UNIVERSIDAD CARLOS III DE MADRID working papers

Working Paper 11-35
Statistics and Econometrics Series 27
November 2011

Departamento de Estadística
Universidad Carlos III de Madrid
Calle Madrid, 126
28903 Getafe (Spain)
Fax (34) 91 624-98-49

HYPOTHESIS TESTING IN A GENERIC NESTING FRAMEWORK WITH GENERAL POPULATION DISTRIBUTIONS

Nirian Martín[1], Narayanaswami Balakrishnan[2]

Abstract

Nested parameter spaces, either in the null or alternative hypothesis, constitute a guarantee for improving the performance of the tests, however in the existing literature on order restricted inference they have been usually skipped for being studied in detail. Divergence based divergence measures provide a flexible tool for creating meaningful test-statistics, which usually contain the likelihood ratio-test statistics as special case. The existing literature on hypothesis testing with inequality constraints using phi-divergence measures, is centered in a very specific models with multinomial sampling. The contribution of this paper consists in extending and unifying widely the existing work: new families of test-statistics are presented, valid for nested parameter spaces containing either equality or inequality constraints and general distributions for either single or multiple populations are considered.

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

[1]Nirian Martín, Department of Statistics, Universidad Carlos III de Madrid, Spain; [2]Narawashami Balakrishnan, Department of Mathematics, McMaster University, Canada; e-mails: nirian.martin@uc3m.es (N. Martín), bala@mcmaster.ca (N. Balakrishnan).
Acknowledgements: This work was carried out in McMaster University during the summer of 2011.

Hypothesis Testing in a Generic Nesting Framework with General Population Distributions

Martín, N.[1]*, Balakrishnan, N.[2]
[1]Dept. Statistics, Carlos III University of Madrid, Spain
[2]Dept. Mathematics, McMaster University, Canada

November 12, 2011

Abstract

Nested parameter spaces, either in the null or alternative hypothesis, constitute a guarantee for improving the performance of the tests, however in the existing literature on order restricted inference they have been usually skipped for being studied in detail. Divergence based divergence measures provide a flexible tool for creating meaningful test-statistics, which usually contain the likelihood ratio-test statistics as special case. The existing literature on hypothesis testing with inequality constraints using phi-divergence measures, is centered in a very specific models with multinomial sampling. The contribution of this paper consists in extending and unifying widely the existing work: new families of test-statistics are presented, valid for nested parameter spaces containing either equality or inequality constraints and general distributions for either single or multiple populations are considered.

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

1 Introduction

We consider samples coming from g populations

$$\boldsymbol{X}_{i1},...,\boldsymbol{X}_{ij},...,\boldsymbol{X}_{in_i},\quad i=1,...,g,$$

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

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

where

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

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

*Corresponding author, E-mail: nirian.martin@uc3m.es.

- $\frac{\partial}{\partial\theta_{iu}} f_{\boldsymbol{\theta}_i}(\boldsymbol{x})$ and $\frac{\partial^2}{\partial\theta_{iu}\partial\theta_{iv}} f_{\boldsymbol{\theta}_i}(\boldsymbol{x})$ exist almost everywhere and are such that $\left|\frac{\partial}{\partial\theta_{iu}} f_{\boldsymbol{\theta}_i}(\boldsymbol{x})\right| \leq G_{i,u}(\boldsymbol{x})$, $\left|\frac{\partial^2}{\partial\theta_{iu}\partial\theta_{iv}} f_{\boldsymbol{\theta}_i}(\boldsymbol{x})\right| \leq G_{i,uv}(\boldsymbol{x})$, with $\int_{\mathbb{R}^{m_i}} G_{i,u}(\boldsymbol{x})d\boldsymbol{x} < \infty$ and $\int_{\mathbb{R}^{m_i}} G_{i,uv}(\boldsymbol{x})d\boldsymbol{x} < \infty$.
- $\frac{\partial}{\partial\theta_{iu}} \log f_{\boldsymbol{\theta}_i}(\boldsymbol{x})$ 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 $\delta \to 0$, $\psi_i(\delta) = E\left[\sup_{\{\boldsymbol{t}:\|\boldsymbol{t}\|\leq\delta\}}\left\|\frac{\partial^2}{\partial\boldsymbol{\theta}_i\partial\boldsymbol{\theta}_i^T}\log f_{\boldsymbol{\theta}_i+\boldsymbol{t}}(\boldsymbol{X}_{i1}) - \frac{\partial^2}{\partial\boldsymbol{\theta}_i\partial\boldsymbol{\theta}_i^T}\log f_{\boldsymbol{\theta}_i}(\boldsymbol{X}_{i1})\right\|\right]$, is such that $\psi_i(\delta) \to 0$.

We would like to make statistical inference with respect to an r-dimensional function $\boldsymbol{h}$ which depends on $\boldsymbol{\theta} = (\boldsymbol{\theta}_1^T, ..., \boldsymbol{\theta}_g^T)^T \in \Theta = \Theta_1 \times \cdots \times \Theta_g \subset \mathbb{R}^k$, with $k = \sum_{i=1}^g k_i > r$. Hypotheses of type $\boldsymbol{h}(\boldsymbol{\theta}) = \boldsymbol{0}_r$, $\boldsymbol{h}(\boldsymbol{\theta}) \neq \boldsymbol{0}_r$, $\boldsymbol{h}(\boldsymbol{\theta}) \lneqq \boldsymbol{0}_r$, $\boldsymbol{h}(\boldsymbol{\theta}) \leq \boldsymbol{0}_r$, $\boldsymbol{h}_1(\boldsymbol{\theta}) = \boldsymbol{0}_{r_1}$, $\boldsymbol{h}_2(\boldsymbol{\theta}) \leq \boldsymbol{0}_{r_2}$, 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, some regularity assumptions are considered:

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

In case of focussing only on an internal comparison of components of $\boldsymbol{\theta}_i$ inside the i-th population, matrix $\boldsymbol{H}(\boldsymbol{\theta})$ is block diagonal. In such a case, if no further comparison is made it is more coherent to make statistical inference separately for each population, that is to take the technics shown in this paper with $g = 1$.

The contribution of this paper consists in extending and unifying widely the existing work in different directions. We shall consider two family of test statistics based on ϕ-divergence measures (S_ϕ and T_ϕ families) for testing not only (15)-(16)-(17) but also (10)-(11)-(12). We consider one or more populations and for the last case when having different sample sizes a different version of the test-statistics must be applied ($\widetilde{S}_\phi$ and $\widetilde{T}_\phi$ families). We do not restrict ourselves to a specific kind of distribution for sampling because we consider general populations. Furthermore, breaking with the previously existing papers our methodology for proving the results is based on the theory developed by Aitchison and Silvey (1958) and Silvey (1959), and we follow the trend initiated by El Barmi and Dykstra (1995) for multinomial sampling, which was extended to a more general kind of populations in some of their posterior works. The paper is organized as follows. Sections 2 and 3 will constitute the basis for developing later the asymptotic theory of the proposed test-statistics. More specifically, in Section 2 well-known results related to the joint asymptotic distribution of maximum likelihood estimators and Lagrange multipliers are presented, and in Section 3 the coverage of hypothesis testing problems treated in this paper is explained, as well as the classical test-statistics and their equivalent test-statistics in term of divergence measures. In Section 4 the new test-statistics are introduced and their asymptotic behavior are meticulously shown. Finally, in Section 5 a simple real data example with two Poisson populations is shown and in Section 6 a simulation study is performed which considers a more complicated case of four Binomial populations.

2 Joint asymptotic distribution of maximum likelihood estimators and Lagrange multipliers

If we consider the likelihood function (2) associated with the i-th population, the following properties of (1) are well-known from the basic statistics (see for instance, Sen and Singer (1993), page 210).

i) The asymptotic distribution of the MLE of $\boldsymbol{\theta}_i$, separately for each population, is

$$\sqrt{n_i}(\widehat{\boldsymbol{\theta}}_i - \boldsymbol{\theta}_{i,0}) \xrightarrow[n_i\to\infty]{\mathcal{L}} \mathcal{N}(\mathbf{0}_{k_i}, \mathcal{I}_F^{-1}(\boldsymbol{\theta}_{i,0})).$$

ii) The asymptotic distribution of the MLE of $\boldsymbol{\theta}$ of all populations assuming that exists $\{\nu_i\}_{i=1}^g$ such that $\nu_i = \lim_{n\to\infty} \frac{n_i}{n}$, with $n = \sum_{i=1}^g n_i$, is

$$\sqrt{n}\left(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\right) \xrightarrow[n\to\infty]{\mathcal{L}} \mathcal{N}(\mathbf{0}_k, \mathcal{I}_F^{-1}(\boldsymbol{\theta}_0)), \tag{3}$$

where

$$\mathcal{I}_F^{(n)}(\boldsymbol{\theta}_0) = \bigoplus_{i=1}^g \nu_i \mathcal{I}_F(\boldsymbol{\theta}_{i,0}),$$

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

$$\mathcal{I}_F^{(n)}(\boldsymbol{\theta}_0) = \frac{1}{g}\bigoplus_{i=1}^g \mathcal{I}_F(\boldsymbol{\theta}_{i,0}).$$

we can consider

$$\sqrt{\frac{n}{g}}\left(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\right) \xrightarrow[n\to\infty]{\mathcal{L}} \mathcal{N}(\mathbf{0}_k, \mathcal{I}_F^{-1}(\boldsymbol{\theta}_0)),$$

where

$$\mathcal{I}_F(\boldsymbol{\theta}_0) = \bigoplus_{i=1}^g \mathcal{I}_F(\boldsymbol{\theta}_{i,0}),$$

that is we can consider directly that we have a population of size $\frac{n}{g}$ with a single parameter $\boldsymbol{\theta}_0$.

The asymptotic behavior of estimators with equality restrictions, was studied in origin by Aitchison and Silvey (1958) and Silvey (1959). If we have the likelihood function

$$\ell_n(\boldsymbol{\theta}) = \sum_{i=1}^g \ell_{n_i}(\boldsymbol{\theta}_i), \tag{4}$$

the equality restrictions define a new parameter space

$$\Theta' = \{\boldsymbol{\theta} \in \Theta : \ \boldsymbol{h}(\boldsymbol{\theta}) = \mathbf{0}_r\}.$$

The restricted maximum likelihood estimator is defined as

$$\widehat{\boldsymbol{\theta}} = \arg\max_{\boldsymbol{\theta}\in\Theta'} \ell_n(\boldsymbol{\theta}), \tag{5}$$

and it is obtained solving

$$\begin{aligned} \frac{\partial}{\partial\boldsymbol{\theta}}\ell_n(\boldsymbol{\theta}) + \boldsymbol{H}^T(\boldsymbol{\theta})\boldsymbol{\lambda} &= \mathbf{0}_k, \\ \boldsymbol{h}(\boldsymbol{\theta}) &= \mathbf{0}_r, \end{aligned}$$

where $\boldsymbol{\lambda} \in \mathbb{R}^r$ is the vector of Lagrange multipliers (Sen et al. (2010), page 267).

The joint asymptotic distribution of maximum likelihood estimators and Lagrange multipliers can be decomposed as

$$\begin{pmatrix} \sqrt{n}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \\ \frac{1}{\sqrt{n}}\widehat{\boldsymbol{\lambda}} \end{pmatrix} = \begin{pmatrix} \boldsymbol{P}(\boldsymbol{\theta}_0) & \boldsymbol{Q}(\boldsymbol{\theta}_0) \\ \boldsymbol{Q}^T(\boldsymbol{\theta}_0) & \boldsymbol{R}(\boldsymbol{\theta}_0) \end{pmatrix} \begin{pmatrix} \frac{1}{\sqrt{n}} \frac{\partial}{\partial\boldsymbol{\theta}}\ell_n(\boldsymbol{\theta})\big|_{\boldsymbol{\theta}_0} \\ \mathbf{0}_r \end{pmatrix} + o_P(\mathbf{1}_{k+r}),$$

where

$$\begin{pmatrix} \boldsymbol{P}(\boldsymbol{\theta}_0) & \boldsymbol{Q}(\boldsymbol{\theta}_0) \\ \boldsymbol{Q}^T(\boldsymbol{\theta}_0) & \boldsymbol{R}(\boldsymbol{\theta}_0) \end{pmatrix} = \begin{pmatrix} \mathcal{I}_F(\boldsymbol{\theta}_{i,0}) & -\boldsymbol{H}^T(\boldsymbol{\theta}_0) \\ -\boldsymbol{H}(\boldsymbol{\theta}_0) & \boldsymbol{0}_{r\times r} \end{pmatrix}^{-1},$$

that is

$$\boldsymbol{P}(\boldsymbol{\theta}_0) = \mathcal{I}_F^{-1}(\boldsymbol{\theta}_0) - \mathcal{I}_F^{-1}(\boldsymbol{\theta}_0)\boldsymbol{H}^T(\boldsymbol{\theta}_0)\left(\boldsymbol{H}(\boldsymbol{\theta}_0)\mathcal{I}_F^{-1}(\boldsymbol{\theta}_0)\boldsymbol{H}^T(\boldsymbol{\theta}_0)\right)^{-1}\boldsymbol{H}(\boldsymbol{\theta}_0)\mathcal{I}_F^{-1}(\boldsymbol{\theta}_0), \tag{6a}$$

$$\boldsymbol{Q}(\boldsymbol{\theta}_0) = -\mathcal{I}_F^{-1}(\boldsymbol{\theta}_0)\boldsymbol{H}^T(\boldsymbol{\theta}_0)\left(\boldsymbol{H}(\boldsymbol{\theta}_0)\mathcal{I}_F^{-1}(\boldsymbol{\theta}_0)\boldsymbol{H}^T(\boldsymbol{\theta}_0)\right)^{-1}, \tag{6b}$$

$$\boldsymbol{R}(\boldsymbol{\theta}_0) = -\left(\boldsymbol{H}(\boldsymbol{\theta}_0)\mathcal{I}_F^{-1}(\boldsymbol{\theta}_0)\boldsymbol{H}^T(\boldsymbol{\theta}_0)\right)^{-1}. \tag{6c}$$

Its asymptotic distribution is

$$\begin{pmatrix} \sqrt{n}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \\ \frac{1}{\sqrt{n}}\widehat{\boldsymbol{\lambda}} \end{pmatrix} \xrightarrow[n\to\infty]{\mathcal{L}} \mathcal{N}(\boldsymbol{0}_{k+r}, \boldsymbol{\Sigma}(\boldsymbol{\theta}_0)), \tag{7}$$

where

$$\boldsymbol{\Sigma}(\boldsymbol{\theta}_0) = \begin{pmatrix} \boldsymbol{\Sigma}_{11}(\boldsymbol{\theta}_0) & \boldsymbol{\Sigma}_{12}(\boldsymbol{\theta}_0) \\ \boldsymbol{\Sigma}_{21}(\boldsymbol{\theta}_0) & \boldsymbol{\Sigma}_{22}(\boldsymbol{\theta}_0) \end{pmatrix},$$

with

$$\begin{aligned} \boldsymbol{\Sigma}_{11}(\boldsymbol{\theta}_0) &= \boldsymbol{P}(\boldsymbol{\theta}_0)\mathcal{I}_F(\boldsymbol{\theta}_0)\boldsymbol{P}^T(\boldsymbol{\theta}_0) = \boldsymbol{P}(\boldsymbol{\theta_0}), \\ \boldsymbol{\Sigma}_{12}(\boldsymbol{\theta}_0) &= \boldsymbol{P}(\boldsymbol{\theta}_0)\mathcal{I}_F(\boldsymbol{\theta}_0)\boldsymbol{Q}(\boldsymbol{\theta}_0) = \boldsymbol{0}_{k\times r}, \\ \boldsymbol{\Sigma}_{21}(\boldsymbol{\theta}_0) &= \boldsymbol{\Sigma}_{12}^T(\boldsymbol{\theta}_0) = \boldsymbol{0}_{r\times k}, \\ \boldsymbol{\Sigma}_{22}(\boldsymbol{\theta}_0) &= \boldsymbol{Q}^T(\boldsymbol{\theta}_0)\mathcal{I}_F(\boldsymbol{\theta}_0)\boldsymbol{Q}(\boldsymbol{\theta}_0) = -\boldsymbol{R}(\boldsymbol{\theta_0}). \end{aligned} \tag{8}$$

When working with g populations, it is very interesting to be able to express $\boldsymbol{P}(\boldsymbol{\theta_0})$ and $\boldsymbol{R}(\boldsymbol{\theta_0})$ in terms of submatrices

$$\begin{aligned} \boldsymbol{R}(\boldsymbol{\theta_0}) &= -\left(\textstyle\sum_{i=1}^{g}\dfrac{1}{\nu_i}\boldsymbol{H}_i(\boldsymbol{\theta})\mathcal{I}_F^{-1}(\boldsymbol{\theta}_{i,0})\boldsymbol{H}_i^T(\boldsymbol{\theta})\right)^{-1}, \\ \boldsymbol{P}(\boldsymbol{\theta_0}) &= (\boldsymbol{P}_{ij}(\boldsymbol{\theta_0}))_{i,j\in\{1,\dots,g\}}, \\ \boldsymbol{P}_{ij}(\boldsymbol{\theta_0}) &= \begin{cases} \frac{1}{\nu_i}\mathcal{I}_F^{-1}(\boldsymbol{\theta}_{i,0}) + \frac{1}{\nu_i^2}\mathcal{I}_F^{-1}(\boldsymbol{\theta}_{i,0})\boldsymbol{H}_i^T(\boldsymbol{\theta})\boldsymbol{R}(\boldsymbol{\theta_0})\boldsymbol{H}_i(\boldsymbol{\theta})\mathcal{I}_F^{-1}(\boldsymbol{\theta}_{i,0}), & \text{if } i=j \\ \frac{1}{\nu_i\nu_j}\mathcal{I}_F^{-1}(\boldsymbol{\theta}_{i,0})\boldsymbol{H}_i^T(\boldsymbol{\theta})\boldsymbol{R}(\boldsymbol{\theta_0})\boldsymbol{H}_j(\boldsymbol{\theta})\mathcal{I}_F^{-1}(\boldsymbol{\theta}_{j,0}), & \text{if } i\neq j \end{cases}. \end{aligned}$$

When working with populations with equal sizes, we can work as we had a single population of size $\frac{n}{g}$, that is

$$\begin{pmatrix} \sqrt{\frac{n}{g}}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \\ \frac{1}{\sqrt{\frac{n}{g}}}\widehat{\boldsymbol{\lambda}} \end{pmatrix} \xrightarrow[n\to\infty]{\mathcal{L}} \mathcal{N}(\boldsymbol{0}_{k+r}, \boldsymbol{\Sigma}(\boldsymbol{\theta}_0)),$$

and the structure of matrix $\boldsymbol{\Sigma}(\boldsymbol{\theta}_0))$ is the same but $\nu_i = 1$, $i = 1, ..., g$, that is

$$\begin{aligned} \boldsymbol{R}(\boldsymbol{\theta_0}) &= -\left(\textstyle\sum_{i=1}^{g}\boldsymbol{H}_i(\boldsymbol{\theta})\mathcal{I}_F^{-1}(\boldsymbol{\theta}_{i,0})\boldsymbol{H}_i^T(\boldsymbol{\theta})\right)^{-1}, \\ \boldsymbol{P}(\boldsymbol{\theta_0}) &= (\boldsymbol{P}_{ij}(\boldsymbol{\theta_0}))_{i,j\in\{1,\dots,g\}}, \\ \boldsymbol{P}_{ij}(\boldsymbol{\theta_0}) &= \begin{cases} \mathcal{I}_F^{-1}(\boldsymbol{\theta}_{i,0}) + \frac{1}{\nu_i^2}\mathcal{I}_F^{-1}(\boldsymbol{\theta}_{i,0})\boldsymbol{H}_i^T(\boldsymbol{\theta})\boldsymbol{R}(\boldsymbol{\theta_0})\boldsymbol{H}_i(\boldsymbol{\theta})\mathcal{I}_F^{-1}(\boldsymbol{\theta}_{i,0}), & \text{if } i=j \\ \mathcal{I}_F^{-1}(\boldsymbol{\theta}_{i,0})\boldsymbol{H}_i^T(\boldsymbol{\theta})\boldsymbol{R}(\boldsymbol{\theta_0})\boldsymbol{H}_j(\boldsymbol{\theta})\mathcal{I}_F^{-1}(\boldsymbol{\theta}_{j,0}), & \text{if } i\neq j \end{cases}. \end{aligned}$$

3 Hypothesis testing formulation

In this section we are going to define a sequence of nested hypotheses which are nested by adding either equality or inequality restrictions. We are going to focus on

$$\Theta^{(1)} = \{\boldsymbol{\theta} \in \Theta : \ \boldsymbol{h}(\boldsymbol{\theta}) = \mathbf{0}_r\},$$
$$\Theta^{(2)} = \{\boldsymbol{\theta} \in \Theta : \ \boldsymbol{h}_1(\boldsymbol{\theta}) = \mathbf{0}_{r_1}\},$$
$$\Omega^{(3)} = \{\boldsymbol{\theta} \in \Theta : \ \boldsymbol{h}(\boldsymbol{\theta}) \leq \mathbf{0}_r\},$$
$$\Omega^{(4)} = \{\boldsymbol{\theta} \in \Theta : \ \boldsymbol{h}_1(\boldsymbol{\theta}) \leq \mathbf{0}_{r_1}\},$$
$$\Theta^{(5)} = \Theta,$$

where $\boldsymbol{h}(\boldsymbol{\theta}) = (\boldsymbol{h}_1(\boldsymbol{\theta}), \boldsymbol{h}_2(\boldsymbol{\theta}))$, with $r = r_1 + r_2$, assuming that the regularity conditions presented in Section 1 hold. Note that $\Theta^{(1)} \subset \Theta^{(2)} \subset \Omega^{(4)} \subset \Theta^{(5)}$, $\Theta^{(1)} \subset \Omega^{(3)} \subset \Omega^{(4)} \subset \Theta^{(5)}$ and $\Theta^{(2)} \not\subset \Omega^{(3)}$, but $\Omega^{(1b)} = \Theta^{(2)} \cap \Omega^{(3)} = \{\boldsymbol{\theta} \in \Theta : \boldsymbol{h}_1(\boldsymbol{\theta}) = \mathbf{0}_{r_1}, \boldsymbol{h}_2(\boldsymbol{\theta}) \leq \mathbf{0}_{r_2}\} \subset \Theta^{(2)}$. Observe also, that parameter spaces denoted by Θ are vector spaces and parameter spaces denoted by Ω are closed and convex cones. We would like to test hypotheses such as

$$H_{Null}: \ \boldsymbol{\theta} \in \Theta^{(1)} \quad \text{vs.} \quad H_{Alt}: \ \boldsymbol{\theta} \in \Theta^{(2)} - \Theta^{(1)}; \tag{9a}$$
$$H_{Null}: \ \boldsymbol{\theta} \in \Theta^{(1)} \quad \text{vs.} \quad H_{Alt}: \ \boldsymbol{\theta} \in \Omega^{(3)} - \Theta^{(1)}; \tag{9b}$$
$$H_{Null}: \ \boldsymbol{\theta} \in \Theta^{(1)} \quad \text{vs.} \quad H_{Alt}: \ \boldsymbol{\theta} \in \Omega^{(4)} - \Theta^{(1)}; \tag{9c}$$
$$H_{Null}: \ \boldsymbol{\theta} \in \Theta^{(1)} \quad \text{vs.} \quad H_{Alt}: \ \boldsymbol{\theta} \in \Theta^{(5)} - \Theta^{(1)}; \tag{9d}$$
$$H_{Null}: \ \boldsymbol{\theta} \in \Omega^{(1b)} \quad \text{vs.} \quad H_{Alt}: \ \boldsymbol{\theta} \in \Theta^{(2)} - \Omega^{(1b)}; \tag{9e}$$
$$H_{Null}: \ \boldsymbol{\theta} \in \Theta^{(2)} \quad \text{vs.} \quad H_{Alt}: \ \boldsymbol{\theta} \in \Omega^{(4)} - \Theta^{(2)}; \tag{9f}$$
$$H_{Null}: \ \boldsymbol{\theta} \in \Theta^{(2)} \quad \text{vs.} \quad H_{Alt}: \ \boldsymbol{\theta} \in \Theta^{(5)} - \Theta^{(2)}; \tag{9g}$$
$$H_{Null}: \ \boldsymbol{\theta} \in \Omega^{(3)} \quad \text{vs.} \quad H_{Alt}: \ \boldsymbol{\theta} \in \Theta^{(5)} - \Omega^{(3)}; \tag{9h}$$
$$H_{Null}: \ \boldsymbol{\theta} \in \Omega^{(4)} \quad \text{vs.} \quad H_{Alt}: \ \boldsymbol{\theta} \in \Theta^{(5)} - \Omega^{(4)}; \tag{9i}$$

We avoided pair $\Omega^{(3)} \subset \Omega^{(4)}$ because we must have at least one vector space either as H_{Null} or H_{Alt}. As in Section 3.2 of Silvapulle and Sen (2004) we classify the tests in three types. Let $E \subset \{1, ..., r\}$ the set of indices such that $h_i(\boldsymbol{\theta})$ is active, that is $E = \{i \in \{1, ..., r\} : h_i(\boldsymbol{\theta}) = 0\}$. We consider test of type O, type A and type B according to the character of the parameter spaces in the null and alternative hypothesis

$$H^O_{Null}: \ \boldsymbol{\theta} \in \Theta(E) \quad \text{vs.} \quad H^O_{Alt}: \ \boldsymbol{\theta} \in \Theta(F) \text{ and } \boldsymbol{\theta} \notin \Theta(E), \tag{10}$$
$$H^A_{Null}: \ \boldsymbol{\theta} \in \Theta(E) \quad \text{vs.} \quad H^A_{Alt}: \ \boldsymbol{\theta} \in \Omega(F) \text{ and } \boldsymbol{\theta} \notin \Theta(E), \tag{11}$$
$$H^B_{Null}: \ \boldsymbol{\theta} \in \Omega(E) \quad \text{vs.} \quad H^B_{Alt}: \ \boldsymbol{\theta} \in \Theta(F) = \Theta \text{ and } \boldsymbol{\theta} \notin \Omega(E), \tag{12}$$

where $F \subset E$. For instance, in (9a) we have $E = \{1, ..., r\}$ and $F = \{1, ..., r_1\}$. Since $\Theta^{(1)}, \Theta^{(2)}, \Theta^{(5)}$ are vector spaces and $\Omega^{(3)}, \Omega^{(4)}$ closed and convex cones, type O tests are (9a), (9d), (9g), type A tests are (9b), (9c), (9f) and type B tests are (9e), (9h), (9i). For MLEs under both hypotheses, null and alternative, and being S the set of active indices, we distinguish

$$\widehat{\boldsymbol{\theta}}(S) = \arg\min_{\Theta(S)} \ell_n(\boldsymbol{\theta}), \tag{13}$$

when $\Theta(S)$ is a vector space, and

$$\widetilde{\boldsymbol{\theta}}(S) = \arg\min_{\Omega(S)} \ell_n(\boldsymbol{\theta}), \tag{14}$$

when $\Omega(S)$ is a closed and convex set. Furthermore, $\widehat{\boldsymbol{\theta}} = \widehat{\boldsymbol{\theta}}(\varnothing)$ denotes the MLE without restrictions.

In the classical perspective of test-statistics the likelihood ratio test-statistic is used to perform (10)-(11)-(12). Moreover, usually their explanation is limited to

$$H_{Null}^{O}: \ \boldsymbol{\theta} \in \Theta(R) \quad \text{vs.} \quad H_{Alt}^{O}: \ \boldsymbol{\theta} \in \Theta(\varnothing) = \Theta \text{ and } \boldsymbol{\theta} \notin \Theta(R), \tag{15}$$

$$H_{Null}^{A}: \ \boldsymbol{\theta} \in \Theta(R) \quad \text{vs.} \quad H_{Alt}^{A}: \ \boldsymbol{\theta} \in \Omega(\varnothing) \text{ and } \boldsymbol{\theta} \notin \Theta(R), \tag{16}$$

$$H_{Null}^{B}: \ \boldsymbol{\theta} \in \Omega(\varnothing) \quad \text{vs.} \quad H_{Alt}^{B}: \ \boldsymbol{\theta} \in \Theta(\varnothing) = \Theta \text{ and } \boldsymbol{\theta} \notin \Omega(\varnothing), \tag{17}$$

where $R = \{1, ..., r\} = E$ is the hypothesis of being active all the constraints, and $F = \varnothing$.

The likelihood ratio test-statistics for testing (10)-(11)-(12) in the case of single population or g populations, are given by

$$G(\widehat{\boldsymbol{\theta}}(F), \widehat{\boldsymbol{\theta}}(E)) = 2\left(\ell_n(\widehat{\boldsymbol{\theta}}(F)) - \ell_n(\widehat{\boldsymbol{\theta}}(E))\right) = 2\left(\sum_{i=1}^{g} \ell_{n_i}(\widehat{\boldsymbol{\theta}}_i(F)) - \sum_{i=1}^{g} \ell_{n_i}(\widehat{\boldsymbol{\theta}}_i(E))\right), \tag{18}$$

$$G(\widetilde{\boldsymbol{\theta}}(F), \widehat{\boldsymbol{\theta}}(E)) = 2\left(\ell_n(\widetilde{\boldsymbol{\theta}}(F)) - \ell_n(\widehat{\boldsymbol{\theta}}(E))\right) = 2\left(\sum_{i=1}^{g} \ell_{n_i}(\widetilde{\boldsymbol{\theta}}_i(F)) - \sum_{i=1}^{g} \ell_{n_i}(\widehat{\boldsymbol{\theta}}_i(E))\right), \tag{19}$$

$$G(\widehat{\boldsymbol{\theta}}(F), \widetilde{\boldsymbol{\theta}}(E)) = 2\left(\ell_n(\widehat{\boldsymbol{\theta}}(F)) - \ell_n(\widetilde{\boldsymbol{\theta}}(E)))\right) = 2\left(\sum_{i=1}^{g} \ell_{n_i}(\widehat{\boldsymbol{\theta}}_i(F)) - \sum_{i=1}^{g} \ell_{n_i}(\widetilde{\boldsymbol{\theta}}_i(E))\right). \tag{20}$$

and the asymptotic distribution of the first one under H_{Null}^{O} is $\chi^2_{\mathrm{card}(F)-\mathrm{card}(E)}$, while for the the other two under H_{Null}^{A} and H_{Null}^{B} are respectively a mixture of $\{\chi_i^2\}_{i=0}^{r}$ random variables, known as chi-bar squared random variable ($\chi_0^2 \equiv 0$). For more details about likelihood ratio test-statistics see Barlow et al. (1972), Robertson et al. (1988) or Silvapulle and Sen (2004). Now we define the Kullback-Leibler divergence based test-statistics and later in Proposition 2 we will show its relationship with the likelihood ratio test-statistics.

Definition 1 *The* ***Kullback divergence*** *based test-statistics for testing (10)-(11)-(12) in the case of single population (or g populations with the same sample size), are given by*

$$T^{O}(\widehat{\boldsymbol{\theta}}, \widehat{\boldsymbol{\theta}}(F), \widehat{\boldsymbol{\theta}}(E)) = 2n\left(\mathrm{d}_{Kull}(f_{\widehat{\boldsymbol{\theta}}}, f_{\widehat{\boldsymbol{\theta}}(E)}) - \mathrm{d}_{Kull}(f_{\widehat{\boldsymbol{\theta}}}, f_{\widehat{\boldsymbol{\theta}}(F)})\right), \tag{21}$$

$$T^{A}(\widehat{\boldsymbol{\theta}}, \widetilde{\boldsymbol{\theta}}(F), \widehat{\boldsymbol{\theta}}(E)) = 2n\left(\mathrm{d}_{Kull}(f_{\widehat{\boldsymbol{\theta}}}, f_{\widehat{\boldsymbol{\theta}}(E)}) - \mathrm{d}_{Kull}(f_{\widehat{\boldsymbol{\theta}}}, f_{\widetilde{\boldsymbol{\theta}}(F)})\right), \tag{22}$$

$$T^{B}(\widehat{\boldsymbol{\theta}}, \widehat{\boldsymbol{\theta}}(F), \widetilde{\boldsymbol{\theta}}(E)) = 2n\left(\mathrm{d}_{Kull}(f_{\widehat{\boldsymbol{\theta}}}, f_{\widetilde{\boldsymbol{\theta}}(E)}) - \mathrm{d}_{Kull}(f_{\widehat{\boldsymbol{\theta}}}, f_{\widehat{\boldsymbol{\theta}}(F)})\right), \tag{23}$$

where

$$\mathrm{d}_{Kull}(f_{\boldsymbol{\theta}_1}, f_{\boldsymbol{\theta}_2}) = E\left[\log\left(\frac{f_{\boldsymbol{\theta}_1}(\boldsymbol{X}_1)}{f_{\boldsymbol{\theta}_2}(\boldsymbol{X}_1)}\right) - \frac{f_{\boldsymbol{\theta}_1}(\boldsymbol{X}_1)}{f_{\boldsymbol{\theta}_2}(\boldsymbol{X}_1)} + 1\right] = \int_{\mathcal{X}} f_{\boldsymbol{\theta}_1}(\boldsymbol{x}) \log\left(\frac{f_{\boldsymbol{\theta}_1}(\boldsymbol{x})}{f_{\boldsymbol{\theta}_2}(\boldsymbol{x})}\right) d\boldsymbol{x}, \tag{24}$$

and with $f_{\boldsymbol{\theta}_1}(\boldsymbol{X}_1)$ being the density function of one individual in the sample and $\mathrm{d}_{Kull}(f_{\boldsymbol{\theta}_1}, f_{\boldsymbol{\theta}_2})$ the Kullback divergence measure among two distributions.

The divergence is also applicable in discrete setting by replacing density function for probability mass function and the integral by the summation.

Proposition 2 *For the exponential family*

$$f_{\boldsymbol{\theta}}(\boldsymbol{x}) = q(\boldsymbol{\theta}) r(\boldsymbol{x}) \exp(\boldsymbol{s}^T(\boldsymbol{\theta}) \boldsymbol{t}(\boldsymbol{x})), \quad \boldsymbol{x} \in \mathcal{X}, \tag{25}$$

the Kullback divergence based test-statistics (21)-(22)-(23) for testing (10)-(11)-(12) in the case of single population, are exactly equal to the likelihood ratio test-statistics (18)-(19)-(20).

Proof. In Pardo (2006, Remark 9.4) we can find the proof for the simple null hypothesis,

$$2n\mathrm{d}_{Kull}(f_{\widehat{\boldsymbol{\theta}}}, f_{\theta_0}) = 2\left(\ell_n(\widehat{\boldsymbol{\theta}}) - \ell_n(\boldsymbol{\theta}_0)\right).$$

From this formula it is straightforward to proof for testing (15) and (12)

$$T^O(\widehat{\boldsymbol{\theta}}, \widehat{\boldsymbol{\theta}}(\varnothing), \widehat{\boldsymbol{\theta}}(R)) = G(\widehat{\boldsymbol{\theta}}(\varnothing), \widehat{\boldsymbol{\theta}}(R)),$$
$$T^B(\widehat{\boldsymbol{\theta}}, \widehat{\boldsymbol{\theta}}(\varnothing), \widetilde{\boldsymbol{\theta}}(\bullet)) = G(\widehat{\boldsymbol{\theta}}(\varnothing), \widetilde{\boldsymbol{\theta}}(\bullet)),$$

where $R = \{1, ..., r\} = E$ is the hypothesis of being active all the constraints, $F = \varnothing$ and $\bullet \subset R$. All the rest of the cases, i.e. with general sets E and F such that $F \subset E$, are immediately obtained from the previous ones, because

$$T^O(\widehat{\boldsymbol{\theta}}, \widehat{\boldsymbol{\theta}}(F), \widehat{\boldsymbol{\theta}}(E)) = T^O(\widehat{\boldsymbol{\theta}}, \widehat{\boldsymbol{\theta}}(\varnothing), \widehat{\boldsymbol{\theta}}(E)) - T^O(\widehat{\boldsymbol{\theta}}, \widehat{\boldsymbol{\theta}}(\varnothing), \widehat{\boldsymbol{\theta}}(F)),$$
$$T^A(\widehat{\boldsymbol{\theta}}, \widetilde{\boldsymbol{\theta}}(F), \widehat{\boldsymbol{\theta}}(E)) = T^O(\widehat{\boldsymbol{\theta}}, \widehat{\boldsymbol{\theta}}(\varnothing), \widehat{\boldsymbol{\theta}}(E)) - T^B(\widehat{\boldsymbol{\theta}}, \widehat{\boldsymbol{\theta}}(\varnothing), \widetilde{\boldsymbol{\theta}}(F)),$$
$$T^B(\widehat{\boldsymbol{\theta}}, \widehat{\boldsymbol{\theta}}(F), \widetilde{\boldsymbol{\theta}}(E)) = T^B(\widehat{\boldsymbol{\theta}}, \widehat{\boldsymbol{\theta}}(\varnothing), \widetilde{\boldsymbol{\theta}}(E)) - T^O(\widehat{\boldsymbol{\theta}}, \widehat{\boldsymbol{\theta}}(\varnothing), \widehat{\boldsymbol{\theta}}(F)).$$

■

Let $\mathcal{M}(n, \boldsymbol{\pi})$ the case of single "multinomial" population, where $\boldsymbol{\pi} = (\pi_1, ..., \pi_k, \pi_{k+1})^T$ is so that the not redundant part of the parameter is $(\pi_1, ..., \pi_k)^T = (\theta_1, ..., \theta_k)^T = \boldsymbol{\theta}$.and $\pi_{k+1} = 1 - \sum_{j=1}^{k} \pi_j = 1 - \boldsymbol{\theta}^T \mathbf{1}$. Looking at pages 239-240 of Robertson et al. (1988), in such a case the chi-square test-statistics for testing (10)-(11)-(12) must be defined as

$$C^O(\widehat{\boldsymbol{\theta}}(F), \widehat{\boldsymbol{\theta}}(E)) = n \sum_{j=1}^{k+1} \frac{(\widehat{\pi}_j(E) - \widehat{\pi}_j(F))^2}{\widehat{\pi}_j(F)},$$
$$C^A(\widetilde{\boldsymbol{\theta}}(F), \widehat{\boldsymbol{\theta}}(E)) = n \sum_{j=1}^{k+1} \frac{(\widehat{\pi}_j(E) - \widetilde{\pi}_j(F))^2}{\widetilde{\pi}_j(F)},$$
$$C^B(\widehat{\boldsymbol{\theta}}(F), \widetilde{\boldsymbol{\theta}}(E)) = n \sum_{j=1}^{k+1} \frac{(\widetilde{\pi}_j(E) - \widehat{\pi}_j(F))^2}{\widehat{\pi}_j(F)}.$$

Now, focussing in general populations, such test-statistics are defined in term of a special divergence measure.

Definition 3 *The **Pearson divergence** based test-statistics for testing (10)-(11)-(12) in the case of single population (or g populations with the same sample size), are given by*

$$C^O(\widehat{\boldsymbol{\theta}}(F), \widehat{\boldsymbol{\theta}}(E)) = 2n\mathrm{d}_{Pearson}(f_{\widehat{\boldsymbol{\theta}}(F)}, f_{\widehat{\boldsymbol{\theta}}(E)}), \tag{26}$$
$$C^A(\widetilde{\boldsymbol{\theta}}(F), \widehat{\boldsymbol{\theta}}(E)) = 2n\mathrm{d}_{Pearson}(f_{\widetilde{\boldsymbol{\theta}}(F)}, f_{\widehat{\boldsymbol{\theta}}(E)}), \tag{27}$$
$$C^B(\widehat{\boldsymbol{\theta}}(F), \widetilde{\boldsymbol{\theta}}(E)) = 2n\mathrm{d}_{Pearson}(f_{\widehat{\boldsymbol{\theta}}(F)}, f_{\widetilde{\boldsymbol{\theta}}(E)}), \tag{28}$$

where

$$\mathrm{d}_{Pearson}(f_{\boldsymbol{\theta}_1}, f_{\boldsymbol{\theta}_2}) = \frac{1}{2} E\left[\left(\frac{f_{\boldsymbol{\theta}}(\mathbf{X}) - f_{\boldsymbol{\theta}_0}(\mathbf{X})}{f_{\boldsymbol{\theta}_0}(\mathbf{X})}\right)^2\right] = \frac{1}{2}\int_{\mathcal{X}} \frac{(f_{\boldsymbol{\theta}}(\boldsymbol{x}) - f_{\boldsymbol{\theta}_0}(\boldsymbol{x}))^2}{f_{\boldsymbol{\theta}_0}(\boldsymbol{x})} d\boldsymbol{x}, \tag{29}$$

and with $f_{\boldsymbol{\theta}_1}(\boldsymbol{X}_1)$ being the density function of one individual in the sample and $\mathrm{d}_{Pearson}(f_{\boldsymbol{\theta}_1}, f_{\boldsymbol{\theta}_2})$ the Pearson divergence measure among two distributions.

The aim of this work is to build new test-statistics for extending the Kullback-Leibler and Pearson divergence to a more general divergence measures, ϕ-divergence measures, which are valid for testing (10)-(11)-(12) in general populations. Let $\phi : \mathbb{R}_+ \longrightarrow \mathbb{R}$ a convex function such that:

- $\phi(1) = \phi'(1) = 0$, $\phi''(1) > 0$;
- $0\phi(\frac{0}{0}) = 0$, $0\phi(\frac{p}{0}) = \lim_{u\to\infty} \frac{\phi(u)}{u}$, for $p \neq 0$.

Definition 4 *Let $\boldsymbol{X}$ be a random variable with distribution function (density function) $F_{\boldsymbol{\theta}}(\boldsymbol{x})$ $(f_{\boldsymbol{\theta}}(\boldsymbol{x}))$, $i = 1, ..., g$, which depends on an unknown parameter $\boldsymbol{\theta} = (\theta_1, ..., \theta_k)^T \in \Theta \subset \mathbb{R}^k$, its support is $\mathcal{X}$ and hold the regularity conditions of Section 1. The ϕ-**divergence measure between $f_{\boldsymbol{\theta}}$ and** $f_{\boldsymbol{\theta}_0}$, with $\boldsymbol{\theta}, \boldsymbol{\theta}_0 \in \Theta \subset \mathbb{R}^k$, is defined as*

$$\mathrm{d}_\phi(f_{\boldsymbol{\theta}}, f_{\boldsymbol{\theta}_0}) = E\left[\phi\left(\frac{f_{\boldsymbol{\theta}}(\boldsymbol{X})}{f_{\boldsymbol{\theta}_0}(\boldsymbol{X})}\right)\right] = \int_{\mathcal{X}} f_{\boldsymbol{\theta}_0}(\boldsymbol{x})\phi\left(\frac{f_{\boldsymbol{\theta}}(\boldsymbol{x})}{f_{\boldsymbol{\theta}_0}(\boldsymbol{x})}\right) d\boldsymbol{x}.$$

It is assumed that function $\mathrm{d}_\phi(\boldsymbol{\theta}) = \mathrm{d}_\phi(f_{\boldsymbol{\theta}}, f_{\boldsymbol{\theta}'})$ is one and two order differentiable under integration sign, that is for $\frac{\partial}{\partial \boldsymbol{\theta}}\mathrm{d}_\phi(\boldsymbol{\theta}) = (\frac{\partial}{\partial \theta_1}\mathrm{d}_\phi(\boldsymbol{\theta}), ..., \frac{\partial}{\partial \theta_1}\mathrm{d}_\phi(\boldsymbol{\theta}))^T$ it holds

$$\frac{\partial}{\partial \theta_i}\mathrm{d}_\phi(\boldsymbol{\theta}) = \int_{\mathcal{X}} \frac{\partial}{\partial \theta_i}\left(f_{\boldsymbol{\theta}_0}(\boldsymbol{x})\phi\left(\frac{f_{\boldsymbol{\theta}}(\boldsymbol{x})}{f_{\boldsymbol{\theta}_0}(\boldsymbol{x})}\right)\right) d\boldsymbol{x} = \int_{\mathcal{X}} \phi'\left(\frac{f_{\boldsymbol{\theta}}(\boldsymbol{x})}{f_{\boldsymbol{\theta}_0}(\boldsymbol{x})}\right)\frac{\partial f_{\boldsymbol{\theta}}(\boldsymbol{x})}{\partial \theta_i} d\boldsymbol{x}, \quad i = 1, ..., k, \tag{30}$$

and for $\frac{\partial^2}{\partial \boldsymbol{\theta}\partial \boldsymbol{\theta}^T}\mathrm{d}_\phi(\boldsymbol{\theta}) = \left(\frac{\partial^2}{\partial \theta_i \partial \theta_j}\mathrm{d}_\phi(\boldsymbol{\theta})\right)_{i,j\in\{1,...,k\}}$

$$\begin{aligned}\frac{\partial^2}{\partial \theta_i \partial \theta_j}\mathrm{d}_\phi(\boldsymbol{\theta}) &= \int_{\mathcal{X}} \frac{\partial^2}{\partial \theta_i \partial \theta_j}\left(f_{\boldsymbol{\theta}_0}(\boldsymbol{x})\phi\left(\frac{f_{\boldsymbol{\theta}}(\boldsymbol{x})}{f_{\boldsymbol{\theta}_0}(\boldsymbol{x})}\right)\right) d\boldsymbol{x} \\ &= \int_{\mathcal{X}} \phi''\left(\frac{f_{\boldsymbol{\theta}}(\boldsymbol{x})}{f_{\boldsymbol{\theta}_0}(\boldsymbol{x})}\right)\frac{1}{f_{\boldsymbol{\theta}_0}(\boldsymbol{x})}\frac{\partial f_{\boldsymbol{\theta}}(\boldsymbol{x})}{\partial \theta_i}\frac{\partial f_{\boldsymbol{\theta}}(\boldsymbol{x})}{\partial \theta_j} d\boldsymbol{x} + \int_{\mathcal{X}} \phi'\left(\frac{f_{\boldsymbol{\theta}}(\boldsymbol{x})}{f_{\boldsymbol{\theta}_0}(\boldsymbol{x})}\right)\frac{\partial^2 f_{\boldsymbol{\theta}}(\boldsymbol{x})}{\partial \theta_i \partial \theta_j} d\boldsymbol{x}.\end{aligned} \tag{31}$$

Remark 5 *The Kullback-Leibler divergence (24) is a particular case of ϕ-divergence measure with $\phi(x) = x\log x - x + 1$ and the Pearson divergence (29) is a particular case of ϕ-divergence measure with $\phi(x) = \frac{1}{2}(x-1)^2$.*

Definition 4 considers a broad family of divergence measures but there is a very well-known subfamily called **power-divergence measures** (Read and Cressie (1988))

$$\mathrm{d}_{\phi_\lambda}(f_{\boldsymbol{\theta}}, f_{\boldsymbol{\theta}_0}) = \frac{1}{\lambda(1+\lambda)}\left(E\left[\left(\frac{f_{\boldsymbol{\theta}}(\boldsymbol{X})}{f_{\boldsymbol{\theta}_0}(\boldsymbol{x})}\right)^{\lambda+1}\right] - 1\right) = \frac{1}{\lambda(1+\lambda)}\left(\int_{\mathcal{X}} \frac{f_{\boldsymbol{\theta}}^{\lambda+1}(\boldsymbol{x})}{f_{\boldsymbol{\theta}_0}^{\lambda}(\boldsymbol{x})} d\boldsymbol{x} - 1\right), \tag{32}$$

for $\lambda \in \mathbb{R} - \{-1, 0\}$ and $\mathrm{d}_{\phi_\lambda}(f_{\boldsymbol{\theta}}, f_{\boldsymbol{\theta}_0}) = \lim_{\ell\to 0} \mathrm{d}_{\phi_\ell}(f_{\boldsymbol{\theta}}, f_{\boldsymbol{\theta}_0})$ for $\lambda \in \{-1, 0\}$. It is a particular case of ϕ-divergence measure with $\phi_\lambda(x) = \frac{1}{\lambda(1+\lambda)}(x^{\lambda+1} - x - \lambda(x-1))$, and it covers the Kullback-Leibler and Pearson divergence as special case, taking $\lambda = 0$ and $\lambda = 1$ respectively. Definition 4 is the basis in the test-statistics based on ϕ-divergence measures that are going to be built in the following sections, for a single population as well as for multiple populations with equal sizes, because joining the random variable of its population it is possible on one hand to consider a common first individual, second one,... and so on for the whole population, and on the other hand to consider its parameter to be (34). The product-multinomial distribution is a case of multiple multinomial populations but how to manage the whole sample is shown in Section 6, even with different sample sizes, as it were a single population. However in general, Definition 4 is not longer valid to construct test-statistics for multiple populations and unequal sample sizes, and this the reason why a different divergence measure must be defined.

Definition 6 *Let* $\mathcal{L}(\boldsymbol{X};\boldsymbol{\theta})$ *the likelihood function on the whole sample of* g *populations, where*

$$\boldsymbol{X} = \boldsymbol{X}_{11}, ..., \boldsymbol{X}_{1n_1}, ..., \boldsymbol{X}_{g1}, ..., \boldsymbol{X}_{gn_g}, \tag{33}$$

$$\boldsymbol{\theta} = (\boldsymbol{\theta}_1^T, ..., \boldsymbol{\theta}_g^T)^T. \tag{34}$$

The ϕ***-divergence measure between*** $\mathcal{L}_{\boldsymbol{\theta}}$ ***and*** $\mathcal{L}_{\boldsymbol{\theta}_0}$*, with* $\boldsymbol{\theta}, \boldsymbol{\theta}_0 \in \Theta \subset \mathbb{R}^k$*, is defined as*

$$\mathrm{d}_\phi(\mathcal{L}_{\boldsymbol{\theta}}, \mathcal{L}_{\boldsymbol{\theta}_0}) = E\left[\phi\left(\frac{\mathcal{L}_{\boldsymbol{\theta}}(\boldsymbol{X})}{\mathcal{L}_{\boldsymbol{\theta}_0}(\boldsymbol{X})}\right)\right] = \int_{\mathcal{X}^n} \mathcal{L}_{\boldsymbol{\theta}_0}(\boldsymbol{x})\phi\left(\frac{\mathcal{L}_{\boldsymbol{\theta}}(\boldsymbol{x})}{\mathcal{L}_{\boldsymbol{\theta}_0}(\boldsymbol{x})}\right) d\boldsymbol{x}. \tag{35}$$

It is assumed that the condition of differentiability under the integral sign for the likelihood function are the same as defined for the density function.

4 New test-statistics and their asymptotic distributions

In the literature papers where ϕ-divergence measures are applied for testing (15) can be encountered (for example, (9d) or (9g)). The idea of using the Kullback-divergence measure among two densities is attributable to Kupperman (1957). However, when less restriction than established by R are taken into account, $\Theta(E) = \{\boldsymbol{\theta} = (\theta_1, ..., \theta_k)^T : h_i(\boldsymbol{\theta}) = 0, i = 1, ..., \mathrm{card}(E)\}$, with $\mathrm{card}(E) < k$, and $\theta_i^0, i = 1, ..., \mathrm{card}(E)$, composite null hypothesis must be considered and in the paper of Zografos and other (1990) was performed this task but only for multinomial populations. Later in Salicrú et al. (1994), even though general populations were taken into account it was only for a very specific restrictions $h_i(\boldsymbol{\theta}) = \theta_i - \theta_i^0, i = 1, ..., \mathrm{card}(E)$, and in Morales et al. (1997) general models and restrictions were analyzed for (15), either for a population or multiple populations. In Menéndez et al. (1997), Zografos (1998), Morales et al. (1998) some problems related to the previous paper with multiple populations were analyzed. In Morales et al. (2001) “likelihood ϕ-divergence test statistics”, based on (35)

$$\widetilde{S}_\phi^O(\widehat{\boldsymbol{\theta}}, \widehat{\boldsymbol{\theta}}(E)) = \frac{2}{\phi''(1)} \mathrm{d}_\phi(\mathcal{L}_{\widehat{\boldsymbol{\theta}}}, \mathcal{L}_{\widehat{\boldsymbol{\theta}}(E)}), \tag{36}$$

were introduced for the first time. In Hobza et al. (2003) the familiar data problem which is for multiple populations was analyzed using (36). Test-statistics of type $\widetilde{T}_\phi$ have never been applied (see Definitions 9, 14, 20). In Menéndez et al. (2002), Menéndez et al. (2003a), Menéndez et al. (2003b), Pardo and Menéndez (2006), Felipe et al. (2007) (16) and (17) hypotheses were studied only for some specific models with multinomial populations. The last two paper are about 2 and k populations respectively, and the rest about a single population. These papers are specially based on the techniques of Barlow et al. (1972), Robertson et al. (1988). We think that it is important to mention the paper of Shapiro (1985), not only due to its contribution to the general theory of statistical inference with inequality constraints, but also because it was the first in using discrepancy measures as test-statistics inside such an area.

In the following subsections new test-statistics based on ϕ-divergence measures are proposed for testing (10)-(11)-(12) for single or multiple populations with very general distributions which satisfy the regularity conditions presented at the beginning of this paper.

4.1 Type O test-statistics based on ϕ-divergence measures

Definition 7 *Let* $F \subset E$*. The* ***type O test-statistics for the case of*** $g = 1$ ***population*** *based on* ϕ*-divergence measures for (10) are given by*

$$S_\phi^O(\widehat{\boldsymbol{\theta}}(F), \widehat{\boldsymbol{\theta}}(E)) = \frac{2n}{\phi''(1)} \mathrm{d}_\phi(f_{\widehat{\boldsymbol{\theta}}(F)}, f_{\widehat{\boldsymbol{\theta}}(E)}),$$

$$T_\phi^O(\widehat{\boldsymbol{\theta}}, \widehat{\boldsymbol{\theta}}(F), \widehat{\boldsymbol{\theta}}(E)) = \frac{2n}{\phi''(1)}\left(\mathrm{d}_\phi(f_{\widehat{\boldsymbol{\theta}}}, f_{\widehat{\boldsymbol{\theta}}(E)}) - \mathrm{d}_\phi(f_{\widehat{\boldsymbol{\theta}}}, f_{\widehat{\boldsymbol{\theta}}(F)})\right) = S_\phi^O(\widehat{\boldsymbol{\theta}}, \widehat{\boldsymbol{\theta}}(E)) - S_\phi^O(\widehat{\boldsymbol{\theta}}, \widehat{\boldsymbol{\theta}}(F)).$$

Theorem 8 *Under H^O_{Null}, the asymptotic distribution of $S^O_\phi(\widehat{\boldsymbol{\theta}}(F),\widehat{\boldsymbol{\theta}}(E))$ and $T^O_\phi(\widehat{\boldsymbol{\theta}},\widehat{\boldsymbol{\theta}}(F),\widehat{\boldsymbol{\theta}}(E))$ for the case of $g=1$, is χ^2_{df}, with $df=\mathrm{card}(E-F)$.*

Proof. See Section A.1. ∎

For $g>1$ populations with $n_1=...=n_g=\frac{n}{g}$ the definition is the same replacing n by n/g in the test-statistic, and in such a case Theorem 8 remains being true.

Definition 9 *Let $F\subset E$. The **type O test-statistics for the case of $g>1$ populations** based on ϕ-divergence measures for (10) are given by*

$$\widetilde{S}^O_\phi(\widehat{\boldsymbol{\theta}}(F),\widehat{\boldsymbol{\theta}}(E))=\frac{2}{\phi''(1)}\mathrm{d}_\phi(\mathcal{L}_{\widehat{\boldsymbol{\theta}}(F)},\mathcal{L}_{\widehat{\boldsymbol{\theta}}(E)}),$$
$$\widetilde{T}^O_\phi(\widehat{\boldsymbol{\theta}},\widehat{\boldsymbol{\theta}}(F),\widehat{\boldsymbol{\theta}}(E))=\frac{2}{\phi''(1)}\left(\mathrm{d}_\phi(\mathcal{L}_{\widehat{\boldsymbol{\theta}}},\mathcal{L}_{\widehat{\boldsymbol{\theta}}(E)})-\mathrm{d}_\phi(\mathcal{L}_{\widehat{\boldsymbol{\theta}}},\mathcal{L}_{\widehat{\boldsymbol{\theta}}(F)})\right)=\widetilde{S}^O_\phi(\widehat{\boldsymbol{\theta}},\widehat{\boldsymbol{\theta}}(E))-\widetilde{S}^O_\phi(\widehat{\boldsymbol{\theta}},\widehat{\boldsymbol{\theta}}(F)).$$

Theorem 10 *Under H^O_{Null}, the asymptotic distribution of $\widetilde{S}^O_\phi(\widehat{\boldsymbol{\theta}}(F),\widehat{\boldsymbol{\theta}}(E))$ and $\widetilde{T}^O_\phi(\widehat{\boldsymbol{\theta}},\widehat{\boldsymbol{\theta}}(F),\widehat{\boldsymbol{\theta}}(E))$ is χ^2_{df}, with $df=\mathrm{card}(E-F)$.*

Proof. The steps to be followed are very similar to the proof of Theorem 8 except for an important detail in the Fisher information matrix which should be clarified. From $\mathrm{d}_\phi(\boldsymbol{\theta})=\mathrm{d}_\phi(\mathcal{L}_{\boldsymbol{\theta}},\mathcal{L}_{\widehat{\boldsymbol{\theta}}(\bullet)})$ we obtain

$$\left.\frac{\partial^2}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^T}\mathrm{d}_\phi(\boldsymbol{\theta})\right|_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}(\bullet)}=\phi''(1)\,\mathcal{I}_F^{(n)}(\widehat{\boldsymbol{\theta}}(\bullet)),$$

where $\mathcal{I}_F^{(n)}(\widehat{\boldsymbol{\theta}}(\bullet))$ is based on "all" the observation. For instance, when $g=1$ when $\mathcal{I}_F^{(n)}(\widehat{\boldsymbol{\theta}}(\bullet))=n\mathcal{I}_F(\widehat{\boldsymbol{\theta}}(\bullet))$, which is the justification of not having "n" in the expression of the test-statistic. ∎

It is important to clarify that:

- When $F=\varnothing$, $S^O_\phi(\widehat{\boldsymbol{\theta}}(F),\widehat{\boldsymbol{\theta}}(E))=T^O_\phi(\widehat{\boldsymbol{\theta}},\widehat{\boldsymbol{\theta}}(F),\widehat{\boldsymbol{\theta}}(E))$ and $\widetilde{S}^O_\phi(\widehat{\boldsymbol{\theta}}(F),\widehat{\boldsymbol{\theta}}(E))=\widetilde{T}^O_\phi(\widehat{\boldsymbol{\theta}},\widehat{\boldsymbol{\theta}}(F),\widehat{\boldsymbol{\theta}}(E))$.
- The degrees of freedom of the asymptotic distribution of $S^O_\phi(\widehat{\boldsymbol{\theta}},\widehat{\boldsymbol{\theta}}(E))$ and $\widetilde{S}^O_\phi(\widehat{\boldsymbol{\theta}},\widehat{\boldsymbol{\theta}}(E))$, under the null hypothesis of (15) is $\mathrm{card}(E)$, where $E=R$, which match the general result of Theorem 8 because $F=\varnothing$.

4.2 Type A test-statistics based on ϕ-divergence measures

Definition 11 *Let $E=\{1,...,r\}$ and $F\subset E$. The **type A test-statistics for the case of $g=1$ population** based on ϕ-divergence measures for (11) are given by*

$$S^A_\phi(\widetilde{\boldsymbol{\theta}}(F),\widehat{\boldsymbol{\theta}}(E))=\frac{2n}{\phi''(1)}\mathrm{d}_\phi(f_{\widetilde{\boldsymbol{\theta}}(F)},f_{\widehat{\boldsymbol{\theta}}(E)}),$$
$$T^A_\phi(\widehat{\boldsymbol{\theta}},\widetilde{\boldsymbol{\theta}}(F),\widehat{\boldsymbol{\theta}}(E))=\frac{2n}{\phi''(1)}\left(\mathrm{d}_\phi(f_{\widehat{\boldsymbol{\theta}}},f_{\widehat{\boldsymbol{\theta}}(E)})-\mathrm{d}_\phi(f_{\widehat{\boldsymbol{\theta}}},f_{\widetilde{\boldsymbol{\theta}}(F)})\right)=S^O_\phi(\widehat{\boldsymbol{\theta}},\widehat{\boldsymbol{\theta}}(E))-S^B_\phi(\widehat{\boldsymbol{\theta}},\widetilde{\boldsymbol{\theta}}(F)).$$

For $g>1$ populations with $n_1=...=n_g=\frac{n}{g}$ the definition is the same replacing n by n/g in the test-statistic.

In what follows, matrix $\boldsymbol{H}(\boldsymbol{\theta}_0,\bullet)$ is a submatrix of $\boldsymbol{H}(\boldsymbol{\theta}_0)$ with row-indices in set $\bullet$.

Theorem 12 *Under H^A_{Null}, the asymptotic distribution of $S^A_\phi(\widetilde{\boldsymbol{\theta}}(F),\widehat{\boldsymbol{\theta}}(E))$ and $T^A_\phi(\widehat{\boldsymbol{\theta}},\widetilde{\boldsymbol{\theta}}(F),\widehat{\boldsymbol{\theta}}(E))$ for the case of $g=1$, is*

$$\lim_{n\to\infty}\Pr\left(S^A_\phi(\widetilde{\boldsymbol{\theta}}(F),\widehat{\boldsymbol{\theta}}(E))\leq x\right)=\lim_{n\to\infty}\Pr\left(T^A_\phi(\widehat{\boldsymbol{\theta}},\widetilde{\boldsymbol{\theta}}(F),\widehat{\boldsymbol{\theta}}(E))\leq x\right)=\sum_{j=0}^{r-\mathrm{card}(F)}w^A_j(\boldsymbol{\theta}_0)\Pr\left(\chi^2_{r-\mathrm{card}(F)-j}\leq x\right)$$

where

$$w_j^A(\boldsymbol{\theta}_0) = \sum_{S \in \mathcal{F}(E-F), \mathrm{card}(S)=j} \Pr\left(\mathbf{Y}_1(S) \geq \mathbf{0}_j\right) \Pr\left(\mathbf{Y}_2(S) \geq \mathbf{0}_{r-\mathrm{card}(F)-j}\right), \tag{37}$$

$\mathbf{Y}_1(S) \sim \mathcal{N}\left(\mathbf{0}_{\mathrm{card}(S)}, \boldsymbol{\Sigma}_1^A(\boldsymbol{\theta}_0, S)\right)$, $\mathbf{Y}_2(S) \sim \mathcal{N}\left(\mathbf{0}_{\mathrm{card}(E-F)-\mathrm{card}(S)}, \boldsymbol{\Sigma}_2^A(\boldsymbol{\theta}_0, S)\right)$, *with*

$$\boldsymbol{\Sigma}_1^A(\boldsymbol{\theta}_0, S) = \left(\boldsymbol{H}(\boldsymbol{\theta}_0, S)\mathcal{I}_F^{-1}(\boldsymbol{\theta}_0)\boldsymbol{H}^T(\boldsymbol{\theta}_0, S)\right)^{-1}, \tag{38}$$

$$\boldsymbol{\Sigma}_2^A(\boldsymbol{\theta}_0, S) = \boldsymbol{H}(\boldsymbol{\theta}_0, S^C)\mathcal{I}_F^{-1}(\boldsymbol{\theta}_0)\boldsymbol{H}^T(\boldsymbol{\theta}_0, S^C) - \boldsymbol{H}(\boldsymbol{\theta}_0, S^C)\mathcal{I}_F^{-1}(\boldsymbol{\theta}_0)\boldsymbol{H}^T(\boldsymbol{\theta}_0, S)\boldsymbol{\Sigma}_1^A(\boldsymbol{\theta}_0, S)\boldsymbol{H}(\boldsymbol{\theta}_0, S)\mathcal{I}_F^{-1}(\boldsymbol{\theta}_0)\boldsymbol{H}^T(\boldsymbol{\theta}_0, S^C), \tag{39}$$

and $S^C = E - F - S$.

Proof. See Section A.3. ■

It is important to clarify these points:
i) χ_0^2 is a degenerate random variable in the origin

$$\chi_0^2 \equiv 0, \text{ hence } \Pr\left(\chi_0^2 \leq x\right) = I(x \geq 0) = \begin{cases} 1, & x \geq 0 \\ 0, & x < 0 \end{cases};$$

ii) the weight of order 0 (in correspondence with $S = \varnothing$) is $w_0^A(\boldsymbol{\theta}_0) = \Pr\left(\mathbf{Y}_2(\varnothing) \geq \mathbf{0}_r\right)$, where $\mathbf{Y}_2(\varnothing) \sim \mathcal{N}\left(\mathbf{0}_r, \boldsymbol{\Sigma}_2^A(\boldsymbol{\theta}_0, \varnothing)\right)$, with

$$\boldsymbol{\Sigma}_2^A(\boldsymbol{\theta}_0, \varnothing) = \boldsymbol{H}(\boldsymbol{\theta}_0, E-F)\mathcal{I}_F^{-1}(\boldsymbol{\theta}_0)\boldsymbol{H}^T(\boldsymbol{\theta}_0, E-F);$$

iii) the weight of order $r - \mathrm{card}(F)$ (in correspondence with $S = E - F$) is $w_{r-\mathrm{card}(F)}(\boldsymbol{\theta}_0) = \Pr\left(\mathbf{Y}_1(E-F) \geq \mathbf{0}_{r-\mathrm{card}(F)}\right)$, where $\mathbf{Y}_1(E-F) \sim \mathcal{N}\left(\mathbf{0}_r, \boldsymbol{\Sigma}_1^A(\boldsymbol{\theta}_0, E-F)\right)$, with

$$\boldsymbol{\Sigma}_1^A(\boldsymbol{\theta}_0, E-F) = \left(\boldsymbol{H}(\boldsymbol{\theta}_0, E-F)\mathcal{I}_F^{-1}(\boldsymbol{\theta}_0)\boldsymbol{H}^T(\boldsymbol{\theta}_0, E-F)\right)^{-1}.$$

iv) if $\mathrm{card}(S) = j = 1$ then $\Pr\left(\mathbf{Y}_1(S) \geq 0\right) = \frac{1}{2}$ and if $\mathrm{card}(S) = j = r - \mathrm{card}(F) - 1$ then $\Pr\left(\mathbf{Y}_2(S) \geq 0\right) = \frac{1}{2}$.
v) for the normal orthant probabilities $\Pr\left(\mathbf{Y}_1(S) \geq \mathbf{0}_{\mathrm{card}(S)}\right)$, $\Pr\left(\mathbf{Y}_2(S) \geq \mathbf{0}_{r-\mathrm{card}(F)-\mathrm{card}(S)}\right)$, the multiplication of the variance-covariance matrix by a positive constant does not affect.
vi) In most of the cases the weights are unknown because depend on $\boldsymbol{\theta}_0$, however in practice very good approximations are usually obtained replacing $w_j^A(\boldsymbol{\theta}_0)$ by its consistent estimator $w_0^A(\widehat{\boldsymbol{\theta}}(E))$.

Corollary 13 *For $g > 1$, with $n_1 = ... = n_g = \frac{n}{g}$, Theorem 12 remains being true but the structure of the variance-covariance matrices are given by*

$$\boldsymbol{\Sigma}_1^A(S) = \left(\textstyle\sum_{i=1}^g \boldsymbol{H}_i(\boldsymbol{\theta}_0, S)\mathcal{I}_F^{-1}(\boldsymbol{\theta}_{i,0})\boldsymbol{H}_i^T(\boldsymbol{\theta}_0, S)\right)^{-1},$$
$$\boldsymbol{\Sigma}_2^A(S) = \textstyle\sum_{i=1}^g \sum_{i=j}^g \boldsymbol{H}_i(\boldsymbol{\theta}_0, S^C)\boldsymbol{P}_{ij}(\boldsymbol{\theta}_{\mathbf{0}}, S)\boldsymbol{H}_j^T(\boldsymbol{\theta}_0, S^C),$$

$$\boldsymbol{P}_{ij}(\boldsymbol{\theta}_{\mathbf{0}}, S) = (1-\delta_{ij})\mathcal{I}_F^{-1}(\boldsymbol{\theta}_{i,0}) - \mathcal{I}_F^{-1}(\boldsymbol{\theta}_{i,0})\boldsymbol{H}_i^T(\boldsymbol{\theta}_0, S)\boldsymbol{\Sigma}_1^A(S)\boldsymbol{H}_j(\boldsymbol{\theta}_0, S)\mathcal{I}_F^{-1}(\boldsymbol{\theta}_{j,0}),$$

and $S^C = E - F - S$.

Proof. From Sections 1 and 2, plugging in $n_i = \frac{n}{g}$, $\nu_i = 1$, we obtain

$$\boldsymbol{\Sigma}_2^A(S) = \boldsymbol{H}(\boldsymbol{\theta}_0, S^C)(\boldsymbol{P}_{ij}(\boldsymbol{\theta}_{\mathbf{0}}, S))_{i,j \in \{1,\ldots,g\}}\boldsymbol{H}^T(\boldsymbol{\theta}_0, S^C) = \left(\textstyle\sum_{i=1}^g \sum_{i=j}^g \boldsymbol{H}_i(\boldsymbol{\theta}_0, S^C)\boldsymbol{P}_{ij}(\boldsymbol{\theta}_{\mathbf{0}}, S)\boldsymbol{H}_j^T(\boldsymbol{\theta}_0, S^C)\right).$$

■

Definition 14 *Let $F \subset E$. The **type A test-statistics for the case of** $g > 1$ **populations** based on ϕ-divergence measures for (11) are given by*

$$\widetilde{S}_{\phi}^{A}(\widetilde{\boldsymbol{\theta}}(F),\widehat{\boldsymbol{\theta}}(E)) = \frac{2}{\phi''(1)}\mathrm{d}_{\phi}(\mathcal{L}_{\widetilde{\boldsymbol{\theta}}(F)},\mathcal{L}_{\widehat{\boldsymbol{\theta}}(E)}),$$
$$\widetilde{T}_{\phi}^{A}(\widehat{\boldsymbol{\theta}},\widetilde{\boldsymbol{\theta}}(F),\widehat{\boldsymbol{\theta}}(E)) = \frac{2}{\phi''(1)}\left(\mathrm{d}_{\phi}(\mathcal{L}_{\widehat{\boldsymbol{\theta}}},\mathcal{L}_{\widehat{\boldsymbol{\theta}}(E)}) - \mathrm{d}_{\phi}(\mathcal{L}_{\widehat{\boldsymbol{\theta}}},\mathcal{L}_{\widetilde{\boldsymbol{\theta}}(F)})\right) = \widetilde{S}_{\phi}^{O}(\widehat{\boldsymbol{\theta}},\widehat{\boldsymbol{\theta}}(E)) - \widetilde{S}_{\phi}^{B}(\widehat{\boldsymbol{\theta}},\widetilde{\boldsymbol{\theta}}(F)).$$

Theorem 15 *Under H_{Null}^{A}, the asymptotic distribution of $\widetilde{S}_{\phi}^{A}(\widetilde{\boldsymbol{\theta}}(F),\widehat{\boldsymbol{\theta}}(E))$ and $\widetilde{T}_{\phi}^{A}(\widehat{\boldsymbol{\theta}},\widetilde{\boldsymbol{\theta}}(F),\widehat{\boldsymbol{\theta}}(E))$ is*

$$\lim_{n\to\infty}\Pr\left(\widetilde{S}_{\phi}^{A}(\widetilde{\boldsymbol{\theta}}(F),\widehat{\boldsymbol{\theta}}(E)) \le x\right) = \lim_{n\to\infty}\Pr\left(\widetilde{T}_{\phi}^{A}(\widehat{\boldsymbol{\theta}},\widetilde{\boldsymbol{\theta}}(F),\widehat{\boldsymbol{\theta}}(E)) \le x\right) = \sum_{j=0}^{r-\mathrm{card}(F)} w_j^{A}(\boldsymbol{\theta}_0)\Pr\left(\chi^2_{r-\mathrm{card}(F)-j} \le x\right)$$

where

$$w_j^{A}(\boldsymbol{\theta}_0) = \sum_{S\in\mathcal{F}(E-F),\mathrm{card}(S)=j}\Pr\left(\widetilde{\boldsymbol{Y}}_1(S) \ge \boldsymbol{0}_j\right)\Pr\left(\widetilde{\boldsymbol{Y}}_2(S) \ge \boldsymbol{0}_{r-\mathrm{card}(F)-j}\right),$$

with $\widetilde{\boldsymbol{Y}}_1(S) \sim \mathcal{N}\left(\boldsymbol{0}_j, \widetilde{\boldsymbol{\Sigma}}_1^{A}(S)\right)$, $\widetilde{\boldsymbol{Y}}_2(S) \sim \mathcal{N}\left(\boldsymbol{0}_{r-\mathrm{card}(F)-j}, \widetilde{\boldsymbol{\Sigma}}_2^{A}(S)\right)$,

$$\widetilde{\boldsymbol{\Sigma}}_1^{A}(S) = \left(\textstyle\sum_{i=1}^{g}\boldsymbol{H}_i(\boldsymbol{\theta}_0,S)\frac{1}{\nu_i}\mathcal{I}_F^{-1}(\boldsymbol{\theta}_{i,0})\boldsymbol{H}_i^T(\boldsymbol{\theta}_0,S)\right)^{-1},$$
$$\widetilde{\boldsymbol{\Sigma}}_2^{A}(S) = \textstyle\sum_{i=1}^{g}\sum_{i=j}^{g}\boldsymbol{H}_i(\boldsymbol{\theta}_0,S^C)\boldsymbol{P}_{ij}(\boldsymbol{\theta}_{\boldsymbol{0}},S)\boldsymbol{H}_j^T(\boldsymbol{\theta}_0,S^C),$$
$$\boldsymbol{P}_{ij}(\boldsymbol{\theta}_{\boldsymbol{0}},S) = (1-\delta_{ij})\frac{1}{\nu_i}\mathcal{I}_F^{-1}(\boldsymbol{\theta}_{i,0}) - \frac{1}{\nu_i}\mathcal{I}_F^{-1}(\boldsymbol{\theta}_{i,0})\boldsymbol{H}_i^T(\boldsymbol{\theta}_0,S)\widetilde{\boldsymbol{\Sigma}}_1^{A}(S)\boldsymbol{H}_j(\boldsymbol{\theta}_0,S)\frac{1}{\nu_j}\mathcal{I}_F^{-1}(\boldsymbol{\theta}_{j,0}),$$

and $S^C = E - F - S$ and δ_{ij} is the Kronecker delta function, that is, its value is 1 if $i = j$ and 0 otherwise.

4.3 Type B test-statistics based on ϕ-divergence measures

Definition 16 *Let $F \subset E \subset R = \{1,...,r\}$, such that $\mathrm{card}(E) < r$. The **type B test-statistics for the case of** $g = 1$ **population** based on ϕ-divergence measures for (12) are given by*

$$S_{\phi}^{B}(\widehat{\boldsymbol{\theta}}(F),\widetilde{\boldsymbol{\theta}}(E)) = \frac{2n}{\phi''(1)}\mathrm{d}_{\phi}(f_{\widehat{\boldsymbol{\theta}}(F)},f_{\widetilde{\boldsymbol{\theta}}(E)}),$$
$$T_{\phi}^{B}(\widehat{\boldsymbol{\theta}},\widehat{\boldsymbol{\theta}}(F),\widetilde{\boldsymbol{\theta}}(E)) = \frac{2n}{\phi''(1)}\left(\mathrm{d}_{\phi}(f_{\widehat{\boldsymbol{\theta}}},f_{\widetilde{\boldsymbol{\theta}}(E)}) - \mathrm{d}_{\phi}(f_{\widehat{\boldsymbol{\theta}}},f_{\widehat{\boldsymbol{\theta}}(F)})\right) = S_{\phi}^{B}(\widehat{\boldsymbol{\theta}},\widetilde{\boldsymbol{\theta}}(E)) - S_{\phi}^{O}(\widehat{\boldsymbol{\theta}},\widehat{\boldsymbol{\theta}}(F)),$$

For $g > 1$ populations with $n_1 = ... = n_g = \frac{n}{g}$ the definition is the same replacing n by n/g in the test-statistic. Let $S^*(\boldsymbol{\theta}_0,E)$ the unknown set of indices in $R - E$ representing the positions where the true value of $\boldsymbol{\theta}_0$ equals zero.

Theorem 17 *Under H_{Null}^{B}, the asymptotic distribution of $S_{\phi}^{B}(\widehat{\boldsymbol{\theta}}(F),\widetilde{\boldsymbol{\theta}}(E))$ and $T_{\phi}^{B}(\widehat{\boldsymbol{\theta}},\widehat{\boldsymbol{\theta}}(F),\widetilde{\boldsymbol{\theta}}(E))$ for the case of $g = 1$, is*

$$\lim_{n\to\infty}\Pr\left(S_{\phi}^{B}(\widehat{\boldsymbol{\theta}}(F),\widetilde{\boldsymbol{\theta}}(E)) \le x\right) = \lim_{n\to\infty}\Pr\left(T_{\phi}^{B}(\widehat{\boldsymbol{\theta}},\widehat{\boldsymbol{\theta}}(F),\widetilde{\boldsymbol{\theta}}(E)) \le x\right) = \sum_{j=0}^{\eta(\boldsymbol{\theta}_0,E)} w_j^{B}(\boldsymbol{\theta}_0)\Pr\left(\chi^2_{j+\mathrm{card}(E)-\mathrm{card}(F)} \le x\right),$$

where $\eta(\boldsymbol{\theta}_0,E) = \mathrm{card}(S^*(\boldsymbol{\theta}_0,E))$, and

$$w_j^{B}(\boldsymbol{\theta}_0) = \sum_{S\in\mathcal{F}(S^*(\boldsymbol{\theta}_0,E)),\mathrm{card}(S)=j}\Pr\left(\boldsymbol{W}_1(S) \ge \boldsymbol{0}_j\right)\Pr\left(\boldsymbol{W}_2(S) \ge \boldsymbol{0}_{\eta(\boldsymbol{\theta}_0,E)-j}\right), \quad (40)$$

with $\boldsymbol{W}_1(S) \sim \mathcal{N}\left(\mathbf{0}_j, \boldsymbol{\Sigma}_1^B(\boldsymbol{\theta}_0, S)\right)$, $\boldsymbol{W}_2(S) \sim \mathcal{N}\left(\mathbf{0}_{\eta(\boldsymbol{\theta}_0,E)-j}, \boldsymbol{\Sigma}_2^B(\boldsymbol{\theta}_0, S)\right)$,

$$\boldsymbol{\Sigma}_1^B(\boldsymbol{\theta}_0, S) = \left(\boldsymbol{H}(\boldsymbol{\theta}_0, S)\mathcal{I}_F^{-1}(\boldsymbol{\theta}_0)\boldsymbol{H}^T(\boldsymbol{\theta}_0, S)\right)^{-1}, \tag{41}$$

$$\begin{aligned}\boldsymbol{\Sigma}_2^B(\boldsymbol{\theta}_0, S) &= \boldsymbol{H}(\boldsymbol{\theta}_0, S^*(\boldsymbol{\theta}_0, E) - S)\mathcal{I}_F^{-1}(\boldsymbol{\theta}_0)\boldsymbol{H}^T(\boldsymbol{\theta}_0, S^*(\boldsymbol{\theta}_0, E) - S) \\ &\quad - \boldsymbol{H}(\boldsymbol{\theta}_0, S^*(\boldsymbol{\theta}_0, E) - S)\mathcal{I}_F^{-1}(\boldsymbol{\theta}_0)\boldsymbol{H}^T(\boldsymbol{\theta}_0, S)\boldsymbol{\Sigma}_1^B(\boldsymbol{\theta}_0, S)\boldsymbol{H}(\boldsymbol{\theta}_0, S)\mathcal{I}_F^{-1}(\boldsymbol{\theta}_0)\boldsymbol{H}^T(\boldsymbol{\theta}_0, S^*(\boldsymbol{\theta}_0, E) - S).\end{aligned} \tag{42}$$

Proof. See Section A.4. ■

Similar clarifications to those given in page 11 for the weights of the type A tests can be also valid for the weights of type B tests.

Let $s_\phi^B(\widehat{\boldsymbol{\theta}}(F), \widetilde{\boldsymbol{\theta}}(E))$ and $t_\phi^B(\widehat{\boldsymbol{\theta}}, \widehat{\boldsymbol{\theta}}(F), \widetilde{\boldsymbol{\theta}}(E))$ be the observed values of $S_\phi^B(\widehat{\boldsymbol{\theta}}(F), \widetilde{\boldsymbol{\theta}}(E))$ and $T_\phi^B(\widehat{\boldsymbol{\theta}}, \widehat{\boldsymbol{\theta}}(F), \widetilde{\boldsymbol{\theta}}(E))$ based on a sample. It is not correct to consider that the p-values of (12) with the proposed test-statistics are respectively $\lim_{n\to\infty} \Pr(S_\phi^B(\widehat{\boldsymbol{\theta}}(F), \widetilde{\boldsymbol{\theta}}(E)) > s_\phi^B(\widehat{\boldsymbol{\theta}}(F), \widetilde{\boldsymbol{\theta}}(E)))$ and $\lim_{n\to\infty} \Pr(T_\phi^B(\widehat{\boldsymbol{\theta}}, \widehat{\boldsymbol{\theta}}(F), \widetilde{\boldsymbol{\theta}}(E)) > t_\phi^B(\widehat{\boldsymbol{\theta}}, \widehat{\boldsymbol{\theta}}(F), \widetilde{\boldsymbol{\theta}}(E)))$. Actually both probabilities depend on the true value of $\boldsymbol{\theta}$, $\boldsymbol{\theta}_0$, which belongs to $\Omega(E) = \{\boldsymbol{\theta} \in \Theta : h_i(\boldsymbol{\theta}) = 0, i \in E; h_i(\boldsymbol{\theta}) \leq 0, i \in R - E\}$. That is, under H_{Null}^B, $\lim_{n\to\infty} \Pr(S_\phi^B(\widehat{\boldsymbol{\theta}}(F), \widetilde{\boldsymbol{\theta}}(E)) > s_\phi^B(\widehat{\boldsymbol{\theta}}(F), \widetilde{\boldsymbol{\theta}}(E)))$ and $\lim_{n\to\infty} \Pr(T_\phi^B(\widehat{\boldsymbol{\theta}}, \widehat{\boldsymbol{\theta}}(F), \widetilde{\boldsymbol{\theta}}(E)) > t_\phi^B(\widehat{\boldsymbol{\theta}}, \widehat{\boldsymbol{\theta}}(F), \widetilde{\boldsymbol{\theta}}(E)))$ are not specific numbers because they depend on $\boldsymbol{\theta}_0$ and do not define p-values. The p-value must be the probability of rejecting a value as extreme or more than the value of the test-statistic obtained with the sample and with the "least favorable" value of the parameter that belongs to $\Omega(E)$. This means that if H_{Null}^B is rejected for the least favorable value of the parameter, then it is rejected for every value of the parameter. Hence,

$$\begin{aligned}\text{p-value}(S_\phi^B(\widehat{\boldsymbol{\theta}}(F), \widetilde{\boldsymbol{\theta}}(E))) &= \lim_{n\to\infty} \sup_{\boldsymbol{\theta}\in\Omega(E)} \Pr(S_\phi^B(\widehat{\boldsymbol{\theta}}(F), \widetilde{\boldsymbol{\theta}}(E)) > s_\phi^B(\widehat{\boldsymbol{\theta}}(F), \widetilde{\boldsymbol{\theta}}(E))), \\ \text{p-value}(T_\phi^B(\widehat{\boldsymbol{\theta}}, \widehat{\boldsymbol{\theta}}(F), \widetilde{\boldsymbol{\theta}}(E)) &= \lim_{n\to\infty} \sup_{\boldsymbol{\theta}\in\Omega(E)} \Pr(T_\phi^B(\widehat{\boldsymbol{\theta}}, \widehat{\boldsymbol{\theta}}(F), \widetilde{\boldsymbol{\theta}}(E)) > t_\phi^B(\widehat{\boldsymbol{\theta}}, \widehat{\boldsymbol{\theta}}(F), \widetilde{\boldsymbol{\theta}}(E))).\end{aligned}$$

From Theorem 3.8.1 in Silvapulle and Sen (2004) the supremum is reached at $\boldsymbol{\theta} = \mathbf{0}_k$, and this justifies the following result.

Proposition 18 *The p-values for the tests of Definition 16, are*

$$\begin{aligned}\text{p-value}(S_\phi^B(\widehat{\boldsymbol{\theta}}(F), \widetilde{\boldsymbol{\theta}}(E))) &= \sum_{j=0}^{r-\text{card}(E)} \bar{w}_j(\boldsymbol{\theta}_0) \Pr\left(\chi^2_{j+\text{card}(E)-\text{card}(F)} > s_\phi^B(\widehat{\boldsymbol{\theta}}(F), \widetilde{\boldsymbol{\theta}}(E))\right), \\ \text{p-value}(T_\phi^B(\widehat{\boldsymbol{\theta}}, \widehat{\boldsymbol{\theta}}(F), \widetilde{\boldsymbol{\theta}}(E)) &= \sum_{j=0}^{r-\text{card}(E)} \bar{w}_j(\boldsymbol{\theta}_0) \Pr\left(\chi^2_{j+\text{card}(E)-\text{card}(F)} > t_\phi^B(\widehat{\boldsymbol{\theta}}, \widehat{\boldsymbol{\theta}}(F), \widetilde{\boldsymbol{\theta}}(E))\right),\end{aligned}$$

where

$$\bar{w}_j(\boldsymbol{\theta}_0) = \sum_{S\in\mathcal{F}(R-E),\text{card}(S)=j} \Pr\left(\bar{\boldsymbol{W}}_1(S) \geq \mathbf{0}_j\right) \Pr\left(\bar{\boldsymbol{W}}_2(S) \geq \mathbf{0}_{r-\text{card}(E)-j}\right),$$

with $\bar{\boldsymbol{W}}_1(S) \sim \mathcal{N}\left(\mathbf{0}_{\text{card}(S)}, \bar{\boldsymbol{\Sigma}}_1^B(\boldsymbol{\theta}_0, S)\right)$, $\bar{\boldsymbol{W}}_2(S) \sim \mathcal{N}\left(\mathbf{0}_{r-\text{card}(E)-\text{card}(S)}, \bar{\boldsymbol{\Sigma}}_2^B(\boldsymbol{\theta}_0, S)\right)$

$$\bar{\boldsymbol{\Sigma}}_1^B(\boldsymbol{\theta}_0, S) = \left(\boldsymbol{H}^T(\boldsymbol{\theta}_0, S)\mathcal{I}_F^{-1}(\boldsymbol{\theta}_0)\boldsymbol{H}(\boldsymbol{\theta}_0, S)\right)^{-1},$$

$\bar{\boldsymbol{\Sigma}}_2^B(\boldsymbol{\theta}_0, S) = \boldsymbol{H}(\boldsymbol{\theta}_0, S^C)\mathcal{I}_F^{-1}(\boldsymbol{\theta}_0)\boldsymbol{H}^T(\boldsymbol{\theta}_0, S^C) - \boldsymbol{H}(\boldsymbol{\theta}_0, S^C)\mathcal{I}_F^{-1}(\boldsymbol{\theta}_0)\boldsymbol{H}^T(\boldsymbol{\theta}_0, S)\bar{\boldsymbol{\Sigma}}_1^B(\boldsymbol{\theta}_0, S)\boldsymbol{H}(\boldsymbol{\theta}_0, S)\mathcal{I}_F^{-1}(\boldsymbol{\theta}_0)\boldsymbol{H}^T(\boldsymbol{\theta}_0, S^C)$,
and $S^C = R - E - S$.

Corollary 19 *For $g>1$, with $n_1=...=n_g=\frac{n}{g}$, Theorem 17 remains being true but the structure of the variance-covariance matrices are given by*

$$\begin{aligned}\boldsymbol{\Sigma}_1^B(\boldsymbol{\theta}_0,S) &= \left(\textstyle\sum_{i=1}^{g}\boldsymbol{H}_i(\boldsymbol{\theta}_0,S)\mathcal{I}_F^{-1}(\boldsymbol{\theta}_{i,0})\boldsymbol{H}_i^T(\boldsymbol{\theta}_0,S)\right)^{-1},\\ \boldsymbol{\Sigma}_2^B(\boldsymbol{\theta}_0,S) &= \textstyle\sum_{i=1}^{g}\sum_{i=j}^{g}\boldsymbol{H}_i(\boldsymbol{\theta}_0,S^*(\boldsymbol{\theta}_0,E)-S)\boldsymbol{P}_{ij}(\boldsymbol{\theta}_{\mathbf{0}},S)\boldsymbol{H}_j^T(\boldsymbol{\theta}_0,S^*(\boldsymbol{\theta}_0,E)-S),\end{aligned}$$

$$\begin{aligned}\boldsymbol{P}_{ij}(\boldsymbol{\theta}_{\mathbf{0}},S) =&(1-\delta_{ij})\mathcal{I}_F^{-1}(\boldsymbol{\theta}_{i,0})\\ &-\mathcal{I}_F^{-1}(\boldsymbol{\theta}_{i,0})\boldsymbol{H}_i^T(\boldsymbol{\theta}_0,S)\left(\textstyle\sum_{h=1}^{g}\boldsymbol{H}_h(\boldsymbol{\theta}_0,S)\mathcal{I}_F^{-1}(\boldsymbol{\theta}_{h,0})\boldsymbol{H}_h^T(\boldsymbol{\theta}_0,S)\right)^{-1}\boldsymbol{H}_j(\boldsymbol{\theta}_0,S)\mathcal{I}_F^{-1}(\boldsymbol{\theta}_{j,0}).\end{aligned}$$

Definition 20 *Let $F\subset E\subset R=\{1,...,r\}$, such that $\mathrm{card}(E)<r$. The* ***type B test-statistics for the case of*** $g>1$ ***populations*** *based on ϕ-divergence measures for (12) are given by*

$$\begin{aligned}\widetilde{S}_{\phi}^B(\widehat{\boldsymbol{\theta}}(F),\widetilde{\boldsymbol{\theta}}(E)) &= \frac{2}{\phi''(1)}\mathrm{d}_{\phi}(\mathcal{L}_{\widehat{\boldsymbol{\theta}}(F)},\mathcal{L}_{\widetilde{\boldsymbol{\theta}}(E)}),\\ \widetilde{T}_{\phi}^B(\widehat{\boldsymbol{\theta}},\widehat{\boldsymbol{\theta}}(F),\widetilde{\boldsymbol{\theta}}(E)) &= \frac{2}{\phi''(1)}\left(\mathrm{d}_{\phi}(\mathcal{L}_{\widehat{\boldsymbol{\theta}}},\mathcal{L}_{\widetilde{\boldsymbol{\theta}}(E)})-\mathrm{d}_{\phi}(\mathcal{L}_{\widehat{\boldsymbol{\theta}}},\mathcal{L}_{\widehat{\boldsymbol{\theta}}(F)})\right)=\widetilde{S}_{\phi}^B(\widehat{\boldsymbol{\theta}},\widetilde{\boldsymbol{\theta}}(E))-\widetilde{S}_{\phi}^O(\widehat{\boldsymbol{\theta}},\widehat{\boldsymbol{\theta}}(F)).\end{aligned}$$

Theorem 21 *Under H_{Null}^B, the asymptotic distribution of $\widetilde{S}_{\phi}^B(\widehat{\boldsymbol{\theta}}(F),\widetilde{\boldsymbol{\theta}}(E))$ and $\widetilde{T}_{\phi}^B(\widehat{\boldsymbol{\theta}},\widehat{\boldsymbol{\theta}}(F),\widetilde{\boldsymbol{\theta}}(E))$ is*

$$\lim_{n\to\infty}\Pr\left(\widetilde{S}_{\phi}^B(\widehat{\boldsymbol{\theta}}(F),\widetilde{\boldsymbol{\theta}}(E))\le x\right)=\lim_{n\to\infty}\Pr\left(\widetilde{T}_{\phi}^B(\widehat{\boldsymbol{\theta}},\widehat{\boldsymbol{\theta}}(F),\widetilde{\boldsymbol{\theta}}(E))\le x\right)=\sum_{j=0}^{\eta(\boldsymbol{\theta}_0,E)}w_j^B(\boldsymbol{\theta}_0)\Pr\left(\chi^2_{j+\mathrm{card}(E)-\mathrm{card}(F)}\le x\right),$$

where

$$w_j^B(\boldsymbol{\theta}_0)=\sum_{S\in\mathcal{F}(S^*(\boldsymbol{\theta}_0,E)),\mathrm{card}(S)=j}\Pr\left(\widetilde{\boldsymbol{W}}_1(S)\ge\mathbf{0}_{\mathrm{card}(S)}\right)\Pr\left(\widetilde{\boldsymbol{W}}_2(S)\ge\mathbf{0}_{\mathrm{card}(S^*(\boldsymbol{\theta}_0,E)-S)}\right),$$

with $\widetilde{\boldsymbol{W}}_1(S)\sim\mathcal{N}\left(\mathbf{0}_{\mathrm{card}(S)},\widetilde{\boldsymbol{\Sigma}}_1^B(\boldsymbol{\theta}_0,S)\right)$, $\widetilde{\boldsymbol{W}}_2(S)\sim\mathcal{N}\left(\mathbf{0}_{\eta(\boldsymbol{\theta}_0,E)-\mathrm{card}(S)},\widetilde{\boldsymbol{\Sigma}}_2^B(\boldsymbol{\theta}_0,S)\right)$,

$$\begin{aligned}\widetilde{\boldsymbol{\Sigma}}_1^B(\boldsymbol{\theta}_0,S) &= \left(\textstyle\sum_{i=1}^{g}\boldsymbol{H}_i(\boldsymbol{\theta}_0,S)\frac{1}{\nu_i}\mathcal{I}_F^{-1}(\boldsymbol{\theta}_{i,0})\boldsymbol{H}_i^T(\boldsymbol{\theta}_0,S)\right)^{-1},\\ \widetilde{\boldsymbol{\Sigma}}_2^B(\boldsymbol{\theta}_0,S) &= \textstyle\sum_{i=1}^{g}\sum_{i=j}^{g}\boldsymbol{H}_i(\boldsymbol{\theta}_0,S^*(\boldsymbol{\theta}_0,E)-S)\boldsymbol{P}_{ij}(\boldsymbol{\theta}_{\mathbf{0}},S)\boldsymbol{H}_j^T(\boldsymbol{\theta}_0,S^*(\boldsymbol{\theta}_0,E)-S),\end{aligned}$$

$$\begin{aligned}\boldsymbol{P}_{ij}(\boldsymbol{\theta}_{\mathbf{0}},S) =&(1-\delta_{ij})\frac{1}{\nu_i}\mathcal{I}_F^{-1}(\boldsymbol{\theta}_{i,0})\\ &-\frac{1}{\nu_i}\mathcal{I}_F^{-1}(\boldsymbol{\theta}_{i,0})\boldsymbol{H}_i^T(\boldsymbol{\theta}_0,S)\left(\textstyle\sum_{h=1}^{g}\boldsymbol{H}_h(\boldsymbol{\theta}_0,S)\frac{1}{\nu_h}\mathcal{I}_F^{-1}(\boldsymbol{\theta}_{h,0})\boldsymbol{H}_h^T(\boldsymbol{\theta}_0,S)\right)^{-1}\boldsymbol{H}_j(\boldsymbol{\theta}_0,S)\frac{1}{\nu_j}\mathcal{I}_F^{-1}(\boldsymbol{\theta}_{j,0}).\end{aligned}$$

A similar result to one given in Proposition 11 for the p-values of the type B tests with a single population can be also valid for multiple populations.

5 Real data example: Divergence based one-sided testing for the mean of two populations with Poisson distribution

In Simpson (1989) we can find an application example with two Poisson populations, different sample sizes, and

$$H_{Null}^{B}:\ \theta_1 \geq \theta_2 \quad \text{vs.} \quad H_{Alt}^{B}:\ \theta_1 < \theta_2. \tag{43}$$

This biological experiment consisted in exposing treated male flies to a specific degree of chemical and to compare their behavior with a control group. It was counted the number of recessive lethal mutations among the daughters of the explored flies (see Table 1) and this number is assumed to be $X_{ij} \overset{ind}{\sim} \mathcal{P}(\theta_i)$, $\theta_i > 0$, $i = 1, 2$. For g populations, (73) given in Section A.5, is equal to

$$\mathrm{d}_{\phi_\lambda}^{*}(f_\theta, f_{\theta_0}) = \frac{\exp\{-(\lambda+1)\theta\}}{\exp\{-\lambda\theta_0\}} \sum_{i=0}^{\infty} \frac{(\exp\{((\lambda+1)\log\theta - \lambda\log\theta_0)\})^i}{i!} = -(\lambda+1)\theta + \lambda\theta_0 + \frac{\theta^{\lambda+1}}{\theta_0^{\lambda}},$$

and hence for $\lambda \notin \{-1, 0\}$, (69)-(72) given in Section A.5, are equal to

$$\widetilde{S}_{\phi_\lambda}^{B}(\widehat{\boldsymbol{\theta}}(F), \widetilde{\boldsymbol{\theta}}(E)) = \frac{2}{\lambda(\lambda+1)} \left(\exp\left\{ \sum_{i=1}^{g} n_i \left(-(\lambda+1)\widehat{\theta}_i(F) + \lambda\widetilde{\theta}_i(E) + \frac{\widehat{\theta}_i^{\lambda+1}(F)}{\widetilde{\theta}_i^{\lambda}(E)} \right) \right\} - 1 \right) \tag{44}$$

$$\begin{aligned} \widetilde{T}_{\phi_\lambda}^{B}(\widehat{\boldsymbol{\theta}}, \widehat{\boldsymbol{\theta}}(F), \widetilde{\boldsymbol{\theta}}(E)) = \frac{2}{\lambda(\lambda+1)} & \left(\exp\left\{ \sum_{i=1}^{g} n_i \left(-(\lambda+1)\widehat{\theta}_i + \lambda\widetilde{\theta}_i(E) + \frac{\widehat{\theta}_i^{\lambda+1}}{\widetilde{\theta}_i^{\lambda}(E)} \right) \right\} \right. \\ & \left. - \exp\left\{ \sum_{i=1}^{g} n_i \left(-(\lambda+1)\widehat{\theta}_i + \lambda\widehat{\theta}_i(F) + \frac{\widehat{\theta}_i^{\lambda+1}}{\widehat{\theta}_i^{\lambda}(F)} \right) \right\} \right), \end{aligned} \tag{45}$$

where $\widehat{\boldsymbol{\theta}} = (\widehat{\theta}_1, ..., \widehat{\theta}_g)^T$, $\widehat{\theta}_i = \frac{1}{n_i}\sum_{j=1}^{n_i} X_{ij}$. Note that the Kullback based test-statistics ($\lambda = 0$) are

$$\begin{aligned} \widetilde{S}_{\mathrm{Kull}}^{B}(\widehat{\boldsymbol{\theta}}(F), \widetilde{\boldsymbol{\theta}}(E)) &= \lim_{\lambda\to 0} \widetilde{S}_{\phi_\lambda}^{B}(\widehat{\boldsymbol{\theta}}(F), \widetilde{\boldsymbol{\theta}}(E)) = 2\sum_{i=1}^{g} n_i \left(-\widehat{\theta}_i(F) + \widetilde{\theta}_i(E) + \widehat{\theta}_i(F) \log \frac{\widehat{\theta}_i(F)}{\widetilde{\theta}_i(E)} \right), \\ \widetilde{T}_{\mathrm{Kull}}^{B}(\widehat{\boldsymbol{\theta}}, \widehat{\boldsymbol{\theta}}(F), \widetilde{\boldsymbol{\theta}}(E)) &= \lim_{\lambda\to 0} \widetilde{T}_{\phi_\lambda}^{B}(\widehat{\boldsymbol{\theta}}, \widehat{\boldsymbol{\theta}}(F), \widetilde{\boldsymbol{\theta}}(E)) = 2\sum_{i=1}^{g} n_i \left(-\widehat{\theta}_i(F) + \widetilde{\theta}_i(E) + \widehat{\theta}_i \log \frac{\widehat{\theta}_i(F)}{\widetilde{\theta}_i(E)} \right), \end{aligned} \tag{46}$$

but $\widetilde{T}_{\mathrm{Kull}}^{B}(\widehat{\boldsymbol{\theta}}, \widehat{\boldsymbol{\theta}}(F), \widetilde{\boldsymbol{\theta}}(E))$ is the only test-statistic which is always equal to the likelihood ratio test-statistic $T^B(\widehat{\boldsymbol{\theta}}(F), \widetilde{\boldsymbol{\theta}}(E))$, even though we have to consider that $\widetilde{S}_{\mathrm{Kull}}^{B}(\widehat{\boldsymbol{\theta}}, \widetilde{\boldsymbol{\theta}}(E)) = \widetilde{T}_{\mathrm{Kull}}^{B}(\widehat{\boldsymbol{\theta}}, \widehat{\boldsymbol{\theta}}, \widetilde{\boldsymbol{\theta}}(E))$, for the case $F = \varnothing$. Furthermore, the Pearson divergence based test-statistics ($\lambda = 1$) are

$$\begin{aligned} \widetilde{S}_{\phi_1}^{B}(\widehat{\boldsymbol{\theta}}(F), \widetilde{\boldsymbol{\theta}}(E)) &= \exp\left\{ \sum_{i=1}^{g} n_i \frac{\left(\widehat{\theta}_i(F) - \widetilde{\theta}_i(E)\right)^2}{\widetilde{\theta}_i(E)} \right\} - 1, \\ \widetilde{T}_{\phi_\lambda}^{B}(\widehat{\boldsymbol{\theta}}, \widehat{\boldsymbol{\theta}}(F), \widetilde{\boldsymbol{\theta}}(E)) &= \exp\left\{ \sum_{i=1}^{g} n_i \frac{\left(\widehat{\theta}_i - \widetilde{\theta}_i(E)\right)^2}{\widetilde{\theta}_i(E)} \right\} - \exp\left\{ \sum_{i=1}^{g} n_i \frac{\left(\widehat{\theta}_i - \widehat{\theta}_i(F)\right)^2}{\widehat{\theta}_i(F)} \right\}, \end{aligned}$$

but $\widetilde{S}_{\phi_1}^{B}(\widehat{\boldsymbol{\theta}}(F), \widetilde{\boldsymbol{\theta}}(E))$ is the only test-statistic which is always equal to the chi-square test-statistic $C^B(\widehat{\boldsymbol{\theta}}(F), \widetilde{\boldsymbol{\theta}}(E))$, even though we have to consider that $\widetilde{S}_{\phi_1}^{B}(\widehat{\boldsymbol{\theta}}, \widetilde{\boldsymbol{\theta}}(E)) = \widetilde{T}_{\phi_\lambda}^{B}(\widehat{\boldsymbol{\theta}}, \widehat{\boldsymbol{\theta}}, \widetilde{\boldsymbol{\theta}}(E))$, for the case $F = \varnothing$. Taking into account that $\kappa \in \{\widetilde{S}_{\phi_\lambda}^{B}(\widehat{\boldsymbol{\theta}}(F), \widetilde{\boldsymbol{\theta}}(E)), \widetilde{T}_{\phi_\lambda}^{B}(\widehat{\boldsymbol{\theta}}, \widehat{\boldsymbol{\theta}}(F), \widetilde{\boldsymbol{\theta}}(E))\}$ under H_{Null}^{B} must be small as n_i increases, it holds

$\log(\kappa+1)\simeq\kappa$, hence we can propose a modification of (44) and (45)

$$\widetilde{\widetilde{S}}_{\phi_\lambda}^{B}(\widehat{\boldsymbol{\theta}}(F),\widetilde{\boldsymbol{\theta}}(E))=\frac{2}{\lambda(\lambda+1)}\sum_{i=1}^{g}n_i\left(-(\lambda+1)\widehat{\theta}_i(F)+\lambda\widetilde{\theta}_i(E)+\frac{\widehat{\theta}_i^{\lambda+1}(F)}{\widetilde{\theta}_i^{\lambda}(E)}\right),\tag{47}$$

$$\widetilde{\widetilde{T}}_{\phi_\lambda}^{B}(\widehat{\boldsymbol{\theta}},\widehat{\boldsymbol{\theta}}(F),\widetilde{\boldsymbol{\theta}}(E))=\frac{2}{\lambda(\lambda+1)}\sum_{i=1}^{g}n_i\left(\lambda\left(\widetilde{\theta}_i(E)-\widehat{\theta}_i(F)\right)+\widehat{\theta}_i^{\lambda+1}\left(\frac{1}{\widetilde{\theta}_i^{\lambda}(E)}-\frac{1}{\widehat{\theta}_i^{\lambda}(F)}\right)\right).$$

Such test-statistics are the so called Rényi-divergence based test-statistics (see Liese and Vajda (1987)), and they contain also the Kullback-divergence based test-statistics as special case when $\lambda=0$.

For the example of Simpson (1989) we propose totally different test-statistics, in fact the power and Rényi divergence based test-statistics for the example of Simpson (1989) are (44) and (47), respectively with $g=2$, $n_1=177$, $n_2=126$, $F=\varnothing$ and $E=\varnothing$, and the likelihood ratio test-statistic we propose, (46), is not the same because the basis of the methodology is not the same (in the paper a one sided test was aimed but on the basis of a two sided test-statistic). The order restricted MLE of $\boldsymbol{\theta}$ is

$$\widetilde{\boldsymbol{\theta}}=(\widetilde{\theta}_1,\widetilde{\theta}_2)^T=\arg\max_{\boldsymbol{\theta}\in\Omega}-n_1\theta_1+\sum_{j=1}^{n_1}X_{1j}\log\theta_1-n_2\theta_2+\sum_{j=1}^{n_2}X_{2j}\log\theta_2,$$

where $\Omega=\{\theta_1,\theta_2\in\mathbb{R}_+:h(\theta_1,\theta_2)=\theta_2-\theta_1\leq 0\}$. Asymptotically, the p-value when x is the value of one of the proposed test-statistic, is given by $\frac{1}{2}\Pr(\chi_1^2\geq x)+\frac{1}{2}I(x\leq 0)$. That is $\frac{1}{2}\Pr(\chi_1^2\geq x)$, if $x>0$ and 1 if $x\leq 0$. These weights, $w_0^B(\theta_0)=w_1^B(\theta_0)=\frac{1}{2}$, are directly obtained taking into account iv) of page 11. This is a classical example for studying robustness. In Table 2 the values of the MLEs, power-divergence and Rényi divergence based test-statistics with $\lambda\in\{-\frac{1}{2},0,\frac{2}{3},1,\frac{3}{2}\}$ and their p-values are summarized. Power-divergence based test-statistics show a quit different behavior depending on the value of λ, while Rényi divergence based ones behave more homogeneously. Hypothesis H_{Null}^B is accepted with 0.05 significance level in all the cases except for $\widetilde{S}_{\phi_{3/2}}^B(\widehat{\boldsymbol{\theta}},\widetilde{\boldsymbol{\theta}})$. It could be concluded that Rényi divergence based test-statistics for g Poisson populations are more robust that the power-divergence based ones.

Number of recessive lethal daughters	0	1	2
Observations in the sample of population 1 = Control group	159	15	3
Observations in the sample of population 2 = Treated group	110	11	5

Table 1: Observed frequencies in the example of Simpson (1989).

$\widehat{\theta}_1$	$\widehat{\theta}_2$	$\widetilde{S}_{\phi_{-\frac{1}{2}}}^B(\widehat{\boldsymbol{\theta}},\widetilde{\boldsymbol{\theta}})$	$\widetilde{S}_{\phi_0}^B(\widehat{\boldsymbol{\theta}},\widetilde{\boldsymbol{\theta}})$	$\widetilde{S}_{\phi_{\frac{2}{3}}}^B(\widehat{\boldsymbol{\theta}},\widetilde{\boldsymbol{\theta}})$	$\widetilde{S}_{\phi_1}^B(\widehat{\boldsymbol{\theta}},\widetilde{\boldsymbol{\theta}})$	$\widetilde{S}_{\phi_{\frac{3}{2}}}^B(\widehat{\boldsymbol{\theta}},\widetilde{\boldsymbol{\theta}})$
0.118644	0.166666	1.114833	1.207049	1.741264	2.402724	4.876806
	p-values	0.145517	0.135959	0.093489	0.060562	0.013610
$\widetilde{\theta}_1$	$\widetilde{\theta}_2$	$\widetilde{\widetilde{S}}_{\phi_{-\frac{1}{2}}}^B(\widehat{\boldsymbol{\theta}},\widetilde{\boldsymbol{\theta}})$	$\widetilde{\widetilde{S}}_{\phi_0}^B(\widehat{\boldsymbol{\theta}},\widetilde{\boldsymbol{\theta}})$	$\widetilde{\widetilde{S}}_{\phi_{\frac{2}{3}}}^B(\widehat{\boldsymbol{\theta}},\widetilde{\boldsymbol{\theta}})$	$\widetilde{\widetilde{S}}_{\phi_1}^B(\widehat{\boldsymbol{\theta}},\widetilde{\boldsymbol{\theta}})$	$\widetilde{\widetilde{S}}_{\phi_{\frac{3}{2}}}^B(\widehat{\boldsymbol{\theta}},\widetilde{\boldsymbol{\theta}})$
0.138614	0.138614	1.200577	1.207049	1.218055	1.224576	1.235671
	p-values	0.136603	0.135959	0.134871	0.134232	0.133153

Table 2: Power-divergence based test-statistics in the example of Simpson (1989).

6 Simulation study: Divergence based testing for isotonic binomial proportions

In multinomial sampling we consider $k_i = k_0$ and hence

$$\begin{aligned}
&\boldsymbol{X}_{ij} \overset{ind}{\sim} \mathcal{M}(1,\boldsymbol{\pi}_i); \quad \boldsymbol{X}_{ij}^T \mathbf{1}_{k_0+1} = 1; \quad (j = 1, ..., n_i)\\
&\boldsymbol{\pi}_i = (\pi_{i1}, ..., \pi_{ik_0}, \pi_{i,k_0+1,})^T; \quad \boldsymbol{\pi}_i^T \mathbf{1}_{k_0+1} = 1;\\
&\sum_{j=1}^{n_i} \boldsymbol{X}_{ij} = (N_{i1}, ..., N_{ik_0}, N_{i,k_0+1,})^T = \boldsymbol{N}_i; \quad \boldsymbol{N}_i^T \mathbf{1}_{k_0+1} = n_i.
\end{aligned}$$

Suppose we know that the probability vectors of the populations are stochastically ordered, that is

$$\sum_{h=1}^{j} \pi_{ih} \le \sum_{h=1}^{j} \pi_{i+1,h}, \quad i = 1, ..., g-1, \quad j = 1, ..., k_0, \tag{48}$$

which means that $h_i(\boldsymbol{\pi}_1, ..., \boldsymbol{\pi}_g) = \sum_{h=1}^{j} (\pi_{ih} - \pi_{i+1,h}) \le 0$ for $i \in R = \{1, ..., r\} = E$, where $r = (g-1)k_0$. We shall denote (48) shortly with $\boldsymbol{\pi}_i \preceq_s \boldsymbol{\pi}_{i+1}$. This topic can be encountered for instance in Dardanoni and Forcina (1998). In addition, suppose we know that there exists a subset $F \subset E$ such that $h_i(\boldsymbol{\pi}_1, ..., \boldsymbol{\pi}_g) = 0$ for $i \in F$. Under these assumptions, we would like to test wether the probability vectors are equal, that is

$$H_{Null}^A: \ \boldsymbol{\pi}_i = \boldsymbol{\pi}_{i+1}, i \in E \quad \text{vs.} \quad H_{Alt}^A: \ \boldsymbol{\pi}_i \preceq_s \boldsymbol{\pi}_{i+1}, i \in E - F; \quad \boldsymbol{\pi}_i = \boldsymbol{\pi}_{i+1}, i \in F,$$

being strict at least one of the inequalities in $R - F$. Because we are inside the exponential family, we can use (73) given in Section A.5,

$$\mathrm{d}_{\phi_\lambda}^*(f_\theta, f_{\theta_0}) = \sum_{j=1}^{k_0+1} \widehat{\pi}_{ij}^{-\lambda}(E) \widetilde{\pi}_{ij}^{\lambda+1}(F),$$

and hence for $\lambda \notin \{-1, 0\}$, (69)-(72) given in Section A.5, are equal to

$$\widetilde{S}_{\phi_\lambda}^A(\widetilde{\boldsymbol{\theta}}(F), \widehat{\boldsymbol{\theta}}(E)) = \frac{2}{\lambda(\lambda+1)} \left(\prod_{i=1}^{g} \left(\sum_{j=1}^{k_0+1} \widehat{\pi}_{ij}^{-\lambda}(E) \widetilde{\pi}_{ij}^{\lambda+1}(F) \right)^{n_i} - 1 \right), \tag{49}$$

$$\widetilde{T}_{\phi_\lambda}^A(\widehat{\boldsymbol{\theta}}, \widetilde{\boldsymbol{\theta}}(F), \widehat{\boldsymbol{\theta}}(E)) = \frac{2}{\lambda(\lambda+1)} \left(\prod_{i=1}^{g} \left(\sum_{j=1}^{k_0+1} \widehat{\pi}_{ij}^{-\lambda}(E) \widehat{\pi}_{ij}^{\lambda+1} \right)^{n_i} - \prod_{i=1}^{g} \left(\sum_{j=1}^{k_0+1} \widetilde{\pi}_{ij}^{-\lambda}(F) \widehat{\pi}_{ij}^{\lambda+1} \right)^{n_i} \right). \tag{50}$$

where

$$\begin{aligned}
&\widehat{\boldsymbol{\theta}}(E) = (\widehat{\boldsymbol{\theta}}_1(E), ..., \widehat{\boldsymbol{\theta}}_g(E))^T; \quad \widehat{\boldsymbol{\theta}}_i(E) = (\widehat{\pi}_{i1}(E), ..., \widehat{\pi}_{ik_0}(E))^T; \quad \widehat{\pi}_{i,k_0+1}(E) = 1 - \widehat{\boldsymbol{\theta}}_i^T(E) \mathbf{1}_{k_0};\\
&\widetilde{\boldsymbol{\theta}}(F) = (\widetilde{\boldsymbol{\theta}}_1(F), ..., \widetilde{\boldsymbol{\theta}}_g(F))^T; \quad \widetilde{\boldsymbol{\theta}}_i(F) = (\widetilde{\pi}_{i1}(F), ..., \widetilde{\pi}_{ik_0}(F))^T; \quad \widetilde{\pi}_{i,k_0+1}(F) = 1 - \widetilde{\boldsymbol{\theta}}_i^T(F) \mathbf{1}_{k_0};\\
&\widehat{\boldsymbol{\theta}} = (\widehat{\boldsymbol{\theta}}_1, ..., \widehat{\boldsymbol{\theta}}_g)^T; \quad \widehat{\boldsymbol{\theta}}_i = (\widehat{\pi}_{i1}, ..., \widehat{\pi}_{ik_0})^T = \left(\tfrac{N_{i1}}{n_i}, ..., \tfrac{N_{ik_0}}{n_i} \right)^T; \quad \widehat{\pi}_{i,k_0+1}(E) = 1 - \widehat{\boldsymbol{\theta}}_i^T \mathbf{1}_{k_0}.
\end{aligned}$$

Based on the same idea explained in Section 5 we can construct the Rényi divergence based test-statistics

$$\widetilde{\widetilde{S}}_{\phi_\lambda}^A(\widetilde{\boldsymbol{\theta}}(F), \widehat{\boldsymbol{\theta}}(E)) = \frac{2}{\lambda(\lambda+1)} \sum_{i=1}^{g} n_i \log \left(\sum_{j=1}^{k_0+1} \widehat{\pi}_{ij}^{-\lambda}(E) \widetilde{\pi}_{ij}^{\lambda+1}(F) \right), \tag{51}$$

$$\widetilde{\widetilde{T}}_{\phi_\lambda}^A(\widehat{\boldsymbol{\theta}}, \widetilde{\boldsymbol{\theta}}(F), \widehat{\boldsymbol{\theta}}(E)) = \frac{2}{\lambda(\lambda+1)} \sum_{i=1}^{g} n_i \log \left(\frac{\sum_{j=1}^{k_0+1} \widehat{\pi}_{ij}^{-\lambda}(E) \widehat{\pi}_{ij}^{\lambda+1}}{\sum_{j=1}^{k_0+1} \widetilde{\pi}_{ij}^{-\lambda}(F) \widehat{\pi}_{ij}^{\lambda+1}} \right). \tag{52}$$

On the other hand, multinomial sample from multiple populations is suitable for the techniques of single population, even for different sample sizes, taking into account that we can construct a global probability vector weighting on the sample sizes

$$\widehat{\boldsymbol{\vartheta}}(E) = (\widehat{\boldsymbol{\vartheta}}_1(E), ..., \widehat{\boldsymbol{\vartheta}}_g(E))^T; \quad \widehat{\boldsymbol{\vartheta}}_i(E) = \left(\tfrac{n_i}{n}\widehat{\pi}_{i1}(E), ..., \tfrac{n_i}{n}\widehat{\pi}_{ik_0}(E)\right)^T; \quad \tfrac{n_i}{n}\widehat{\pi}_{i,k_0+1}(E) = \tfrac{n_i}{n} - \widehat{\boldsymbol{\vartheta}}_i^T(E)\mathbf{1}_{k_0};$$
$$\widetilde{\boldsymbol{\vartheta}}(F) = (\widetilde{\boldsymbol{\vartheta}}_1(F), ..., \widetilde{\boldsymbol{\vartheta}}_g(F))^T; \quad \widetilde{\boldsymbol{\vartheta}}_i(F) = \left(\tfrac{n_i}{n}\widetilde{\pi}_{i1}(F), ..., \tfrac{n_i}{n}\widetilde{\pi}_{ik_0}(F)\right)^T; \quad \tfrac{n_i}{n}\widetilde{\pi}_{i,k_0+1}(F) = \tfrac{n_i}{n} - \widetilde{\boldsymbol{\vartheta}}_i^T(F)\mathbf{1}_{k_0};$$
$$\widehat{\boldsymbol{\vartheta}} = (\widehat{\boldsymbol{\vartheta}}_1, ..., \widehat{\boldsymbol{\vartheta}}_g)^T; \quad \widehat{\boldsymbol{\vartheta}}_i = \left(\tfrac{n_i}{n}\widehat{\pi}_{i1}, ..., \tfrac{n_i}{n}\widehat{\pi}_{ik_0}\right)^T = \left(\tfrac{N_{i1}}{n}, ..., \tfrac{N_{ik_0}}{n}\right)^T; \quad \tfrac{n_i}{n}\widehat{\pi}_{i,k_0+1}(E) = \tfrac{n_i}{n} - \widehat{\boldsymbol{\vartheta}}_i^T\mathbf{1}_{k_0}.$$

Therefore,

$$S_{\phi_\lambda}^A(\widetilde{\boldsymbol{\vartheta}}(F), \widehat{\boldsymbol{\vartheta}}(E)) = \frac{2}{\lambda(\lambda+1)}\left(\sum_{i=1}^{g} n_i \sum_{j=1}^{k_0+1} \widehat{\pi}_{ij}^{-\lambda}(E)\widetilde{\pi}_{ij}^{\lambda+1}(F) - n\right), \tag{53}$$

$$T_{\phi_\lambda}^A(\widehat{\boldsymbol{\vartheta}}, \widetilde{\boldsymbol{\vartheta}}((F), \widehat{\boldsymbol{\vartheta}}(E)) = \frac{2}{\lambda(\lambda+1)}\left(\sum_{i=1}^{g} n_i \sum_{j=1}^{k_0+1} \widehat{\pi}_{ij}^{-\lambda}(E)\widehat{\pi}_{ij}^{\lambda+1} - \sum_{i=1}^{g} n_i \sum_{j=1}^{k_0+1} \widetilde{\pi}_{ij}^{-\lambda}(F)\widehat{\pi}_{ij}^{\lambda+1}\right). \tag{54}$$

Note that the Kullback based test-statistics ($\lambda = 0$) are

$$S_{\text{Kull}}^A(\widetilde{\boldsymbol{\vartheta}}(F), \widehat{\boldsymbol{\vartheta}}(E)) = 2\sum_{i=1}^{g} n_i \sum_{j=1}^{k_0+1} \widetilde{\pi}_{ij}(F) \log \frac{\widetilde{\pi}_{ij}(F)}{\widehat{\pi}_{ij}(E)},$$
$$T_{\text{Kull}}^A(\widehat{\boldsymbol{\vartheta}}, \widetilde{\boldsymbol{\vartheta}}(F), \widehat{\boldsymbol{\vartheta}}(E)) = 2\sum_{i=1}^{g} n_i \sum_{j=1}^{k_0+1} \widehat{\pi}_{ij} \log \frac{\widetilde{\pi}_{ij}(F)}{\widehat{\pi}_{ij}(E)}. \tag{55}$$

with $S_{\text{Kull}}^A(\widetilde{\boldsymbol{\vartheta}}(F), \widehat{\boldsymbol{\vartheta}}(E))) = \widetilde{S}_{\text{Kull}}^A(\widetilde{\boldsymbol{\theta}}(F), \widehat{\boldsymbol{\theta}}(E)) = \widetilde{\widetilde{S}}_{\text{Kull}}^A(\widetilde{\boldsymbol{\theta}}(F), \widehat{\boldsymbol{\theta}}(E))$ and $T_{\text{Kull}}^A(\widehat{\boldsymbol{\vartheta}}, \widetilde{\boldsymbol{\vartheta}}(F), \widehat{\boldsymbol{\vartheta}}(E)) = \widetilde{T}_{\text{Kull}}^A(\widehat{\boldsymbol{\theta}}, \widetilde{\boldsymbol{\theta}}(F), \widehat{\boldsymbol{\theta}}(E)) = \widetilde{\widetilde{T}}_{\text{Kull}}^A(\widehat{\boldsymbol{\theta}}, \widetilde{\boldsymbol{\theta}}(F), \widehat{\boldsymbol{\theta}}(E))$. The likelihood ratio test-statistic is just $T_{\text{Kull}}^A(\widehat{\boldsymbol{\vartheta}}, \widetilde{\boldsymbol{\vartheta}}(F), \widehat{\boldsymbol{\vartheta}}(E))$, and even though is not exactly equal to $S_{\text{Kull}}^A(\widetilde{\boldsymbol{\theta}}(F), \widehat{\boldsymbol{\theta}}(E))$, in practice their accepting and rejecting probabilities are equal with high precision (it can be seen in the simulation study). It is also remarkable that among the Pearson divergence based test-statistics ($\lambda = 1$) what is called usually called chi-square test-statistic is

$$S_{\phi_1}^A(\widetilde{\boldsymbol{\theta}}(F), \widehat{\boldsymbol{\theta}}(E)) = \sum_{i=1}^{g} n_i \sum_{j=1}^{k_0+1} \frac{\left(\widetilde{\pi}_{ij}(F) - \widehat{\pi}_{ij}(E)\right)^2}{\widehat{\pi}_{ij}(E)}$$

(see expressions for one population in page 240 of Robertson et al. (1988)).

This general case of $k_0 \in \mathbb{N}$, was discussed for instance in Dardanoni and Forcina (1998) for testing (16) with the likelihood ratio test-statistic. The case of $k_0 = 1$ with small probabilities of success is known for being problematic because the test-statistics have not a good behaviour (it can be seen in the plots shown at the end of this section taking $s = 1$). In Tebbs and Bilder (2006) some tests were analyzed with a "pooling design" ($s \geq 2$), useful to overcome this problem. Taking into account that in the aforementioned paper, $F = \varnothing$, $E = R$ and the likelihood ratio test-statistic $T_{\phi_0}^A(\widetilde{\boldsymbol{\theta}}(\varnothing), \widehat{\boldsymbol{\theta}}(R))$ and the chi-square test-statistic (or "Bartholomew's statistic") $S_{\phi_1}^A(\widehat{\boldsymbol{\vartheta}}, \widetilde{\boldsymbol{\vartheta}}(\varnothing), \widehat{\boldsymbol{\vartheta}}(R))$ were considered to be the best test-statistics among other alternative test-statistics, it is of common sense to analyze what happens with power divergence test-statistics. Furthermore, we consider in this paper hypothesis testing (11), which is more general than (16). The pooled testing for small proportions coming from g binomial populations considers a prefixed number of s individuals, which are independently pooled within each population and independently from other populations. An event of an individual of the sample in a specific population is considered to be successful if at least one of the s pools associated with it is successful, that is calling p_i (which is suppose to be small) the probability of having one successful pool in the i-th population, $i = 1, ..., g$, we have $X_{i1} \sim \mathcal{B}er(\theta_i)$, $N_{i1} \sim \mathcal{B}in(n_i, \theta_i)$ for $i = 1, ..., g$, where $\pi_{i1} = \theta_i$, $\pi_{i2} = 1 - \theta_i$ and $\theta_i = 1 - (1 - p_i)^s$. Our aim is to study

$$H_{Null}^A : \ p_1 = p_2 = p_3 = p_4 \quad \text{vs.} \quad H_{Alt}^A : \ p_1 \leq p_2 = p_3 \leq p_4, \text{ and } (p_1 < p_2 \text{ or } p_3 < p_4),$$

but this is equivalent to (56). The order restricted MLE of $\boldsymbol{\theta} = (\theta_1, \theta_2, \theta_3, \theta_4)^T = (\pi_{11}, \pi_{21}, \pi_{31}, \pi_{41})^T$ is

$$\widetilde{\boldsymbol{\theta}}(F) = (\widetilde{\theta}_1(F), \widetilde{\theta}_2(F), \widetilde{\theta}_3(F), \widetilde{\theta}_4(F))^T = \arg \max_{\boldsymbol{\theta} \in \Omega(F)} \sum_{i=1}^{g} \left(N_{i1} \log \theta_i + (n_i - N_{i1}) \log(1 - \theta_i) \right),$$

where $F = \{2\}$, $\Omega(F) = \{\theta_1, \theta_2, \theta_3, \theta_4 \in (0,1) : h_1(\boldsymbol{\theta}) = \theta_1 - \theta_2 \leq 0, h_2(\boldsymbol{\theta}) = \theta_2 - \theta_3 = 0, h_3(\boldsymbol{\theta}) = \theta_3 - \theta_4 \leq 0\}$, and the estimators under equality restrictions, $\widehat{\boldsymbol{\theta}}(E) = (\widehat{\theta}_1(E), \widehat{\theta}_2(E), \widehat{\theta}_3(E), \widehat{\theta}_4(E))^T$ where $\widehat{\theta}_i(E) = \sum_{i=1}^{g} N_{i1}/n$. Different sample sizes for each population are considered, $n_1 = 20$, $n_2 = 25$, $n_3 = 30$, $n_4 = 35$. We shall perform a simulation study to illustrate that the performance of the test is improved when nesting the models under H_{Alt}^A and moreover that for small sample sizes the likelihood ratio test (LRT) can be improved. It can be seen in Section A.7 that the behaviour of T and S test statistics is quite similar, and (49)-(50) are non.recommendable test-statistics. We have studied all the proposed test-statistics (see Section A.7) but are going to show plots only for two families of test-statistics, $\widetilde{\widetilde{S}}_{\phi_\lambda}^A(\widetilde{\boldsymbol{\theta}}(F), \widehat{\boldsymbol{\theta}}(E))$ and $S_{\phi_\lambda}^A(\widetilde{\boldsymbol{\vartheta}}(F), \widehat{\boldsymbol{\vartheta}}(E))$ with $\lambda \in \{0, \frac{2}{3}, 1\}$, for testing

$$H_{Null}^A : \ \pi_{11} = \pi_{21} = \pi_{31} = \pi_{41} \quad \text{vs.} \quad H_{Alt}^A : \ \pi_{11} \leq \pi_{21} = \pi_{31} \leq \pi_{41}, \text{ and } (\pi_{11} < \pi_{21} \text{ or } \pi_{31} < \pi_{41}), \tag{56}$$

with different pool sizes, $s \in \{1, 5, 10, 15, 20\}$. These test-statistics are simpler to compute and the conclusions for plots of $\widetilde{\widetilde{T}}_{\phi_\lambda}^A(\widehat{\boldsymbol{\theta}}, \widetilde{\boldsymbol{\theta}}(F), \widehat{\boldsymbol{\theta}}(E))$ and $T_{\phi_\lambda}^A(\widehat{\boldsymbol{\vartheta}}, \widetilde{\boldsymbol{\vartheta}}(F), \widehat{\boldsymbol{\vartheta}}(E))$ are very similar Notation LRT, CRT(2/3), CRT(1), LRT, RT(2/3), RT(1) simplifies respectively the notation of these test-statistics $\widetilde{\widetilde{S}}_{\phi_\lambda}^A(\widetilde{\boldsymbol{\theta}}(F), \widehat{\boldsymbol{\theta}}(E))$ and $S_{\phi_\lambda}^A(\widetilde{\boldsymbol{\vartheta}}(F), \widehat{\boldsymbol{\vartheta}}(E))$ with $\lambda \in \{0, \frac{2}{3}, 1\}$. Asymptotically, the p-value when x is the value of one of the 12 test-statistics for (56), is given by

$$w_0^A(\theta_0) \Pr(\chi_2^2 \geq x) + w_1^A(\theta_0) \Pr(\chi_1^2 \geq x) + w_2^A(\theta_0) \Pr(\chi_0^2 \geq x),$$

where $w_0^A(\theta_0) = \frac{1}{4}$, $w_1^A(\theta_0) = \frac{1}{2}$, $w_2^A(\theta_0) = \frac{1}{4}$ (see details in Section A.6). The exact sizes with nominal size 0.05 are calculated by simulation with 30 000 replication, and three cases are distinguished, in scenario A $(p_1, p_2, p_3, p_4) = (0.1, 0.1, 0.1, 0.1)$, in scenario B $(p_1, p_2, p_3, p_4) = (0.05, 0.05, 0.05, 0.05)$, in scenario C $(p_1, p_2, p_3, p_4) = (0.01, 0.01, 0.01, 0.01)$. Exact powers are also calculated for scenario A with $(p_1, p_2, p_3, p_4) = (0.05, 0.1, 0.1, 0.15)$, for scenario B with $(p_1, p_2, p_3, p_4) = (0.01, 0.04, 0.04, 0.07)$, for scenario C with $(p_1, p_2, p_3, p_4) = (0.01, 0.025, 0.025, 0.04)$. In order to illustrate in what degree is the test improved when using (56) rather than

$$H_{Null}^A : \ \pi_{11} = \pi_{21} = \pi_{31} = \pi_{41} \quad \text{vs.} \quad H_{Alt}^A : \ \pi_{11} \leq \pi_{21} \leq \pi_{31} \leq \pi_{41}, \text{ and } (\pi_{11} < \pi_{21} \text{ or } \pi_{21} < \pi_{31} \text{ or } \pi_{31} < \pi_{41}), \tag{57}$$

we shall consider the same scenarios with the same values of (p_1, p_2, p_3, p_4). In this case, the order restricted MLEs, $\widetilde{\boldsymbol{\theta}}$, are defined in the same way but the parametric space is $\Omega(\varnothing) = \{\theta_1, \theta_2, \theta_3, \theta_4 \in (0,1) : h_1(\boldsymbol{\theta}) = \theta_1 - \theta_2 \leq 0, h_2(\boldsymbol{\theta}) = \theta_2 - \theta_3 \leq 0, h_3(\boldsymbol{\theta}) = \theta_3 - \theta_4 \leq 0\}$. Asymptotically, the p-value when x is the value of one of the analyzed test-statistics for (57), is given by

$$w_0^A(\theta_0) \Pr(\chi_3^2 \geq x) + w_1^A(\theta_0) \Pr(\chi_2^2 \geq x) + w_2^A(\theta_0) \Pr(\chi_1^2 \geq x) + w_3^A(\theta_0) \Pr(\chi_0^2 \geq x),$$

where $w_0^A(\theta_0) = 0.04229179$, $w_1^A(\theta_0) = 0.2515227$, $w_2^A(\theta_0) = 0.4577082$, $w_3^A(\theta_0) = 0.2484773$ (see details in Section A.6). In Figures 1, 2, 3, 4 the results all of these scenarios are shown. As expected, it can be seen that as the pool size, s, is greater, the the power of the test increase and the approximation of the simulated size to the nominal size is much better (pooled-testing experiments use a larger number of individuals than individual testing). When comparing all the tests-statistics what is very clear is that the estimated sizes are smaller for test 56 than for test 57, and furthermore the LRT for test 57 is "liberal" because its simulated sizes tend to be greater than the nominal size. This is the main reason to support the model with $\Omega(\{2\}) - \Theta(\{1,2,3\})$ of test (56) which is contained in $\Omega(\varnothing) - \Theta(\{1,2,3\})$ of test (57). For scenario C, it is not easy to conclude something specific because the sample sizes are not large enough, the LRT tend to be "liberal" (simulated sizes above the nominal size, 0.05) which is not convenient, but the powers are much greater. In scenarios A and B for test 57 either R(2/3) or R(1) are good choices since tend to be "conservative" (their simulated sizes are usually below the nominal size, 0.05) and the powers are not much worse than for the LRT, while for test 56 none of them can be considered clearly better than others.

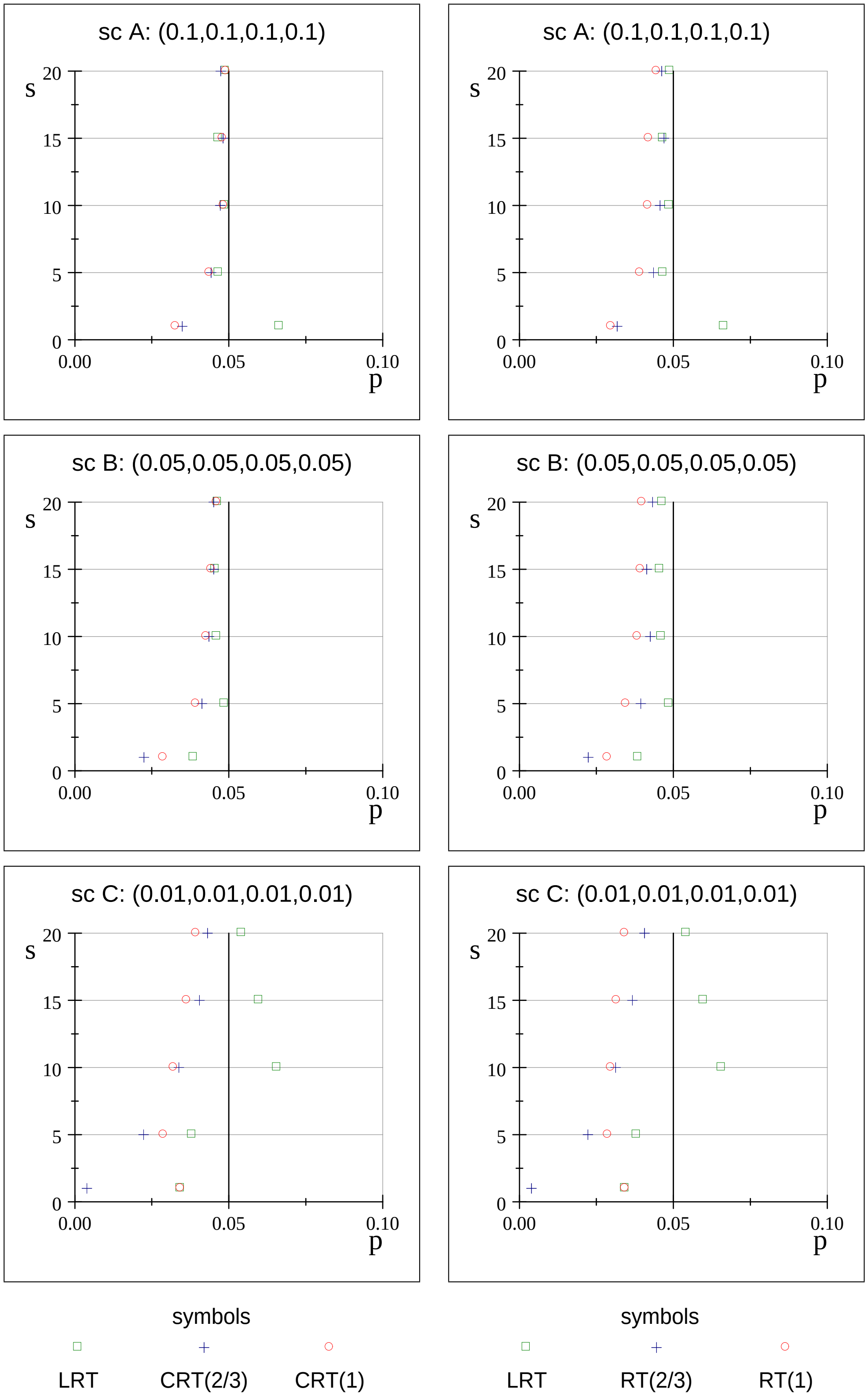

Figure 1: Simulated sizes for test (56) with different statistics (symbols) and pooling sizes (s) in three scenarios.

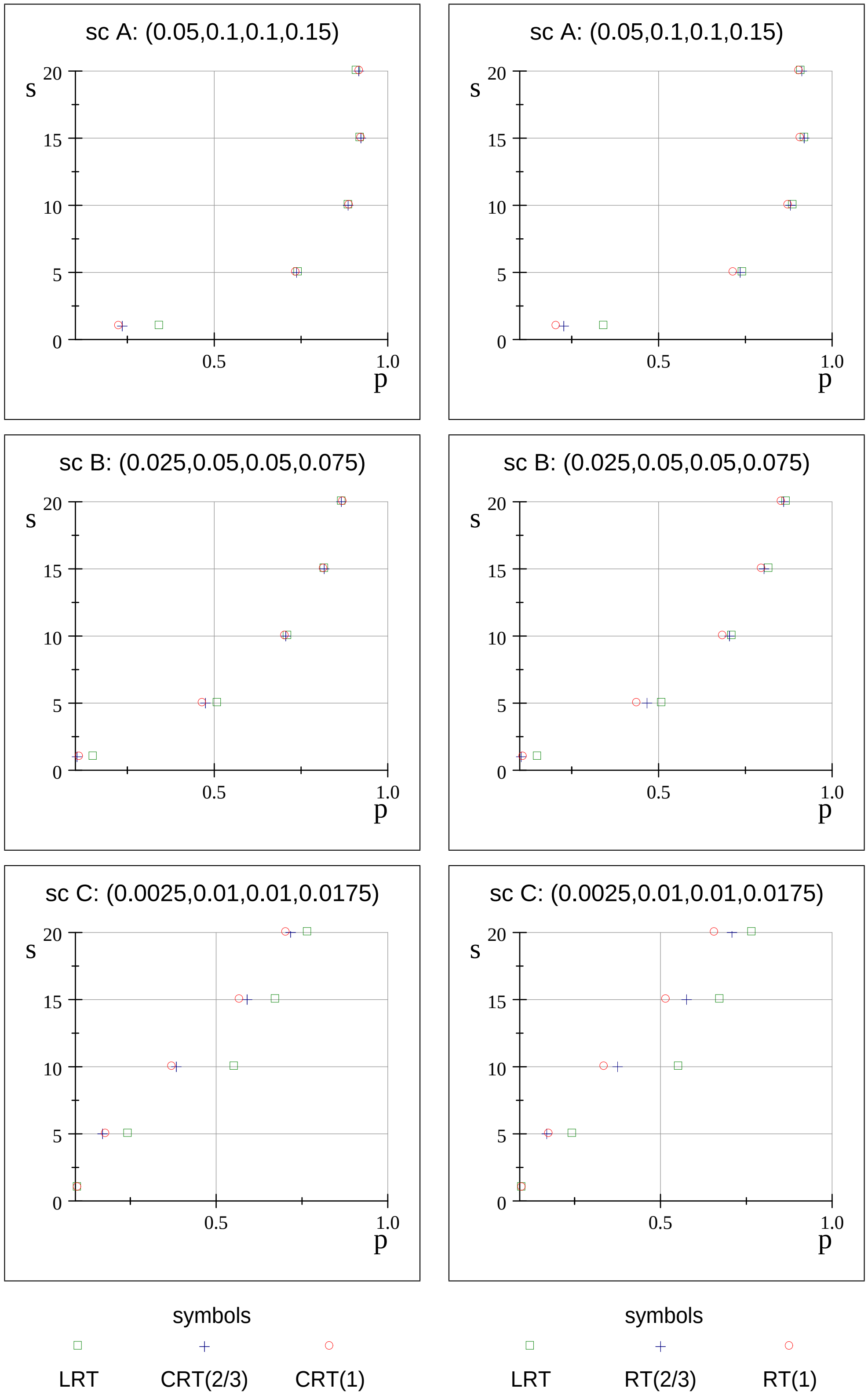

Figure 2: Simulated powers for test (56) with different statistics (symbols) and pooling sizes (s) in three scenarios.

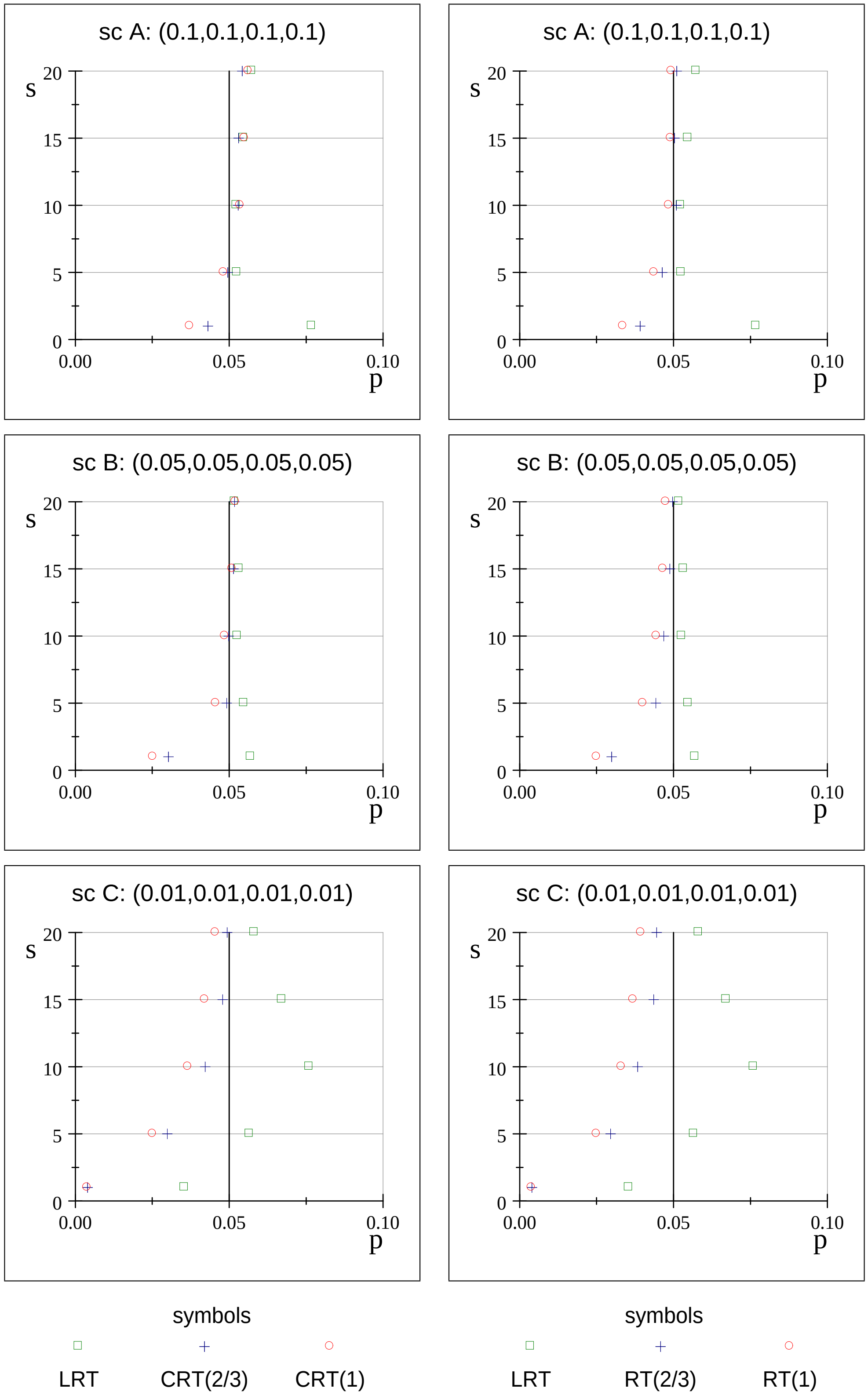

Figure 3: Simulated sizes for test (57) with different statistics (symbols) and pooling sizes (s) in three scenarios.

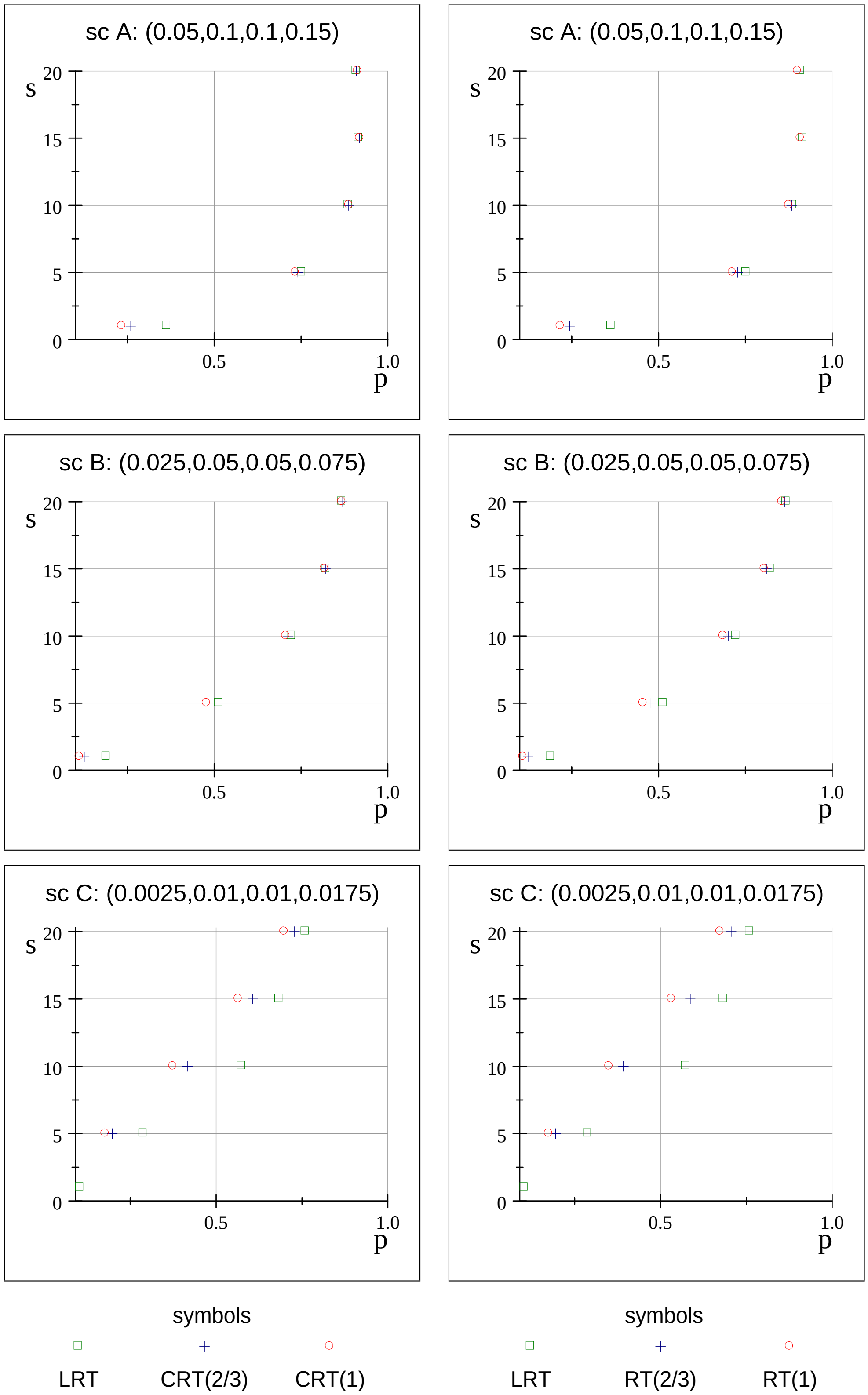

Figure 4: Simulated powers for test (57) with different statistics (symbols) and pooling sizes (s) in three scenarios.

References

[1] Aitchison, J. and Silvey, S. D. (1958). Maximum-Likelihood Estimation of Parameters Subject to Restraints. *Annals of Mathematical Statistics*, **29**, 813-828.

[2] Barlow, R. E., Bartholomew, D. J., Bremmer, J.M. and Brunk, H.D. (1972). *Statistical inference under order restrictions : the theory and application of isotonic regression.* Wiley.

[3] Bartholomew, D. J. (1959). A tests for homogeneity for ordered alternatives. *Biometrika*, **46**, 36-48.

[4] Bazaraa, M. S., Sherali, H. D. and Shetty, C. M. (2006). *Nonlinear Programming: Theory and Algorithms* (3rd Edition). John Wiley and Sons.

[5] Dardanoni, V. and Forcina, A. (1998). A unified approach to likelihood inference on stochastic orderings in a nonparametric context. Journal of the American Statistical Association, 93, 1112–1123.

[6] El Barmi, H. and Dykstra, R. (1995). Testing for and against a set of linear inequality constraints in a multinomial setting. *The Canadian Journal of Statistics*, **23**, 131–143.

[7] Felipe, A., Menéndez, M. L. and Pardo, L. (2007). Order-restricted Dose-related Trend Phi-divergence Tests for Generalized Linear Models. *Journal of Applied Statistics*, **34**, 611–623.

[8] Ferguson, T. S. (1996). *A Course in Large Sample Theory.* Chapman & Hall.

[9] Harville, D. A. (2008). *Matrix algebra from a statistician's perspective.* Springer.

[10] 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.

[11] Kudô, A. (1963). A multivariate analogue of the one-sided test. *Biometrika*, **50**, 403-418.

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

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

[14] Menéndez, M. L., Morales, D., Pardo, L. and Salicrú, M. (1997). Divergence measures between s populations: Statistical applications in the exponential model. *Communications in Statistic. Theory and Methods*, **26**, 1099–1117.

[15] Menéndez, M. L., Pardo, L. and Zografos, K. (2002). Tests of hypotheses for and against order restrictions on multinomial parameters based on ϕ-divergences. *Utílitas Mathematica*, **61**, 209-223

[16] Menéndez, M. L., Pardo, J.A. and Pardo, L. (2003a). Tests for Bivariate Symmetry Against Ordered Alternatives in Square Contingency Tables. *Australian & New Zealand Journal of Statistics*, **45**, 115–123.

[17] Menéndez, M. L., Morales, D. and Pardo, L. (2003b). Tests based on divergences for and against ordered alternatives in cubic contingency tables. *Applied Mathematics and Computation*, **134**, 207–216.

[18] 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.

[19] Morales, D., Pardo, L. and Zografos, K. (1998) Informational distances and related statistics in mixed continuous and categorical variables. *Journal of Statistical Planning and Inference*, **75**, 47–63.

[20] 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.

[21] Pardo, L. and Menéndez, M. L. (2006). Phi-Divergence-Type Test for Positive Dependence Alternatives in $2 \times k$ Contingency Tables. In: Advances in Distribution Theory, Order Statistics, and Inference. Ed.: N. Balakrishnan, J. M. Sarabia and E. Castillo. Pages 417–431.

[22] Read, T. and Cressie, N. (1988). Goodness of fit statistics for discrete multivariate data. Springer, New York.

[23] Robertson, T.,Wright, F.T. and Dykstra, R.L. (1988). *Order Restricted Statistical Inference*. Wiley, New York.

[24] 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.

[25] Sen, P. K. and Singer, J. M. (1993). *Large Sample Methods in Statistics: An Introduction with Applications.* Chapman & Hall.

[26] Sen, P. K., Singer, J. M. and Pedroso de Lima, A. C. (2010). *From Finite Sample to Asymptotic Methods in Statistics*. Cambridge University Press.

[27] Shapiro, A. (1985). Asymptotic Distribution of Test Statistics in the Analysis of Moment Structures Under Inequality Constraints. *Biometrika*, **72**, 133–144.

[28] Silvapulle, M. J. and Sen, P. K. (2004). *Constrained Statistical Inference: Order, Inequality, and Shape Constraints*, Wiley, New York.

[29] Silvey, S.D. (1959). The Lagrange-multiplier test. *Annals of Mathematical Statistics*, **30**, 389–407.

[30] Simpson, D. G. (1989). Deviance Tests: Efficiency, Breakdown Points, and Examples. *Journal of the American Statistical Association*, **84**, 107-113.

[31] Tebbs, J. M. and Bilder, C. R. (2006). Hypothesis Tests for and against a Simple Order among Proportions Estimated by Pooled Testing. *Biometrical Journal*, **48**, 792–804.

[32] Zografos, K., Ferentinos, K. and Papaioannou, T. (1990). Divergence statistics: sampling properties, multinomial goodness of fit and divergence tests. *Communications in Statistics. Theory and Methods*, **18**, 1785–1802.

[33] Zografos, K. (1998) f-Dissimilarity of several distributions in testing statistical hypotheses. *Annals of the Institute of Statistical Mathematics*, **50**, 295–310.

A Appendix

A.1 Proof of Theorem 8

The second order Taylor expansion of function $\mathrm{d}_\phi(\boldsymbol{\theta}) = \mathrm{d}_\phi(f_{\boldsymbol{\theta}}, f_{\widehat{\boldsymbol{\theta}}(\bullet)})$ about $\widehat{\boldsymbol{\theta}}(\bullet)$ is

$$\mathrm{d}_\phi(\boldsymbol{\theta}) = \mathrm{d}_\phi(\widehat{\boldsymbol{\theta}}(\bullet)) + (\boldsymbol{\theta} - \widehat{\boldsymbol{\theta}}(\bullet))^T \left.\frac{\partial}{\partial \boldsymbol{\theta}} \mathrm{d}_\phi(\boldsymbol{\theta})\right|_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}(\bullet)} + \frac{1}{2}(\boldsymbol{\theta} - \widehat{\boldsymbol{\theta}}(\bullet))^T \left.\frac{\partial^2}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T} \mathrm{d}_\phi(\boldsymbol{\theta})\right|_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}(\bullet)} (\boldsymbol{\theta} - \widehat{\boldsymbol{\theta}}(\bullet)) + o\left(\left\|\boldsymbol{\theta} - \widehat{\boldsymbol{\theta}}(\bullet)\right\|^2\right), \tag{58}$$

where $\bullet$ a general term to refer to the set of indices that are active in $\boldsymbol{h}$, $\mathrm{d}_\phi(\widehat{\boldsymbol{\theta}}(\bullet)) = 0$, and according to (30) and (31) we have

$$\left.\frac{\partial}{\partial \boldsymbol{\theta}} \mathrm{d}_\phi(\boldsymbol{\theta})\right|_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}(\bullet)} = \mathbf{0}_k,$$
$$\left.\frac{\partial^2}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T} \mathrm{d}_\phi(\boldsymbol{\theta})\right|_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}(\bullet)} = \phi''(1)\, \mathcal{I}_F(\widehat{\boldsymbol{\theta}}(\bullet)).$$

That is, in particular for $\boldsymbol{\theta} = \widehat{\boldsymbol{\theta}}$ we have

$$\mathrm{d}_\phi(\widehat{\boldsymbol{\theta}}, \widehat{\boldsymbol{\theta}}(E)) = \frac{\phi''(1)}{2} (\widehat{\boldsymbol{\theta}} - \widehat{\boldsymbol{\theta}}(E))^T \mathcal{I}_F(\widehat{\boldsymbol{\theta}}(E))(\widehat{\boldsymbol{\theta}} - \widehat{\boldsymbol{\theta}}(E)) + \mathrm{o}\left(\left\|\widehat{\boldsymbol{\theta}} - \widehat{\boldsymbol{\theta}}(E)\right\|^2\right),$$
$$\mathrm{d}_\phi(\widehat{\boldsymbol{\theta}}, \widehat{\boldsymbol{\theta}}(F)) = \frac{\phi''(1)}{2} (\widehat{\boldsymbol{\theta}} - \widehat{\boldsymbol{\theta}}(F))^T \mathcal{I}_F(\widehat{\boldsymbol{\theta}}(F))(\widehat{\boldsymbol{\theta}} - \widehat{\boldsymbol{\theta}}(F)) + \mathrm{o}\left(\left\|\widehat{\boldsymbol{\theta}} - \widehat{\boldsymbol{\theta}}(F)\right\|^2\right).$$

Multiplying both sides of the equality by $\frac{2n}{\phi''(1)}$ and taking the difference in both sides of the equality

$$\begin{aligned} T_\phi^O(\widehat{\boldsymbol{\theta}}, \widehat{\boldsymbol{\theta}}(F), \widehat{\boldsymbol{\theta}}(E)) &= \frac{2n}{\phi''(1)} \left(\mathrm{d}_\phi(f_{\widehat{\boldsymbol{\theta}}}, f_{\widehat{\boldsymbol{\theta}}(E)}) - \mathrm{d}_\phi(f_{\widehat{\boldsymbol{\theta}}}, f_{\widehat{\boldsymbol{\theta}}(F)}) \right) \\ &= \sqrt{n}(\widehat{\boldsymbol{\theta}} - \widehat{\boldsymbol{\theta}}(E))^T \mathcal{I}_F(\widehat{\boldsymbol{\theta}}(E)) \sqrt{n}(\widehat{\boldsymbol{\theta}} - \widehat{\boldsymbol{\theta}}(E)) + \mathrm{o}\left(\left\|\sqrt{n}\left(\widehat{\boldsymbol{\theta}} - \widehat{\boldsymbol{\theta}}(E)\right)\right\|^2\right) \\ &\quad - \sqrt{n}(\widehat{\boldsymbol{\theta}} - \widehat{\boldsymbol{\theta}}(F))^T \mathcal{I}_F(\widehat{\boldsymbol{\theta}}(F)) \sqrt{n}(\widehat{\boldsymbol{\theta}} - \widehat{\boldsymbol{\theta}}(F)) + \mathrm{o}\left(\left\|\sqrt{n}\left(\widehat{\boldsymbol{\theta}} - \widehat{\boldsymbol{\theta}}(F)\right)\right\|^2\right). \end{aligned}$$

It is well-known that

$$\sqrt{n}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) = \mathcal{I}_F^{-1}(\boldsymbol{\theta}_0) \sqrt{n} \left.\frac{\partial}{\partial \boldsymbol{\theta}} \ell_1(\boldsymbol{\theta})\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} + \mathrm{o}_P(\mathbf{1}_k), \tag{59}$$
$$\sqrt{n}(\widehat{\boldsymbol{\theta}}(\bullet) - \boldsymbol{\theta}_0) = \boldsymbol{P}(\boldsymbol{\theta}_0, \bullet) \sqrt{n} \left.\frac{\partial}{\partial \boldsymbol{\theta}} \ell_1(\boldsymbol{\theta})\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} + \mathrm{o}_P(\mathbf{1}_k), \tag{60}$$

where $\boldsymbol{\theta}_0$ is the true and unknown value of the parameter,

$$\boldsymbol{P}(\boldsymbol{\theta}_0, \bullet) = \mathcal{I}_F^{-1}(\boldsymbol{\theta}_0) - \mathcal{I}_F^{-1}(\boldsymbol{\theta}_0) \boldsymbol{H}^T(\boldsymbol{\theta}_0, \bullet) \left(\boldsymbol{H}(\boldsymbol{\theta}_0, \bullet) \mathcal{I}_F^{-1}(\boldsymbol{\theta}_0) \boldsymbol{H}^T(\boldsymbol{\theta}_0, \bullet) \right)^{-1} \boldsymbol{H}(\boldsymbol{\theta}_0, \bullet) \mathcal{I}_F^{-1}(\boldsymbol{\theta}_0),$$

is the variance covariance matrix of $\widehat{\boldsymbol{\theta}}(\bullet)$ according to (7)-(8)-(6a), and $\sqrt{n} \left.\frac{\partial}{\partial \boldsymbol{\theta}} \ell_1(\boldsymbol{\theta})\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} \xrightarrow[n\to\infty]{\mathcal{L}} \mathcal{N}(\mathbf{0}_k, \mathcal{I}_F(\boldsymbol{\theta}_0))$ by the Central Limit Theorem. Taking the differences of both sides of the equality in (59) and (60), we obtain

$$\sqrt{n}(\widehat{\boldsymbol{\theta}} - \widehat{\boldsymbol{\theta}}(\bullet)) = \left(\mathcal{I}_F^{-1}(\boldsymbol{\theta}_0) - \boldsymbol{P}(\boldsymbol{\theta}_0, \bullet) \right) \sqrt{n} \left.\frac{\partial}{\partial \boldsymbol{\theta}} \ell_1(\boldsymbol{\theta})\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} + \mathrm{o}_P(\mathbf{1}_k),$$

and taking into account $\mathcal{I}_F(\widehat{\boldsymbol{\theta}}(E)) \xrightarrow[n\to\infty]{P} \mathcal{I}_F(\boldsymbol{\theta}_0)$,

$$\begin{aligned} &T_\phi^O(\widehat{\boldsymbol{\theta}}, \widehat{\boldsymbol{\theta}}(F), \widehat{\boldsymbol{\theta}}(E)) \\ &= \sqrt{n} \left.\frac{\partial}{\partial \boldsymbol{\theta}^T} \ell_1(\boldsymbol{\theta})\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} (\boldsymbol{P}(\boldsymbol{\theta}_0, F) - \boldsymbol{P}(\boldsymbol{\theta}_0, E))^T \mathcal{I}_F(\boldsymbol{\theta}_0) (\boldsymbol{P}(\boldsymbol{\theta}_0, F) - \boldsymbol{P}(\boldsymbol{\theta}_0, E)) \sqrt{n} \left.\frac{\partial}{\partial \boldsymbol{\theta}} \ell_1(\boldsymbol{\theta})\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} + \mathrm{o}_P(1) \\ &= \mathbf{Y}^T \mathbf{Y} + \mathrm{o}_P(1), \end{aligned}$$

where

$$\boldsymbol{Y} = \boldsymbol{A}(\boldsymbol{\theta}_0)\left(\boldsymbol{P}(\boldsymbol{\theta}_0, F) - \boldsymbol{P}(\boldsymbol{\theta}_0, E)\right)\boldsymbol{A}(\boldsymbol{\theta}_0)^T \boldsymbol{Z},$$

with $\boldsymbol{Z} \sim \mathcal{N}(\mathbf{0}_k, \boldsymbol{I}_k)$ and $\boldsymbol{A}(\boldsymbol{\theta}_0)$ is Cholesky's factorization matrix for a non singular matrix such a Fisher information matrix, that is $\mathcal{I}_F(\boldsymbol{\theta}_0) = \boldsymbol{A}(\boldsymbol{\theta}_0)^T \boldsymbol{A}(\boldsymbol{\theta}_0)$. In other words

$$\boldsymbol{Y} \sim \mathcal{N}(\mathbf{0}_k, \boldsymbol{A}(\boldsymbol{\theta}_0)\left(\boldsymbol{P}(\boldsymbol{\theta}_0, F) - \boldsymbol{P}(\boldsymbol{\theta}_0, E)\right)\boldsymbol{A}(\boldsymbol{\theta}_0)^T),$$

where the variance covariance matrix is idempotent and symmetric. Following Lemma 3 in Ferguson (1996, page 57), $\boldsymbol{A}(\boldsymbol{\theta}_0)\left(\boldsymbol{P}(\boldsymbol{\theta}_0, F) - \boldsymbol{P}(\boldsymbol{\theta}_0, E)\right)\boldsymbol{A}(\boldsymbol{\theta}_0)^T$ is idempotent and symmetric, if only if $T_\phi^O(\widehat{\boldsymbol{\theta}}(E), \widehat{\boldsymbol{\theta}}(F))$ is a chi-square random variable with degrees of freedom

$$df = \operatorname{rank}(\boldsymbol{A}(\boldsymbol{\theta}_0)\left(\boldsymbol{P}(\boldsymbol{\theta}_0, F) - \boldsymbol{P}(\boldsymbol{\theta}_0, E)\right)\boldsymbol{A}(\boldsymbol{\theta}_0)^T) = \operatorname{trace}(\boldsymbol{A}(\boldsymbol{\theta}_0)\left(\boldsymbol{P}(\boldsymbol{\theta}_0, F) - \boldsymbol{P}(\boldsymbol{\theta}_0, E)\right)\boldsymbol{A}(\boldsymbol{\theta}_0)^T).$$

Since

$$\left(\boldsymbol{P}(\boldsymbol{\theta}_0, F) - \boldsymbol{P}(\boldsymbol{\theta}_0, E)\right)^T \mathcal{I}_F(\boldsymbol{\theta}_0)\left(\boldsymbol{P}(\boldsymbol{\theta}_0, F) - \boldsymbol{P}(\boldsymbol{\theta}_0, E)\right) = \boldsymbol{P}(\boldsymbol{\theta}_0, F) - \boldsymbol{P}(\boldsymbol{\theta}_0, E),$$

the condition is reached. The effective degrees of freedom are given by

$$\begin{aligned} df &= \operatorname{trace}(\boldsymbol{P}(\boldsymbol{\theta}_0, F)\boldsymbol{A}(\boldsymbol{\theta}_0)^T\boldsymbol{A}(\boldsymbol{\theta}_0)) - \operatorname{trace}(\boldsymbol{P}(\boldsymbol{\theta}_0, E)\boldsymbol{A}(\boldsymbol{\theta}_0)^T\boldsymbol{A}(\boldsymbol{\theta}_0)) = \operatorname{trace}(\boldsymbol{P}(\boldsymbol{\theta}_0, F)\mathcal{I}_F(\boldsymbol{\theta}_0)) - \operatorname{trace}(\boldsymbol{P}(\boldsymbol{\theta}_0, E)\mathcal{I}_F(\boldsymbol{\theta}_0)) \\ &= \operatorname{trace}(-\left(\boldsymbol{H}(\boldsymbol{\theta}_0, F)\mathcal{I}_F^{-1}(\boldsymbol{\theta}_0)\boldsymbol{H}^T(\boldsymbol{\theta}_0, F)\right)^{-1}\boldsymbol{H}(\boldsymbol{\theta}_0, F)\mathcal{I}_F^{-1}(\boldsymbol{\theta}_0)\boldsymbol{H}^T(\boldsymbol{\theta}_0, F)) \\ &- \operatorname{trace}(-\left(\boldsymbol{H}(\boldsymbol{\theta}_0, E)\mathcal{I}_F^{-1}(\boldsymbol{\theta}_0)\boldsymbol{H}^T(\boldsymbol{\theta}_0, E)\right)^{-1}\boldsymbol{H}(\boldsymbol{\theta}_0, E)\mathcal{I}_F^{-1}(\boldsymbol{\theta}_0)\boldsymbol{H}^T(\boldsymbol{\theta}_0, E)) \\ &= \operatorname{card}(E - F). \end{aligned}$$

Regarding the other test-statistic $S_\phi^O(\widehat{\boldsymbol{\theta}}(F), \widehat{\boldsymbol{\theta}}(E))$, observe that if we take (58), in particular for $\boldsymbol{\theta} = \widehat{\boldsymbol{\theta}}(F)$ and $\bullet = \widehat{\boldsymbol{\theta}}(E)$ it is directly obtained

$$\mathrm{d}_\phi(\widehat{\boldsymbol{\theta}}(F), \widehat{\boldsymbol{\theta}}(E)) = \frac{\phi''(1)}{2}(\widehat{\boldsymbol{\theta}}(F) - \widehat{\boldsymbol{\theta}}(E))^T \mathcal{I}_F(\widehat{\boldsymbol{\theta}}(E))(\widehat{\boldsymbol{\theta}}(F) - \widehat{\boldsymbol{\theta}}(E)) + \mathrm{o}\left(\left\|\widehat{\boldsymbol{\theta}}(F) - \widehat{\boldsymbol{\theta}}(E)\right\|^2\right),$$

and the rest of the steps to reach the final result are very similar compared with the other test-statistic.

A.2 Lemma

Let $\boldsymbol{Y}$ be a k-dimensional random variable with normal distribution $\mathcal{N}(\mathbf{0}_k, \boldsymbol{P})$ with $\boldsymbol{P}$ being a projection matrix, that is idempotent and symmetric, and let fixed k-dimensional vectors $\boldsymbol{d}_i$ such that for them either $\boldsymbol{P}\boldsymbol{d}_i = \mathbf{0}_k$ or $\boldsymbol{P}\boldsymbol{d}_i = \boldsymbol{d}_i$, $i = 1, ..., k$, is true. Then $\left(\boldsymbol{Y}^T\boldsymbol{Y} \middle| \boldsymbol{d}_i^T\boldsymbol{Y} \leq 0, i = 1, ..., k\right) \sim \chi^2_{df}$, where $df = \operatorname{rank}(\boldsymbol{P})$.

Proof. This result can be found in several sources, for instance in Kudô (1963, page 414), Barlow et al. (1972, page 128) and Shapiro (1985, page 139). ■

Without any loss of generality we shall consider $E = \{1, ..., r\}$ since otherwise the null hypothesis is not a vector space. In case of $E = \{1, ..., r_2\}$ with $r_2 < r$ we are assuming that there are no restriction on $h_i(\boldsymbol{\theta})$ with $i \notin E$, nor under the null hypothesis and neither for the alternative, which is essentially the same as taking $E = \{1, ..., r\}$.

A.3 Proof of Theorem 12

We shall perform the proof for $S_\phi^A(\widetilde{\boldsymbol{\theta}}(F), \widehat{\boldsymbol{\theta}}(E))$. Under H_{Alt}^A, $h_i(\boldsymbol{\theta}_0) = 0$ for $i \in F$ is conditionally established. Hence, either $h_i(\boldsymbol{\theta}_0) = 0$ or $h_i(\boldsymbol{\theta}_0) < 0$ can be true for $i \in E - F$ and we want to test $h_i(\boldsymbol{\theta}_0) = 0$, $i \in E$ (H_{Null}^A). Since $F \subset E$, it is clear that if H_{Null}^A is not true is because there exists $i \in E - F$ such that $h_i(\boldsymbol{\theta}_0) < 0$. With respect to the estimators, under H_{Alt}^{A} we know that $h_i(\widetilde{\boldsymbol{\theta}}(F)) = 0$ for $i \in F$, but if $i \in E - F$ then either $h_i(\widetilde{\boldsymbol{\theta}}(F)) = 0$ or $h_i(\widetilde{\boldsymbol{\theta}}(F)) < 0$ can be true. Let us consider the family of all possible subsets in $E - F$, denoted by $\mathcal{F}(E - F)$, then $S \in \mathcal{F}(E - F)$ represents $h_i(\widetilde{\boldsymbol{\theta}}(F)) = 0$ for $i \in S$ (by assumption $h_i(\widetilde{\boldsymbol{\theta}}(F)) = 0$ for $i \in F$) and $h_i(\widetilde{\boldsymbol{\theta}}(F)) < 0$ for $i \in S^C$, that is $\widetilde{\boldsymbol{\theta}}(F) = \widetilde{\boldsymbol{\theta}}(S \cup F)$. It is clear that for a sample $\widetilde{\boldsymbol{\theta}}(F) = \widetilde{\boldsymbol{\theta}}(S \cup F)$ can be true only for a unique set of indices $S \in \mathcal{F}(E - F)$, and thus by applying the Theorem of Total Probability

$$\Pr\left(S_{\phi}^{A}(\widetilde{\boldsymbol{\theta}}(F), \widehat{\boldsymbol{\theta}}(E)) \le x\right) = \sum_{S \in \mathcal{F}(E-F)} \Pr\left(S_{\phi}^{A}(\widetilde{\boldsymbol{\theta}}(F), \widehat{\boldsymbol{\theta}}(E)) \le x, \widetilde{\boldsymbol{\theta}}(F) = \widetilde{\boldsymbol{\theta}}(S \cup F)\right),$$

where $\widetilde{\boldsymbol{\theta}}(S \cup F)$ was defined in (14). From the complementary slackness condition in the Karush-Khun-Tucker Theorem (see for instance Theorem 4.2.13 in Bazaraa et al. (2006))), it holds for all $S \in \mathcal{F}(E - F)$

$$\begin{aligned}&\Pr\left(S_{\phi}^{A}(\widetilde{\boldsymbol{\theta}}(F), \widehat{\boldsymbol{\theta}}(E)) \le x, \widetilde{\boldsymbol{\theta}}(F) = \widetilde{\boldsymbol{\theta}}(S \cup F)\right) = \\ &\Pr\left(S_{\phi}^{A}(\widetilde{\boldsymbol{\theta}}(F), \widehat{\boldsymbol{\theta}}(E)) \le x, \widetilde{\boldsymbol{\lambda}}(S) > \mathbf{0}_{\mathrm{card}(S)}, h_i(\widetilde{\boldsymbol{\theta}}(S \cup F)) < 0, i \in S^C\right),\end{aligned}$$

where $S^C = E - F - S$ and $\widetilde{\boldsymbol{\lambda}}(S)$ is the subvector of the vector of Karush-Khun-Tucker multipliers $\widetilde{\boldsymbol{\lambda}}(S \cup F)$ associated with estimator $\widetilde{\boldsymbol{\theta}}(S \cup F)$ which only considers indices in S. Furthermore, under H_{Null}^{A}, $h_i(\widetilde{\boldsymbol{\theta}}(S \cup F)) = h_i(\widetilde{\boldsymbol{\theta}}(S \cup F)) - h_i(\boldsymbol{\theta}_0)$, because $h_i(\boldsymbol{\theta}_0) = 0$, $i = 1, ..., r$, hence

$$\Pr\left(S_{\phi}^{A}(\widetilde{\boldsymbol{\theta}}(F), \widehat{\boldsymbol{\theta}}(E)) \le x\right) = \sum_{S \in \mathcal{F}(E-F)} \Pr\left(S_{\phi}^{A}(\widetilde{\boldsymbol{\theta}}(F), \widehat{\boldsymbol{\theta}}(E)) \le x, \widetilde{\boldsymbol{\lambda}}(S) > \mathbf{0}_{\mathrm{card}(S)}, \boldsymbol{h}(\widetilde{\boldsymbol{\theta}}(S \cup F), S^C) - \boldsymbol{h}(\boldsymbol{\theta}_0, S^C) < \mathbf{0}_{\mathrm{card}(S^C)}\right),$$

where $\boldsymbol{h}(\boldsymbol{\theta}, S^C) = (h_i(\boldsymbol{\theta}))_{i \in S^C}$ is the subvector of $\boldsymbol{h}(\boldsymbol{\theta})$ which only considers indices in S^C. The Taylor series expansion of $S_{\phi}^{A}(\widetilde{\boldsymbol{\theta}}(F), \widehat{\boldsymbol{\theta}}(E))$ is obtained in a similar way followed for the proof of Theorem 8, and its expression is

$$S_{\phi}^{A}(\widetilde{\boldsymbol{\theta}}(F), \widehat{\boldsymbol{\theta}}(E)) = (\sqrt{n}(\widetilde{\boldsymbol{\theta}}(F) - \widehat{\boldsymbol{\theta}}(E)))^T \mathcal{I}_F(\boldsymbol{\theta}_0)(\sqrt{n}(\widetilde{\boldsymbol{\theta}}(F) - \widehat{\boldsymbol{\theta}}(E))) + o\left(\left\|\sqrt{n}(\widetilde{\boldsymbol{\theta}}(F) - \widehat{\boldsymbol{\theta}}(E))\right\|^2\right). \tag{61}$$

The first order Taylor series expansion of $\boldsymbol{h}(\boldsymbol{\theta}, S^C)$ about $\boldsymbol{\theta}_0$ taking $\boldsymbol{\theta} = \widetilde{\boldsymbol{\theta}}(S \cup F)$, leads to

$$\sqrt{n}\left(\boldsymbol{h}(\widetilde{\boldsymbol{\theta}}(S \cup F), S^C) - \boldsymbol{h}(\boldsymbol{\theta}_0, S^C)\right) = \sqrt{n}\boldsymbol{H}(\boldsymbol{\theta}_0, S^C)(\widetilde{\boldsymbol{\theta}}(S \cup F) - \boldsymbol{\theta}_0) + o\left(\left\|\sqrt{n}(\widetilde{\boldsymbol{\theta}}(S \cup F) - \boldsymbol{\theta}_0)\right\|\right), \tag{62}$$

where $\boldsymbol{H}(\boldsymbol{\theta}, S^C) = \frac{\partial}{\partial \boldsymbol{\theta}^T} \boldsymbol{h}(\boldsymbol{\theta}, S^C)$. On the other hand, from the Karush-Kuhn-Tucker Theorem it holds for $\left(\widetilde{\boldsymbol{\theta}}^T(S \cup F), \widetilde{\boldsymbol{\lambda}}^T(S)\right)^T$

$$\begin{aligned}\left.\frac{\partial}{\partial \boldsymbol{\theta}^T}\ell_n(\boldsymbol{\theta})\right|_{\boldsymbol{\theta}=\widetilde{\boldsymbol{\theta}}(S\cup F)} + \boldsymbol{H}(\widetilde{\boldsymbol{\theta}}(S \cup F), S^C)\widetilde{\boldsymbol{\lambda}}(S) &= \mathbf{0}_{\mathrm{card}(S^C)} \\ \boldsymbol{h}(\widetilde{\boldsymbol{\theta}}(S \cup F), S^C) &= \mathbf{0}_{\mathrm{card}(S^C)} \\ \widetilde{\boldsymbol{\lambda}}(S) &\ge \mathbf{0}_{\mathrm{card}(S)}\end{aligned}$$

and the first two equations are also true for $(\widehat{\boldsymbol{\theta}}^T(S \cup F), \widehat{\boldsymbol{\lambda}}^T(S))^T$ according to the Lagrange multipliers method. Hence, $\widetilde{\boldsymbol{\theta}}(S \cup F) = \widehat{\boldsymbol{\theta}}(S \cup F)$ and $\widetilde{\boldsymbol{\lambda}}(S) = \widehat{\boldsymbol{\lambda}}(S)$. From it and (60) it follows that:

- (62) leads to

$$\begin{aligned}\sqrt{n}\left(\boldsymbol{h}(\widehat{\boldsymbol{\theta}}(S \cup F), S^C) - \boldsymbol{h}(\boldsymbol{\theta}_0, S^C)\right) &= \sqrt{n}\boldsymbol{H}(\boldsymbol{\theta}_0, S^C)\boldsymbol{P}(\boldsymbol{\theta}_0, S) \left.\frac{\partial}{\partial \boldsymbol{\theta}}\ell_1(\boldsymbol{\theta})\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} + o_P(\mathbf{1}_{\mathrm{card}(S^C)}) \\ &= \boldsymbol{H}(\boldsymbol{\theta}_0, S^C)\boldsymbol{P}(\boldsymbol{\theta}_0, S)\boldsymbol{A}(\boldsymbol{\theta}_0)^T \boldsymbol{Z} + o_P(\mathbf{1}_{\mathrm{card}(S^C)}),\end{aligned}$$

where $\boldsymbol{Z} \sim \mathcal{N}(\mathbf{0}_k, \boldsymbol{I}_k)$ and

$$\begin{aligned}\boldsymbol{P}(\boldsymbol{\theta}_0, S) &= \mathcal{I}_F^{-1}(\boldsymbol{\theta}_0) + \mathcal{I}_F^{-1}(\boldsymbol{\theta}_0)\boldsymbol{H}^T(\boldsymbol{\theta}_0, S)\boldsymbol{R}(\boldsymbol{\theta}_0, S)\boldsymbol{H}(\boldsymbol{\theta}_0, S)\mathcal{I}_F^{-1}(\boldsymbol{\theta}_0),\\ \boldsymbol{R}(\boldsymbol{\theta}_0, S) &= -\left(\boldsymbol{H}(\boldsymbol{\theta}_0, S)\mathcal{I}_F^{-1}(\boldsymbol{\theta}_0)\boldsymbol{H}^T(\boldsymbol{\theta}_0, S)\right)^{-1};\end{aligned}$$

- from Sen et al. (2010, page 267)

$$\begin{aligned}\frac{1}{\sqrt{n}}\widehat{\boldsymbol{\lambda}}(S) &= \sqrt{n}\boldsymbol{Q}^T(\boldsymbol{\theta}_0, S)\left.\frac{\partial}{\partial\boldsymbol{\theta}}\ell_1(\boldsymbol{\theta})\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} + o_P(\mathbf{1}_{\mathrm{card}(S)})\\ &= \boldsymbol{Q}^T(\boldsymbol{\theta}_0, S)\mathbf{A}(\boldsymbol{\theta}_0)^T\boldsymbol{Z} + o_P(\mathbf{1}_{\mathrm{card}(S)}),\end{aligned}$$

where

$$\boldsymbol{Q}(\boldsymbol{\theta}_0, S) = \mathcal{I}_F^{-1}(\boldsymbol{\theta}_0)\boldsymbol{H}^T(\boldsymbol{\theta}_0, S)\boldsymbol{R}(\boldsymbol{\theta}_0, S);$$

- under $\widetilde{\boldsymbol{\theta}}(F) = \widehat{\boldsymbol{\theta}}(S \cup F)$ (61) leads to

$$\begin{aligned}&S_\phi^A(\widehat{\boldsymbol{\theta}}(S \cup F), \widehat{\boldsymbol{\theta}}(E)) = T_\phi^O(\widehat{\boldsymbol{\theta}}, \widehat{\boldsymbol{\theta}}(S \cup F), \widehat{\boldsymbol{\theta}}(E)) + o_P(1)\\ &= \left(\mathbf{A}(\boldsymbol{\theta}_0)\left(\boldsymbol{P}(\boldsymbol{\theta}_0, S \cup F) - \boldsymbol{P}(\boldsymbol{\theta}_0, E)\right)\mathbf{A}(\boldsymbol{\theta}_0)^T\boldsymbol{Z}\right)^T\left(\mathbf{A}(\boldsymbol{\theta}_0)\left(\boldsymbol{P}(\boldsymbol{\theta}_0, S \cup F) - \boldsymbol{P}(\boldsymbol{\theta}_0, E)\right)\mathbf{A}(\boldsymbol{\theta}_0)^T\boldsymbol{Z}\right) + o_P(1),\\ &= \boldsymbol{Z}^T\mathbf{A}(\boldsymbol{\theta}_0)\left(\boldsymbol{P}(\boldsymbol{\theta}_0, S \cup F) - \boldsymbol{P}(\boldsymbol{\theta}_0, E)\right)\mathbf{A}(\boldsymbol{\theta}_0)^T\boldsymbol{Z} + o_P(1),\end{aligned}$$

where matrix $\mathbf{A}(\boldsymbol{\theta}_0)$ is defined in the proof of Theorem 8.
That is,

$$\begin{aligned}&\lim_{n\to\infty}\Pr\left(S_\phi^A(\widetilde{\boldsymbol{\theta}}(F), \widehat{\boldsymbol{\theta}}(E)) \le x\right) = \sum_{S\in\mathcal{F}(E-F)}\Pr\left(\mathbf{Y}_3^T(S)\mathbf{Y}_3(S) \le x, \mathbf{Y}_1(S) \ge \mathbf{0}_{\mathrm{card}(S)}, \mathbf{Y}_2(S) \ge \mathbf{0}_{\mathrm{card}(S^C)}\right)\\ &= \sum_{S\in\mathcal{F}(E-F)}\Pr\left(\mathbf{Y}_3^T(S)\mathbf{Y}_3(S) \le x | \mathbf{Y}_1(S) \ge \mathbf{0}_{\mathrm{card}(S)}, \mathbf{Y}_2(S) \ge \mathbf{0}_{\mathrm{card}(S^C)}\right)\Pr\left(\mathbf{Y}_1(S) \ge \mathbf{0}_{\mathrm{card}(S)}, \mathbf{Y}_2(S) \ge \mathbf{0}_{\mathrm{card}(S^C)}\right)\\ &= \sum_{S\in\mathcal{F}(E-F)}\Pr\left(\mathbf{Y}_3^T(S)\mathbf{Y}_3(S) \le x \,\middle|\, \left(\mathbf{Y}_1^T(S), \mathbf{Y}_2^T(S)\right)^T \ge \mathbf{0}_k\right)\Pr\left(\mathbf{Y}_1(S) \ge \mathbf{0}_{\mathrm{card}(S)}, \mathbf{Y}_2(S) \ge \mathbf{0}_{\mathrm{card}(S^C)}\right),\end{aligned}$$

where

$$\begin{aligned}\mathbf{Y}_1(S) &= \boldsymbol{M}_1(\boldsymbol{\theta}_0, S)\boldsymbol{Z}, & \boldsymbol{M}_1(\boldsymbol{\theta}_0, S) &= \boldsymbol{Q}^T(\boldsymbol{\theta}_0, S)\mathbf{A}(\boldsymbol{\theta}_0)^T,\\ \mathbf{Y}_2(S) &= \boldsymbol{M}_2(\boldsymbol{\theta}_0, S)\boldsymbol{Z}, & \boldsymbol{M}_2(\boldsymbol{\theta}_0, S) &= -\boldsymbol{H}(\boldsymbol{\theta}_0, S^C)\boldsymbol{P}(\boldsymbol{\theta}_0, S)\mathbf{A}(\boldsymbol{\theta}_0)^T,\\ \mathbf{Y}_3(S) &= \boldsymbol{M}_3(\boldsymbol{\theta}_0, S)\boldsymbol{Z}, & \boldsymbol{M}_3(\boldsymbol{\theta}_0, S) &= \mathbf{A}(\boldsymbol{\theta}_0)\left(\boldsymbol{P}(\boldsymbol{\theta}_0, S \cup F) - \boldsymbol{P}(\boldsymbol{\theta}_0, E)\right)\mathbf{A}(\boldsymbol{\theta}_0)^T.\end{aligned}$$

Taking into account properties (8) it holds $\boldsymbol{M}_3(\boldsymbol{\theta}_0, S)\boldsymbol{M}_2^T(\boldsymbol{\theta}_0, S) = \boldsymbol{M}_2^T(\boldsymbol{\theta}_0, S)$ and $\boldsymbol{M}_3(\boldsymbol{\theta}_0, S)\boldsymbol{M}_1^T(\boldsymbol{\theta}_0, S) = \mathbf{0}_{k\times\mathrm{card}(S)}$, hence by applying the lemma given in Section A.2

$$\Pr\left(\mathbf{Y}_3^T(S)\mathbf{Y}_3(S) \le x \,\middle|\, \left(\mathbf{Y}_1^T(S), \mathbf{Y}_2^T(S)\right)^T \ge \mathbf{0}_k\right) = \Pr\left(\chi_{df}^2 \le x\right)$$

where

$$\begin{aligned}df &= \mathrm{rank}\left(\mathbf{A}(\boldsymbol{\theta}_0)\left(\boldsymbol{P}(\boldsymbol{\theta}_0, S \cup F) - \boldsymbol{P}(\boldsymbol{\theta}_0, E)\right)\mathbf{A}(\boldsymbol{\theta}_0)^T\right) = \mathrm{trace}\left(\mathbf{A}(\boldsymbol{\theta}_0)\left(\boldsymbol{P}(\boldsymbol{\theta}_0, S \cup F) - \boldsymbol{P}(\boldsymbol{\theta}_0, E)\right)\mathbf{A}(\boldsymbol{\theta}_0)^T\right)\\ &= \mathrm{trace}(\boldsymbol{P}(\boldsymbol{\theta}_0, S \cup F)\mathbf{A}(\boldsymbol{\theta}_0)^T\mathbf{A}(\boldsymbol{\theta}_0)) - \mathrm{trace}(\boldsymbol{P}(\boldsymbol{\theta}_0, E)\mathbf{A}(\boldsymbol{\theta}_0)^T\mathbf{A}(\boldsymbol{\theta}_0))\\ &= \mathrm{trace}(-\left(\boldsymbol{H}(\boldsymbol{\theta}_0, S \cup F)\mathcal{I}_F^{-1}(\boldsymbol{\theta}_0)\boldsymbol{H}^T(\boldsymbol{\theta}_0, S \cup F)\right)^{-1}\boldsymbol{H}(\boldsymbol{\theta}_0, S \cup F)\mathcal{I}_F^{-1}(\boldsymbol{\theta}_0)\boldsymbol{H}^T(\boldsymbol{\theta}_0, S \cup F))\\ &- \mathrm{trace}(-\left(\boldsymbol{H}(\boldsymbol{\theta}_0, E)\mathcal{I}_F^{-1}(\boldsymbol{\theta}_0)\boldsymbol{H}^T(\boldsymbol{\theta}_0, E)\right)^{-1}\boldsymbol{H}(\boldsymbol{\theta}_0, E)\mathcal{I}_F^{-1}(\boldsymbol{\theta}_0)\boldsymbol{H}^T(\boldsymbol{\theta}_0, E))\\ &= (-\mathrm{card}(S \cup F)) + \mathrm{card}(E) = r - \mathrm{card}(F) - \mathrm{card}(S).\end{aligned}$$

Finally,

$$\begin{aligned}&\lim_{n\to\infty}\Pr\left(S_{\phi}^{A}(\widetilde{\boldsymbol{\theta}}(F),\widehat{\boldsymbol{\theta}}(E))\leq x\right)\\&=\sum_{S\in\mathcal{F}(E-F)}\Pr\left(\chi^2_{r-\mathrm{card}(F)-\mathrm{card}(S)}\leq x\right)\Pr\left(\mathbf{Y}_1(S)\geq\mathbf{0}_{\mathrm{card}(S)},\mathbf{Y}_2(S)\geq\mathbf{0}_{\mathrm{card}(S^C)}\right)\\&=\sum_{j=0}^{r-\mathrm{card}(F)}\Pr\left(\chi^2_{r-\mathrm{card}(F)-j}\leq x\right)\sum_{S\in\mathcal{F}(E-F),\mathrm{card}(S)=j}\Pr\left(\mathbf{Y}_1(S)\geq\mathbf{0}_{\mathrm{card}(S)},\mathbf{Y}_2(S)\geq\mathbf{0}_{\mathrm{card}(S^C)}\right),\end{aligned}$$

and since $\boldsymbol{M}_1(\boldsymbol{\theta}_0,S)\boldsymbol{M}_2^T(\boldsymbol{\theta}_0,S)=\mathbf{0}_{\mathrm{card}(S)\times\mathrm{card}(S^C)}$ (see the second expression of (8)), $\mathbf{Y}_1(S)$ and $\mathbf{Y}_2(S)$ are independent, that is

$$\lim_{n\to\infty}\Pr\left(S_{\phi}^{A}(\widetilde{\boldsymbol{\theta}}(F),\widehat{\boldsymbol{\theta}}(E))\leq x\right)=\sum_{j=0}^{r-\mathrm{card}(F)}\Pr\left(\chi^2_{r-\mathrm{card}(F)-j}\leq x\right)w_j^A(\boldsymbol{\theta}_0)$$

where the expression of $w_j^A(\boldsymbol{\theta}_0)$ is (37) because

$$\boldsymbol{\Sigma}_1^A(\boldsymbol{\theta}_0,S)=\boldsymbol{M}_1(\boldsymbol{\theta}_0,S)\boldsymbol{M}_1^T(\boldsymbol{\theta}_0,S)=\boldsymbol{Q}^T(\boldsymbol{\theta}_0,S)\mathcal{I}_F(\boldsymbol{\theta}_0)\boldsymbol{Q}(\boldsymbol{\theta}_0,S)=-\mathbf{R}(\boldsymbol{\theta}_0,S),$$

$$\begin{aligned}\boldsymbol{\Sigma}_2^A(\boldsymbol{\theta}_0,S)&=\boldsymbol{M}_2(\boldsymbol{\theta}_0,S)\boldsymbol{M}_2^T(\boldsymbol{\theta}_0,S)=\boldsymbol{H}(\boldsymbol{\theta}_0,S^C)\boldsymbol{P}(\boldsymbol{\theta}_0,S)\mathcal{I}_F(\boldsymbol{\theta}_0)\boldsymbol{P}^T(\boldsymbol{\theta}_0,S)\boldsymbol{H}^T(\boldsymbol{\theta}_0,S^C)\\&=\boldsymbol{H}(\boldsymbol{\theta}_0,S^C)\boldsymbol{P}(\boldsymbol{\theta}_0,S)\boldsymbol{H}^T(\boldsymbol{\theta}_0,S^C).\end{aligned}$$

The proof of $T_{\phi}^{A}(\widehat{\boldsymbol{\theta}},\widetilde{\boldsymbol{\theta}}(F),\widehat{\boldsymbol{\theta}}(E))$ is omitted because it is almost immediate from the proof for $S_{\phi}^{A}(\widetilde{\boldsymbol{\theta}}(F),\widehat{\boldsymbol{\theta}}(E))$ and taking into account that for some $S\in\mathcal{F}(E-F)$

$$T_{\phi}^{A}(\widehat{\boldsymbol{\theta}},\widetilde{\boldsymbol{\theta}}(F),\widehat{\boldsymbol{\theta}}(E))=T_{\phi}^{O}(\widehat{\boldsymbol{\theta}},\widehat{\boldsymbol{\theta}}(S),\widehat{\boldsymbol{\theta}}(E))+o_P(1)=S_{\phi}^{A}(\widetilde{\boldsymbol{\theta}}(F),\widehat{\boldsymbol{\theta}}(E)).$$

A.4 Proof of Theorem 17

Under H_{Alt}^B it is conditionally established that $h_i(\boldsymbol{\theta}_0)=0$ for $i\in F$. No condition is established for $i\in E-F$ and we want to test $h_i(\boldsymbol{\theta}_0)=0$, $i\in E$ (H_{Null}^B). Hence, either $h_i(\boldsymbol{\theta}_0)=0$ or $h_i(\boldsymbol{\theta}_0)<0$ can be true for $i\in R-E$. Since $F\subset E$, it is clear that if $H_{Null}^B$is not true is because there exists $i\in R-F$ such that $h_i(\boldsymbol{\theta}_0)\neq 0$. With respect to the estimators, under H_{Null}^A we know that it holds $h_i(\widetilde{\boldsymbol{\theta}}(E))=0$ for $i\in E$, but if $i\in R-E$ then either $h_i(\widetilde{\boldsymbol{\theta}}(E))=0$ or $h_i(\widetilde{\boldsymbol{\theta}}(E))<0$ can be true. Let $S^*(\boldsymbol{\theta}_0,E)\in\mathcal{F}(R-E)$ the "unknown" set of indices such that $h_i(\boldsymbol{\theta}_0)=0$ if $i\in S^*(\boldsymbol{\theta}_0,E)$ and $h_i(\boldsymbol{\theta}_0)<0$ if $i\in R-S^*(\boldsymbol{\theta}_0,E)\cup E$. Taking into account the consistency of the MLEs, for n large enough:
a) if $S\not\subset S^*(\boldsymbol{\theta}_0,E)$ then $\widetilde{\boldsymbol{\theta}}(E)\neq\widetilde{\boldsymbol{\theta}}(S\cup E)$ with probability 1;
b) if $S\subset S^*(\boldsymbol{\theta}_0,E)$ then $\widetilde{\boldsymbol{\theta}}(E)=\widetilde{\boldsymbol{\theta}}(S\cup E)$ with probability 1.
Thus, instead of taking all of the possible subsets of $R-E$, we shall consider the family of all of the possible subsets of $S^*(\boldsymbol{\theta}_0,E)$, denoted by $\mathcal{F}(S^*(\boldsymbol{\theta}_0,E))$, where $S\in\mathcal{F}(S^*(\boldsymbol{\theta}_0,E))$ is such that $h_i(\widetilde{\boldsymbol{\theta}}(E))=0$ for $i\in S$ ($h_i(\widetilde{\boldsymbol{\theta}}(E))=0$ for $i\in E$ is true by assumption) and $h_i(\widetilde{\boldsymbol{\theta}}(E))<0$ for $i\in S^*(\boldsymbol{\theta}_0,E)-S\cup E$, or equivalently $\widetilde{\boldsymbol{\theta}}(E)=\widetilde{\boldsymbol{\theta}}(S\cup E)$. It is clear that for a particular sample of size n large enough, $\widetilde{\boldsymbol{\theta}}(E)=\widetilde{\boldsymbol{\theta}}(S\cup E)$ can be true only for a unique set of indices $S\in\mathcal{F}(S^*(\boldsymbol{\theta}_0,E))$, and thus by applying the Theorem of Total Probability

$$\lim_{n\to\infty}\Pr\left(S_{\phi}^{B}(\widehat{\boldsymbol{\theta}}(F),\widetilde{\boldsymbol{\theta}}(E))\leq x\right)=\lim_{n\to\infty}\sum_{S\in\mathcal{F}(S^*(\boldsymbol{\theta}_0,E))}\Pr\left(S_{\phi}^{B}(\widehat{\boldsymbol{\theta}}(F),\widetilde{\boldsymbol{\theta}}(E))\leq x,\widetilde{\boldsymbol{\theta}}(E)=\widetilde{\boldsymbol{\theta}}(S\cup E)\right),$$

where $\widetilde{\boldsymbol{\theta}}(S \cup E)$ was defined in (14). From the complementary slackness condition in the Karush-Khun-Tucker Theorem and a similar procedure to one followed in Theorem 12, we have

$$\lim_{n\to\infty} \Pr\left(S_{\phi}^{B}(\widehat{\boldsymbol{\theta}}(F), \widetilde{\boldsymbol{\theta}}(E)) \leq x\right) = \lim_{n\to\infty} \sum_{S\in\mathcal{F}(S^*(\boldsymbol{\theta}_0,E))} \Pr\left(S_{\phi}^{B}(\widehat{\boldsymbol{\theta}}(F), \widetilde{\boldsymbol{\theta}}(E)) \leq x, \widetilde{\boldsymbol{\lambda}}(S) \geq \mathbf{0}_{\mathrm{card}(S)}, \boldsymbol{h}(\widetilde{\boldsymbol{\theta}}(S\cup E),\right.$$
$$\left. S^*(\boldsymbol{\theta}_0, E) - S) - \boldsymbol{h}(\boldsymbol{\theta}_0, S^*(\boldsymbol{\theta}_0, E) - S) \leq \mathbf{0}_{\mathrm{card}(S^*(\boldsymbol{\theta}_0,E)-S)}\right)$$

where $\boldsymbol{h}(\boldsymbol{\theta}, S^*(\boldsymbol{\theta}_0, E)-S) = (h_i(\boldsymbol{\theta}))_{i\in S^*(\boldsymbol{\theta}_0,E)-S}$ is the subvector of $\boldsymbol{h}(\boldsymbol{\theta})$ that considers only indices in $S^*(\boldsymbol{\theta}_0, E)-S$, and $\boldsymbol{h}(\boldsymbol{\theta}_0, S^*(\boldsymbol{\theta}_0, E) - S) = \mathbf{0}_{\mathrm{card}(S^*(\boldsymbol{\theta}_0,E)-S)}$. The Taylor series expansion of $S_{\phi}^{B}(\widehat{\boldsymbol{\theta}}(F), \widetilde{\boldsymbol{\theta}}(E))$ is obtained in a similar way followed for the proof of Theorem 8, and its expression is

$$S_{\phi}^{B}(\widehat{\boldsymbol{\theta}}(F), \widetilde{\boldsymbol{\theta}}(E)) = (\sqrt{n}(\widehat{\boldsymbol{\theta}}(F) - \widetilde{\boldsymbol{\theta}}(E)))^T \mathcal{I}_F(\boldsymbol{\theta}_0)(\sqrt{n}(\widehat{\boldsymbol{\theta}}(F) - \widetilde{\boldsymbol{\theta}}(E))) + \mathrm{o}\left(\left\|\sqrt{n}(\widehat{\boldsymbol{\theta}}(F) - \widetilde{\boldsymbol{\theta}}(E))\right\|^2\right).$$

The first order Taylor series expansion of $\boldsymbol{h}^{(S)}(\boldsymbol{\theta})$ about $\boldsymbol{\theta}_0$ taking $\boldsymbol{\theta} = \widetilde{\boldsymbol{\theta}}(S \cup F)$, leads to

$$\begin{aligned}&\sqrt{n}\left(\boldsymbol{h}(\widetilde{\boldsymbol{\theta}}(S\cup E), S^*(\boldsymbol{\theta}_0, E) - S) - \boldsymbol{h}(\boldsymbol{\theta}_0, S^*(\boldsymbol{\theta}_0, E) - S)\right)\\ &= \sqrt{n}\boldsymbol{H}(\boldsymbol{\theta}_0, S^*(\boldsymbol{\theta}_0, E) - S)(\widetilde{\boldsymbol{\theta}}(S\cup E) - \boldsymbol{\theta}_0) + \mathrm{o}\left(\left\|\sqrt{n}(\widetilde{\boldsymbol{\theta}}(S\cup E) - \boldsymbol{\theta}_0)\right\|\right),\end{aligned}$$

where $\boldsymbol{H}(\boldsymbol{\theta}, S^*(\boldsymbol{\theta}_0, E) - S) = \frac{\partial}{\partial \boldsymbol{\theta}^T}\boldsymbol{h}(\boldsymbol{\theta}, S^*(\boldsymbol{\theta}_0, E) - S)$. On the other hand, from the Karush-Kuhn-Tucker Theorem it holds for $\left(\widetilde{\boldsymbol{\theta}}^T(S\cup E), \widetilde{\boldsymbol{\lambda}}^T(S)\right)^T$

$$\begin{aligned}\frac{\partial}{\partial \boldsymbol{\theta}^T}\ell_n(\boldsymbol{\theta})\Big|_{\boldsymbol{\theta}=\widetilde{\boldsymbol{\theta}}(S\cup F)} + \boldsymbol{H}^T(\widetilde{\boldsymbol{\theta}}(S\cup E), S^*(\boldsymbol{\theta}_0, E) - S)\widetilde{\boldsymbol{\lambda}}(S) &= 0\\ \boldsymbol{h}(\widetilde{\boldsymbol{\theta}}(S\cup E), S^*(\boldsymbol{\theta}_0, E) - S) &= 0\\ \widetilde{\boldsymbol{\lambda}}(S) &\geq 0\end{aligned}$$

and the first two equations are also true for $(\widehat{\boldsymbol{\theta}}^T(S\cup E), \widehat{\boldsymbol{\lambda}}^T(S))^T$ according to the Lagrange multipliers. Hence, $\widetilde{\boldsymbol{\theta}}(S\cup E) = \widehat{\boldsymbol{\theta}}(S\cup E)$ and $\widetilde{\boldsymbol{\lambda}}(S\cup E) = \widehat{\boldsymbol{\lambda}}(S\cup E)$. From it and (60) it follows that:

- (62) leads to

$$\begin{aligned}&\sqrt{n}\left(\boldsymbol{h}(\widehat{\boldsymbol{\theta}}(S\cup E), S^*(\boldsymbol{\theta}_0, E) - S) - \boldsymbol{h}(\boldsymbol{\theta}_0, S^*(\boldsymbol{\theta}_0, E) - S)\right)\\ &= \sqrt{n}\boldsymbol{H}(\boldsymbol{\theta}_0, S^*(\boldsymbol{\theta}_0, E) - S)\boldsymbol{P}(\boldsymbol{\theta}_0, S)\ \frac{\partial}{\partial \boldsymbol{\theta}}\ell_1(\boldsymbol{\theta})\Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} + \mathrm{o}_P(\mathbf{1}_{\eta(\boldsymbol{\theta}_0,E)-\mathrm{card}(S)})\\ &= \boldsymbol{H}(\boldsymbol{\theta}_0, S^*(\boldsymbol{\theta}_0, E) - S)\boldsymbol{P}(\boldsymbol{\theta}_0, S)\boldsymbol{A}(\boldsymbol{\theta}_0)^T\boldsymbol{Z} + \mathrm{o}_P(\mathbf{1}_{\eta(\boldsymbol{\theta}_0,E)-\mathrm{card}(S)}),\end{aligned}$$

where $\boldsymbol{Z} \sim \mathcal{N}(\mathbf{0}_k, \boldsymbol{I}_k)$ and

$$\begin{aligned}&\boldsymbol{P}(\boldsymbol{\theta}_0, S) = \mathcal{I}_F^{-1}(\boldsymbol{\theta}_0) + \mathcal{I}_F^{-1}(\boldsymbol{\theta}_0)\boldsymbol{H}^T(\boldsymbol{\theta}_0, S)\boldsymbol{R}(\boldsymbol{\theta}_0, S)\boldsymbol{H}(\boldsymbol{\theta}_0, S)\mathcal{I}_F^{-1}(\boldsymbol{\theta}_0),\\ &\boldsymbol{R}(\boldsymbol{\theta}_0, S) = -\left(\boldsymbol{H}(\boldsymbol{\theta}_0, S)\mathcal{I}_F^{-1}(\boldsymbol{\theta}_0)\boldsymbol{H}^T(\boldsymbol{\theta}_0, S)\right)^{-1};\end{aligned}$$

- from Sen et al. (2010, page 267)

$$\begin{aligned}\frac{1}{\sqrt{n}}\widehat{\boldsymbol{\lambda}}(S) &= \sqrt{n}\boldsymbol{Q}^T(\boldsymbol{\theta}_0, S)\ \frac{\partial}{\partial \boldsymbol{\theta}}\ell_1(\boldsymbol{\theta})\Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} + \mathrm{o}_P(\mathbf{1}_{\mathrm{card}(S)})\\ &= \boldsymbol{Q}^T(\boldsymbol{\theta}_0, S)\boldsymbol{A}(\boldsymbol{\theta}_0)^T\boldsymbol{Z} + \mathrm{o}_P(\mathbf{1}_{\mathrm{card}(S)}),\end{aligned}$$

where

$$\boldsymbol{Q}(\boldsymbol{\theta}_0, S)=\mathcal{I}_F^{-1}(\boldsymbol{\theta}_0)\boldsymbol{H}^T(\boldsymbol{\theta}_0, S)\mathbf{R}(\boldsymbol{\theta}_0, S);$$

- under $\widetilde{\boldsymbol{\theta}}(E) = \widehat{\boldsymbol{\theta}}(S \cup E)$ (61) leads to

$$\begin{aligned}
&S_\phi^B(\widehat{\boldsymbol{\theta}}(F), \widetilde{\boldsymbol{\theta}}(E)) = T_\phi^O(\widehat{\boldsymbol{\theta}}, \widehat{\boldsymbol{\theta}}(S \cup E), \widehat{\boldsymbol{\theta}}(F)) + o_P(1) \\
&= \left(\boldsymbol{A}(\boldsymbol{\theta}_0)\left(\boldsymbol{P}(\boldsymbol{\theta}_0, F) - \boldsymbol{P}(\boldsymbol{\theta}_0, S \cup E)\right)\boldsymbol{A}(\boldsymbol{\theta}_0)^T\boldsymbol{Z}\right)^T \left(\boldsymbol{A}(\boldsymbol{\theta}_0)\left(\boldsymbol{P}(\boldsymbol{\theta}_0, S \cup E) - \boldsymbol{P}(\boldsymbol{\theta}_0, F)\right)\boldsymbol{A}(\boldsymbol{\theta}_0)^T\boldsymbol{Z}\right) + o_P(1), \\
&= \boldsymbol{Z}^T\boldsymbol{A}(\boldsymbol{\theta}_0)\left(\boldsymbol{P}(\boldsymbol{\theta}_0, F) - \boldsymbol{P}(\boldsymbol{\theta}_0, S \cup E)\right)\boldsymbol{A}(\boldsymbol{\theta}_0)^T\boldsymbol{Z} + o_P(1).
\end{aligned}$$

That is,

$$\begin{aligned}
&\lim_{n\to\infty} \Pr\left(S_\phi^B(\widehat{\boldsymbol{\theta}}(F), \widetilde{\boldsymbol{\theta}}(E)) \leq x\right) \\
&= \sum_{S \in \mathcal{F}(S^*(\boldsymbol{\theta}_0, E))} \Pr\left(\boldsymbol{W}_3^T(S)\boldsymbol{W}_3(S) \leq x, \boldsymbol{W}_1(S) \geq \mathbf{0}_{\mathrm{card}(S)}, \boldsymbol{W}_2(S) \geq \mathbf{0}_{\eta(\boldsymbol{\theta}_0, E) - \mathrm{card}(S)}\right) \\
&= \sum_{S \in \mathcal{F}(S^*(\boldsymbol{\theta}_0, E))} \left(\Pr \boldsymbol{W}_3^T(S)\boldsymbol{W}_3(S) \leq x \middle| \boldsymbol{W}_1(S) \geq \mathbf{0}_{\mathrm{card}(S)}, \boldsymbol{W}_2(S) \geq \mathbf{0}_{\eta(\boldsymbol{\theta}_0, E) - \mathrm{card}(S)}\right) \\
&\times \Pr\left(\boldsymbol{W}_1(S) \geq \mathbf{0}_{\mathrm{card}(S)}, \boldsymbol{W}_2(S) \geq \mathbf{0}_{\eta(\boldsymbol{\theta}_0, E) - \mathrm{card}(S)}\right) \\
&= \sum_{S \in \mathcal{F}(S^*(\boldsymbol{\theta}_0, E))} \Pr\left(\boldsymbol{W}_3^T(S)\boldsymbol{W}_3(S) \leq x \middle| \left(\boldsymbol{W}_1^T(S), \boldsymbol{W}_2^T(S)\right)^T \geq \mathbf{0}_{\eta(\boldsymbol{\theta}_0, E)}\right) \\
&\times \Pr\left(\boldsymbol{W}_1(S) \geq \mathbf{0}_{\mathrm{card}(S)}, \boldsymbol{W}_2(S) \geq \mathbf{0}_{\eta(\boldsymbol{\theta}_0, E) - \mathrm{card}(S)}\right),
\end{aligned}$$

where

$$\begin{aligned}
\boldsymbol{W}_1(S) &= \boldsymbol{N}_1(\boldsymbol{\theta}_0, S)\boldsymbol{Z}, &\quad \boldsymbol{N}_1(\boldsymbol{\theta}_0, S) &= \boldsymbol{Q}^T(\boldsymbol{\theta}_0, S)\boldsymbol{A}(\boldsymbol{\theta}_0)^T, \\
\boldsymbol{W}_2(S) &= \boldsymbol{N}_2(\boldsymbol{\theta}_0, S)\boldsymbol{Z}, &\quad \boldsymbol{N}_2(\boldsymbol{\theta}_0, S) &= \boldsymbol{H}(\boldsymbol{\theta}_0, S^*(\boldsymbol{\theta}_0, E) - S)\boldsymbol{P}(\boldsymbol{\theta}_0, S)\boldsymbol{A}(\boldsymbol{\theta}_0)^T, \\
\boldsymbol{W}_3(S) &= \boldsymbol{N}_3(\boldsymbol{\theta}_0, S)\boldsymbol{Z}, &\quad \boldsymbol{N}_3(\boldsymbol{\theta}_0, S) &= \boldsymbol{A}(\boldsymbol{\theta}_0)\left(\boldsymbol{P}(\boldsymbol{\theta}_0, F) - \boldsymbol{P}(\boldsymbol{\theta}_0, S \cup E)\right)\boldsymbol{A}(\boldsymbol{\theta}_0)^T.
\end{aligned}$$

Taking into account properties (8) it holds $\boldsymbol{N}_3(\boldsymbol{\theta}_0, S)\boldsymbol{N}_2^T(\boldsymbol{\theta}_0, S) = \boldsymbol{N}_2^T(\boldsymbol{\theta}_0, S)$ and $\boldsymbol{N}_3(\boldsymbol{\theta}_0, S)\boldsymbol{N}_1^T(\boldsymbol{\theta}_0, S) = \mathbf{0}_{k\times\mathrm{card}(S)}$, hence by applying the lemma given in Section A.2

$$\Pr\left(\boldsymbol{W}_3^T(S)\boldsymbol{W}_3(S) \leq x \middle| \left(\boldsymbol{W}_1^T(S), \boldsymbol{W}_2^T(S)\right)^T \geq \mathbf{0}_k\right) = \Pr\left(\chi_{df}^2 \leq x\right)$$

where

$$\begin{aligned}
df &= \mathrm{rank}\left(\boldsymbol{A}(\boldsymbol{\theta}_0)\left(\boldsymbol{P}(\boldsymbol{\theta}_0, F) - \boldsymbol{P}(\boldsymbol{\theta}_0, S \cup E)\right)\boldsymbol{A}(\boldsymbol{\theta}_0)^T\right) = \mathrm{trace}\left(\boldsymbol{A}(\boldsymbol{\theta}_0)\left(\boldsymbol{P}(\boldsymbol{\theta}_0, F) - \boldsymbol{P}(\boldsymbol{\theta}_0, S \cup E)\right)\boldsymbol{A}(\boldsymbol{\theta}_0)^T\right) \\
&= \mathrm{trace}(\boldsymbol{P}(\boldsymbol{\theta}_0, F)\boldsymbol{A}(\boldsymbol{\theta}_0)^T\boldsymbol{A}(\boldsymbol{\theta}_0)) - \mathrm{trace}(\boldsymbol{P}(\boldsymbol{\theta}_0, S \cup E)\boldsymbol{A}(\boldsymbol{\theta}_0)^T\boldsymbol{A}(\boldsymbol{\theta}_0)) \\
&= \mathrm{trace}(-\left(\boldsymbol{H}(\boldsymbol{\theta}_0, F)\mathcal{I}_F^{-1}(\boldsymbol{\theta}_0)\boldsymbol{H}^T(\boldsymbol{\theta}_0, F)\right)^{-1}\boldsymbol{H}(\boldsymbol{\theta}_0, F)\mathcal{I}_F^{-1}(\boldsymbol{\theta}_0)\boldsymbol{H}^T(\boldsymbol{\theta}_0, F)) \\
&- \mathrm{trace}(-\left(\boldsymbol{H}(\boldsymbol{\theta}_0, S \cup E)\mathcal{I}_F^{-1}(\boldsymbol{\theta}_0)\boldsymbol{H}^T(\boldsymbol{\theta}_0, S)\right)^{-1}\boldsymbol{H}(\boldsymbol{\theta}_0, S \cup E)\mathcal{I}_F^{-1}(\boldsymbol{\theta}_0)\boldsymbol{H}^T(\boldsymbol{\theta}_0, S \cup E)) \\
&= -\mathrm{card}(F) + \mathrm{card}(S) + \mathrm{card}(E).
\end{aligned}$$

Finally,

$$
\begin{aligned}
&\lim_{n\to\infty}\Pr\left(S_{\phi}^{B}(\widehat{\boldsymbol{\theta}}(F),\widetilde{\boldsymbol{\theta}}(E))\leq x\right)\\
&=\sum_{S\in\mathcal{F}(S^{*}(\boldsymbol{\theta}_{0},E))}\Pr\left(\chi^{2}_{\mathrm{card}(S)+\mathrm{card}(E)-\mathrm{card}(F)}\leq x\right)\Pr\left(\boldsymbol{W}_{1}(S)\geq\boldsymbol{0}_{\mathrm{card}(S)},\boldsymbol{W}_{2}(S)\geq\boldsymbol{0}_{\eta(\boldsymbol{\theta}_{0},E)-\mathrm{card}(S)}\right)\\
&=\sum_{j=0}^{\mathrm{card}(S^{*}(\boldsymbol{\theta}_{0},E))}\Pr\left(\chi^{2}_{j+\mathrm{card}(E)-\mathrm{card}(F)}\leq x\right)\sum_{S\in\mathcal{F}(S^{*}(\boldsymbol{\theta}_{0},E)),\mathrm{card}(S)=j}\Pr\left(\boldsymbol{W}_{1}(S)\geq\boldsymbol{0}_{\mathrm{card}(S)},\boldsymbol{W}_{2}(S)\geq\boldsymbol{0}_{\eta(\boldsymbol{\theta}_{0},E)-\mathrm{card}(S)}\right),
\end{aligned}
$$

and since $\boldsymbol{N}_{1}(\boldsymbol{\theta}_{0},S)\boldsymbol{N}_{2}^{T}(\boldsymbol{\theta}_{0},S)=\boldsymbol{0}_{\mathrm{card}(S)\times\mathrm{card}(S^{*}(\boldsymbol{\theta}_{0},E)-S)}$ (see the second expression of (8)), $\boldsymbol{W}_{1}(S)$ and $\boldsymbol{W}_{2}(S)$ are independent, that is

$$
\lim_{n\to\infty}\Pr\left(S_{\phi}^{B}(\widehat{\boldsymbol{\theta}}(F),\widetilde{\boldsymbol{\theta}}(E))\leq x\right)=\sum_{j=0}^{\eta(\boldsymbol{\theta}_{0},E)}\Pr\left(\chi^{2}_{j+\mathrm{card}(E)-\mathrm{card}(F)}\leq x\right)w_{j}^{B}(\boldsymbol{\theta}_{0})
$$

where the expression of $w_{j}^{B}(\boldsymbol{\theta}_{0})$ is (40) because

$$
\boldsymbol{\Sigma}_{1}^{B}(\boldsymbol{\theta}_{0},S)=\boldsymbol{N}_{1}(\boldsymbol{\theta}_{0},S)\boldsymbol{N}_{1}^{T}(\boldsymbol{\theta}_{0},S)=\boldsymbol{Q}^{T}(\boldsymbol{\theta}_{0},S)\mathcal{I}_{F}(\boldsymbol{\theta}_{0})\boldsymbol{Q}(\boldsymbol{\theta}_{0},S)=-\boldsymbol{R}(\boldsymbol{\theta}_{0},S),
$$

$$
\begin{aligned}
\boldsymbol{\Sigma}_{2}^{B}(\boldsymbol{\theta}_{0},S)=\boldsymbol{N}_{2}(\boldsymbol{\theta}_{0},S)\boldsymbol{N}_{2}^{T}(\boldsymbol{\theta}_{0},S)&=\boldsymbol{H}(\boldsymbol{\theta}_{0},S^{*}(\boldsymbol{\theta}_{0},E)-S)\boldsymbol{P}(\boldsymbol{\theta}_{0},S)\mathcal{I}_{F}(\boldsymbol{\theta}_{0})\boldsymbol{P}^{T}(\boldsymbol{\theta}_{0},S)\boldsymbol{H}^{T}(\boldsymbol{\theta}_{0},S^{*}(\boldsymbol{\theta}_{0},E)-S)\\
&=\boldsymbol{H}(\boldsymbol{\theta}_{0},S^{*}(\boldsymbol{\theta}_{0},E)-S)\boldsymbol{P}(\boldsymbol{\theta}_{0},S)\boldsymbol{H}^{T}(\boldsymbol{\theta}_{0},S^{*}(\boldsymbol{\theta}_{0},E)-S).
\end{aligned}
$$

The proof of $T_{\phi}^{B}(\widehat{\boldsymbol{\theta}},\widehat{\boldsymbol{\theta}}(F),\widetilde{\boldsymbol{\theta}}(E))$ is omitted because it is almost immediate from the proof for $S_{\phi}^{B}(\widehat{\boldsymbol{\theta}}(F),\widetilde{\boldsymbol{\theta}}(E))$ and taking into account

$$
T_{\phi}^{B}(\widehat{\boldsymbol{\theta}},\widehat{\boldsymbol{\theta}}(F),\widetilde{\boldsymbol{\theta}}(E))=T_{\phi}^{O}(\widehat{\boldsymbol{\theta}},\widehat{\boldsymbol{\theta}}(S\cup F),\widehat{\boldsymbol{\theta}}(E))+o_{P}(1)=S_{\phi}^{B}(\widehat{\boldsymbol{\theta}}(F),\widetilde{\boldsymbol{\theta}}(E)).
$$

A.5 Power divergence based test-statistics with populations in the exponential family

Taking into account (32), the expressions of the power divergence based test-statistics for testing (10)-(11)-(12) with populations in the exponential family (25), are as follows for $\lambda\notin\{-1,0\}$

$$
\widetilde{S}_{\phi_{\lambda}}^{O}(\widehat{\boldsymbol{\theta}}(F),\widehat{\boldsymbol{\theta}}(E))=\tfrac{2}{\lambda(1+\lambda)}\left(\prod_{i=1}^{g}\mathrm{d}_{\phi_{\lambda}}^{*}(f_{\widehat{\boldsymbol{\theta}}_{i}(F)},f_{\widehat{\boldsymbol{\theta}}_{i}(E)})-1\right), \tag{67}
$$

$$
\widetilde{S}_{\phi_{\lambda}}^{A}(\widetilde{\boldsymbol{\theta}}(F),\widehat{\boldsymbol{\theta}}(E))=\tfrac{2}{\lambda(1+\lambda)}\left(\prod_{i=1}^{g}\mathrm{d}_{\phi_{\lambda}}^{*}(f_{\widetilde{\boldsymbol{\theta}}_{i}(F)},f_{\widehat{\boldsymbol{\theta}}_{i}(E)})-1\right), \tag{68}
$$

$$
\widetilde{S}_{\phi_{\lambda}}^{B}(\widehat{\boldsymbol{\theta}}(F),\widetilde{\boldsymbol{\theta}}(E))=\tfrac{2}{\lambda(1+\lambda)}\left(\prod_{i=1}^{g}\mathrm{d}_{\phi_{\lambda}}^{*}(f_{\widehat{\boldsymbol{\theta}}_{i}(F)},f_{\widetilde{\boldsymbol{\theta}}_{i}(E)})-1\right), \tag{69}
$$

$$
\widetilde{T}_{\phi}^{O}(\widehat{\boldsymbol{\theta}},\widehat{\boldsymbol{\theta}}(F),\widehat{\boldsymbol{\theta}}(E))=\tfrac{2}{\lambda(1+\lambda)}\left(\prod_{i=1}^{g}\mathrm{d}_{\phi_{\lambda}}^{*}(f_{\widehat{\boldsymbol{\theta}}_{i}},f_{\widehat{\boldsymbol{\theta}}_{i}(E)})-\prod_{i=1}^{g}\mathrm{d}_{\phi_{\lambda}}^{*}(f_{\widehat{\boldsymbol{\theta}}_{i}},f_{\widehat{\boldsymbol{\theta}}_{i}(F)})\right), \tag{70}
$$

$$
\widetilde{T}_{\phi}^{A}(\widehat{\boldsymbol{\theta}},\widetilde{\boldsymbol{\theta}}(F),\widehat{\boldsymbol{\theta}}(E))=\tfrac{2}{\lambda(1+\lambda)}\left(\prod_{i=1}^{g}\mathrm{d}_{\phi_{\lambda}}^{*}(f_{\widehat{\boldsymbol{\theta}}_{i}},f_{\widehat{\boldsymbol{\theta}}_{i}(E)})-\prod_{i=1}^{g}\mathrm{d}_{\phi_{\lambda}}^{*}(f_{\widehat{\boldsymbol{\theta}}_{i}},f_{\widetilde{\boldsymbol{\theta}}_{i}(F)})\right), \tag{71}
$$

$$
\widetilde{T}_{\phi}^{B}(\widehat{\boldsymbol{\theta}},\widehat{\boldsymbol{\theta}}(F),\widetilde{\boldsymbol{\theta}}(E))=\tfrac{2}{\lambda(1+\lambda)}\left(\prod_{i=1}^{g}\mathrm{d}_{\phi_{\lambda}}^{*}(f_{\widehat{\boldsymbol{\theta}}_{i}},f_{\widetilde{\boldsymbol{\theta}}_{i}(E)})-\prod_{i=1}^{g}\mathrm{d}_{\phi_{\lambda}}^{*}(f_{\widehat{\boldsymbol{\theta}}_{i}},f_{\widehat{\boldsymbol{\theta}}_{i}(F)})\right), \tag{72}
$$

where

$$
\mathrm{d}_{\phi_{\lambda}}^{*}(f_{\boldsymbol{\theta}},f_{\boldsymbol{\theta}_{0}})=\tfrac{q^{\lambda+1}(\boldsymbol{\theta})}{q^{\lambda}(\boldsymbol{\theta}_{0})}\int_{\mathcal{X}}r(\boldsymbol{x})\exp\left\{((\lambda+1)\boldsymbol{s}(\boldsymbol{\theta})-\lambda\boldsymbol{s}(\boldsymbol{\theta}_{0}))^{T}\,\boldsymbol{t}(\boldsymbol{x})\right\}d\boldsymbol{x},\quad\boldsymbol{x}\in\mathcal{X}. \tag{73}
$$

A.6 Weights computation for Section 6

For testing the isotonic binomial proportions we have

$$\boldsymbol{H}(\boldsymbol{\theta}_0) = (\boldsymbol{H}_1(\boldsymbol{\theta}_0), \boldsymbol{H}_2(\boldsymbol{\theta}_0), \boldsymbol{H}_3(\boldsymbol{\theta}_0), \boldsymbol{H}_4(\boldsymbol{\theta}_0)) = \begin{pmatrix} 1 & -1 & 0 & 0 \\ 0 & 1 & -1 & 0 \\ 0 & 0 & 1 & -1 \end{pmatrix},$$

$$\mathcal{I}_F^{-1}(\boldsymbol{\theta}_0) = \theta_0(1-\theta_0)\begin{pmatrix} \frac{1}{\nu_1} & 0 & 0 & 0 \\ 0 & \frac{1}{\nu_2} & 0 & 0 \\ 0 & 0 & \frac{1}{\nu_3} & 0 \\ 0 & 0 & 0 & \frac{1}{\nu_4} \end{pmatrix} = \begin{pmatrix} 5 & 0 & 0 & 0 \\ 0 & \frac{30}{7} & 0 & 0 \\ 0 & 0 & \frac{15}{4} & 0 \\ 0 & 0 & 0 & \frac{10}{3} \end{pmatrix}.$$

In particular, for testing (56), we have $F = \{2\}$, $E = \{1,2,3\}$, $\mathcal{F}(E-F) = \{\varnothing, \{1\}, \{3\}, \{1,3\}\}$,

$$w_0^A(\theta_0) = \Pr\left(\mathcal{N}\left(\mathbf{0}_2, \boldsymbol{\Sigma}_2^A(\boldsymbol{\theta}_0, \varnothing)\right) \geq \mathbf{0}_2\right) = \Pr\left(\mathcal{N}(0, \theta_0(1-\theta_0)\tfrac{85}{12})\right)\Pr\left(\mathcal{N}(0, \theta_0(1-\theta_0)\tfrac{65}{7})\right) = \frac{1}{2}\frac{1}{2} = \frac{1}{4},$$

where

$$\boldsymbol{\Sigma}_2^A(\boldsymbol{\theta}_0, \varnothing) = \theta_0(1-\theta_0)\begin{pmatrix} \frac{\nu_1+\nu_2}{\nu_1\nu_2} & 0 \\ 0 & \frac{\nu_3+\nu_4}{\nu_3\nu_4} \end{pmatrix} = \theta_0(1-\theta_0)\begin{pmatrix} \frac{65}{7} & 0 \\ 0 & \frac{85}{12} \end{pmatrix},$$

$$w_2^A(\theta_0) = \Pr\left(\mathcal{N}\left(\mathbf{0}_2, \boldsymbol{\Sigma}_1^A(\boldsymbol{\theta}_0, \{1,3\})\right) \geq \mathbf{0}_2\right) = \Pr\left(\mathcal{N}(0, \frac{1}{\theta_0(1-\theta_0)}\tfrac{12}{85})\right)\Pr\left(\mathcal{N}(0, \frac{1}{\theta_0(1-\theta_0)}\tfrac{7}{65})\right) = \frac{1}{2}\frac{1}{2} = \frac{1}{4},$$

where

$$\boldsymbol{\Sigma}_1^A(\boldsymbol{\theta}_0, \{1,3\}) = \frac{1}{\theta_0(1-\theta_0)}\begin{pmatrix} \frac{\nu_1\nu_2}{\nu_1+\nu_2} & 0 \\ 0 & \frac{\nu_3\nu_4}{\nu_3+\nu_4} \end{pmatrix} = \frac{1}{\theta_0(1-\theta_0)}\begin{pmatrix} \frac{7}{65} & 0 \\ 0 & \frac{12}{85} \end{pmatrix},$$

and

$$\begin{aligned} w_1^A(\theta_0) &= \Pr\left(\mathcal{N}(0, \boldsymbol{\Sigma}_1^A(\boldsymbol{\theta}_0, \{1\}))\right)\Pr\left(\mathcal{N}(0, \boldsymbol{\Sigma}_2^A(\boldsymbol{\theta}_0, \{1\}))\right) + \Pr\left(\mathcal{N}(0, \boldsymbol{\Sigma}_1^A(\boldsymbol{\theta}_0, \{3\}))\right)\Pr\left(\mathcal{N}(0, \boldsymbol{\Sigma}_2^A(\boldsymbol{\theta}_0, \{3\}))\right) \\ &= \frac{1}{2}\frac{1}{2} + \frac{1}{2}\frac{1}{2} = \frac{1}{2}. \end{aligned}$$

On the other hand, for testing (57), we have $F = \varnothing$, $E = \{1,2,3\}$, $\mathcal{F}(E) = \{\varnothing, \{1\}, \{2\}, \{3\}, \{1,2\}, \{1,3\}, \{2,3\}, E\}$, and $w_0^A(\theta_0) = \Pr\left(\mathcal{N}(\mathbf{0}_3, \boldsymbol{\Sigma}_2^A(\boldsymbol{\theta}_0, \varnothing)) \geq \mathbf{0}_3\right) = 0.04232627$, $w_3^A(\theta_0) = \Pr\left(\mathcal{N}(\mathbf{0}_3, \boldsymbol{\Sigma}_1^A(\boldsymbol{\theta}_0, E)) \geq \mathbf{0}_3\right) = 0.2484738$, $w_1^A(\theta_0) = \frac{1}{2}\left(\Pr\left(\mathcal{N}\left(\mathbf{0}_2, \boldsymbol{\Sigma}_2^A(\boldsymbol{\theta}_0, \{1\})\right) \geq \mathbf{0}_2\right) + \Pr(\mathcal{N}(\mathbf{0}_2, \boldsymbol{\Sigma}_2^A(\boldsymbol{\theta}_0, \{2\}) \geq \mathbf{0}_2) + \Pr\left(\mathcal{N}(\mathbf{0}_2, \boldsymbol{\Sigma}_2^A(\boldsymbol{\theta}_0, \{3\}) \geq \mathbf{0}_2\right)\right) = 0.2550019$, $w_2^A(\theta_0) = 1 - w_0^A(\theta_0) - w_1^A(\theta_0) - w_3^A(\theta_0) = 0.454198$, where

$$\boldsymbol{\Sigma}_2^A(\boldsymbol{\theta}_0, \varnothing) = \theta_0(1-\theta_0)\begin{pmatrix} \frac{\nu_1+\nu_2}{\nu_1\nu_2} & -\frac{1}{\nu_2} & 0 \\ -\frac{1}{\nu_2} & \frac{\nu_2+\nu_3}{\nu_2\nu_3} & -\frac{1}{\nu_3} \\ 0 & -\frac{1}{\nu_3} & \frac{\nu_3+\nu_4}{\nu_3\nu_4} \end{pmatrix} = \theta_0(1-\theta_0)\begin{pmatrix} \frac{65}{7} & -\frac{30}{7} & 0 \\ -\frac{30}{7} & \frac{225}{28} & -\frac{15}{4} \\ 0 & -\frac{15}{4} & \frac{85}{12} \end{pmatrix}, \tag{74}$$

$$\boldsymbol{\Sigma}_1^A(\boldsymbol{\theta}_0, E) = \frac{1}{\theta_0(1-\theta_0)}\begin{pmatrix} \frac{\nu_1(\nu_2+\nu_3+\nu_4)}{\nu_1+\nu_2+\nu_3+\nu_4} & \frac{\nu_1(\nu_3+\nu_4)}{\nu_1+\nu_2+\nu_3+\nu_4} & \frac{\nu_1\nu_4}{\nu_1+\nu_2+\nu_3+\nu_4} \\ \frac{\nu_1(\nu_3+\nu_4)}{\nu_1+\nu_2+\nu_3+\nu_4} & \frac{(\nu_3+\nu_4)(\nu_1+\nu_2)}{\nu_1+\nu_2+\nu_3+\nu_4} & \frac{\nu_4(\nu_1+\nu_2)}{\nu_1+\nu_2+\nu_3+\nu_4} \\ \frac{\nu_1\nu_4}{\nu_1+\nu_2+\nu_3+\nu_4} & \frac{\nu_4(\nu_1+\nu_2)}{\nu_1+\nu_2+\nu_3+\nu_4} & \frac{\nu_4(\nu_1+\nu_2+\nu_3)}{\nu_1+\nu_2+\nu_3+\nu_4} \end{pmatrix} = \frac{1}{\theta_0(1-\theta_0)}\begin{pmatrix} \frac{4}{25} & \frac{17}{150} & \frac{3}{50} \\ \frac{17}{150} & \frac{221}{900} & \frac{13}{100} \\ \frac{3}{50} & \frac{13}{100} & \frac{21}{100} \end{pmatrix},$$

$$\boldsymbol{\Sigma}_2^A(\boldsymbol{\theta}_0, \{1\}) = \theta_0(1-\theta_0)\begin{pmatrix} \frac{\nu_1+\nu_2+\nu_3}{\nu_3(\nu_1+\nu_2)} & -\frac{1}{\nu_3} \\ -\frac{1}{\nu_3} & \frac{\nu_3+\nu_4}{\nu_3\nu_4} \end{pmatrix} = \theta_0(1-\theta_0)\begin{pmatrix} \frac{315}{52} & -\frac{15}{4} \\ -\frac{15}{4} & \frac{85}{12} \end{pmatrix},$$

$$\boldsymbol{\Sigma}_2^A(\boldsymbol{\theta}_0, \{2\}) = \theta_0(1-\theta_0)\begin{pmatrix} \frac{\nu_1+\nu_2+\nu_3}{\nu_1(\nu_2+\nu_3)} & -\frac{1}{\nu_2+\nu_3} \\ -\frac{1}{\nu_2+\nu_3} & \frac{\nu_2+\nu_3+\nu_4}{\nu_4(\nu_2+\nu_3)} \end{pmatrix} = \theta_0(1-\theta_0)\begin{pmatrix} 7 & -2 \\ -2 & \frac{16}{3} \end{pmatrix},$$

$$\boldsymbol{\Sigma}_2^A(\boldsymbol{\theta}_0, \{3\}) = \theta_0(1-\theta_0)\begin{pmatrix} \frac{\nu_1+\nu_2}{\nu_1\nu_2} & -\frac{1}{\nu_2} \\ -\frac{1}{\nu_2} & \frac{\nu_2+\nu_3+\nu_4}{\nu_2(\nu_3+\nu_4)} \end{pmatrix} = \theta_0(1-\theta_0)\begin{pmatrix} \frac{65}{7} & -\frac{30}{7} \\ -\frac{30}{7} & \frac{720}{119} \end{pmatrix}.$$

There is available an R package called 'mvtnorm' for computing normal orthants via numerical integrals (http://CRAN.R-project.org/package=mvtnorm). Note that from property v) in page 11 we can avoid the constants ($\theta_0(1-\theta_0)$ and $\frac{1}{\theta_0(1-\theta_0)}$) for computing normal orthants.The specific R-commands and outputs for the simulation study are:

```
> library(mvtnorm)
> m <- 3
> cov <- matrix(c(65/7,-30/7,0,-30/7,225/28,-15/4,0,-15/4,85/12),nrow=3)
> pmvnorm(mean = rep(0,m), sigma=cov, lower = rep(0,m), upper = rep(Inf,m))
[1] 0.04232627
attr(,"error")
[1] 8.327814e-05
attr(,"msg")
[1] "Normal Completion"
> m <- 3
> cov <- matrix(c(4/25,17/150,3/50,17/150,221/900,13/100,3/50,13/100,21/100),nrow=3)
> pmvnorm(mean = rep(0,m), sigma=cov, lower = rep(0,m), upper = rep(Inf,m))
[1] 0.2484738
attr(,"error")
[1] 0.0001767959
attr(,"msg")
[1] "Normal Completion"
> library(mvtnorm)
> m <- 2
> cov <- matrix(c(315/52,-15/4,-15/4,85/12),nrow=2)
> pmvnorm(mean = rep(0,m), sigma=cov, lower = rep(0,m), upper = rep(Inf,m))
[1] 0.1529911
attr(,"error")
[1] 1e-15
attr(,"msg")
[1] "Normal Completion"
> m <- 2
> cov <- matrix(c(7,-2,-2,7),nrow=2)
> pmvnorm(mean = rep(0,m), sigma=cov, lower = rep(0,m), upper = rep(Inf,m))
[1] 0.2038846
attr(,"error")
[1] 1e-15
attr(,"msg")
[1] "Normal Completion"
> m <- 2
> cov <- matrix(c(65/7,-30/7,-30/7,720/119),nrow=2)
> pmvnorm(mean = rep(0,m), sigma=cov, lower = rep(0,m), upper = rep(Inf,m))
[1] 0.1531281
attr(,"error")
[1] 1e-15
attr(,"msg")
[1] "Normal Completion"
```

We can compare these values with the exact values of the weights obtained from the explicitly that are available (only for $r = 3$ at most)

$$
\begin{aligned}
w_0^A(\theta_0) &= \tfrac{1}{4\pi}\left(2\pi - \arccos\rho_{12} - \arccos\rho_{13} - \arccos\rho_{23}\right) = 0.04229179,\\
w_2^A(\theta_0) &= \tfrac{1}{2} - w_0^A(\theta_0) = 0.4577082,\\
w_1^A(\theta_0) &= \tfrac{1}{4\pi}\left(3\pi - \arccos\rho_{12\cdot 3} - \arccos\rho_{13\cdot 2} - \arccos\rho_{23\cdot 1}\right) = 0.2515227,\\
w_3^A(\theta_0) &= \tfrac{1}{2} - w_1^A(\theta_0) = 0.2484773,
\end{aligned}
$$

which depend on the marginal and conditional correlations

$$
\begin{aligned}
\rho_{12} &= \tfrac{\sigma_{12}}{\sqrt{\sigma_{11}\sigma_{22}}} = -\tfrac{4}{\sqrt{65}}, \qquad \rho_{12\cdot 3} = \tfrac{\rho_{12}-\rho_{13}\rho_{32}}{\sqrt{(1-\rho_{13}^2)(1-\rho_{32}^2)}} = -\tfrac{\sqrt{221}}{26},\\
\rho_{13} &= \tfrac{\sigma_{13}}{\sqrt{\sigma_{11}\sigma_{33}}} = 0, \qquad \rho_{13\cdot 2} = \tfrac{\rho_{13}-\rho_{12}\rho_{23}}{\sqrt{(1-\rho_{12}^2)(1-\rho_{23}^2)}} = -\tfrac{\sqrt{21}}{14},\\
\rho_{23} &= \tfrac{\sigma_{23}}{\sqrt{\sigma_{22}\sigma_{33}}} = -\tfrac{\sqrt{1785}}{85}, \qquad \rho_{23\cdot 1} = \tfrac{\rho_{23}-\rho_{21}\rho_{13}}{\sqrt{(1-\rho_{21}^2)(1-\rho_{13}^2)}} = -\tfrac{\sqrt{4641}}{119},
\end{aligned}
$$

associated with the variance-covariance matrix $\boldsymbol{\Sigma}_2^A(\boldsymbol{\theta}_0, \varnothing) = (\sigma_{ij})_{i,j\in E}$, given by (74). The simulation procedure based on generating normal multivariate random variables and counting the proportions of times that the MLE of the mean vector has exactly a specific quantity of non negative components, is also an accurate method for weights computation (see Section 3.5 in Sen and Silvapulle (2005)).

A.7 Tables of the Simulation study

s	λ	$\widetilde{S}_\lambda$	$\widetilde{T}_\lambda$	$\widetilde{\widetilde{S}}_\lambda$	$\widetilde{\widetilde{T}}_\lambda$	S_λ	T_λ
1	0	0.034	0.034	0.034	0.034	0.034	0.034
1	2/3	0.037	0.123	0.004	0.004	0.004	0.004
1	1	0.238	0.586	0.034	0.034	0.034	0.034
5	0	0.038	0.038	0.038	0.038	0.038	0.038
5	2/3	0.093	0.282	0.022	0.020	0.022	0.020
5	1	0.179	0.373	0.029	0.025	0.029	0.027
10	0	0.066	0.066	0.066	0.066	0.066	0.066
10	2/3	0.141	0.255	0.031	0.029	0.034	0.031
10	1	0.199	0.341	0.030	0.023	0.032	0.028
15	0	0.060	0.060	0.060	0.060	0.060	0.060
15	2/3	0.137	0.246	0.037	0.036	0.040	0.038
15	1	0.188	0.339	0.032	0.029	0.036	0.034
20	0	0.054	0.054	0.054	0.054	0.054	0.054
20	2/3	0.136	0.245	0.041	0.039	0.043	0.041
20	1	0.186	0.338	0.034	0.033	0.039	0.039

s	λ	$\widetilde{S}_\lambda$	$\widetilde{T}_\lambda$	$\widetilde{\widetilde{S}}_\lambda$	$\widetilde{\widetilde{T}}_\lambda$	S_λ	T_λ
1	0	0.098	0.098	0.098	0.098	0.098	0.098
1	2/3	0.110	0.223	0.018	0.018	0.018	0.018
1	1	0.419	0.731	0.098	0.098	0.098	0.098
5	0	0.245	0.245	0.245	0.245	0.245	0.245
5	2/3	0.423	0.673	0.169	0.162	0.169	0.163
5	1	0.579	0.763	0.177	0.155	0.180	0.171
10	0	0.555	0.555	0.555	0.555	0.555	0.555
10	2/3	0.726	0.811	0.375	0.371	0.384	0.382
10	1	0.799	0.871	0.338	0.320	0.373	0.354
15	0	0.675	0.675	0.675	0.675	0.675	0.675
15	2/3	0.833	0.888	0.576	0.576	0.590	0.592
15	1	0.881	0.929	0.518	0.529	0.570	0.568
20	0	0.768	0.768	0.768	0.768	0.768	0.768
20	2/3	0.901	0.933	0.708	0.705	0.717	0.716
20	1	0.932	0.961	0.659	0.667	0.705	0.705

Table 3: Simulated sizes (left) and powers (right) for scenario A (nominal size=0.05).

s	λ	$\widetilde{S}_\lambda$	$\widetilde{T}_\lambda$	$\widetilde{\widetilde{S}}_\lambda$	$\widetilde{\widetilde{T}}_\lambda$	S_λ	T_λ
1	0	0.039	0.039	0.039	0.039	0.039	0.039
1	2/3	0.095	0.283	0.022	0.020	0.022	0.020
1	1	0.180	0.372	0.029	0.025	0.029	0.027
5	0	0.049	0.049	0.049	0.049	0.049	0.049
5	2/3	0.133	0.244	0.039	0.038	0.041	0.041
5	1	0.186	0.339	0.035	0.032	0.039	0.039
10	0	0.046	0.046	0.046	0.046	0.046	0.046
10	2/3	0.140	0.245	0.043	0.040	0.044	0.043
10	1	0.185	0.345	0.038	0.036	0.043	0.042
15	0	0.046	0.046	0.046	0.046	0.046	0.046
15	2/3	0.143	0.254	0.041	0.041	0.045	0.045
15	1	0.190	0.355	0.039	0.038	0.044	0.044
20	0	0.046	0.046	0.046	0.046	0.046	0.046
20	2/3	0.142	0.247	0.043	0.042	0.045	0.046
20	1	0.183	0.347	0.040	0.039	0.046	0.046

s	λ	$\widetilde{S}_\lambda$	$\widetilde{T}_\lambda$	$\widetilde{\widetilde{S}}_\lambda$	$\widetilde{\widetilde{T}}_\lambda$	S_λ	T_λ
1	0	0.153	0.153	0.153	0.153	0.153	0.153
1	2/3	0.299	0.536	0.105	0.098	0.105	0.098
1	1	0.439	0.631	0.112	0.097	0.113	0.106
5	0	0.511	0.511	0.511	0.511	0.511	0.511
5	2/3	0.712	0.793	0.467	0.461	0.475	0.472
5	1	0.787	0.858	0.439	0.435	0.467	0.464
10	0	0.713	0.713	0.713	0.713	0.713	0.713
10	2/3	0.880	0.917	0.704	0.700	0.706	0.706
10	1	0.915	0.949	0.686	0.680	0.705	0.704
15	0	0.819	0.819	0.819	0.819	0.819	0.819
15	2/3	0.937	0.958	0.804	0.803	0.817	0.817
15	1	0.957	0.975	0.798	0.797	0.817	0.817
20	0	0.869	0.869	0.869	0.869	0.869	0.869
20	2/3	0.957	0.970	0.860	0.859	0.866	0.867
20	1	0.970	0.983	0.855	0.853	0.872	0.871

Table 4: Simulated sizes (left) and powers (right) for scenario B (nominal size=0.05).

s	λ	$\widetilde{S}_\lambda$	$\widetilde{T}_\lambda$	$\widetilde{\widetilde{S}}_\lambda$	$\widetilde{\widetilde{T}}_\lambda$	S_λ	T_λ
1	0	0.067	0.067	0.067	0.067	0.067	0.067
1	2/3	0.142	0.254	0.032	0.030	0.035	0.032
1	1	0.198	0.339	0.030	0.023	0.033	0.029
5	0	0.047	0.047	0.047	0.047	0.047	0.047
5	2/3	0.138	0.243	0.044	0.041	0.044	0.044
5	1	0.182	0.343	0.039	0.037	0.044	0.043
10	0	0.049	0.049	0.049	0.049	0.049	0.049
10	2/3	0.143	0.248	0.046	0.044	0.047	0.047
10	1	0.186	0.347	0.042	0.041	0.048	0.048
15	0	0.047	0.047	0.047	0.047	0.047	0.047
15	2/3	0.139	0.254	0.047	0.045	0.048	0.048
15	1	0.192	0.350	0.042	0.042	0.048	0.049
20	0	0.049	0.049	0.049	0.049	0.049	0.049
20	2/3	0.140	0.258	0.046	0.046	0.047	0.050
20	1	0.184	0.357	0.045	0.043	0.049	0.052

s	λ	$\widetilde{S}_\lambda$	$\widetilde{T}_\lambda$	$\widetilde{\widetilde{S}}_\lambda$	$\widetilde{\widetilde{T}}_\lambda$	S_λ	T_λ
1	0	0.344	0.344	0.344	0.344	0.344	0.344
1	2/3	0.520	0.632	0.227	0.220	0.235	0.229
1	1	0.609	0.720	0.207	0.188	0.227	0.211
5	0	0.744	0.744	0.744	0.744	0.744	0.744
5	2/3	0.897	0.931	0.735	0.732	0.737	0.737
5	1	0.930	0.960	0.717	0.711	0.737	0.736
10	0	0.889	0.889	0.889	0.889	0.889	0.889
10	2/3	0.965	0.976	0.880	0.879	0.886	0.887
10	1	0.965	0.976	0.880	0.879	0.886	0.887
15	0	0.922	0.922	0.922	0.922	0.922	0.922
15	2/3	0.976	0.984	0.919	0.918	0.922	0.923
15	1	0.985	0.991	0.910	0.911	0.924	0.925
20	0	0.911	0.911	0.911	0.911	0.911	0.911
20	2/3	0.975	0.983	0.913	0.911	0.916	0.918
20	1	0.981	0.990	0.906	0.905	0.919	0.923

Table 5: Simulated sizes (left) and powers (right) for scenario C (nominal size=0.05).