Brazilian Journal of Probability and Statistics
2009, Vol. 23, No. 2, 196–215
DOI: 10.1214/08-BJPS008

Multi-sample Rényi test statistics

Tomáš Hobza[a], Isabel Molina[b] and Domingo Morales[c]

[a]*Czech Technical University and Institute of Information Theory and Automation of the ASCR*
[b]*Universidad Carlos III de Madrid*
[c]*Miguel Hernández University of Elche*

Abstract. This paper focuses on testing composite hypotheses about parameters of s independent samples of different sizes. With this purpose, it introduces test statistics based on the family of Rényi divergences between likelihoods. The asymptotic distributions of the proposed test statistics and of the likelihood ratio statistic are derived under standard regularity assumptions. An application to test the homogeneity of variances in data from families belonging to different populations is described and, under this setup, a simulation experiment compares the small sample performance of the likelihood ratio test and some members of the Rényi family of tests. The experiment indicates that some of the Rényi tests perform better under null hypothesis.

1 Introduction

The likelihood ratio test is a standard tool for testing a general hypothesis about parameters of one population. It works by calculating a measure of deviation between the maximum likelihood achieved under the null hypothesis and the maximum achieved over the whole parameter space. Following similar philosophy but using different measures of deviation such as divergences, different tests can be obtained. Some tests based on divergences have already been proposed, and the literature indicates that in many cases these tests represent good competitors to classical tests. Our first available reference on this issue is Kupperman (1957), who suggested to test a simple null hypothesis using the Kullback–Leibler divergence (Kullback (1959)), providing its asymptotic distribution. Salicrú et al. (1994), Morales, Pardo and Vajda (1997, 2000) and Morales et al. (2004) extended these results to the problem of testing composite hypotheses using families of divergences such as Csiszár's ϕ-divergence (Csiszár (2006)) or the Rényi family of divergences (Rényi, 1961).

This work proposes to construct statistics based on the Rényi family of divergences, for testing general composite hypotheses about parameters of s populations. Comparing characteristics of several populations is a problem that appears in many practical applications. An interesting example is the familial data problem, in which families coming from different populations follow multivariate normal

Key words and phrases. Rényi divergence, divergence statistics, testing composite hypotheses, homogeneity of variances.

Received November 2007; accepted April 2008.

distributions with a specified covariance structure common for all families of the same population. Under this setup, it is of interest to test whether some parameters are common for all populations or not. Section 4 introduces this application and describes a simulation experiment for testing the homogeneity of variances across populations.

Under a general setup, this paper defines the family of Rényi test statistics depending on a parameter that can be moved to provide different members of the family. This family includes some well-known test statistics such as the Kullback–Leibler statistic (KLS) and also the likelihood ratio statistic (LRS) when the distributions at hand belong to the exponential family. The widespread use of the LRS may not be so justified since for finite samples it may not perform better than other members. In fact, in the simulation experiments described in Section 4.2, other Rényi tests performed better than the LR test under null hypothesis. Thus, the family of Rényi tests allows flexibility in the selection of a particular test for the problem at hand. Moreover, spelled-out formulas for the divergences can be easily obtained for probability distributions belonging to the exponential family (Morales, Pardo and Vajda (2000)), and these formulas may not be much more complicated than the expression of the LRS, see for example, formulas (4.4) and (4.5) of this paper.

The problem of testing under multiple populations was considered before by Morales, Pardo and Pardo (2001), but their results are not applicable when there is some equality restriction on the parameters of the s populations. Hobza, Molina and Morales (2003) treated a particular case of this situation. They defined the family of Rényi statistics for testing the equality of intraclass correlations of several multivariate normal populations under the setup of the familial data problem. Under this particular data structure, they obtained the asymptotic distributions of the Rényi test statistics. Here we extend the results of Hobza, Molina and Morales (2003) to general hypotheses and general populations, in which equality restrictions on some of the parameters of the s populations are allowed. In order to define the test statistics under the setup of several populations (Section 2), we shall consider the product statistical space, the joint sample and the product of likelihoods of each sample. The null hypothesis will be formulated in terms of a vector γ containing all the different parameters of the s populations. With these considerations, the statistics for testing a statement about γ are defined analogously to the one-sample case. In Section 3, the asymptotic distribution of the proposed multi-sample test statistics is obtained under regular null hypotheses. The proofs are based on the asymptotic equivalence of test statistics to quadratic forms and some of them are presented in the Appendix. In Section 4, an application is given for testing the homogeneity of variances of several multivariate normal populations having the structure of the familial data problem, and simulations are developed for studying small-sample properties of the proposed tests.

2 Definition of the problem and test statistics

In this section we introduce notation and formulate the problem of testing hypotheses about several populations with a common set of parameters. Under this setup, the Rényi family of test statistics is introduced.

Let $(\mathcal{X}_i, \mathcal{B}_{\mathcal{X}_i}, P_{i,\theta_i})_{\theta_i \in \Theta_i}$, $i = 1, \ldots, s$, be statistical spaces associated with independent populations, where $\mathcal{X}_i \subset \mathbb{R}^{p_i}$ is the sample space, $\mathcal{B}_{\mathcal{X}_i}$ is the Borel σ-field of subsets of $\mathcal{X}_i$, $\Theta_i \subset \mathbb{R}^{k_i}$ is an open set, and f_{i,θ_i} is the p.d.f. of P_{i,θ_i} with respect to a σ-finite measure μ_i, $i = 1, \ldots, s$. Assume that from the ith population, a sample $(X_{i1}, \ldots, X_{in_i})$ of independent and identically distributed random variables with common p.d.f. f_{i,θ_i}, is extracted, $i = 1, \ldots, s$, and that the s samples are independent. Let $\mu \triangleq \mu_1^{n_1} \otimes \cdots \otimes \mu_s^{n_s}$ be the product measure and $\mathcal{X} \triangleq \mathcal{X}_1^{n_1} \times \cdots \times \mathcal{X}_s^{n_s}$ be the product sample space. Suppose that the sample sizes n_i tend to infinity at the same rate, that is, if $n = \sum_{i=1}^s n_i$, then

$$\frac{n_i}{n} \underset{\substack{n_1 \longrightarrow \infty \\ \vdots \\ n_s \longrightarrow \infty}}{\longrightarrow} \lambda_i \in (0,1), \qquad i = 1, \ldots, s, \tag{2.1}$$

where $\sum_{i=1}^s \lambda_i = 1$. Unless otherwise explicitly stated, in this paper all convergence results and symbols $o_P(1)$ and $O_P(1)$ are referred to $n_1 \to \infty, \ldots, n_s \to \infty$ satisfying (2.1).

Suppose that the parameters $\theta_i = (\theta_{i,1}, \ldots, \theta_{i,k_i})^t$, $i = 1, \ldots, s$, from the s populations have the same k first components, that is,

$$\theta_{1,\ell} = \theta_{2,\ell} = \cdots = \theta_{s,\ell}, \qquad \ell = 1, \ldots, k, \tag{2.2}$$

where $k \le \min\{k_1, \ldots, k_s\}$. Let $(x_{i1}, \ldots, x_{in_i})$ be a realization of the sample $(X_{i1}, \ldots, X_{in_i})$, $i = 1, \ldots, s$. We define the *joint sample* as

$$x = (x_{11}, \ldots, x_{1n_1}; x_{21}, \ldots, x_{2n_2}; \ldots; x_{s1}, \ldots, x_{sn_s})$$

and the *joint parameter* as the vector with all different parameters of the s populations,

$$\gamma = (\theta_{1,1}, \ldots, \theta_{1,k_1}; \theta_{2,k+1}, \ldots, \theta_{2,k_2}; \ldots; \theta_{s,k+1}, \ldots, \theta_{s,k_s})^t.$$

We assume that $\gamma \in \Gamma$, where the parameter space Γ is an open subset of $\mathbb{R}^M$ and $M = \sum_{i=1}^s k_i - (s-1)k$. This work deals with testing a composite null hypothesis about γ, that is, with testing

$$H_0 : \gamma \in \Gamma_0 \quad \text{versus} \quad H_1 : \gamma \in \Gamma_1, \tag{2.3}$$

where $\Gamma_0 \subset \Gamma$ and $\Gamma_1 = \Gamma - \Gamma_0$. The set Γ_0 defining the null hypothesis is assumed to satisfy the regularity condition (A3) stated in the Appendix. The likelihood and log-likelihood of θ_i based on the ith sample $x_i = (x_{i1}, \ldots, x_{in_i})$ are given by

$$f_{i,\theta_i}(x_i) = \prod_{j=1}^{n_i} f_{i,\theta_i}(x_{ij}) \quad \text{and} \quad l_i(\theta_i) = \sum_{j=1}^{n_i} \log f_{i,\theta_i}(x_{ij}).$$

Since the populations are independent, the likelihood and log-likelihood of the joint parameter $\gamma = (\gamma_1, \ldots, \gamma_M)^t$ based on the joint sample x are, respectively,

$$f_\gamma(x) = \prod_{i=1}^{s} f_{i,\theta_i}(x_i) \quad \text{and} \quad l(\gamma) = \sum_{i=1}^{s} l_i(\theta_i).$$

Let $I^i(\theta_i)$ denote the Fisher Information matrix of f_{i,θ_i}. We split $I^i(\theta_i)$ into blocks as follows:

$$I^i(\theta_i) = \left(\begin{array}{c|c} I^i_{k,k}(\theta_i) & I^i_{k,k_i}(\theta_i) \\ \hline I^i_{k_i,k}(\theta_i) & I^i_{k_i,k_i}(\theta_i) \end{array}\right)_{k_i \times k_i},$$

where $I^i_{k,k}(\theta_i)$, $I^i_{k,k_i}(\theta_i)$, $I^i_{k_i,k}(\theta_i)$, and $I^i_{k_i,k_i}(\theta_i)$ are the submatrices with sizes $k \times k$, $k \times (k_i - k)$, $(k_i - k) \times k$, and $(k_i - k) \times (k_i - k)$, respectively. The following matrix, constructed using the s Fisher information matrices, will play the fundamental role of the Fisher information matrix in this multi-sample problem

$$V(\gamma) = \left(\begin{array}{c|c|c|c} \sum_{i=1}^{s} \lambda_i I^i_{k,k}(\theta_i) & \lambda_1 I^1_{k,k_1}(\theta_1) & \cdots & \lambda_s I^s_{k,k_s}(\theta_s) \\ \hline \lambda_1 I^1_{k_1,k}(\theta_1) & \lambda_1 I^1_{k_1,k_1}(\theta_1) & \mathbf{0} & \mathbf{0} \\ \hline \vdots & \mathbf{0} & \ddots & \mathbf{0} \\ \hline \lambda_s I^s_{k_s,k}(\theta_s) & \mathbf{0} & \mathbf{0} & \lambda_s I^s_{k_s,k_s}(\theta_s) \end{array}\right)_{M \times M}. \tag{2.4}$$

Maximum likelihood estimators are generally obtained by solving the likelihood equations

$$\partial l(\gamma)/\partial \gamma_p = \sum_{i=1}^{s} \sum_{j=1}^{n_i} \partial \log f_{i,\theta_i}(x_{ij})/\partial \gamma_p = 0, \qquad p = 1, \ldots, M. \tag{2.5}$$

Let $\widehat{\gamma} = \widehat{\gamma}(x)$ denote a consistent sequence of solutions of (2.5). Observe that in virtue of assumption (A3) of the Appendix, H_0 can be expressed as $H_0 : \gamma = g(\beta)$, for $\beta = (\beta_1, \ldots, \beta_{M_0})^t$. Then, the likelihood equations for the model restricted to H_0 are

$$\partial l(g(\beta))/\partial \beta_p = 0, \qquad p = 1, \ldots, M_0. \tag{2.6}$$

A consistent sequence of solutions of (2.6) will be denoted $\widehat{\beta} = \widehat{\beta}(x)$.

In this paper we consider several statistics for testing (2.3). The first one is the *likelihood ratio statistic* (LRS), defined as

$$\lambda(\widehat{\gamma}, g(\widehat{\beta})) = -2 \log\big(f_{g(\widehat{\beta})}(x)/f_{\widehat{\gamma}}(x)\big).$$

In Theorem 3 we prove that the asymptotic distribution of the LRS for testing (2.3) is $\chi^2_{M-M_0}$. Thus, the *likelihood ratio test* (LRT) with asymptotic significance level α is the decision rule

$$\text{Reject } H_0 \text{ if } \lambda(\widehat{\gamma}, g(\widehat{\beta})) > \mathcal{X}^2_{M-M_0, 1-\alpha},$$

where $\mathcal{X}^2_{M-M_0,1-\alpha}$ is the $(1-\alpha)$-quantile of a chi-squared distribution with $M-M_0$ degrees of freedom.

We consider also test statistics based on the family of Rényi divergences. Following the definition given by Liese and Vajda (1987), the Rényi divergence of order a between $f_{\widehat{\gamma}}(x)$ and $f_{g(\widehat{\beta})}(x)$ is obtained by the formula

$$D_a(\widehat{\gamma}, g(\widehat{\beta})) = \frac{1}{a(a-1)} \log \int_{\mathcal{X}} f^a_{\widehat{\gamma}}(x) f^{1-a}_{g(\widehat{\beta})}(x)\, d\mu(x), \qquad a \in \mathbb{R} - \{0, 1\}.$$

Using these divergences, the *Rényi statistic* (RS) of order a is defined as

$$2D_a(\widehat{\gamma}, g(\widehat{\beta})), \qquad a \in \mathbb{R} - \{0, 1\}.$$

In Theorem 1 we show that, under regularity assumptions, the asymptotic distribution of the RS of order $a \in \mathbb{R} - \{0, 1\}$ is also $\chi^2_{M-M_0}$. Therefore, the *Rényi test* (RT) of order $a \in \mathbb{R} - \{0, 1\}$ and asymptotic significance level α is given by the decision rule

$$\text{Reject } H_0 \text{ if } 2D_a(\widehat{\gamma}, g(\widehat{\beta})) > \mathcal{X}^2_{M-M_0,1-\alpha}. \tag{2.7}$$

Finally, we consider the statistic based on the Kullback–Leibler divergence between $f_{\widehat{\gamma}}(x)$ and $f_{g(\widehat{\beta})}(x)$. This divergence is obtained by taking limit as $a \to 1$ in the Rényi divergence of order a, and is equal to

$$D_1(\widehat{\gamma}, g(\widehat{\beta})) = \int_{\mathcal{X}} f_{\widehat{\gamma}}(x) \log\big(f_{\widehat{\gamma}}(x)/f_{g(\widehat{\beta})}(x)\big)\, d\mu(x).$$

Thus, the *Kullback–Leibler statistic* (KLS) for testing (2.3) is defined as

$$2D_1(\widehat{\gamma}, g(\widehat{\beta})) \tag{2.8}$$

and the corresponding test is obtained by taking $a = 1$ in (2.7).

3 Asymptotic distribution of test statistics

In this section we derive the asymptotic distributions of the test statistics for testing (2.3) that were introduced in Section 2, namely the likelihood ratio and the Rényi family of test statistics, which includes the Kullback–Leibler statistic. This is achieved by taking second-order Taylor expansions that lead, up to terms of order $o_P(1)$, to particular quadratic forms. Theorem 5 of the Appendix gives the asymptotic distribution of these quadratic forms. The regularity assumptions (A1)–(A3) required for the results of this section are listed at the beginning of the Appendix. Hereafter γ^0 and β^0 denote the true values of γ and β.

Theorem 1. *Let the null hypothesis $H_0 : \gamma \in \Gamma_0$ be true, where Γ_0 satisfies* (A3). *For each $i = 1, \dots, s$, let $(X_{i1}, \dots, X_{in_i})$ be independent samples of i.i.d. random variables with common p.d.f. $f_{i\theta_i}(x)$ satisfying* (2.1), (A1) *and* (A2). *Assume further that there exist measurable and μ-integrable functions $P_1, P_2, P_3 : \mathcal{X} \to [0, \infty)$, possibly depending on γ^0, such that for each γ^1, γ^2 in a neighborhood $N(\gamma^0)$ and for each $1 \le p, q \le M$, it holds*

(H1) $|f^a_{\gamma^1}(x) f^{1-a}_{\gamma^2}(x)| \le P_1(x), x \in \mathcal{X}$,
(H2) $|\frac{\partial}{\partial \gamma^1_p} f^a_{\gamma^1}(x) f^{1-a}_{\gamma^2}(x)| \le P_2(x), x \in \mathcal{X}$,
(H3) $|\frac{\partial^2}{\partial \gamma^1_p \partial \gamma^1_q} f^a_{\gamma^1}(x) f^{1-a}_{\gamma^2}(x)| \le P_3(x), x \in \mathcal{X}$.

Then

$$2D_a(\widehat{\gamma}, g(\widehat{\beta})) \xrightarrow{L} \chi^2_{M-M_0}, \qquad a \in \mathbb{R} - \{0, 1\}.$$

Proof. Let us fix $g(\widehat{\beta})$ and define the function $h(\widehat{\gamma}) \triangleq D_a(\widehat{\gamma}, g(\widehat{\beta}))$. A second-order Taylor expansion of $h(\widehat{\gamma})$ around $g(\widehat{\beta})$ gives

$$h(\widehat{\gamma}) = h(g(\widehat{\beta})) + \sum_{p=1}^{M} \frac{\partial h(\widehat{\gamma})}{\partial \widehat{\gamma}_p}\Big|_{\widehat{\gamma}=g(\widehat{\beta})} (\widehat{\gamma}_p - g_p(\widehat{\beta}))$$
$$+ \frac{1}{2} \sum_{p=1}^{M} \sum_{q=1}^{M} \frac{\partial^2 h(\widehat{\gamma})}{\partial \widehat{\gamma}_p \partial \widehat{\gamma}_q}\Big|_{\widehat{\gamma}=\gamma_*} (\widehat{\gamma}_p - g_p(\widehat{\beta}))(\widehat{\gamma}_q - g_q(\widehat{\beta})),$$

where $\|\gamma^* - g(\widehat{\beta})\| < \|\widehat{\gamma} - g(\widehat{\beta})\|$. Obviously, it holds $h(g(\widehat{\beta})) = D_a(g(\widehat{\beta}), g(\widehat{\beta})) = 0$. From (H2), derivatives can be introduced into the integral and then the first-order derivatives are zero. Using (H3), the second-order partial derivative is

$$\frac{\partial^2 h(\widehat{\gamma})}{\partial \widehat{\gamma}_p \partial \widehat{\gamma}_q} = \frac{1}{a(a-1)[\int_{\mathcal{X}} F^2_{\widehat{\gamma},g(\widehat{\beta})}(x)\, d\mu(x)]^2}$$
$$\times \left\{ \int_{\mathcal{X}} F^1_{\widehat{\gamma},g(\widehat{\beta})}(x)\, d\mu(x) \int_{\mathcal{X}} F^2_{\widehat{\gamma},g(\widehat{\beta})}(x)\, d\mu(x) \right.$$
$$\left. - \int_{\mathcal{X}} F^3_{\widehat{\gamma},g(\widehat{\beta})}(x)\, d\mu(x) \int_{\mathcal{X}} F^4_{\widehat{\gamma},g(\widehat{\beta})}\, d\mu(x) \right\},$$

where, for simplicity, the dependence of $F^i_{\widehat{\gamma},g(\widehat{\beta})}(x)$ on p and q has been omitted, and

$$F^1_{\widehat{\gamma},g(\widehat{\beta})}(x) \triangleq a f^{a-2}_{\widehat{\gamma}}(x) \left[(a-1) \frac{\partial f_{\widehat{\gamma}}(x)}{\partial \widehat{\gamma}_p} \frac{\partial f_{\widehat{\gamma}}(x)}{\partial \widehat{\gamma}_q} + f_{\widehat{\gamma}}(x) \frac{\partial^2 f_{\widehat{\gamma}}(x)}{\partial \widehat{\gamma}_p \partial \widehat{\gamma}_q} \right] f^{1-a}_{g(\widehat{\beta})}(x),$$

$$F^2_{\widehat{\gamma},g(\widehat{\beta})}(x) \triangleq f^a_{\widehat{\gamma}}(x) f^{1-a}_{g(\widehat{\beta})}(x),$$

$$F^3_{\widehat{\gamma},g(\widehat{\beta})}(x) \triangleq a f^{a-1}_{\widehat{\gamma}}(x) \frac{\partial f_{\widehat{\gamma}}(x)}{\partial \widehat{\gamma}_q} f^{1-a}_{g(\widehat{\beta})}(x),$$

$$F^4_{\widehat{\gamma},g(\widehat{\beta})}(x) \triangleq a f^{a-1}_{\widehat{\gamma}}(x) \frac{\partial f_{\widehat{\gamma}}(x)}{\partial \widehat{\gamma}_p} f^{1-a}_{g(\widehat{\beta})}(x).$$

But under H_0, it holds that $\gamma_* \xrightarrow{\text{a.s.}} \gamma^0$ and $g(\widehat{\beta}) \xrightarrow{\text{a.s.}} \gamma^0$. Furthermore, since $F^i_{\gamma^0,\gamma^0}(x)$ is continuous in γ^0, $i = 1, 2, 3, 4$, we get

$$F^i_{\gamma_*,g(\widehat{\beta})}(x) - F^i_{\gamma^0,\gamma^0}(x) \xrightarrow{\text{a.s.}} 0, \qquad i = 1, 2, 3, 4.$$

Applying (H1)–(H3) and the dominated convergence theorem, under H_0 it holds

$$\int_{\mathcal{X}} F^i_{\gamma_*,g(\widehat{\beta})}(x)\, d\mu(x) - \int_{\mathcal{X}} F^i_{\gamma^0,\gamma^0}(x)\, d\mu(x) \xrightarrow{\text{a.s.}} 0, \qquad i = 1, 2, 3, 4,$$

where it is immediate to see that

$$\int_{\mathcal{X}} F^2_{\gamma^0,\gamma^0}(x)\, d\mu(x) = 1, \qquad \int_{\mathcal{X}} F^i_{\gamma^0,\gamma^0}(x)\, d\mu(x) = 0, \qquad i = 3, 4.$$

These results imply

$$\frac{1}{n}\frac{\partial^2 h(\widehat{\gamma})}{\partial\widehat{\gamma}_p\,\partial\widehat{\gamma}_q}\bigg|_{\widehat{\gamma}=\gamma_*} - \frac{1}{na(a-1)}\int_{\mathcal{X}} F^1_{\gamma^0,\gamma^0}\, d\mu(x) \xrightarrow{P} 0. \tag{3.1}$$

Moreover,

$$\frac{1}{na(a-1)}\int_{\mathcal{X}} F^1_{\gamma^0,\gamma^0}(x)\, d\mu(x) = \frac{1}{n}E\left[\frac{\partial \log f_\gamma(x)}{\partial\gamma_p}\frac{\partial \log f_\gamma(x)}{\partial\gamma_q}\right]\bigg|_{\gamma=\gamma^0},$$

and applying Lemma 2 we get

$$\frac{1}{n}E\left[\frac{\partial \log f_\gamma(x)}{\partial\gamma_p}\frac{\partial \log f_{\partial\gamma}(x)}{\partial\gamma_q}\right] \longrightarrow V_{p,q}(\gamma) \qquad \text{for all } \gamma \in \Gamma.$$

Therefore,

$$\frac{1}{na(a-1)}\int_{\mathcal{X}} F^1_{\gamma^0,\gamma^0}(x)\, d\mu(x) - V_{p,q}(\gamma^0) \xrightarrow{P} 0. \tag{3.2}$$

Formulas (3.1) and (3.2) together lead to

$$\frac{1}{n}\frac{\partial^2 h(\widehat{\gamma})}{\partial\widehat{\gamma}_p\,\partial\widehat{\gamma}_q}\bigg|_{\widehat{\gamma}=\gamma_*} - V_{pq}(\gamma^0) \xrightarrow{P} 0,$$

and this in turn implies

$$2h(\widehat{\gamma}) - n\big(\widehat{\gamma} - g(\widehat{\beta})\big)^t V(\gamma^0)\big(\widehat{\gamma} - g(\widehat{\beta})\big) \xrightarrow{P} 0.$$

Finally, applying Theorem 5 of the Appendix we obtain the desired result. □

Theorem 2. *Let the null hypothesis $H_0 : \gamma \in \Gamma_0$ be true, where Γ_0 satisfies* (A3). *For each $i = 1, \ldots, s$, let $(X_{i1}, \ldots, X_{in_i})$ be independent samples of i.i.d. random variables with common p.d.f. $f_{i\theta_i}(x)$ satisfying* (2.1), (A1), *and* (A2). *Assume further that there exist measurable and μ-integrable functions $Q_1, Q_2 : \mathcal{X} \to [0, \infty)$, possibly depending on γ^0, such that for each γ^1, γ^2 in a neighborhood $N(\gamma^0)$ and for each $1 \le p, q \le M$*

(H4) $|\frac{\partial}{\partial \gamma_p^1}(f_{\gamma^1}(x)\log\frac{f_{\gamma^1}(x)}{f_{\gamma^2}(x)})| \le Q_1(x), x \in \mathcal{X}$,

(H5) $|\frac{\partial^2}{\partial \gamma_p^1 \partial \gamma_q^1}(f_{\gamma^1}(x)\log\frac{f_{\gamma^1}(x)}{f_{\gamma^2}(x)})| \le Q_2(x), x \in \mathcal{X}$.

Then

$$2D_1(\widehat{\gamma}, g(\widehat{\beta})) \xrightarrow{L} \chi^2_{M-M_0}.$$

Proof. The proof follows the same arguments as that of Theorem 1. □

Theorem 3. *Let the null hypothesis* $H_0 : \gamma \in \Gamma_0$ *be true*, *where* Γ_0 *satisfies* (A3). *For each* $i = 1, \ldots, s$, *let* $(X_{i1}, \ldots, X_{in_i})$ *be independent samples of i.i.d. random variables with common p.d.f.* $f_{i\theta_i}(x)$ *satisfying* (2.1), (A1), *and* (A2). *Then*

$$\lambda(\widehat{\gamma}, g(\widehat{\beta})) \xrightarrow{L} \chi^2_{M-M_0}.$$

Proof. Let us fix $g(\widehat{\beta})$ and define the function $h(\gamma) = \lambda(\gamma, g(\widehat{\beta}))$. A second-order Taylor expansion of $h(\gamma)$ around $\widehat{\gamma}$ evaluated at point $\gamma = g(\widehat{\beta})$ and the facts that $h(g(\widehat{\beta})) = 0$ and that $\widehat{\gamma}$ is solution of likelihood equations (2.5) lead to

$$h(\widehat{\gamma}) = -n\big(\widehat{\gamma} - g(\widehat{\beta})\big)^t n^{-1} B(\gamma_*)\big(\widehat{\gamma} - g(\widehat{\beta})\big),$$

where $B(\gamma)$ is defined in (A.3). From assumption (A1), it follows that for each $p, q \in \{1, \ldots, M\}$, the (p, q)th element of $n^{-1}B(\gamma)$, is continuous in θ_i, $i = 1, \ldots, s$. Since $\gamma_* \xrightarrow{\text{a.s.}} \gamma^0$ holds under H_0, we deduce that

$$n^{-1}B(\gamma_*) - n^{-1}B(\gamma^0) \xrightarrow{P} 0.$$

Applying part (B) of the proof of Theorem 4 and the Slutsky theorem, we get

$$n^{-1}B(\gamma_*) + V(\gamma^0) \xrightarrow{P} 0.$$

Then Lemma 1 implies

$$\begin{aligned} &h(\widehat{\gamma}) - n\big(\widehat{\gamma} - g(\widehat{\beta})\big)V(\gamma^0)\big(\widehat{\gamma} - g(\widehat{\beta})\big)^t \\ &\quad = -n\big(\widehat{\gamma} - g(\widehat{\beta})\big)^t[n^{-1}B(\gamma_*) + V(\gamma^0)]\big(\widehat{\gamma} - g(\widehat{\beta})\big) \xrightarrow{P} 0, \end{aligned}$$

and the desired result follows from Theorem 5. □

4 Application to familial data

4.1 Description of the problem

Suppose that some biometric or anthropometric characteristic such as blood pressure, cholesterol, weight, height, stature, lung capacity, etc. has been measured to randomly selected n_i families from different populations $i = 1, \dots, s$; for example, in different geographical areas. Let $\mathbf{X}_i = (\mathbf{X}_{i1}, \dots, \mathbf{X}_{in_i})$ be the vector of observations from population i, $i = 1, \dots, s$, where $\mathbf{X}_{ij}$ are random vectors whose p_i coordinates are the values that the measured characteristic takes at each of the p_i members of jth family in ith sample. Suppose that $\{\mathbf{X}_{ij} : i = 1, \dots, s, j = 1, \dots, n_i\}$ are independent and multivariate normal, that is,

$$\mathbf{X}_{ij} \sim N_{p_i}(\boldsymbol{\mu}_i, \Sigma_i), \qquad j = 1, \dots, n_i, \tag{4.1}$$

with mean vector and covariance matrix given by

$$\begin{aligned} \boldsymbol{\mu}_i &= (\mu_i, \dots, \mu_i)_{1 \times p_i}, \\ \Sigma_i &= \sigma_i^2 \begin{pmatrix} 1 & \varrho_i & \cdots & \varrho_i \\ \varrho_i & 1 & \cdots & \varrho_i \\ \vdots & \vdots & \ddots & \vdots \\ \varrho_i & \varrho_i & \cdots & 1 \end{pmatrix}_{p_i \times p_i}, \qquad i = 1, \dots, s. \end{aligned} \tag{4.2}$$

The parameter ϱ_i is the correlation between two members of the same family extracted from population i and it is called *intraclass correlation coefficient* of population i.

For $s = 1$ (one population), Srivastava (1984) developed estimators for the intraclass correlation coefficient and the variance that are easier to calculate than MLE's. Srivastava and Katapa (1986) studied their asymptotic properties. We will use similar estimators as initial values for solving maximum likelihood equations. Bhandary and Alam (2006) proposed the likelihood ratio statistic for testing the equality of intraclass correlation coefficients between three populations ($s = 3$), under the assumption $\sigma_1^2 = \sigma_2^2 = \sigma_3^2$. Here we use the family of Rényi divergences for testing this last assumption. Since previous works deal with the case $s = 3$ and generalization to any natural s is straightforward, we restrict ourselves to $s = 3$ as well. Thus, we are concerned with testing the following hypotheses

$$H_0 : \sigma_1^2 = \sigma_2^2 = \sigma_3^2 \quad \text{versus} \quad H_1 : \sigma_i^2 \neq \sigma_j^2 \qquad \text{for some } i \neq j. \tag{4.3}$$

There exists a one-to-one transformation of $\mathbf{X}_{ij}$ that leads to a model with diagonal covariance matrix (see Hobza, Molina and Morales (2003)). Under the transformed model, explicit expressions for the test statistics can be easily obtained. Let us denote by $\hat{\mu}_i$, $\hat{\sigma}_i^2$ and $\hat{\varrho}_i$ the unrestricted MLE's of μ_i, σ_i^2, and ϱ_i, respectively, $i = 1, 2, 3$, and by $\hat{\mu}_{i,0}$, $\hat{\sigma}_0^2$, and $\hat{\varrho}_{i,0}$ the H_0-restricted MLE's of the corresponding parameters. The Rényi test statistic of order $a \in \mathbb{R} - \{0, 1\}$ for testing (4.3) is given by

$$\begin{aligned} &2D_a(\widehat{\gamma}, g(\widehat{\beta})) \\ &\quad = -\frac{1}{a(a-1)} \left\{ \sum_{i=1}^{3} n_i \log [a \widehat{\sigma}_0^2 \widehat{\eta}_{i,0} + (1-a) \widehat{\sigma}_i^2 \widehat{\eta}_i] \right. \end{aligned} \tag{4.4}$$

$$+\sum_{i=1}^{3} n_i(p_i-1)\log[a\widehat{\sigma}_0^2(1-\widehat{\varrho}_{i,0})+(1-a)\widehat{\sigma}_i^2(1-\widehat{\varrho}_i)]$$

$$-\sum_{i=1}^{3} n_i p_i[a\log\widehat{\sigma}_0^2+(1-a)\log\widehat{\sigma}_i^2]$$

$$-\sum_{i=1}^{3} n_i[a\log\widehat{\eta}_{i,0}+(1-a)\log\widehat{\eta}_i]$$

$$-\sum_{i=1}^{3} n_i(p_i-1)[a\log(1-\widehat{\varrho}_{i,0})+(1-a)\log(1-\widehat{\varrho}_i)]\Bigg\},$$

where $\widehat{\eta}_i$ and $\widehat{\eta}_{i,0}$ are defined as

$$\widehat{\eta}_i = p_i^{-1}\{1+(p_i-1)\widehat{\varrho}_i\}, \qquad \widehat{\eta}_{i,0} = p_i^{-1}\{1+(p_i-1)\widehat{\varrho}_{i,0}\}, \qquad i=1,2,3.$$

Similarly, the Kullback–Leibler statistic for testing (4.3) is

$$\begin{aligned}2D_1(\widehat{\gamma}, g(\widehat{\beta})) = {} & \sum_{i=1}^{3} n_i \frac{\widehat{\sigma}_i^2}{\widehat{\sigma}_0^2}\frac{1+(p_i-1)\widehat{\varrho}_i}{1+(p_i-1)\widehat{\varrho}_{i,0}} + \sum_{i=1}^{3} n_i(p_i-1)\frac{\widehat{\sigma}_i^2}{\widehat{\sigma}_0^2}\frac{1-\widehat{\varrho}_i}{1-\widehat{\varrho}_{i,0}} \\ & + \sum_{i=1}^{3} n_i p_i \log\frac{\widehat{\sigma}_0^2}{\widehat{\sigma}_i^2} + \sum_{i=1}^{3} n_i \log\frac{1+(p_i-1)\widehat{\varrho}_{i,0}}{1+(p_i-1)\widehat{\varrho}_i} - \sum_{i=1}^{3} n_i p_i \\ & + \sum_{i=1}^{3} n_i(p_i-1)\log\frac{1-\widehat{\varrho}_{i,0}}{1-\widehat{\varrho}_i}.\end{aligned} \tag{4.5}$$

It is not difficult to see that, under the model defined by (4.1) and (4.2), the likelihood ratio statistic coincides with the Kullback–Leibler statistic, that is, $\lambda(\widehat{\gamma}, g(\widehat{\beta})) = 2D_1(\widehat{\gamma}, g(\widehat{\beta}))$. Sufficient conditions for this equality can be found in Morales, Pardo and Vajda (1997, 2000). Furthermore, since all regularity assumptions (A1) and (A2) and (H1)–(H5) are satisfied in this case, Theorems 1–3 imply that Rényi, Kullback–Leibler, and likelihood ratio statistics for testing (4.3) are asymptotically chi-squared distributed with two degrees of freedom.

4.2 Simulation study

Previous section established that the asymptotic distributions of all members of the Rényi family of statistics and of the LRS are the same. Then, the selection of a particular member should be done on the basis of small sample properties. Thus, this section describes a simulation study designed to compare test sizes and powers of the Rényi tests and the likelihood ratio test and to support the asymptotic results. We will also analyze the robustness of the tests under particular departures from the assumed probability distribution; concretely, under the heavy tails distributions

Table 1 *Sets of parameters*

Set	n_1	n_2	n_3	ϱ_1	ϱ_2	ϱ_3
I	20	20	20	0.5	0.5	0.5
II	50	50	50	0.5	0.5	0.5
III	10	20	30	0.5	0.5	0.5
IV	20	20	20	0.25	0.5	0.75

Student t with 10 degrees of freedom (t_{10}) and Logistic. For this, the following Monte Carlo simulation procedure was implemented:

(1) Generate data from the transformed model using as initial values the parameters listed in the first row of Table 1 and taking $p_i = 3$, $\mu_i = 0$, $\varrho_i = 0.5$, and $\sigma_i^2 = 1$, for $i = 1, 2, 3$. Calculate the restricted and unrestricted MLE's. Plug these MLE's in the formulas of Rényi test statistics for $a \in A = \{0.5, 0.75, 1, 1.25, 1.5, 1.75, 2, 2.25\}$. Remember that $a = 1$ gives the likelihood ratio test statistic.

(2) Repeat step (1) independently 10^4 times and calculate, for $\alpha = 0.05$ and for $a \in A$, the test sizes as

$$\widehat{\alpha}_a = 10^{-4} \#\{2D_a(\widehat{\gamma}, g(\widehat{\beta})) > \chi^2_{2,0.95}\}, \tag{4.6}$$

where #{*condition*} denotes the number of replications in which *condition* is true.

(3) Repeat steps (1) and (2) for a grid of values $\sigma_3^2 = 1 + 0.05\nu$, where $\nu \in \{-10, \dots, -1, 1, \dots, 10\}$ instead of $\sigma_3^2 = 1$ and estimate the corresponding powers $\widehat{\beta}_{a,\nu}$ by using formula (4.6).

(4) Repeat steps (1)–(3) for each set of parameters in Table 1.

(5) Repeat steps (1)–(4) by generating the data from Logistic and Student's t_{10} distributions.

Table 2 lists the resulting test sizes for the considered test statistics, the underlying probability distributions and the sets of parameters specified in Table 1. Figures 1–4 plot the power functions for sets I–IV respectively. Sets I and II are useful for observing the asymptotic behavior of power functions when increasing sample sizes. See how the power functions in Figure 2 are very close to each other in contrast to Figure 1. Sets I and III allow to compare test powers in the cases of equal and different sample sizes. Figure 1 shows that on the right-hand side of null hypothesis ($\sigma_3^2 \geq 1$), power functions are closer than on the left side. The opposite can be observed in Figure 3. Finally, sets I and IV allow the comparison of equal and unequal intraclass correlation coefficients. Note that Figure 4 is slightly more asymmetric than Figure 1. Thus, in Figure 4, the tests with higher power on one side of null hypothesis are not necessarily those with higher power on the other side. This asymmetry is typical for the Rényi family of tests.

Table 2 *Estimated type I error probabilities* $\widehat{\alpha}_a$, $a \in A$ *and* $\widehat{\alpha}_{LR}$

F	Set	$\widehat{\alpha}_{0.5}$	$\widehat{\alpha}_{0.75}$	$\widehat{\alpha}_{LR}$	$\widehat{\alpha}_{1.25}$	$\widehat{\alpha}_{1.5}$	$\widehat{\alpha}_{1.75}$	$\widehat{\alpha}_{2}$	$\widehat{\alpha}_{2.25}$
Normal	I	0.087	0.075	0.065	0.058	0.054	0.052	0.052	0.055
	II	0.064	0.059	0.056	0.054	0.052	0.051	0.051	0.052
	III	0.096	0.081	0.069	0.062	0.057	0.057	0.058	0.059
	IV	0.088	0.075	0.066	0.059	0.054	0.052	0.053	0.056
Logistic	I	0.190	0.172	0.158	0.148	0.139	0.134	0.133	0.137
	II	0.172	0.165	0.159	0.156	0.153	0.151	0.151	0.153
	III	0.191	0.171	0.156	0.146	0.138	0.134	0.130	0.123
	IV	0.186	0.169	0.154	0.145	0.138	0.137	0.136	0.138
t_{10}	I	0.164	0.147	0.134	0.122	0.117	0.114	0.117	0.123
	II	0.143	0.136	0.130	0.126	0.123	0.121	0.122	0.123
	III	0.170	0.152	0.136	0.128	0.121	0.117	0.115	0.108
	IV	0.166	0.150	0.137	0.128	0.120	0.117	0.117	0.118

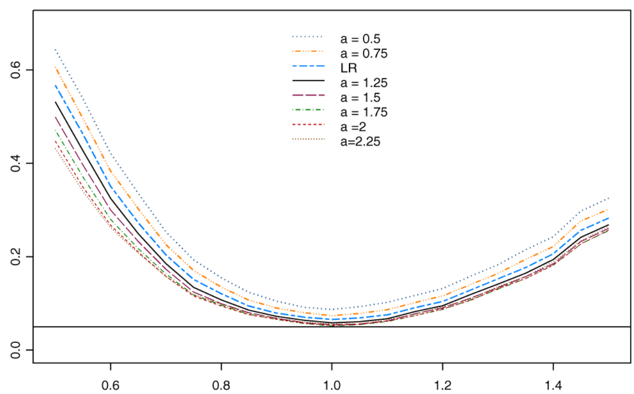

Figure 1 *Powers of LRT and RT's for* $a \in A$ *with* $n_1 = n_2 = n_3 = 20$ *and* $\varrho_1 = \varrho_2 = \varrho_3 = 0.5$.

Now we focus on the comparison of tests for each set of parameters. A fair comparison of powers should be based on tests with the same sizes, but the only way of obtaining tests with similar sizes would be to increase the sample sizes until a level in which all powers would be also very similar due to the asymptotic equivalence of test statistics (see Figure 2). For sample sizes above 50 all test are approximately equivalent. Therefore, we restrict ourselves to the sets of parameters I, III, and IV with smaller sample sizes. From the tests with sizes closer to the desired level $\alpha = 0.05$ we choose those whose maximum difference in size is at most 0.01 and we call them acceptable tests. Then, from the set of acceptable tests, we prefer the ones with higher power. Applying this rule to Table 2 we obtain that the RT's with $a \in \{1.25, 1.5, 1.75, 2, 2.25\}$ belong to the set of acceptable tests for the three sets of parameters I, III, and IV. In addition, when the true (but unknown) distribu-

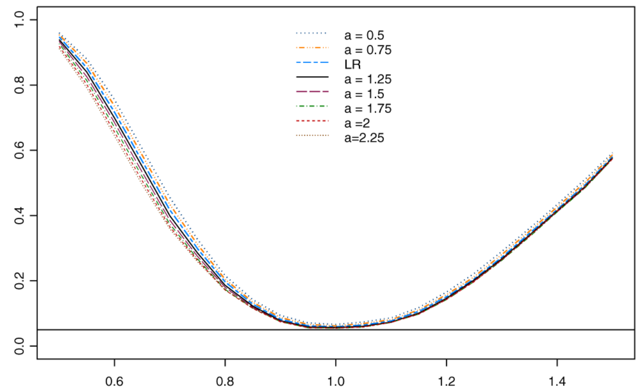

Figure 2 *Powers of LRT and RT's for* $a \in A$ *with* $n_1 = n_2 = n_3 = 50$ *and* $\varrho_1 = \varrho_2 = \varrho_3 = 0.5$.

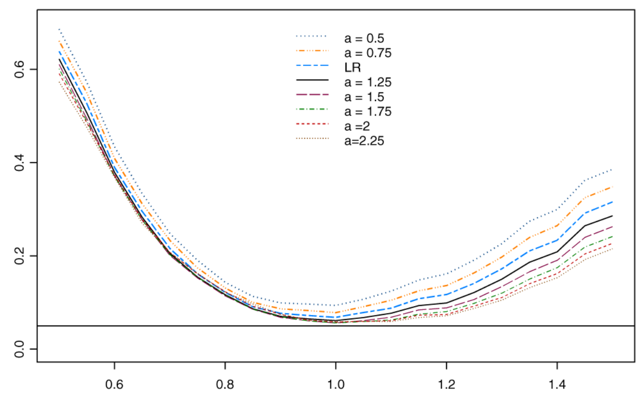

Figure 3 *Powers of LRT and RT's for* $a \in A$ *with* $n_1 = 10$, $n_2 = 20$, $n_3 = 30$, *and* $\varrho_1 = \varrho_2 = \varrho_3 = 0.5$.

tion has heavier tails than normal such as the Logistic and Student's t_{10}, these tests seem to be more robust under H_0. Comparing powers, Figures 1 and 3 show that among the acceptable tests (LRT is not included), the one with the highest power is RT of order $a = 1.25$, followed by the one with $a = 1.5$. In Figure 4 we observe the mentioned asymmetry, namely that some tests with better power on one side of null hypothesis are those with worse power on the other side. Since on the right side all power functions are very close to each other, a reasonable choice is to select, among tests that do not compromise the type I error, those with better power on the left side. Thus, among the *acceptable* tests, the RT's with $a \in \{1.25, 1.5\}$ have the highest powers.

The conclusion of this simulation study is that, for small sample sizes, RT's with $a \in [1.25, 1.5]$ have sizes closer to the nominal value $\alpha = 0.05$ than the rest of members of the family including LRT, and at the same time they have the highest

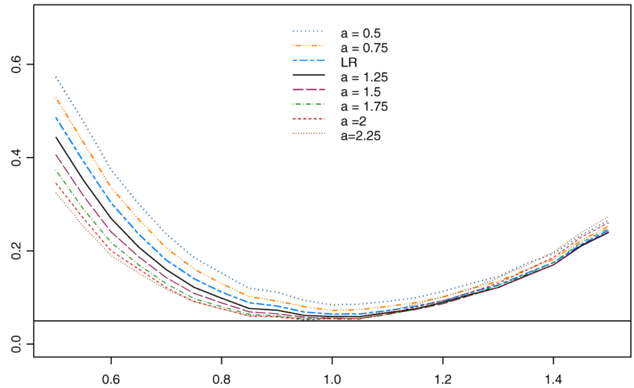

Figure 4 *Powers of LRT and RT's for* $a \in A$ *with* $n_1 = n_2 = n_3 = 20$, $\varrho_1 = 0.25$, $\varrho_2 = 0.5$, *and* $\varrho_3 = 0.75$.

powers among tests with acceptable size. Moreover, they seem to have a more robust test size under extreme values. In cases where the type I error has serious consequences, we recommend to use the RT of order $a = 1.5$.

Appendix

This appendix states the regularity conditions assumed throughout the paper and includes some auxiliary results needed in the proofs of Theorems 1–3.

Regularity assumptions:

(A1) Assume that each statistical space $(\mathcal{X}_i, \mathcal{B}_{\mathcal{X}_i}, P_{i,\theta_i})_{\theta_i \in \Theta_i}$, with its corresponding sample $(X_{i1}, \dots, X_{in_i})$ and p.d.f. f_{i,θ_i} satisfies regularity assumptions (M1)–(M8) stated in Lehmann (1999, pages 499–501), and additionally all the partial derivatives of f_{i,θ_i} till third order are continuous and finite on a (open) neighborhood of the true parameter θ_i, $i = 1, \dots, s$.

(A2) The matrix $V(\gamma)$ defined in (2.4) is positive definite.

(A3) The set Γ_0 defining H_0 in (2.3) can be expressed as

$$\Gamma_0 = \{\gamma \in \Gamma : R_i(\gamma) = 0, i = 1, \dots, M - M_0\},$$

or alternatively as

$$\Gamma_0 = \{\gamma \in \Gamma : \gamma_i = g_i(\beta), i = 1, \dots, M\},$$

where $\beta = (\beta_1, \dots, \beta_{M_0})^t \in B$ and $B \subset \mathbb{R}^{M_0}$ is open, and where functions g_i and R_j have continuous first-order partial derivatives and the ranks of matrices

$$T_\gamma = \left(\partial R_i(\gamma)/\partial \gamma_j\right)_{\substack{i=1,\dots,M-M_0\\ j=1,\dots,M}} \quad \text{and} \quad M_\beta = \left(\partial g_i(\beta)/\partial \beta_j\right)_{\substack{i=1,\dots,M\\ j=1,\dots,M_0}}$$

are $M - M_0$ and M_0, respectively.

In order to derive the asymptotic distribution of the multi-sample test statistics proposed in Section 2, the first step is to obtain the asymptotic distribution of MLE's. But they may not exist or not be solutions of likelihood equations, lying on the border of the parameter space. However, for the asymptotic results of the present paper it is enough to have consistent estimators that are solution of likelihood equations. Thus, the following theorem gives the asymptotic distribution of consistent solutions of the likelihood equations.

Theorem 4. *For each* $i = 1, \dots, s$, *let* $(X_{i1}, \dots, X_{in_i})$ *be independent samples of i.i.d. random variables with common p.d.f.* $f_{i\theta_i}(x)$ *satisfying* (A1) *and* (A2). *Suppose that* $(n_1, \dots, n_s)$ *is a sequence of sample sizes satisfying* (2.1). *Then, any consistent sequence* $\widehat{\gamma} = \widehat{\gamma}(x)$ *of solutions of likelihood equations* (2.5) *satisfies*

$$\sqrt{n}(\widehat{\gamma} - \gamma^0) \xrightarrow{L} \mathcal{N}(0, V^{-1}(\gamma^0)). \tag{A.1}$$

Proof. The proof is based on the facts that $\widehat{\gamma}$ satisfies (2.5) and $\widehat{\gamma}$ is known to be close to γ^0. Taking a second-order Taylor expansion of $\partial l(\widehat{\gamma})/\partial \gamma_p$ around γ^0, for all $p = 1, \dots, M$, and multiplying by $1/\sqrt{n}$, we obtain in matrix notation

$$0 = \frac{1}{\sqrt{n}} \nabla l(\gamma^0) + \left[\frac{1}{n} B(\gamma^0) + \frac{1}{2n} \begin{pmatrix} (\widehat{\gamma} - \gamma^0)^t D_1(\gamma^1) \\ \cdots \\ (\widehat{\gamma} - \gamma^0)^t D_M(\gamma^M) \end{pmatrix} \right] \sqrt{n}(\widehat{\gamma} - \gamma^0), \tag{A.2}$$

where

$$\nabla l(\gamma) = \left(\frac{\partial l(\gamma)}{\partial \gamma_1}, \dots, \frac{\partial l(\gamma)}{\partial \gamma_M} \right)^t, \qquad B(\gamma) = \left(\frac{\partial^2 l(\gamma)}{\partial \gamma_p \, \partial \gamma_q} \right)_{p,q=1,\dots,M},$$
$$D_p(\gamma) = \left(\frac{\partial^3 l(\gamma)}{\partial \gamma_r \, \partial \gamma_q \, \partial \gamma_p} \right)_{q,r=1,\dots,M}, \qquad p = 1, \dots, M, \tag{A.3}$$

and γ^p satisfies $\|\gamma^p - \gamma^0\| \le \|\widehat{\gamma} - \gamma^0\|$, $p = 1, \dots, M$.

The proof is based on the following three results:

(A) $n^{-1/2} \nabla l(\gamma^0) \xrightarrow{L} \mathcal{N}(0, V(\gamma^0))$, which follows simply by expressing the vector $n^{-1/2} \nabla l(\gamma^0)$ as a double sum, where the first sum is over the s samples, and then applying the central limit theorem s times, one for each sample.

(B) $n^{-1} B(\gamma^0) \xrightarrow{P} -V(\gamma^0)$, which follows similarly, by writing each element of the matrix $n^{-1} B(\gamma^0)$ as a double sum, then expressing the derivatives of the log-densities in terms of the derivatives of the densities, and finally applying the weak law of large numbers and assumption (A1).

(C) $n^{-1} D_p(\gamma^p)$ is bounded in probability, $p = 1, \dots, M$. Again, this result can be proved by expressing the elements of $n^{-1} D_p(\gamma^p)$ as a double sum of third-order derivatives, and combining the consistency of $\widehat{\gamma}$ to γ^0 with assumption

(A1), so that

$$\lim_{n\to\infty} P\left(\left|\frac{1}{n}\frac{\partial^3 l(\gamma)}{\partial\gamma_r\,\partial\gamma_q\,\partial\gamma_p}\bigg|_{\gamma=\gamma^p}\right| < \frac{1}{n}\sum_{i=1}^{s}\sum_{j=1}^{n_i} H_i(x_{ij})\right) = 1.$$

From the law of large numbers and (A1), we have

$$\frac{1}{n}\sum_{i=1}^{s}\sum_{j=1}^{n_i} H_i(x_{ij}) \xrightarrow{P} \sum_{i=1}^{s}\lambda_i E[H_i(x_{ij})] < \infty,$$

which implies the result.

Now, let us define

$$Z_n \triangleq \frac{1}{n}B(\gamma^0) + \frac{1}{2n}\begin{pmatrix} (\widehat{\gamma}-\gamma^0)^t D_1(\gamma^1) \\ \cdots \\ (\widehat{\gamma}-\gamma^0)^t D_M(\gamma^M)\end{pmatrix}.$$

From part (A) and (A.2), we know that

$$n^{-1/2}\nabla l(\gamma^0) = -Z_n\sqrt{n}(\widehat{\gamma}-\gamma^0) \xrightarrow{L} \mathcal{N}(0, V(\gamma^0)). \tag{A.4}$$

Now parts (B) and (C) and $\widehat{\gamma} \xrightarrow{P} \gamma^0$ imply $-Z_n - V(\gamma^0) = o_P(1)$. This fact together with (A.4) imply $\sqrt{n}(\widehat{\gamma}-\gamma^0) = O_P(1)$. Then

$$-Z_n\sqrt{n}(\widehat{\gamma}-\gamma^0) - V(\gamma^0)\sqrt{n}(\widehat{\gamma}-\gamma^0) = (-Z_n - V(\gamma^0))\sqrt{n}(\widehat{\gamma}-\gamma^0) \xrightarrow{P} 0.$$

Therefore, by Slutsky's theorem (see, e.g., Ferguson (1996), page 39) $V(\gamma^0) \times \sqrt{n}(\widehat{\gamma}-\gamma^0)^t$ is asymptotically equivalent to $-Z_n\sqrt{n}(\widehat{\gamma}-\gamma^0)$. Thus, (A.4) implies

$$V(\gamma^0)\sqrt{n}(\widehat{\gamma}-\gamma^0) \xrightarrow{L} \mathcal{N}(0, V(\gamma^0))$$

and under assumption (A2) we get (A.1). □

Let us now assume that the s submodels $(\mathcal{X}_i, \mathcal{B}_{\mathcal{X}_i}, P_{i,\theta_i})_{\theta_i\in\Theta_i}$, $i = 1,\dots,s$, restricted to null hypothesis (i.e., with $\gamma \in \Gamma_0$) satisfy assumptions (A1) and (A2), with derivatives taken with respect to the new parameter β. The following lemma sets up a relation between the matrix T_γ, defined in the statement of assumption (A3), and $V(\gamma)$.

Lemma 1. *If $\gamma \in \Gamma_0$, then $V(\gamma) = T_\gamma^t(T_\gamma V^{-1}(\gamma)T_\gamma^t)^{-1}T_\gamma$.*

Proof. It follows the same steps as the proof on page 159 in Serfling (1980). □

Let us define $R_\gamma = (R_1(\gamma),\dots,R_{M-M_0}(\gamma))^t$, for $R_i(\gamma)$ defined in (A3). The following proposition provides the asymptotic distribution of $R_{\widehat{\gamma}}$.

Proposition 1. *Under the assumptions of Theorem* 4, *if* $H_0 : \gamma \in \Gamma_0$ *satisfies* (A3), *it holds*

$$\sqrt{n}(R_{\widehat{\gamma}} - R_{\gamma^0}) \xrightarrow{L} \mathcal{N}(0, T_{\gamma^0} V^{-1}(\gamma^0) T^t_{\gamma^0}).$$

Proof. The first-order Taylor expansion of $R_{\widehat{\gamma}}$ around γ^0 is

$$R_{\widehat{\gamma}} = R_{\gamma^0} + T_*(\widehat{\gamma} - \gamma^0),$$

where

$$T_* = \left(\partial R_i(\gamma)/\partial \gamma_j|_{\gamma=\gamma^i}\right)_{\substack{i=1,\dots,M-M_0\\ j=1,\dots,M}} \tag{A.5}$$

and $\|\gamma^i - \gamma^0\| < \|\widehat{\gamma} - \gamma^0\|$, $i = 1, \dots, M - M_0$. The consistency of $\widehat{\gamma}$ and the continuity of the elements of T_* imply

$$T_* - T_{\gamma^0} \xrightarrow{P} 0.$$

By Theorem 4, we know that $\sqrt{n}(\widehat{\gamma} - \gamma^0) = O_p(1)$, and so

$$\sqrt{n} T_*(\widehat{\gamma} - \gamma^0) - \sqrt{n} T_{\gamma^0}(\widehat{\gamma} - \gamma^0) \xrightarrow{P} 0.$$

Consequently, Slutsky's theorem and Theorem 4 lead to the desired result. □

Proposition 2. *Under the assumptions of Theorem* 4, *if* $H_0 : \gamma \in \Gamma_0$ *is true*, *where* Γ_0 *satisfies* (A3), *then*

$$n R^t_{\widehat{\gamma}} (T_{\gamma^0} V^{-1}(\gamma^0) T^t_{\gamma^0})^{-1} R_{\widehat{\gamma}} \xrightarrow{L} \chi^2_{M-M_0}.$$

Proof. It is immediate by applying Proposition 1 under $H_0 : R_{\gamma^0} = 0$. □

Suppose that $H_0 : \gamma \in \Gamma_0$ is true (that is, $\gamma^0 = g(\beta^0)$). By Theorem 4, we have

$$\sqrt{n}(\widehat{\beta} - \beta^0) \xrightarrow{L} \mathcal{N}(0, V^{-1}(\beta^0)),$$

where $V(\beta^0)$ is defined in (2.4) and from Crámer's theorem (see, e.g., Ferguson (1996), page 45) it follows that

$$\sqrt{n}\big(g(\widehat{\beta}) - \gamma^0\big) \xrightarrow{L} \mathcal{N}(0, M_\beta V^{-1}(\beta^0) M^t_\beta)$$

for M_β defined in (A3). As $\widehat{\beta} \xrightarrow{P} \beta^0$ and g is continuous, it holds $g(\widehat{\beta}) \xrightarrow{P} \gamma^0$. Now the following result can be proved.

Theorem 5. *Under the assumptions of Theorem 4, if $H_0 : \gamma \in \Gamma_0$ is true, where Γ_0 satisfies* (A3), *then for any consistent sequences $\widehat{\gamma} = \widehat{\gamma}(x)$ and $\widehat{\beta} = \widehat{\beta}(x)$ of solutions of the likelihood equations* (2.5) *and* (2.6), *it holds*

$$n(\widehat{\gamma} - g(\widehat{\beta}))^t V(\gamma^0)(\widehat{\gamma} - g(\widehat{\beta})) \xrightarrow{L} \chi^2_{M-M_0},$$

where $V(\gamma)$ is defined in (2.4).

Proof. From Lemma 1, we obtain

$$\begin{aligned} &n(\widehat{\gamma} - g(\widehat{\beta}))^t V(\gamma^0)(\widehat{\gamma} - g(\widehat{\beta})) \\ &\quad = n(\widehat{\gamma} - g(\widehat{\beta}))^t T^t_{\gamma^0}(T_{\gamma^0} V^{-1}(\gamma^0) T^t_{\gamma^0})^{-1} T_{\gamma^0}(\widehat{\gamma} - g(\widehat{\beta})). \end{aligned} \tag{A.6}$$

A first-order Taylor expansion of $R_{\widehat{\gamma}}$ around $g(\widehat{\beta})$ leads to

$$R_{\widehat{\gamma}} = R_{g(\widehat{\beta})} + T_*(\widehat{\gamma} - g(\widehat{\beta})),$$

where the matrix T_* is defined in (A.5) and γ^i satisfies

$$\|\gamma^i - g(\widehat{\beta})\| < \|\widehat{\gamma} - g(\widehat{\beta})\| < \|\widehat{\gamma} - \gamma^0\| + \|g(\widehat{\beta}) - \gamma^0\|, \qquad i = 1, \ldots, M - M_0.$$

As $R_{g(\beta)} = (0, \ldots, 0)^t$ for each β, in particular for $\widehat{\beta}$, we have

$$R_{\widehat{\gamma}} = T_*(\widehat{\gamma} - g(\widehat{\beta})).$$

The consistency of $\widehat{\gamma}$ and $g(\widehat{\beta})$ and the continuity of the elements of T_* lead to $T_* \xrightarrow{P} T_{\gamma^0}$. By Theorem 4, it holds $\sqrt{n}(\widehat{\gamma} - g(\widehat{\beta})) = \sqrt{n}(\widehat{\gamma} - \gamma^0) + \sqrt{n}(\gamma^0 - g(\widehat{\beta})) = O_P(1)$. From this fact we deduce

$$\sqrt{n} T_*(\widehat{\gamma} - g(\widehat{\beta})) - \sqrt{n} T_{\gamma^0}(\widehat{\gamma} - g(\widehat{\beta})) \xrightarrow{P} 0.$$

Consequently

$$\sqrt{n} R_{\widehat{\gamma}} - \sqrt{n} T_{\gamma^0}(\widehat{\gamma} - g(\widehat{\beta})) \xrightarrow{P} 0. \tag{A.7}$$

Using (A.6), (A.7), and Slutsky's theorem, we get

$$n(\widehat{\gamma} - g(\widehat{\beta}))^t V(\gamma^0)(\widehat{\gamma} - g(\widehat{\beta})) - n R^t_{\widehat{\gamma}}(T_{\gamma^0} V^{-1}(\gamma^0) T^t_{\gamma^0})^{-1} R_{\widehat{\gamma}} \xrightarrow{P} 0.$$

Finally, applying Proposition 2 and Slutsky's theorem again, we get the desired result. □

Lemma 2. *For each $\gamma \in \Gamma$ it holds that*

$$\frac{1}{n} E\left[\frac{\partial \log f_\gamma(X)}{\partial \gamma_p} \frac{\partial \log f_\gamma(X)}{\partial \gamma_q}\right] \longrightarrow V_{pq}(\gamma),$$

where $V_{pq}(\gamma)$ is the (p, q)th element of $V(\gamma)$.

Proof. We must consider three cases depending on the relative position of p and q with respect to k (see (2.2)): (i) $p > k$, $q > k$, (ii) $p \le k$, $q \le k$, and (iii) $p \le k$, $q > k$. Since the proofs of the three cases follow similar steps, here we give only the proof of (i). Assume that γ_q is one of the components of θ_r. Then the value of the second-order partial derivative of $\log f_\gamma(X)$ with respect to γ_q and γ_p depends on whether γ_p belongs also to the same parameter θ_r or not, that is,

$$\frac{\partial^2 \log f_\gamma(X)}{\partial \gamma_p \, \partial \gamma_q} = \begin{cases} \displaystyle\sum_{j=1}^{n_r} \frac{\partial^2 \log f_{r,\theta_r}(X_{rj})}{\partial \gamma_p \, \partial \gamma_q}, & \text{if } \gamma_p \text{ and } \gamma_q \text{ are both components of } \theta_r, \\ 0, & \text{otherwise.} \end{cases}$$

Since the random variables X_{rj}, $j = 1, \dots, n_r$ are i.i.d., then

$$\begin{aligned} \frac{1}{n} E\left[\frac{\partial \log f_\gamma(X)}{\partial \gamma_p} \frac{\partial \log f_\gamma(X)}{\partial \gamma_q}\right] &= -\frac{n_r}{n} E\left[\frac{\partial^2 \log f_{r,\theta_r}(X_{rj})}{\partial \gamma_p \, \partial \gamma_q}\right] \\ &= \frac{n_r}{n} E\left[\frac{\partial \log f_{r,\theta_r}(X_{rj})}{\partial \gamma_p} \frac{\partial \log f_{r,\theta_r}(X_{rj})}{\partial \gamma_q}\right] \\ &\longrightarrow \lambda_r V_{p,q}^r(\theta_r) = V_{pq}(\gamma). \end{aligned}$$

□

Acknowledgments

This work has been supported by Grants MTM2006-05693, SEJ2007-64500, ASCR Grant A1075403, and MSMT Grant 1M0572.

References

Bhandary, M. and Alam, M. K. (2000). Test for the equality of intraclass correlation coefficients under unequal family sizes for several populations. *Communications in Statistics. Theory and Methods* **29** 755–768. MR1769323

Csiszár, I. (1963). Eine Informationstheoretische Ungleidung und ihre Anwendung auf den Bewis del Ergodizität on Markhoffschen Ketten. In *Publications of the Mathematical Institute, Hungarian Academy of Sciences* **8** 85–108. Budapest, Hungary. MR0164374

Ferguson, T. S. (1988). *A Course in Large Sample Theory*. Chapman & Hall, New York. MR1699953

Hobza, T., Molina, I. and Morales, D. (2003). Likelihood divergence statistics for testing hypotheses in familial data. *Communications in Statistics. Theory and Methods* **32** 415–434. MR1958521

Kullback, S. (1959). *Information Theory and Statistics*. Wiley, New York. MR0103557

Kupperman, M. (1957). Further applications of information theory to multivariate analysis and statistical inference. Ph.D. dissertation, The George Washington University.

Lehmann, E. L. (1999). *Elements of Large-Sample Theory*. Springer, New York. MR1663158

Liese, F. and Vajda, I. (1987). *Convex Statistical Distances*. Teubner, Leipzig. MR0926905

Morales, D., Pardo, L. and Pardo, M. C. (2001). Likelihood divergence statistics for testing hypotheses about multiple populations. *Communications in Statistics. Simulation and Computation* **30** 867–884. MR1878458

Morales, D., Pardo, L. and Vajda, I. (1997). Some new statistics for testing composite hypotheses in parametric models. *Journal of Multivariate Analysis* **62** 137–168. MR1467878

Morales, D., Pardo, L. and Vajda, I. (2000). Rényi statistics in directed families of exponential experiments. *Statistics* **34** 151–174. MR1768246

Morales, D., Pardo, L., Pardo, M. C. and Vajda, I. (2004). Rényi statistics for testing composite hypotheses in general exponential models. *Statistics* **38** 133–147. MR2061151

Rényi, A. (1961). On measures of entropy and information. In *Fourth Berkeley Symposium on Mathematics, Statistics and Probability* **1** 547–561. Univ. California Press, Berkeley. MR0132570

Salicrú, M., Menéndez, M. L., Morales, D. and Pardo, L. (1994). On the applications of divergence type measures in testing statistical hypotheses. *Journal of Multivariate Analysis* **51** 372–391. MR1321304

Serfling, R. J. (1980). *Approximation Theorems of Mathematical Statistics*. Wiley, New York. MR0595165

Srivastava, M. S. (1984). Estimation of interclass correlations in familial data. *Biometrika* **71** 177–185. MR0738339

Srivastava, M. S. and Katapa, R. S. (1986). Comparison of estimators of interclass and intraclass correlations from familial data. *La Revue Canadienne de Statistique* **14** 29–42. MR0839288

T. Hobza
Department of Mathematics
Czech Technical University
Prague
Czech Republic
E-mail: tomas.hobza@fjfi.cvut.cz

I. Molina
Department of Statistics
Universidad Carlos III de Madrid
Getafe (Madrid)
Spain
E-mail: isabel.molina@uc3m.es

D. Morales
Operation Research Center
Miguel Hernández University of Elche
Elche
Spain
E-mail: d.morales@umh.es