Journal of Multivariate Analysis 118 (2013) 1–23

Contents lists available at SciVerse ScienceDirect

Journal of Multivariate Analysis

journal homepage: www.elsevier.com/locate/jmva

Hypothesis testing in a generic nesting framework for general distributions

N. Martín [a,*], N. Balakrishnan [b]

[a] Department of Statistics, Carlos III University of Madrid, Spain
[b] Department of Mathematics & Statistics, McMaster University, Canada

ARTICLE INFO

Article history:
Received 6 December 2011
Available online 21 March 2013

AMS subject classifications:
62F03
62G10
62E20

Keywords:
Chi-bar-square statistic
Chi-square statistic
Divergence based test statistics
Equality constraints
Exponential family of distributions
Inequality constraints

ABSTRACT

Nested parameter spaces, either in the null or alternative hypothesis, often enable an improvement in the performance of the tests. In this context, order restricted inference has not been studied in detail. Divergence based measures provide a flexible tool for proposing some useful test statistics, which usually contain the likelihood ratio-test statistics as a special case. The existing literature on hypothesis testing under inequality constraints, based on phi-divergence measures, is concentrated on specific models with multinomial sampling. In this paper the existing results are extended and unified through new families of test-statistics that are valid for nested parameter spaces containing either equality or inequality constraints and general distributions for either single or multiple populations.

1. Introduction

We consider samples coming from g populations $\boldsymbol{X}_{i1}, \ldots, \boldsymbol{X}_{ij}, \ldots, \boldsymbol{X}_{in_i}$, $i = 1, \ldots, g$, with n_i being the sample size and $\boldsymbol{X}_{ij} = (X_{ij1}, \ldots, X_{ijm_i})^T$ being m_i-dimensional independent and identically distributed random variables. The sampling units have the same distribution function (density function) $F_{\boldsymbol{\theta}_i}(\boldsymbol{x})(f_{\boldsymbol{\theta}_i}(\boldsymbol{x}))$, $i = 1, \ldots, g$, which depend on an unknown parameter $\boldsymbol{\theta}_i = (\theta_{i1}, \ldots, \theta_{ik_i})^T \in \Theta_i \subset \mathbb{R}^{k_i}$. For the i-th population, the maximum likelihood estimator (MLE) of parameter $\boldsymbol{\theta}_i$ is given by

$$\widehat{\boldsymbol{\theta}}_i = \arg\max_{\boldsymbol{\theta}_i \in \Theta_i} \ell_{n_i}(\boldsymbol{\theta}_i), \tag{1}$$

where

$$\ell_{n_i}(\boldsymbol{\theta}_i) = \log \mathcal{L}(\boldsymbol{X}_{i1}, \ldots, \boldsymbol{X}_{in_i}; \boldsymbol{\theta}_i) \tag{2}$$

and $\mathcal{L}(\boldsymbol{X}_{i1}, \ldots, \boldsymbol{X}_{in_i}; \boldsymbol{\theta}_i) = \prod_{j=1}^{n_i} f_{\boldsymbol{\theta}_i}(\boldsymbol{X}_{ij})$ is the likelihood function associated with the i-th population. For each population $i = 1, \ldots, g$, we shall assume the following regularity conditions with respect to the distributions:

- $\frac{\partial}{\partial\theta_{iu}} f_{\boldsymbol{\theta}_i}(\boldsymbol{x})$ and $\frac{\partial^2}{\partial\theta_{iu}\partial\theta_{iv}} f_{\boldsymbol{\theta}_i}(\boldsymbol{x})$ exist almost everywhere and are such that $\left|\frac{\partial}{\partial\theta_{iu}} f_{\boldsymbol{\theta}_i}(\boldsymbol{x})\right| \le G_{i,u}(\boldsymbol{x})$, $\left|\frac{\partial^2}{\partial\theta_{iu}\partial\theta_{iv}} f_{\boldsymbol{\theta}_i}(\boldsymbol{x})\right| \le G_{i,uv}(\boldsymbol{x})$, with $\int_{\mathbb{R}^{m_i}} G_{i,u}(\boldsymbol{x}) d\boldsymbol{x} < \infty$ and $\int_{\mathbb{R}^{m_i}} G_{i,uv}(\boldsymbol{x}) d\boldsymbol{x} < \infty$;

* Corresponding author.
E-mail addresses: nirian.martin@uc3m.es, nirian.martin@gmail.com (N. Martín).

http://dx.doi.org/10.1016/j.jmva.2013.03.012

- $\frac{\partial}{\partial\theta_{iu}}\log f_{\boldsymbol{\theta}_i}(\boldsymbol{x})$ and $\frac{\partial^2}{\partial\theta_{iu}\partial\theta_{iv}}\log f_{\boldsymbol{\theta}_i}(\boldsymbol{x})$ exist almost everywhere and
  - the Fisher information matrix
$$\mathcal{I}_F(\boldsymbol{\theta}_i) = E\left[\left(\frac{\partial}{\partial\boldsymbol{\theta}_i}\log f_{\boldsymbol{\theta}_i}(\boldsymbol{X}_{i1})\right)\left(\frac{\partial}{\partial\boldsymbol{\theta}_i}\log f_{\boldsymbol{\theta}_i}(\boldsymbol{X}_{i1})\right)^T\right],$$
    is finite positive definite;
  - as $\delta \to 0$, $\psi_i(\delta) = E\left[\sup_{\{\boldsymbol{t}:\|\boldsymbol{t}\|\le\delta\}}\left\|\frac{\partial^2}{\partial\boldsymbol{\theta}_i\partial\boldsymbol{\theta}_i^T}\log f_{\boldsymbol{\theta}_i+\boldsymbol{t}}(\boldsymbol{X}_{i1}) - \frac{\partial^2}{\partial\boldsymbol{\theta}_i\partial\boldsymbol{\theta}_i^T}\log f_{\boldsymbol{\theta}_i}(\boldsymbol{X}_{i1})\right\|\right]$ is such that $\psi_i(\delta) \to 0$.

Our interest is in developing statistical inference with respect to an r-dimensional function $\boldsymbol{h}$ which depends on $\boldsymbol{\theta} = (\boldsymbol{\theta}_1^T, \ldots, \boldsymbol{\theta}_g^T)^T \in \Theta = \Theta_1 \times \cdots \times \Theta_g \subset \mathbb{R}^k$, with $k = \sum_{i=1}^g k_i > r$. Hypotheses of type $\boldsymbol{h}(\boldsymbol{\theta}) = \boldsymbol{0}_r$, $\boldsymbol{h}(\boldsymbol{\theta}) \neq \boldsymbol{0}_r$, $\boldsymbol{h}(\boldsymbol{\theta}) \lneqq \boldsymbol{0}_r$, $\boldsymbol{h}(\boldsymbol{\theta}) \le \boldsymbol{0}_r$, $\boldsymbol{h}_1(\boldsymbol{\theta}) = \boldsymbol{0}_{r_1}$, $\boldsymbol{h}_2(\boldsymbol{\theta}) \le \boldsymbol{0}_{r_2}$, are established on $\boldsymbol{h}(\boldsymbol{\theta}) = (\boldsymbol{h}_1(\boldsymbol{\theta}), \boldsymbol{h}_2(\boldsymbol{\theta}))$, with $r = r_1 + r_2$. For this purpose, the following regularity assumptions are made:

- function $\boldsymbol{h}$ is convex and first-order differentiable in Θ_i, $i = 1, \ldots, g$;
- the $r \times k$ Jacobian matrix associated with $\boldsymbol{h}$, namely, $\boldsymbol{H}(\boldsymbol{\theta}) = \frac{\partial}{\partial\boldsymbol{\theta}^T}\boldsymbol{h}(\boldsymbol{\theta})$, has the form $\boldsymbol{H}(\boldsymbol{\theta}) = (\boldsymbol{H}_1(\boldsymbol{\theta}), \ldots, \boldsymbol{H}_g(\boldsymbol{\theta}))$, where each $r \times k_i$ submatrix $\boldsymbol{H}_i(\boldsymbol{\theta}) = \frac{\partial}{\partial\boldsymbol{\theta}_i^T}\boldsymbol{h}(\boldsymbol{\theta})$, $i = 1, \ldots, g$, is of full rank.

In the case when only an internal comparison of components of $\boldsymbol{\theta}_i$ inside the i-th population is considered, matrix $\boldsymbol{H}(\boldsymbol{\theta})$ is block diagonal. In such a case, if no further comparison is made, it is convenient to make the inference separately for each population, that is, to take the results developed here with $g = 1$.

This paper extends and unifies the existing results in different directions. We shall propose two families of test statistics based on ϕ-divergence measures (S_ϕ and T_ϕ families) for testing the hypotheses in (16)–(18) as well as those in (11)–(13). We consider one or more populations and for the latter when having different sample sizes a different version of test statistics must be applied ($\widetilde{S}_\phi$ and $\widetilde{T}_\phi$ families). We do not restrict ourselves to a specific form of distribution for sampling and we consider general populations. Even though most of the procedures in the literature are developed based on the theory of Aitchison and Silvey [1] and Silvey [28], here we follow the work of El Barmi and Dykstra [5] for multinomial sampling, which was further extended to more general of populations in their subsequent work. The paper is organized as follows. Sections 2 and 3 provide the basis for the asymptotic theory of the proposed test statistics to be developed later. More specifically, in Section 2, the well-known results on the joint asymptotic distribution of maximum likelihood estimators and Lagrange multipliers are detailed, and in Section 3 the hypothesis testing problems discussed in this paper are explained along with the classical test statistics and the corresponding test statistics in terms of divergence measures. In Section 4, the new test statistics are introduced and their asymptotic properties are formally established. In Section 5, a real data with two Poisson populations is used to illustrate the results developed here. Finally, in Section 6, we focus on multinomial distributions to perform a simulation study, more thoroughly binomial populations with small probabilities of success are analyzed.

2. Joint asymptotic distribution of maximum likelihood estimators and Lagrange multipliers

If we consider the likelihood function in (2) associated with the i-th population, the following properties of the MLE in (1) are well-known; see for example Sen and Singer [24, p. 210].

(i) The asymptotic distribution of the MLE of $\boldsymbol{\theta}_i$, separately for each population, is
$$\sqrt{n_i}(\widehat{\boldsymbol{\theta}}_i - \boldsymbol{\theta}_{i,0}) \xrightarrow[n_i\to\infty]{\mathcal{L}} \mathcal{N}(\boldsymbol{0}_{k_i}, \mathcal{I}_F^{-1}(\boldsymbol{\theta}_{i,0}));$$

(ii) Assuming that $\{\nu_i\}_{i=1}^g$ exist such that $\nu_i = \lim_{n\to\infty}\frac{n_i}{n} \in (0, 1)$, with $n = \sum_{i=1}^g n_i$ and $\sum_{i=1}^g \nu_i = 1$, the asymptotic distribution of the MLE of $\boldsymbol{\theta}$ of all populations is
$$\sqrt{n}\left(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\right) \xrightarrow[n\to\infty]{\mathcal{L}} \mathcal{N}(\boldsymbol{0}_k, \mathcal{I}_F^{-1}(\boldsymbol{\theta}_0)), \tag{3}$$
where
$$\mathcal{I}_F(\boldsymbol{\theta}_0) = \bigoplus_{i=1}^g \nu_i \mathcal{I}_F(\boldsymbol{\theta}_{i,0}) \tag{4}$$
is the information matrix based on "all" the observations and $\oplus$ is the direct sum of matrices;

(iii) In particular, when $n_1 = \cdots = n_g = \frac{n}{g}$, apart from (3) with $\nu_i = \frac{1}{g}$, we can consider
$$\sqrt{\frac{n}{g}}\left(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\right) \xrightarrow[n\to\infty]{\mathcal{L}} \mathcal{N}\left(\boldsymbol{0}_k, \bigoplus_{i=1}^g \mathcal{I}_F^{-1}(\boldsymbol{\theta}_{i,0})\right),$$
in other words, we can consider artificially that we have a population of size $\frac{n}{g}$ with a single parameter $\boldsymbol{\theta}_0$ and the Fisher information matrix is (4) with $\nu_i = 1$, $i = 1, \ldots, g$.