table 9.

6 Concluding remarks

The density power divergence family has already demonstrated its worth in robust parametric estimation. The application of the same divergence for parametric hypothesis testing problems is more tricky since usual drop-in-divergence type tests are difficult to adapt to this scenario. Here we have proposed a statistic based on the above divergence which generates a structure that is asymptotically described by a linear combination of independent χ^2 variables. Within the domain of its applicability, our numerical results show that the tests based on moderate values of γ (such as those between 0.1 and 0.4) can work as very useful robust alternatives to the likelihood ratio test.

In order to keep a clear focus in our presentations, the discussion, derivation and implications of this paper has been kept limited to the case of the simple null. While our examples and simulation results have demonstrated the worth of our method in these situations, for greater scope of application it will be important to extend these results to the case of the composite null. We hope to undertake this extension in a sequel paper.

Acknowledgments This work was partially supported by Grant MTM 2009-10072. The authors gratefully acknowledge the suggestions of the Associate Editor and two anonymous referees which led to an improved version of the paper.

References

Basu, A., Harris, I. R., Hjort, N. L., Jones, M. C. (1998). Robust and efficient estimation by minimizing a density power divergence. *Biometrika, 85*, 549–559.

Basu, A., Shioya, H., Park, C. (2011). *Statistical inference: the minimum distance approach*. Boca Raton: Chapman Hall/CRC.

Csiszár, I. (1991). Why least squares and maximum entropy? An axiomatic approach to inference for linear inverse problems. *The Annal of Statistics, 19*, 2032–2066.

De Angelis, D., Young, G. (1992). Smoothing the bootstrap. *International Statistical Review, 60*, 45–56.

Dik, J. J., de Gunst, M. C. M. (1985). The distribution of general quadratic forms in normal variables. *Statistica Neerlandica, 39*, 14–26.

Eckler, A. R. (1969). A survey of coverage problems associated with point and area targets. *Technometrics, 11*, 561–589.

Fraser, D. A. S. (1956). Sufficient statistics with nuisance parameters. *The Annal of Mathematical Statistics, 27*, 838–842.

Gupta, S. S. (1963). Bibliography on the multivariate normal integrals and related topics. *The Annal of Mathematical Statistics, 34*, 829–838.

Harville, D. A. (2008). *Matrix Algebra from a statistician's perspective*. New York: Springer.

Heritier, S., Ronchetti, E. (1994). Robust Bounded-influence Tests in General Parametric Models. *Journal of the American Statistical Association, 89*, 897–904.

Jensen, D. R., Solomon, H. (1972). A Gaussian approximation to the distribution of a definite quadratic form. *Journal of the American Statistical Association, 67*(340), 898–902.

Johnson, N. L., Kotz, S. (1968). Tables of distributions of positive definite quadratic forms in central normal variables. *Sankhya B—The Indian Journal of Statistics, 30*, 303–314.

Jones, M. C., Hjort, N. L., Harris, I. R., Basu, A. (2001). A comparison of related density-based minimum divergence methods. *Biometrika, 88*, 865–873

Lindsay, B. G. (1994). Efficiency versus robustness: the case for minimum Hellinger distance and related methods. *The Annals of Statistics, 22*, 1081–1114.

Modarres, R., Jernigan, R. W. (1992). Testing the equality of correlation matrices. *Communications in Statistics—Theory and Methods, 21*, 2107–2125.

Pardo, L. (2006). *Statistical inference based on divergence measures*. Boca Raton: Chapman Hall/CRC.

Rao, J. N. K., Scott, A. J. (1981). The analysis of categorical data from complex sample surveys: Chi-squared tests for goodness of fit and independence in two-way tables. *Journal of the American Statistical Association, 76*, 221–230.

Satterthwaite, F. E. (1946). An approximate distribution of estimates of variance components. *Biometrics, 2*, 110–114.

Simpson, D. G. (1989). Hellinger deviance tests: Efficiency, breakdown points and examples. *Journal of the American Statistical Association, 84*, 107–113.

Solomon, H. (1960). Distribution of quadratic forms: tables and applications. Applied Mathematics and Statistics Laboratories, Technical Report 45, Stanford University, Stanford.

Welch, W. J. (1987). Rerandomizing the median in matched-pair designs. *Biometrika, 74*, 609–614.

Woodruff, R. C., Mason, J. M., Valencia, R., Zimmering, S. (1984). Chemical mutagenesis testing in Drosophila: I. Comparison of positive and negative control data for sex-linked recessive lethal mutations and reciprocal translocations in three laboratories. *Environmental mutagenesis, 6*, 189–202.