



Article

Contingency Table Analysis and Inference via Double Index Measures

Christos Meselidis *  and Alex Karagrigoriou 

Laboratory of Statistics and Data Analysis, Department of Statistics and Actuarial-Financial Mathematics, University of the Aegean, Karlovasi, GR-83200 Samos, Greece; alex.karagrigoriou@aegean.gr

* Correspondence: meselidis@aegean.gr

Abstract: In this work, we focus on a general family of measures of divergence for estimation and testing with emphasis on conditional independence in cross tabulations. For this purpose, a restricted minimum divergence estimator is used for the estimation of parameters under constraints and a new double index (dual) divergence test statistic is introduced and thoroughly examined. The associated asymptotic theory is provided and the advantages and practical implications are explored via simulation studies.

Keywords: double index divergence test statistic; multivariate data analysis; conditional independence; cross tabulations

1. Introduction

The concept of distance or divergence is known since at least the time of Pearson, who, in 1900, considered the classical goodness-of-fit (gof) problem by considering the distance between observed and expected frequencies. The problem for both discrete and discretized continuous distributions have been in the center of attention for the last 100+ years. The classical set-up is the one considered by Pearson where a hypothesized m -dimensional multinomial distribution, say $Multi(N, p_1, \dots, p_m)$ is examined as being the underlying distributional mechanism for producing a given sample of size N . The problem can be extended to examine the homogeneity (in terms of the distributional mechanisms) among two independent samples or the independence among two population characteristics. In all such problems we are dealing with cross tabulations or crosstabs (or contingency tables). Problems of such nature appear frequently in a great variety of fields including biosciences, socio-economic and political sciences, actuarial science, finance, business, accounting, and marketing. The need to establish for instance, whether the mechanisms producing two phenomena are the same or not is vital for altering economic policies, preventing socio-economic crises or enforcing the same economic or financial decisions to groups with similar underlying mechanisms (e.g., retaining the insurance premium in case of similarity or having different premiums in case of diversity). It is important to note that divergence measures play a pivotal role also in statistical inference in continuous settings. Indeed, for example, in [1] the authors investigate the multivariate normal case while in a recent work [2], the modified skew-normal-Cauchy (MSNC) distribution is considered, against normality.

Let us consider the general case of two m -dimensional multinomial distributions for which each probability depends on an s -dimensional unknown parameter, say $\theta = (\theta_1, \dots, \theta_s)^\top$. A general family of measures introduced by [3] is the d_Φ^α family defined by

$$d_\Phi^\alpha(\mathbf{p}(\theta), \mathbf{q}(\theta)) = \sum_{i=1}^m q_i(\theta)^{1+\alpha} \Phi\left(\frac{p_i(\theta)}{q_i(\theta)}\right); \alpha > 0, \Phi \in F^* \quad (1)$$



Citation: Meselidis, C.; Karagrigoriou, A. Contingency Table Analysis and Inference via Double Index Measures. *Entropy* **2022**, *24*, 477. <https://doi.org/10.3390/e24040477>

Academic Editor: Narayanaswamy Balakrishnan

Received: 20 February 2022

Accepted: 28 March 2022

Published: 29 March 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

where α is a positive indicator (index) value, $\mathbf{p}(\boldsymbol{\theta}) = (p_1(\boldsymbol{\theta}), \dots, p_m(\boldsymbol{\theta}))^\top$ and $\mathbf{q}(\boldsymbol{\theta}) = (q_1(\boldsymbol{\theta}), \dots, q_m(\boldsymbol{\theta}))^\top$, F^* is a class of functions s.t. $F^* = \{\Phi(\cdot) : \Phi(x) \text{ strictly convex, } x \in \mathbb{R}^+, \Phi(1) = \Phi'(1) = 0, \Phi''(1) \neq 0 \text{ and by convention, } \Phi(0/0) = 0 \text{ and } 0\Phi(p/0) = \lim_{x \rightarrow \infty} [\Phi(x)/x]\}$.

Note that the well known Csiszar family of measures [4] is obtained for the special case where the indicator is taken to be equal to 0 while the classical Kullback–Leibler (KL) distance [5] is obtained if the indicator α is equal to 0 and at the same time the function $\Phi(\cdot)$ is taken to be $\Phi(x) \equiv \Phi_{KL}(x) = x \log(x)$ or $x \log(x) - x + 1$.

The function

$$\Phi_\lambda(x) = \frac{1}{\lambda(\lambda + 1)} [x(x^\lambda - 1) - \lambda(x - 1)] \in F^*, \lambda \neq 0, -1$$

is associated with the Freeman–Tukey test when $\lambda = -1/2$, with the recommended Cressie and Read (CR) power divergence [6] when $\lambda = 2/3$, with the Pearson’s chi-squared divergence [7] when $\lambda = 1$ and with the classical KL distance when $\lambda \rightarrow 0$.

Finally, the function

$$\Phi_\alpha(x) \equiv (\lambda + 1)\Phi_\lambda(x)|_{\lambda=\alpha} = \frac{1}{\alpha} [x(x^\alpha - 1) - \alpha(x - 1)], \alpha \neq 0$$

produces the BHHJ or Φ_α -power divergence [8] given by

$$d_{\Phi_\alpha}^\alpha(\mathbf{p}(\boldsymbol{\theta}), \mathbf{q}(\boldsymbol{\theta})) = \sum_{i=1}^m q_i^\alpha(\boldsymbol{\theta}) \{q_i(\boldsymbol{\theta}) - p_i(\boldsymbol{\theta})\} + \frac{1}{\alpha} \sum_{i=1}^m p_i(\boldsymbol{\theta}) \{p_i^\alpha(\boldsymbol{\theta}) - q_i^\alpha(\boldsymbol{\theta})\}.$$

Assume that the underlying true distribution of an m -dimensional multinomial random variable with N experiments, is

$$\mathbf{X} = (X_1, \dots, X_m)^\top \sim \text{Multi}(N, \mathbf{p} = (p_1, \dots, p_m)^\top)$$

where \mathbf{p} is, in general, unknown, belonging to the parametric family

$$\mathcal{P} = \left\{ \mathbf{p}(\boldsymbol{\theta}) = (p_1(\boldsymbol{\theta}), \dots, p_m(\boldsymbol{\theta}))^\top : \boldsymbol{\theta} = (\theta_1, \dots, \theta_s)^\top \in \Theta \subset \mathbb{R}^s \right\}. \tag{2}$$

The sample estimate $\hat{\mathbf{p}} = (\hat{p}_1, \dots, \hat{p}_m)^\top$ of \mathbf{p} is easily obtained by $\hat{p}_i = x_i/N$ where x_i is the observed frequency for the i -th category (or class).

Divergence measures can be used for estimating purposes by minimizing the associated measure. The classical estimating technique is the one where (1) we take $\alpha = 0$ and $\Phi(x) = \Phi_{KL}(x)$. Then, the resulting KL minimization is equivalent to the classical maximization of the likelihood producing the well-known Maximum Likelihood Estimator (MLE, see ([9], Section 5.2)). In general, the minimization with respect to the parameter of interest of the divergence measure, gives rise to the corresponding minimum divergence estimator (see, e.g., [6,10,11]). For the case where constraints are involved the case associated with Csiszar’s family of measures was recently investigated [12]. For further references, please refer to [13–21].

Consider the hypothesis

$$H_0: \mathbf{p} = \mathbf{p}(\boldsymbol{\theta}_0) \text{ vs. } H_1: \mathbf{p} \neq \mathbf{p}(\boldsymbol{\theta}_0), \boldsymbol{\theta}_0 = (\theta_{01}, \dots, \theta_{0s})^\top \in \Theta \subset \mathbb{R}^s \tag{3}$$

where \mathbf{p} is the vector of the true but unknown probabilities of the underlying distribution and $\mathbf{p}(\boldsymbol{\theta}_0)$ the vector of the corresponding probabilities of the hypothesized distribution which is unknown and falls within the family of \mathcal{P} with the unknown parameters satisfying in general, certain constraints, e.g., of the form $c(\boldsymbol{\theta}) = 0$, under which the estimation of the parameter will be performed. The purpose of this work is twofold: having as a reference the divergence measure given in (1), we will first propose a general double

index divergence class of measures and make inference regarding the parameter estimators involved. Then, we proceed with the hypothesis problem with the emphasis given to the concept of conditional independence. The innovative idea proposed in this work is the duality in choosing among the members of the general class of divergences, one for estimating and one for testing purposes which may not be necessarily, the same. In that sense, we propose a double index divergence test statistic offering the greatest possible range of options, both for the strictly convex function Φ and the indicator value $\alpha > 0$.

Thus, the estimation problem can be examined considering expression (1) using a function $\Phi_2 \in F^*$ and an indicator $\alpha_2 > 0$:

$$d_{\Phi_2}^{\alpha_2}(\mathbf{p}, \mathbf{p}(\theta)) = \sum_{i=1}^m p_i^{1+\alpha_2}(\theta) \Phi_2\left(\frac{p_i}{p_i(\theta)}\right) \tag{4}$$

the minimization of which with respect to the unknown parameter, will produce the restricted minimum (Φ_2, α_2) divergence (rMD) estimator

$$\hat{\theta}_{(\Phi_2, \alpha_2)}^r = \arg \inf_{\theta \in \Theta: c(\theta)=0} d_{\Phi_2}^{\alpha_2}(\hat{\mathbf{p}}, \mathbf{p}(\theta)) \tag{5}$$

for some constraints $c(\theta) = 0$. Observe that the unknown vector of underlying probabilities has been replaced by the vector of the corresponding sample frequencies $\hat{\mathbf{p}}$. Then, the testing problem will be based on

$$d_{\Phi_1}^{\alpha_1}(\hat{\mathbf{p}}, \mathbf{p}(\hat{\theta}_{(\Phi_2, \alpha_2)}^r)) = \sum_{i=1}^m p_i^{1+\alpha_1}(\hat{\theta}_{(\Phi_2, \alpha_2)}^r) \Phi_1\left(\frac{\hat{p}_i}{p_i(\hat{\theta}_{(\Phi_2, \alpha_2)}^r)}\right) \tag{6}$$

where $\Phi_1(\cdot)$ and α_1 may be different from the corresponding quantities used for the estimation problem in (4). Finally, the duality of the proposed methodology surfaces when the testing problem is explored via the dual divergence test statistic formulated on the basis of the double- α -double- Φ divergence given by

$$d_{\Phi_1}^{\alpha_1}(\hat{\mathbf{p}}, \mathbf{p}(\hat{\theta}_{(\Phi_2, \alpha_2)}^r)) \tag{7}$$

where $\Phi_1, \Phi_2 \in F^*$ and $\alpha_1, \alpha_2 > 0$.

The remaining parts of this work are: Section 2 presents the formal definition and the asymptotic properties of the rMD estimator (rMDE). Section 3 deals with the general testing problem with the use of rMDE. The associated set up for the case of three-way contingency tables is developed in Section 4 with a simulation section emphasizing on the conditional independence of three random variables. We close this work with some conclusions.

2. Restricted Minimum (Φ, α) -Power Divergence Estimator

In what follows, we will provide the formal definition and the expansion of the rMD estimator and prove its asymptotic normality. The assumptions required for establishing the results of this section for the rMD estimator under constraints, are provided below:

Assumption 1.

(A₀) $f_1(\theta), \dots, f_v(\theta)$ are the constrained functions on the s -dimensional parameter θ , $f_k(\theta) = 0$, $k = 1, \dots, v$ and $v < s < m - 1$;

(A₁) There exists a value $\theta_0 \in \Theta$, such that $\mathbf{X} = (X_1, \dots, X_m)^\top \sim \text{Multi}(N, \mathbf{p}(\theta_0))$;

(A₂) Each constraint function $f_k(\theta)$ has continuous second partial derivatives;

(A₃) The $v \times s$ and $m \times s$ matrices

$$\mathbf{Q}(\theta_0) = \left(\frac{\partial f_k(\theta_0)}{\partial \theta_j} \right)_{\substack{k=1, \dots, v \\ j=1, \dots, s}} \quad \text{and} \quad \mathbf{J}(\theta_0) = \left(\frac{\partial p_i(\theta_0)}{\partial \theta_j} \right)_{\substack{i=1, \dots, m \\ j=1, \dots, s}}$$

are of full rank;

(A₄) $\mathbf{p}(\boldsymbol{\theta})$ has continuous second partial derivatives in a neighbourhood of $\boldsymbol{\theta}_0$;

(A₅) $\boldsymbol{\theta}_0$ satisfies the Birch regularity conditions (see Appendix A and [22]).

Definition 1. Under assumptions (A₀)–(A₃) the rMD estimator of $\boldsymbol{\theta}_0$ is any vector in Θ , such that

$$\hat{\boldsymbol{\theta}}_{(\Phi, \alpha)}^r = \arg \inf_{\{\boldsymbol{\theta} \in \Theta \subset \mathbb{R}^s: f_k(\boldsymbol{\theta})=0, k=1, \dots, \nu\}} d_{\Phi}^{\alpha}(\hat{\mathbf{p}}, \mathbf{p}(\boldsymbol{\theta})). \tag{8}$$

In order to derive the decomposition of $\hat{\boldsymbol{\theta}}_{(\Phi, \alpha)}^r$ the Implicit Function Theorem (IFT) is exploited according to which if a function has an invertible derivative at a point then itself is invertible in a neighbourhood of this point but it cannot be expressed in closed form [23].

Theorem 1. Under Assumptions (A₀)–(A₅), the rMD estimator of $\boldsymbol{\theta}_0$ is such that

$$\begin{aligned} \hat{\boldsymbol{\theta}}_{(\Phi, \alpha)}^r &= \boldsymbol{\theta}_0 + \mathbf{H}(\boldsymbol{\theta}_0) \left(\mathbf{B}(\boldsymbol{\theta}_0)^{\top} \mathbf{B}(\boldsymbol{\theta}_0) \right)^{-1} \mathbf{B}(\boldsymbol{\theta}_0)^{\top} \text{diag}(\mathbf{p}(\boldsymbol{\theta}_0)^{\alpha/2}) \times \\ &\quad \times \text{diag}(\mathbf{p}(\boldsymbol{\theta}_0)^{-1/2}) (\hat{\mathbf{p}} - \mathbf{p}(\boldsymbol{\theta}_0)) + o(\|\hat{\mathbf{p}} - \mathbf{p}(\boldsymbol{\theta}_0)\|) \end{aligned} \tag{9}$$

where $\hat{\boldsymbol{\theta}}_{(\Phi, \alpha)}^r$ is unique in a neighbourhood of $\boldsymbol{\theta}_0$ and

$$\begin{aligned} \mathbf{H}(\boldsymbol{\theta}_0) &= \mathbf{I} - \left(\mathbf{B}(\boldsymbol{\theta}_0)^{\top} \mathbf{B}(\boldsymbol{\theta}_0) \right)^{-1} \mathbf{Q}(\boldsymbol{\theta}_0)^{\top} \times \\ &\quad \times \left(\mathbf{Q}(\boldsymbol{\theta}_0) \left(\mathbf{B}(\boldsymbol{\theta}_0)^{\top} \mathbf{B}(\boldsymbol{\theta}_0) \right)^{-1} \mathbf{Q}(\boldsymbol{\theta}_0)^{\top} \right)^{-1} \mathbf{Q}(\boldsymbol{\theta}_0), \\ \mathbf{B}(\boldsymbol{\theta}_0) &= \text{diag}(\mathbf{p}(\boldsymbol{\theta}_0)^{\alpha/2}) \mathbf{A}(\boldsymbol{\theta}_0), \text{ while } \mathbf{A}(\boldsymbol{\theta}_0) = \text{diag}(\mathbf{p}(\boldsymbol{\theta}_0)^{-1/2}) \mathbf{J}(\boldsymbol{\theta}_0). \end{aligned}$$

Proof. Let V be a neighbourhood of $\boldsymbol{\theta}_0$ on which $\mathbf{p}(\cdot) : \Theta \rightarrow \mathcal{P} \subset l^m$ has continuous second partial derivatives where l^m is the interior of the unit cube of dimension m . Let

$$\mathbf{F} = (F_1, \dots, F_{\nu+s}) : l^m \times \mathbb{R}^{\nu+s} \rightarrow \mathbb{R}^{\nu+s}$$

with

$$F_j(\mathbf{p}, \boldsymbol{\lambda}, \boldsymbol{\theta}) = \begin{cases} f_j(\boldsymbol{\theta}), & j = 1, \dots, \nu \\ \frac{\partial d_{\Phi}^{\alpha}(\mathbf{p}, \mathbf{p}(\boldsymbol{\theta}))}{\partial \theta_{j-\nu}} + \sum_{k=1}^{\nu} \lambda_k \frac{\partial f_k(\boldsymbol{\theta})}{\partial \theta_{j-\nu}}, & j = \nu + 1, \dots, \nu + s. \end{cases}$$

where $(\mathbf{p}, \boldsymbol{\lambda}, \boldsymbol{\theta}) = (p_1, \dots, p_m, \lambda_1, \dots, \lambda_{\nu}, \theta_1, \dots, \theta_s)$ and $\lambda_k, k = 1, \dots, \nu$ are the coefficients of the constraints.

It holds that

$$F_j(p_1(\boldsymbol{\theta}_0), \dots, p_m(\boldsymbol{\theta}_0), 0, \dots, 0, \theta_{01}, \dots, \theta_{0s}) = 0, j = 1, \dots, \nu + s$$

and by denoting $\boldsymbol{\gamma} = (\gamma_1, \dots, \gamma_{\nu+s}) = (\lambda_1, \dots, \lambda_{\nu}, \theta_1, \dots, \theta_s)$, the matrix

$$\frac{\partial \mathbf{F}}{\partial \boldsymbol{\gamma}} = \left(\frac{\partial F_j}{\partial \gamma_k} \right)_{\substack{j=1, \dots, \nu+s \\ k=1, \dots, \nu+s}} = \begin{pmatrix} \mathbf{0}_{\nu \times \nu} & \mathbf{Q}(\boldsymbol{\theta}_0) \\ \mathbf{Q}(\boldsymbol{\theta}_0)^{\top} & \Phi''(1) \mathbf{B}(\boldsymbol{\theta})^{\top} \mathbf{B}(\boldsymbol{\theta}) \end{pmatrix}$$

is nonsingular at $(\mathbf{p}, \boldsymbol{\lambda}, \boldsymbol{\theta}) = (\mathbf{p}(\boldsymbol{\theta}_0), \boldsymbol{\gamma}_0) = (p_1(\boldsymbol{\theta}_0), \dots, p_m(\boldsymbol{\theta}_0), 0, \dots, 0, \theta_{01}, \dots, \theta_{0s})$ with $\boldsymbol{\gamma}_0 = (\mathbf{0}_{\nu}, \boldsymbol{\theta}_0)$.

Using the IFT a neighbourhood U of $(\mathbf{p}(\boldsymbol{\theta}_0), \boldsymbol{\gamma}_0)$ exists, such that $\partial \mathbf{F} / \partial \boldsymbol{\gamma}$ is nonsingular and a unique differentiable function $\boldsymbol{\gamma}^* = (\boldsymbol{\lambda}^*, \boldsymbol{\theta}^*) : A \subset l^m \rightarrow \mathbb{R}^{\nu+s}$, such that

$\mathbf{p}(\theta_0) \in A$ and $\{(\mathbf{p}, \gamma) \in U: \mathbf{F}(\mathbf{p}, \gamma) = 0\} = \{(\mathbf{p}, \gamma^*(\mathbf{p})) : \mathbf{p} \in A\}$ and $\gamma^*(\mathbf{p}(\theta_0)) = (\lambda^*(\mathbf{p}(\theta_0)), \theta^*(\mathbf{p}(\theta_0))) = \gamma_0$. By the chain rule and for $\mathbf{p} = \mathbf{p}(\theta_0)$ we obtain

$$\frac{\partial \mathbf{F}}{\partial \mathbf{p}(\theta_0)} + \frac{\partial \mathbf{F}}{\partial \gamma_0} \frac{\partial \gamma_0}{\partial \mathbf{p}(\theta_0)} = 0.$$

Then

$$\frac{\partial \theta_0}{\partial \mathbf{p}(\theta_0)} = \begin{pmatrix} \mathbf{E}(\theta_0) \\ \mathbf{W}(\theta_0) \end{pmatrix}$$

where

$$\begin{aligned} \mathbf{E}(\theta_0) &= \Phi''(1) \left(\mathbf{Q}(\theta_0) \left(\mathbf{B}(\theta_0)^\top \mathbf{B}(\theta_0) \right)^{-1} \mathbf{Q}(\theta_0)^\top \right)^{-1} \times \\ &\quad \times \mathbf{Q}(\theta_0) \left(\mathbf{B}(\theta_0)^\top \mathbf{B}(\theta_0) \right)^{-1} \mathbf{B}(\theta_0)^\top \text{diag}(\mathbf{p}(\theta_0)^{\alpha/2}) \text{diag}(\mathbf{p}(\theta_0)^{-1/2}) \end{aligned}$$

and

$$\mathbf{W}(\theta_0) = \mathbf{H}(\theta_0) \left(\mathbf{B}(\theta_0)^\top \mathbf{B}(\theta_0) \right)^{-1} \mathbf{B}(\theta_0)^\top \text{diag}(\mathbf{p}(\theta_0)^{\alpha/2}) \text{diag}(\mathbf{p}(\theta_0)^{-1/2}) \tag{10}$$

since

$$\frac{\partial \mathbf{F}}{\partial \mathbf{p}(\theta_0)} = \begin{pmatrix} \mathbf{0}_{v \times m} \\ -\Phi''(1) \mathbf{B}(\theta_0)^\top \text{diag}(\mathbf{p}(\theta_0)^{\alpha/2}) \text{diag}(\mathbf{p}(\theta_0)^{-1/2}) \end{pmatrix}.$$

Expanding $\theta^*(\mathbf{p})$ around $\mathbf{p}(\theta_0)$ and using (10) gives, for $\theta^*(\mathbf{p}(\theta_0)) = \theta_0$,

$$\begin{aligned} \theta^*(\mathbf{p}) &= \theta_0 + \mathbf{H}(\theta_0) \left(\mathbf{B}(\theta_0)^\top \mathbf{B}(\theta_0) \right)^{-1} \mathbf{B}(\theta_0)^\top \text{diag}(\mathbf{p}(\theta_0)^{\alpha/2}) \times \\ &\quad \times \text{diag}(\mathbf{p}(\theta_0)^{-1/2}) (\hat{\mathbf{p}} - \mathbf{p}(\theta_0)) + o(\|\hat{\mathbf{p}} - \mathbf{p}(\theta_0)\|). \end{aligned}$$

Since $\hat{\mathbf{p}} \xrightarrow{P} \mathbf{p}(\theta_0)$ eventually $\hat{\mathbf{p}} \in A$ and then $\gamma^*(\hat{\mathbf{p}}) = (\lambda^*(\hat{\mathbf{p}}), \theta^*(\hat{\mathbf{p}}))$ is the unique solution of the system

$$\begin{aligned} f_k(\theta) &= 0, & k &= 1, \dots, v \\ \frac{\partial d_\Phi^\alpha(\mathbf{p}, \mathbf{p}(\theta))}{\partial \theta_j} + \sum_{k=1}^v \lambda_k \frac{\partial f_k(\theta)}{\partial \theta_j} &= 0, & j &= 1, \dots, s \end{aligned}$$

and $(\hat{\mathbf{p}}, \gamma^*(\hat{\mathbf{p}})) \in U$. Hence, $\theta^*(\hat{\mathbf{p}})$ coincides with rMDE $\hat{\theta}_{(\Phi, \alpha)}^r$ given in (9). \square

The theorem below establishes the asymptotic normality of rMDE which is a straightforward extension of Theorem 2.4 [11] since by the Central Limit Theorem we know that

$$\sqrt{N}(\hat{\mathbf{p}} - \mathbf{p}(\theta_0)) \xrightarrow[N \rightarrow \infty]{L} \mathbf{N}(\mathbf{0}, \Sigma_{\mathbf{p}(\theta_0)}) \tag{11}$$

with the asymptotic variance-covariance matrix $\Sigma_{\mathbf{p}(\theta_0)}$ given by $\text{diag}(\mathbf{p}(\theta_0)) - \mathbf{p}(\theta_0)\mathbf{p}(\theta_0)^\top$.

Theorem 2. Under Assumptions (A₀)–(A₅), by (11) and for $\mathbf{W}(\theta_0)$ given in (10), the asymptotic distribution of rMDE is the s -dimensional Normal distribution given by

$$\sqrt{N}(\hat{\theta}_{(\Phi, \alpha)}^r - \theta_0) \xrightarrow[N \rightarrow \infty]{L} \mathbf{N}_s(\mathbf{0}, \mathbf{W}(\theta_0) \Sigma_{\mathbf{p}(\theta_0)} \mathbf{W}(\theta_0)^\top).$$

Remark 1. The proposed class of estimators forms a family of estimators that goes beyond the indicator α since it is easy to see that estimators obtained for the Csiszar’s φ family are given for $\alpha = 0$ in (1) and also the standard equiprobable model.

3. Statistical Inference

In this section, we introduce the double index divergence test statistic

$$T_{\Phi_1}^{\alpha_1}(\hat{\theta}_{(\Phi_2, \alpha_2)}^r) = \frac{2N}{\Phi_1''(1)} d_{\Phi_1}^{\alpha_1}(\hat{\mathbf{p}}, \mathbf{p}(\hat{\theta}_{(\Phi_2, \alpha_2)}^r)) \tag{12}$$

with $\Phi_1, \Phi_2 \in F^*$ and $\alpha_1, \alpha_2 > 0$ and make the additional assumptions by which we focus on the Csiszar’s family of measures for testing purposes (the notation φ is used for clarity) and the equiprobable model:

Assumption 2.

(A₆) $p_i = 1/m, \forall i$

(A₇) $\Phi_1 = \varphi, \alpha_1 = 0$.

The Theorem below provides the asymptotic distribution of (12) under Assumptions (A₀)–(A₇). Assumption (A₇) will be later relaxed and a general asymptotic result will be presented in the next subsection. A discussion about Assumption A₆ will also be made in the sequel.

Theorem 3. Under Assumptions (A₀)–(A₇) and for the hypothesis in (3) we have

$$T_{\varphi}^0(\hat{\theta}_{(\Phi_2, \alpha_2)}^r) = \frac{2N}{\varphi''(1)} d_{\varphi}(\hat{\mathbf{p}}, \mathbf{p}(\hat{\theta}_{(\Phi_2, \alpha_2)}^r)) \xrightarrow{N \rightarrow \infty} \chi_{m-1-s-\nu}^2$$

with $\hat{\theta}_{(\Phi_2, \alpha_2)}^r$ given in (9).

Proof. It is straightforward that

$$\mathbf{p}(\hat{\theta}_{(\Phi_2, \alpha_2)}^r) = \mathbf{p}(\theta_0) + \mathbf{J}(\theta_0)(\hat{\theta}_{(\Phi_2, \alpha_2)}^r - \theta_0) + o(\|\hat{\theta}_{(\Phi_2, \alpha_2)}^r - \theta_0\|)$$

which by Theorem 2, expression (11), and for $\mathbf{M}(\theta_0) = \mathbf{J}(\theta_0)\mathbf{W}(\theta_0)$ reduces to

$$\mathbf{p}(\hat{\theta}_{(\Phi_2, \alpha_2)}^r) - \mathbf{p}(\theta_0) = \mathbf{M}(\theta_0)(\hat{\mathbf{p}} - \mathbf{p}(\theta_0)) + o_p(N^{-1/2})$$

which implies that

$$\sqrt{N}(\mathbf{p}(\hat{\theta}_{(\Phi_2, \alpha_2)}^r) - \mathbf{p}(\theta_0)) \xrightarrow{N \rightarrow \infty} \mathbf{N}(\mathbf{0}, \mathbf{M}(\theta_0)\Sigma_{\mathbf{p}(\theta_0)}\mathbf{M}(\theta_0)^\top). \tag{13}$$

Combining the above we obtain

$$\sqrt{N} \begin{pmatrix} \hat{\mathbf{p}} - \mathbf{p}(\theta_0) \\ \mathbf{p}(\hat{\theta}_{(\Phi_2, \alpha_2)}^r) - \mathbf{p}(\theta_0) \end{pmatrix} \xrightarrow{N \rightarrow \infty} \mathbf{N} \left(\mathbf{0}, \begin{pmatrix} I \\ \mathbf{M}(\theta_0) \end{pmatrix} \Sigma_{\mathbf{p}(\theta_0)} (I, \mathbf{M}(\theta_0)^\top) \right)$$

and

$$\sqrt{N}(\hat{\mathbf{p}} - \mathbf{p}(\hat{\theta}_{(\Phi_2, \alpha_2)}^r)) \xrightarrow{N \rightarrow \infty} \mathbf{N}(\mathbf{0}, \mathbf{L}(\theta_0))$$

where

$$\mathbf{L}(\theta_0) = \Sigma_{\mathbf{p}(\theta_0)} - \mathbf{M}(\theta_0)\Sigma_{\mathbf{p}(\theta_0)} - \Sigma_{\mathbf{p}(\theta_0)}\mathbf{M}(\theta_0)^\top + \mathbf{M}(\theta_0)\Sigma_{\mathbf{p}(\theta_0)}\mathbf{M}(\theta_0)^\top. \tag{14}$$

The expansion of $d_{\varphi}(\mathbf{p}, \mathbf{q})$ around $(\mathbf{p}(\theta_0), \mathbf{p}(\theta_0))$ yields

$$T_{\varphi}^0(\hat{\theta}_{(\Phi_2, \alpha_2)}^r) = \sum_{i=1}^m \frac{N}{p_i(\theta_0)} (\hat{p}_i - p_i(\hat{\theta}_{(\Phi_2, \alpha_2)}^r))^2 + o_p(1) = \mathbf{X}^\top \mathbf{X} + o_p(1)$$

where

$$\mathbf{X} = \sqrt{N} \text{diag}(\mathbf{p}(\boldsymbol{\theta}_0)^{-1/2})(\hat{\mathbf{p}} - \mathbf{p}(\hat{\boldsymbol{\theta}}^r_{(\Phi_2, \alpha_2)})) \xrightarrow[N \rightarrow \infty]{L} \mathbf{N}(\mathbf{0}, \mathbf{T}(\boldsymbol{\theta}_0)).$$

Then, under A_7 , $\mathbf{T}(\boldsymbol{\theta}_0)$ (see (14)) is a projection matrix of rank $m - 1 - s + \nu$ since the trace of the matrices $\mathbf{A}(\boldsymbol{\theta}_0)(\mathbf{A}(\boldsymbol{\theta}_0)^\top \mathbf{A}(\boldsymbol{\theta}_0))^{-1} \mathbf{A}(\boldsymbol{\theta}_0)^\top$ and $\mathbf{A}(\boldsymbol{\theta}_0)(\mathbf{A}(\boldsymbol{\theta}_0)^\top \mathbf{A}(\boldsymbol{\theta}_0))^{-1} \mathbf{Q}(\boldsymbol{\theta}_0)^\top (\mathbf{Q}(\boldsymbol{\theta}_0)(\mathbf{A}(\boldsymbol{\theta}_0)^\top \mathbf{A}(\boldsymbol{\theta}_0))^{-1} \mathbf{Q}(\boldsymbol{\theta}_0)^\top)^{-1} \mathbf{Q}(\boldsymbol{\theta}_0) (\mathbf{A}(\boldsymbol{\theta}_0)^\top \mathbf{A}(\boldsymbol{\theta}_0))^{-1} \mathbf{A}(\boldsymbol{\theta}_0)^\top$ is equal to s and ν , respectively.

Then, the result follows from the fact (see ([24], p. 57)) that $\mathbf{X}^\top \mathbf{X}$ has a chi-squared distribution with degrees of freedom equal to the rank of the variance-covariance matrix of the random vector \mathbf{X} as long as it is a projection matrix. \square

Remark 2. Relaxation of Assumption (A_6): Arguing as in [11], when the true model is not the equiprobable the result of Theorem 3 holds true as long as $\alpha_2 = 0$ and approximately true when $\alpha_2 \rightarrow 0$.

Asymptotic Theory of the Dual Divergence Test Statistic

Having established the two main results of the work, namely the decomposition of the proposed restricted estimator (Theorem 1) together with its asymptotic properties (Theorem 2), as well as the asymptotic distribution of the associated test statistic under the class of Csiszar φ -functions (Theorem 3) we continue below extended in a natural way the results of [11] for the dual divergence test statistic. The extensions presented in this section are considered vital due to their practical impication on cross tabulations discussed in Section 4. The proofs will be omitted since both results (Theorems 4 and 5) follow along the lines of previous results (see Theorems 3.4 and 3.9 of [11]). In what follows we adopt the following notation:

$$b = m^{-\alpha_1}, p_{(1)}^{\alpha_1} = \min_{i \in \{1, \dots, m\}} p_i(\boldsymbol{\theta}_0)^{\alpha_1}, p_{(m)}^{\alpha_1} = \max_{i \in \{1, \dots, m\}} p_i(\boldsymbol{\theta}_0)^{\alpha_1}, k = m - 1 - s + \nu.$$

Theorem 4. Under Assumptions (A_0)–(A_7) we have

$$T_{\Phi_1}^{\alpha_1}(\hat{\boldsymbol{\theta}}^r_{(\Phi_2, \alpha_2)}) \xrightarrow[N \rightarrow \infty]{L} b \chi_k^2.$$

Remark 3. Consider the case where Assumption (A_6) is relaxed. Then, the asymptotic distribution of the test statistic $T_{\Phi_1}^{\alpha_1}(\hat{\boldsymbol{\theta}}^r_{(\Phi_2, \alpha_2)})$ is estimated to be approximately $b \chi_k^2$ where

$$b = \frac{p_{(1)}^{\alpha_1} + p_{(m)}^{\alpha_1}}{2} \tag{15}$$

as long as $\alpha_2 = 0$ or $\alpha_2 \rightarrow 0$. For further elaboration of this remark we refer to [11].

Remark 4. Observe that if $\alpha_1 \rightarrow 0$ then $b \rightarrow 1$ and the asymptotic distribution becomes χ_k^2 , while for α_1 away from 0 the distribution is proportional to χ_k^2 with proportionality index $b \neq 1$. However, for not equiprobable models these statements hold true as long as α_2 is close to zero.

Consider now the hypothesis with contiguous alternatives [25,26]

$$H_0: \mathbf{p} = \mathbf{p}(\boldsymbol{\theta}_0) \text{ vs. } H_{1,N}: \mathbf{p} = \mathbf{p}(\boldsymbol{\theta}_0) + \frac{\mathbf{d}}{\sqrt{N}} \tag{16}$$

where \mathbf{d} is an m -dimensional vector of known real values with components d_i satisfying the assumption $\sum_{i=1}^m d_i = 0$.

Observe that as N tends to infinity, the local contiguous alternative converges to the null hypothesis at the rate $O(N^{1/2})$. Alternatives, such as those in (16), are known as Pitman

transition alternatives or Pitman (local) alternatives or local contiguous alternatives to the null hypothesis H_0 [25].

Theorem 5. Under Assumptions (A_0) – (A_7) and for the hypothesis (16) we have

$$T_{\Phi_1}^{\alpha_1}(\hat{\theta}_{(\Phi_2, \alpha_2)}^r) \xrightarrow[N \rightarrow \infty]{L} b\chi_k^2(\xi^\top \xi)$$

which represents a non-central chi-squared distribution with k degrees of freedom and non-centrality parameter $\xi^\top \xi$ for which $\xi = \text{diag}(\mathbf{p}(\theta_0)^{-1/2})(I - \mathbf{J}(\theta_0)\mathbf{W}(\theta_0))\mathbf{d}$.

Remark 5. Observe that under Assumption (A_6) ($p_i = 1/m$) the asymptotic distribution is independent of Φ , α_1 and α_2 . As a result the associated power of the test is $\Pr(\chi_k^2(\xi^\top \xi) \geq \chi_{k,a}^2)$ where a the $100(1 - a)\%$ percentile of the distribution. If assumption A_6 is relaxed then the distribution is approximately non-central chi-squared with proportionality index $b = \frac{p_{(1)}^{\alpha_1} + p_{(m)}^{\alpha_1}}{2}$.

4. Cross Tabulations and Dual Divergence Test Statistic

In this section, we try to take advantage of the methodology proposed earlier for the analysis of cross tabulations. In particular we focus on the case of three categorical variables, say X, Y , and Z with corresponding, I, J , and K . Then, assume that the probability mass of a realization of a randomly selected subject is denoted by $p_{ijk}(\theta) = \Pr(X = i, Y = j, Z = k) > 0$, where here and in what follows $i = 1, \dots, I, j = 1, \dots, J, k = 1, \dots, K$ unless otherwise stated. The associated probability vector is given as $\mathbf{p}(\theta) = \{p_{ijk}(\theta)\}$ where

$$p_{ijk}(\theta) = \begin{cases} \theta_{ijk}, & (i, j, k) \neq (I, J, K) \\ 1 - \sum_{\substack{i=1 \\ (i,j,k) \neq (I,J,K)}}^I \sum_{j=1}^J \sum_{k=1}^K \theta_{ijk}, & (i, j, k) = (I, J, K) \end{cases}$$

and the parameter space as $\Theta = \{\theta_{ijk}, (i, j, k) \neq (I, J, K)\}$. The sample estimator of $p_{ijk}(\theta)$ is $\hat{p}_{ijk} = n_{ijk}/N$, where n_{ijk} is the frequency of the corresponding (i, j, k) cell.

In this set up the dual divergence test statistics is given as

$$T_{\Phi_1}^{\alpha_1}(\hat{\theta}_{(\Phi_2, \alpha_2)}^r) = \frac{2N}{\Phi_1''(1)} \sum_{i=1}^I \sum_{j=1}^J \sum_{k=1}^K p_{ijk}(\hat{\theta}_{(\Phi_2, \alpha_2)}^r)^{1+\alpha} \Phi_1\left(\frac{\hat{p}_{ijk}}{p_{ijk}(\hat{\theta}_{(\Phi_2, \alpha_2)}^r)}\right) \tag{17}$$

where \hat{p}_{ijk} as above and the rMD estimator as

$$\hat{\theta}_{(\Phi_2, \alpha_2)}^r = \arg \inf_{\{\theta \in \Theta \subset \mathbb{R}^s: f_k(\theta)=0, k=1, \dots, v\}} \sum_{i=1}^I \sum_{j=1}^J \sum_{k=1}^K p_{ijk}(\theta)^{1+\alpha_2} \Phi_2\left(\frac{\hat{p}_{ijk}}{p_{ijk}(\theta)}\right). \tag{18}$$

For $\alpha_1, \alpha_2 = 0$ and special cases of the functions Φ_1 and Φ_2 , classical restricted minimum divergence estimators and associated test statistics can be derived from (18) and (17), respectively. For example, for $\alpha_1, \alpha_2 = 0$, and $\Phi_1, \Phi_2 = \Phi_{KL}$ the likelihood ratio test statistic with the restricted maximum likelihood estimator ($G^2(\hat{\theta}^r)$) can be derived, while for $\Phi_1, \Phi_2 = \Phi_\lambda$ and $\lambda = 1$ we obtain the chi-squared test statistic with the restricted minimum chi-squared estimator ($X^2(\hat{\theta}_{X^2}^r)$). For $\Phi_1, \Phi_2 = \Phi_\lambda$ and $\lambda = 2/3$ the dual divergence test statistic reduces to the power divergence test statistic with the restricted minimum power divergence estimator ($CR(\hat{\theta}_{CR}^r)$) whereas for $\lambda = -1/2$ reduces to the Freeman–Tukey test statistic with the restricted minimum Freeman–Tukey estimator ($FT(\hat{\theta}_{FT}^r)$).

The hypothesis of conditional independence between X , Y , and Z is given for any triplet i, j, k by

$$H_0 : p_{ijk}(\boldsymbol{\theta}_0) = \frac{p_{i*k}(\boldsymbol{\theta}_0)p_{*jk}(\boldsymbol{\theta}_0)}{p_{**k}(\boldsymbol{\theta}_0)}, \boldsymbol{\theta}_0 \in \Theta \text{ unknown}$$

where

$$p_{i*k}(\boldsymbol{\theta}_0) = \sum_{j=1}^J p_{ijk}(\boldsymbol{\theta}_0), \quad p_{*jk}(\boldsymbol{\theta}_0) = \sum_{i=1}^I p_{ijk}(\boldsymbol{\theta}_0) \quad \text{and} \quad p_{**k}(\boldsymbol{\theta}_0) = \sum_{i=1}^I \sum_{j=1}^J p_{ijk}(\boldsymbol{\theta}_0).$$

Under the $(I-1)(J-1)K$ constrained functions

$$f_{ijk}(\boldsymbol{\theta}) = p_{11k}(\boldsymbol{\theta})p_{ijk}(\boldsymbol{\theta}) - p_{1jk}(\boldsymbol{\theta})p_{i1k}(\boldsymbol{\theta}) = 0$$

$i = 2, \dots, I, j = 2, \dots, J, k = 1, \dots, K$ the above H_0 hypothesis with $\boldsymbol{\theta}_0$ unknown, becomes

$$H_0 : \mathbf{p} = \mathbf{p}(\boldsymbol{\theta}_0), \text{ for } \boldsymbol{\theta}_0 \in \Theta_0,$$

where $\Theta_0 = \{\boldsymbol{\theta} \in \Theta : f_{ijk}(\boldsymbol{\theta}) = 0, i = 2, \dots, I, j = 2, \dots, J, k = 1, \dots, K\}$.

Remark 6. For practical purposes, the choice of the values of the indices is motivated by the work of [8] where, in an attempt to achieve a compromise between robustness and efficiency of estimators, they recommended the use of small values in the $(0, 1)$ region. In the following subsection, our analysis will reconfirm their findings since as it will be seen, values of both indices close to (0) (than to one (1)) will be found to be associated with a good performance not only in terms of estimation but also in terms of goodness of fit as it will be reflected in the size and the power of the test.

Simulation Study

In this simulation study, we use the rMD estimator and the associated dual divergence test statistic for the analysis of cross tabulations. Specifically, we are going to compare in terms of size and power classical tests with those that can be derived through the proposed methodology, for the problem of conditional independence of three random variables in contingency tables. We test the hypothesis of conditional independence for a $2 \times 2 \times 2$ contingency table, thus in this case we have $m = 8$ probabilities of the multinomial model, $s = 7$ unknown parameters to estimate and two constraint functions ($\nu = 2$) which are given by

$$f_{221}(\boldsymbol{\theta}) = \theta_{111}\theta_{221} - \theta_{121}\theta_{211} \quad \text{and} \quad f_{222}(\boldsymbol{\theta}) = \theta_{112} \left(1 - \sum_{i=1}^2 \sum_{j=1}^2 \sum_{k=1}^2 \theta_{ijk} \right) - \theta_{122}\theta_{212}.$$

$(i,j,k) \neq (2,2,2)$

For a better understanding of the behaviour of the dual divergence test statistic given in (17) we compare it with the four classical tests-of-fit mentioned earlier in Section 4, namely with the $G^2(\hat{\boldsymbol{\theta}}^r)$, $X^2(\hat{\boldsymbol{\theta}}_{X^2}^r)$, $CR(\hat{\boldsymbol{\theta}}_{CR}^r)$ and $FT(\hat{\boldsymbol{\theta}}_{FT}^r)$. The proposed test $T_{\Phi_1}^{\alpha_1}(\hat{\boldsymbol{\theta}}_{(\Phi_2, \alpha_2)}^r)$ is applied for $\Phi_1 = \Phi_{\alpha_1}$, $\Phi_2 = \Phi_{\alpha_2}$ and six different values of α_1 and α_2 , $\alpha_1, \alpha_2 = 10^{-7}, 0.01, 0.05, 0.10, 0.50$, and 1.50 . Note that, the critical values used in this simulation study, are the asymptotic critical values based on the asymptotic distribution $b\chi_2^2$ with b as in (15) for the double index family of test statistics, and the χ_2^2 for the classical test statistics. For the analysis we used 100,000 simulations and sample sizes equal to $n = 20, 25$ (small sample sizes) and $n = 40, 45$ (moderate sample sizes).

In this study, we have used the model previously considered by [27] given by

$$\begin{aligned}
 p_{111} &= \pi_{111} - \pi_{111}w & p_{211} &= \pi_{211} + \pi_{222}w - \pi_{111}w \\
 p_{112} &= \pi_{112} + \pi_{111}w - \pi_{222}w & p_{212} &= \pi_{212} + \pi_{111}w - \pi_{222}w \\
 p_{121} &= \pi_{121} + \pi_{222}w & p_{221} &= \pi_{221} + \pi_{222}w - \pi_{111}w \\
 p_{122} &= \pi_{122} + \pi_{111}w & p_{222} &= \pi_{222} - \pi_{222}w
 \end{aligned}$$

where $0 \leq w < 1$ and $\pi_{ijk} = p_{i**} \times p_{*j*} \times p_{**k}$ $i, j, k = 1, 2$ with

$$\begin{aligned}
 \pi_{111} &= 0.036254 & \pi_{112} &= 0.164994 & \pi_{121} &= 0.092809 & \pi_{122} &= 0.133645 \\
 \pi_{211} &= 0.092809 & \pi_{212} &= 0.133645 & \pi_{221} &= 0.237591 & \pi_{222} &= 0.108253.
 \end{aligned}$$

For $w = 0$ we take the model under the null hypothesis of conditional independence while for values $w \neq 0$ we take the models under the alternative hypotheses. We considered the following values of $w = 0.00, 0.30, 0.60,$ and 0.90 . Note that the larger the value of w the more we deviate from the null model. For the simulation study, we used the R software [28], while for the constrained optimization the **auglag** function from the **nloptr** package [29].

From Table 1, we can observe that in terms of size the performance of the $T_{\Phi_1}^{\alpha_1}(\hat{\theta}_{(\Phi_2, \alpha_2)}^r)$ is adequate for values of $\alpha_1, \alpha_2 \leq 0.5$ both for small and moderate sample sizes. In addition, we can see that for $\alpha_1 \leq 0.10$, $T_{\Phi_1}^{\alpha_1}(\hat{\theta}_{(\Phi_2, \alpha_2)}^r)$ appears to be liberal while for $\alpha_1 \geq 0.5$ appears to be conservative. We also note that the size becomes smaller as α_1 and α_2 increase with $\alpha_1 \geq \alpha_2$. Table 2 provides the size of the classical tests-of-fit from where we can observe that $CR(\hat{\theta}_{CR}^r)$ has the best performance among all competing tests for every sample size. In contrast, $FT(\hat{\theta}_{FT}^r)$ has the worst performance among all competing tests and appears to be very liberal. Furthermore, $X^2(\hat{\theta}_{X^2}^r)$ appears to be conservative while $G^2(\hat{\theta}^r)$ appears to be liberal. Note that for $\alpha_1 \in [0.01, 0.5]$ and $\alpha_2 \leq 0.10$, $T_{\Phi_1}^{\alpha_1}(\hat{\theta}_{(\Phi_2, \alpha_2)}^r)$ behaves better than the $G^2(\hat{\theta}^r)$ test statistic and its performance is quite close to the performance of the $X^2(\hat{\theta}_{X^2}^r)$.

Table 1. Size ($w = 0.00$) calculations (%) of the $T_{\Phi_1}^{\alpha_1}(\hat{\theta}_{(\Phi_2, \alpha_2)}^r)$ test statistic for sample sizes $n = 20, 25, 40, 45$. Sizes that satisfy Dale’s criterion are presented in bold.

α_1	α_2											
	10^{-7}	0.01	0.05	0.10	0.50	1.50	10^{-7}	0.01	0.05	0.10	0.50	1.50
	$n = 20$						$n = 25$					
10^{-7}	8.256	8.257	8.260	8.263	9.216	13.856	7.863	7.865	7.878	7.920	8.927	13.192
0.01	8.207	8.206	8.209	8.224	9.224	13.623	7.753	7.754	7.763	7.817	8.797	12.930
0.05	7.896	7.849	7.879	7.886	8.719	12.916	7.340	7.334	7.327	7.350	8.313	12.277
0.10	7.403	7.404	7.378	7.356	8.046	11.994	6.965	6.959	6.940	6.934	7.675	11.364
0.50	3.873	3.850	3.769	3.612	3.023	4.050	3.857	3.819	3.722	3.604	3.191	4.304
1.50	0.920	0.893	0.807	0.758	0.509	0.202	1.046	1.019	0.948	0.885	0.602	0.203
	$n = 40$						$n = 45$					
10^{-7}	7.016	7.016	7.027	7.055	7.887	11.362	6.858	6.858	6.870	6.908	7.732	11.099
0.01	6.933	6.933	6.940	6.957	7.778	11.183	6.760	6.760	6.770	6.805	7.601	10.941
0.05	6.590	6.589	6.580	6.593	7.342	10.505	6.427	6.422	6.415	6.426	7.153	10.340
0.10	6.246	6.239	6.228	6.222	6.794	9.758	6.082	6.070	6.053	6.043	6.612	9.586
0.50	3.854	3.832	3.762	3.661	3.367	4.362	3.813	3.789	3.716	3.635	3.331	4.269
1.50	1.172	1.160	1.115	1.066	0.760	0.383	1.183	1.170	1.119	1.068	0.773	0.437

In order to examine the closeness of the estimated (true) size to the nominal size $\alpha = 0.05$ we consider the criterion given by Dale [30]. The criterion involves the following inequality

$$|\text{logit}(1 - \hat{\alpha}_n) - \text{logit}(1 - \alpha)| \leq d \tag{19}$$

where $\text{logit}(p) = \log(p/(1 - p))$ and $\hat{\alpha}_n$ is the estimated (true) size. The estimated (true) size is considered to be close to the nominal size if (19) is satisfied with $d = 0.35$. Note that in this situation the estimated (true) size is close to the nominal one if $\hat{\alpha}_n \in [0.0357, 0.0695]$ and is presented in Tables 1 and 2 in bold. This criterion has been used previously among others by [27,31].

Regarding the proposed test we can see that for small sample sizes the estimated (true) size is close to the nominal for $\alpha_1 \in [0.10, 0.50]$ and $\alpha_2 \leq 0.10$ while for moderate sample sizes for $\alpha_1 \in [10^{-7}, 0.50]$ and $\alpha_2 \leq 0.10$. With reference to the classical tests-of-fit we can observe that the size of the $CR(\hat{\theta}_{CR}^r)$ is close to the nominal for every sample size whereas the size of $G^2(\hat{\theta}^r)$ and $X^2(\hat{\theta}_{X^2}^r)$ is close only for moderate sample sizes. Finally, we note that the estimated (true) size of $FT(\hat{\theta}_{FT}^r)$ fails to be close to the nominal both for small and moderate sample sizes.

In Tables 3–5, we provide the results regarding the power of the proposed family of test statistics for the three alternatives and sample sizes $n = 20, 25, 40, 45$, while Table 2 provides the results regarding the power of the classical tests-of-fit. The performance tends to be better as we deviate from the null model and as the sample size increases both for the classical and the proposed tests.

Table 2. Size ($w = 0.00$) and power ($w = 0.30, 0.60, 0.90$) calculations (%) for the classical tests-of-fit. Sizes that satisfy Dale’s criterion are presented in bold.

Sample size	FT	G ²	CR	X ²	FT	G ²	CR	X ²
	$w = 0.00$				$w = 0.30$			
$n = 20$	14.715	8.261	4.219	3.140	18.366	9.072	4.200	2.966
$n = 25$	13.664	7.865	4.333	3.477	19.674	9.846	4.783	3.646
$n = 40$	11.154	7.016	4.722	4.059	21.920	12.192	6.935	5.548
$n = 45$	10.787	6.858	4.703	4.082	22.467	12.992	7.471	6.081
	$w = 0.40$				$w = 0.45$			
$n = 20$	29.707	14.936	7.096	4.910	47.859	26.721	13.789	9.704
$n = 25$	35.768	18.966	9.469	7.118	62.810	38.023	20.147	15.296
$n = 40$	48.366	31.513	18.780	15.030	85.773	69.599	47.644	39.481
$n = 45$	50.821	35.381	22.367	18.217	89.108	76.685	57.000	48.451

Table 3. Power ($w = 0.30$) calculations (%) of the $T_{\Phi_1}^{\alpha_1}(\hat{\theta}_{(\Phi_2, \alpha_2)}^r)$ test statistic for sample sizes $n = 20, 25, 40, 45$.

α_1	α_2											
	10^{-7}	0.01	0.05	0.10	0.50	1.50	10^{-7}	0.01	0.05	0.10	0.50	1.50
	$n = 20$						$n = 25$					
10^{-7}	9.073	9.072	9.071	9.076	9.993	15.062	9.846	9.846	9.868	9.895	10.924	15.729
0.01	8.990	8.989	8.988	9.006	9.948	14.724	9.630	9.630	9.651	9.727	10.712	15.343
0.05	8.350	8.278	8.340	8.357	9.231	13.819	9.033	9.008	8.990	9.022	9.876	14.332
0.10	7.694	7.696	7.626	7.616	8.273	12.656	8.225	8.216	8.194	8.188	8.890	13.111
0.50	3.751	3.717	3.607	3.418	2.889	4.199	3.797	3.761	3.656	3.581	3.252	4.620
1.50	0.793	0.764	0.676	0.630	0.415	0.163	0.820	0.810	0.756	0.718	0.479	0.158
	$n = 40$						$n = 45$					
10^{-7}	12.192	12.193	12.207	12.231	13.142	17.775	12.992	12.992	13.003	13.052	14.014	18.490
0.01	11.935	11.934	11.942	11.979	12.853	17.387	12.724	12.724	12.730	12.764	13.721	18.148
0.05	11.075	11.075	11.069	11.074	11.844	16.046	11.799	11.786	11.760	11.768	12.628	16.815
0.10	10.072	10.060	10.039	10.022	10.565	14.549	10.747	10.729	10.688	10.669	11.218	15.183
0.50	4.863	4.842	4.743	4.648	4.342	5.815	5.214	5.179	5.078	4.977	4.648	6.116
1.50	0.979	0.970	0.928	0.890	0.662	0.379	1.032	1.019	0.978	0.928	0.693	0.412

Table 4. Power ($w = 0.60$) calculations (%) of the $T_{\Phi_1}^{\alpha_1}(\hat{\theta}_{(\Phi_2, \alpha_2)}^r)$ test statistic for sample sizes $n = 20, 25, 40, 45$.

α_1	α_2											
	10^{-7}	0.01	0.05	0.10	0.50	1.50	10^{-7}	0.01	0.05	0.10	0.50	1.50
	$n = 20$						$n = 25$					
10^{-7}	14.928	14.937	14.932	14.944	16.186	22.900	18.965	18.964	19.004	19.042	20.607	27.684
0.01	14.807	14.813	14.808	14.833	16.117	22.486	18.565	18.564	18.598	18.702	20.235	27.069
0.05	13.711	13.583	13.726	13.735	14.939	21.143	17.436	17.383	17.360	17.422	18.733	25.365
0.10	12.612	12.619	12.529	12.525	13.217	19.545	15.794	15.767	15.743	15.726	16.869	23.368
0.50	6.088	5.994	5.811	5.416	4.553	6.403	6.879	6.821	6.656	6.473	5.912	8.458
1.50	1.118	1.077	0.944	0.889	0.553	0.215	1.275	1.240	1.152	1.081	0.729	0.260
	$n = 40$						$n = 45$					
10^{-7}	31.513	31.518	31.533	31.608	33.469	40.799	35.381	35.381	35.404	35.465	37.411	44.556
0.01	30.904	30.903	30.925	30.999	32.868	40.221	34.848	34.845	34.863	34.941	36.744	43.942
0.05	28.949	28.946	28.938	28.956	30.509	37.756	32.727	32.716	32.697	32.715	34.310	41.510
0.10	26.504	26.485	26.434	26.398	27.631	34.747	30.146	30.110	30.051	30.014	31.289	38.456
0.50	11.949	11.867	11.598	11.409	10.830	14.703	14.052	13.966	13.632	13.321	12.731	16.901
1.50	1.797	1.761	1.692	1.578	1.142	0.716	1.973	1.945	1.870	1.776	1.295	0.838

Table 5. Power ($w = 0.90$) calculations (%) of the $T_{\Phi_1}^{\alpha_1}(\hat{\theta}_{(\Phi_2, \alpha_2)}^r)$ test statistic for sample sizes $n = 20, 25, 40, 45$.

α_1	α_2											
	10^{-7}	0.01	0.05	0.10	0.50	1.50	10^{-7}	0.01	0.05	0.10	0.50	1.50
	$n = 20$						$n = 25$					
10^{-7}	26.712	26.710	26.707	26.711	28.495	37.924	38.017	38.016	38.132	38.191	40.982	50.954
0.01	26.589	26.586	26.585	26.613	28.718	37.421	37.365	37.364	37.456	37.645	40.482	50.206
0.05	25.437	25.267	25.531	25.502	27.170	35.979	35.674	35.559	35.526	35.643	38.260	48.187
0.10	24.287	24.284	24.232	24.172	24.868	33.946	33.014	32.939	32.867	32.854	35.184	45.569
0.50	12.003	11.780	11.424	10.772	8.807	11.665	14.353	14.226	13.870	13.560	12.312	16.886
1.50	1.731	1.662	1.489	1.422	0.904	0.298	2.268	2.226	2.026	1.916	1.387	0.506
	$n = 40$						$n = 45$					
10^{-7}	69.599	69.605	69.637	69.755	72.196	79.363	76.685	76.685	76.731	76.805	78.802	84.683
0.01	68.923	68.923	68.954	69.049	71.518	79.003	76.177	76.173	76.192	76.264	78.143	84.344
0.05	66.310	66.309	66.306	66.365	68.576	77.069	73.760	73.745	73.732	73.766	75.748	82.751
0.10	62.500	62.455	62.372	62.343	64.660	74.161	70.295	70.264	70.144	70.131	72.172	80.319
0.50	30.094	29.904	29.349	28.848	27.895	36.902	36.612	36.465	35.792	35.073	34.056	43.732
1.50	3.748	3.678	3.472	3.210	2.269	1.562	4.349	4.274	4.017	3.747	2.665	1.927

As general comments regarding the behaviour of the proposed and the classical tests-of-fit in terms of power we state that the best results for the $T_{\Phi_1}^{\alpha_1}(\hat{\theta}_{(\Phi_2, \alpha_2)}^r)$ are obtained for small values of α_1 in the range $(0, 0.1]$ and large values of α_2 with $\alpha_1 \leq \alpha_2$. Note that although in terms of power results become better as α_2 increases in terms of size these are adequate only for $\alpha_2 \leq 0.5$. In addition, we can observe that the performance of $T_{\Phi_1}^{\alpha_1}(\hat{\theta}_{(\Phi_2, \alpha_2)}^r)$ is better than the $CR(\hat{\theta}_{CR}^r)$ and $X^2(\hat{\theta}_{X^2}^r)$ for every alternative and every sample size for $\alpha_1 \leq 0.1$ and $\alpha_2 \leq 0.5$ and slightly better than $G^2(\hat{\theta}^r)$ for small values of α_1 and large values of α_2 , for example for $\alpha_1 = 0.01$ and $\alpha_2 = 0.50$. Furthermore, we can observe that for $\alpha_1 = 0.1$ and $\alpha_2 \leq 0.1$ the size of the test is better than the size of the $G^2(\hat{\theta}^r)$ and slightly worst form the size of the $CR(\hat{\theta}_{CR}^r)$ and $X^2(\hat{\theta}_{X^2}^r)$ test statistics while its power is quite better than the power of the $CR(\hat{\theta}_{CR}^r)$ and $X^2(\hat{\theta}_{X^2}^r)$ and slightly worst than the $G^2(\hat{\theta}^r)$. Additionally, we can see that as α_1 and α_2 tend to 0 the behaviour of the $T_{\Phi_1}^{\alpha_1}(\hat{\theta}_{(\Phi_2, \alpha_2)}^r)$ test statistic coincides with the $G^2(\hat{\theta}^r)$ test both in terms of size and power as it was expected.

In order to attain a better insight about the behaviour of the test statistics, we apply Dale’s criterion, not only for the nominal size $\alpha = 0.05$, but also for a range of nominal sizes that are of interest. Based on the previous analysis, beside the classical tests, we will focus

our interest on the $T_{\Phi_1}^{0.05}(\hat{\theta}_{(\Phi_2,0.05)}^r)$, $T_{\Phi_1}^{0.10}(\hat{\theta}_{(\Phi_2,0.10)}^r)$, and $T_{\Phi_1}^{0.20}(\hat{\theta}_{(\Phi_2,0.20)}^r)$. The following simplified notation is used in every Figure, $FT \equiv FT(\hat{\theta}_{FT}^r)$, $ML \equiv G^2(\hat{\theta}^r)$, $CR \equiv CR(\hat{\theta}_{CR}^r)$, $Pe \equiv X^2(\hat{\theta}_{X^2}^r)$, $T1 \equiv T_{\Phi_1}^{0.05}(\hat{\theta}_{(\Phi_2,0.05)}^r)$, $T2 \equiv T_{\Phi_1}^{0.10}(\hat{\theta}_{(\Phi_2,0.10)}^r)$, and $T3 = T_{\Phi_1}^{0.20}(\hat{\theta}_{(\Phi_2,0.20)}^r)$. From Figure 1a, we can see that for small sample sizes ($n = 25$) $T_{\Phi_1}^{0.20}(\hat{\theta}_{(\Phi_2,0.20)}^r)$ and $CR(\hat{\theta}_{CR}^r)$ satisfy Dale’s criterion for every nominal size while $T_{\Phi_1}^{0.10}(\hat{\theta}_{(\Phi_2,0.10)}^r)$ and $X^2(\hat{\theta}_{X^2}^r)$ for nominal sizes greater than 0.03 and 0.06, respectively. Note that the dashed line in Figure 1 denotes the situation in which the estimated (true) size equals to the nominal size and thus lines that lie above this reference line refer to liberal tests while those that lie below to conservative ones. On the other hand, for moderate sample sizes ($n = 45$) all chosen test statistics satisfy Dale’s criterion except $FT(\hat{\theta}_{FT}^r)$.

Taking into account the fact that the actual size of each test differs from the targeted nominal size, we have to make an adjustment in order to proceed further with the comparison of the tests in terms of power. We focus our interest in those tests that satisfy Dale’s criterion and follow the method proposed in [32] which involves the so-called receiver operating characteristic (ROC) curves. In particular, let $G(t) = Pr(T \geq t)$ be the survivor function of a general test statistic T , and $c = \inf\{t : G(t) \leq \alpha\}$ be the critical value, then ROC curves can be formulated by plotting the power $G_1(c)$ against the size $G_0(c)$ for various values of the critical value c . Note that with $G_0(t)$ we denote the distribution of the test statistic under the null hypothesis and with $G_1(t)$ under the alternative.

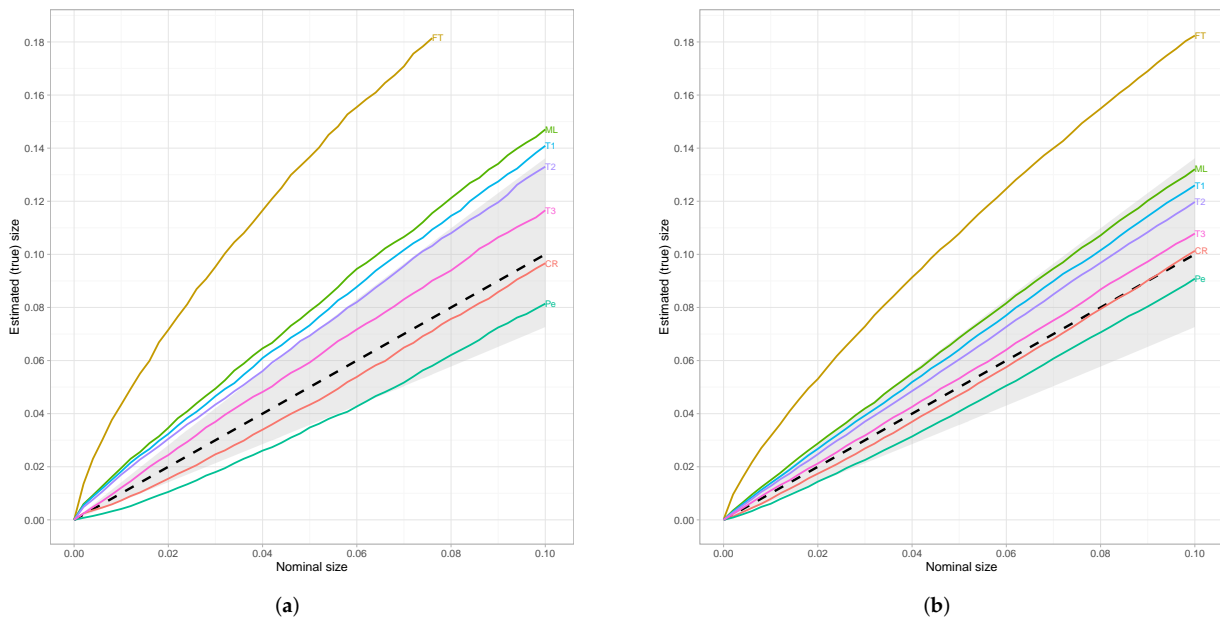


Figure 1. Estimated (true) sizes against nominal sizes. The shaded area refers to Dale’s criterion. (a) $n = 25$. (b) $n = 45$.

Since results are similar for every alternative we restrict ourselves to $w = 0.60$ which refers to an alternative that is neither too close nor too far from the null. For small sample sizes ($n = 25$) results are presented in Figure 2, where we can see that the proposed test is superior from the classical tests-of-fit in terms of power. However, for moderate sample sizes ($n = 45$) we can observe in Figure 3 that $G^2(\hat{\theta}^r)$ has the best performance among all competing tests followed by the proposed test-of-fit.

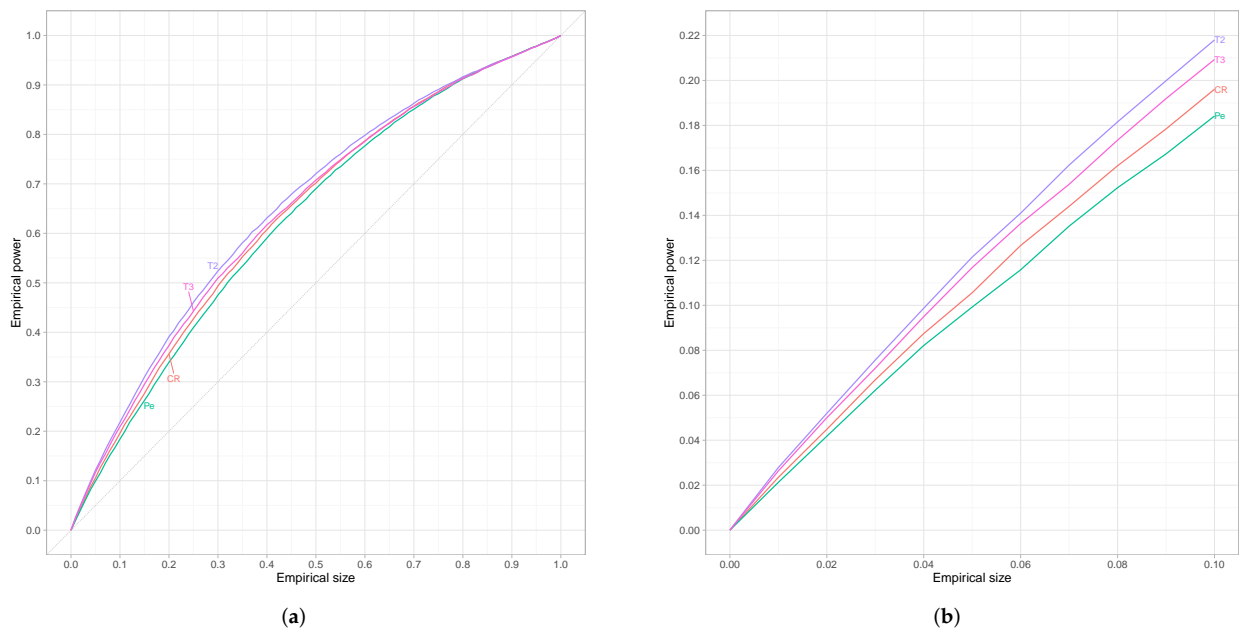


Figure 2. (a) Empirical ROC curves for $n = 25$. (b) The same curves magnified over a relevant range of empirical sizes.

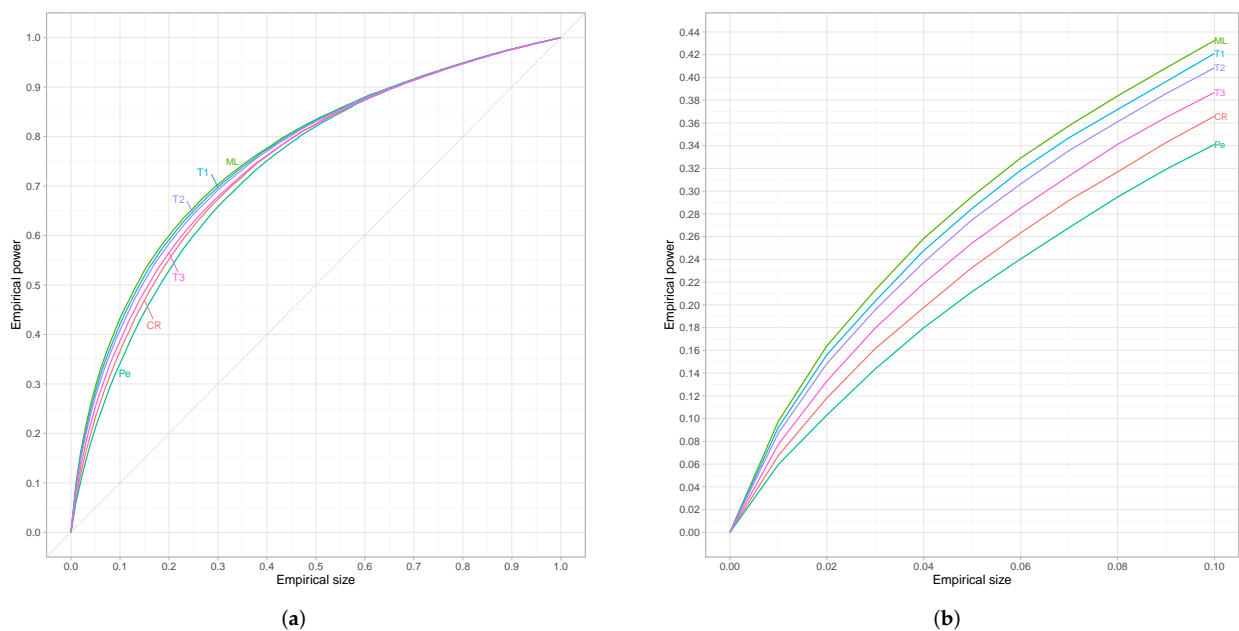


Figure 3. (a) Empirical ROC curves for $n = 45$. (b) The same curves magnified over a relevant range of empirical sizes.

From the conducted analysis we conclude that regarding the proposed test there is a trade off between size and power for different choices of the indices α_1 and α_2 . In particular, we can see that as α_1 increases the size becomes smaller in the expense of smaller power, while as α_2 increases the power becomes better and the tests more liberal. In conclusion, we could state that for values of α_1 and α_2 in the range $(0.05, 0.25)$ the resulting test statistic provides a fair balance between size and power which makes it an attractive alternative to the classical tests-of-fit where for small sample sizes larger values of the indices are preferable whereas for moderate sample sizes, smaller ones are recommended.

5. Conclusions

In this work, a general divergence family of test statistics is presented for hypothesis testing problems as in (3), under constraints. For estimating purposes, we introduce, discuss and use the rMD (restricted minimum divergence) estimator presented in (8). The proposed double index (dual) divergence test statistic involves two pairs of elements, namely (Φ_2, α_2) to be used for the estimation problem and (Φ_1, α_1) to be used for the testing problem. The duality refers to the fact that the two pairs may or may not be the same providing the researcher with the greatest possible flexibility.

The asymptotic distribution of the dual divergence test statistic is found to be proportional to the chi-squared distribution irrespectively of the nature of the multinomial model, as long as the values of the two indicators involved are relative close to zero (less than 0.5). Such values are known to provide a satisfactory balance between efficiency and robustness (see, for instance, [8] or [3]).

The methodology developed in this work can be used in the analysis of contingency tables which is applicable in various scientific fields: biosciences, such as genetics [33] and epidemiology [34]; finance, such as the evaluation of investment effectiveness or business performance [35]; insurance science [36]; or socioeconomics [37]. This work concludes with a comparative simulation study between classical test statistics and members of the proposed family, where the focus is placed on the conditional independence of three random variables. Results indicate that, by selecting wisely the values of the α_1 and α_2 indices, we can derive a test statistic that can be thought of as a powerful and reliable alternative to the classical tests-of-fit especially for small sample sizes.

Author Contributions: Conceptualization, A.K. and C.M.; data curation, C.M.; methodology, A.K. and C.M.; software, C.M.; formal analysis, A.K. and C.M.; writing—original draft preparation, C.M.; writing—review and editing, A.K. and C.M.; supervision, A.K. All authors have read and agreed to the published version of the manuscript.

Funding: The research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Acknowledgments: The authors wish to express their appreciation to the anonymous referees and the Associated Editor for their valuable comments and suggestions. The authors wish also to express their appreciation to the professor A. Batsidis of the University of Ioannina for bringing to their attention citation [31] which helped greatly the comparative analysis performed in this work. This work was completed as part of the first author PhD thesis and falls within the research activities of the Laboratory of Statistics and Data Analysis of the University of the Aegean.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A

The Birch regularity conditions mentioned in Assumption (A5) of Section 2 are stated below (for details see [22])

1. The point θ_0 is an interior point of Θ ;
2. $p_i = p_i(\theta_0) > 0$ for $i = 1, \dots, m$;
3. The mapping $\mathbf{p}(\theta) : \Theta \rightarrow \mathcal{P}$ is totally differentiable at θ_0 so that the partial derivatives of $p_i(\theta_0)$ with respect to each θ_j exist at θ_0 and $\mathbf{p}(\theta)$ has a linear approximation at θ_0 given by

$$p_i(\theta) = p_i(\theta_0) + \sum_{j=1}^s (\theta_j - \theta_{0j}) \frac{\partial p_i(\theta_0)}{\partial \theta_j} + o(\|\theta - \theta_0\|), \quad i = 1, \dots, m$$

as $\theta \rightarrow \theta_0$.

4. The Jacobian matrix

$$J(\theta_0) = \left(\frac{\partial \mathbf{p}(\theta)}{\partial \theta} \right)_{\theta=\theta_0} = \left(\frac{\partial p_i(\theta_0)}{\partial \theta_j} \right)_{\substack{i=1,\dots,m \\ j=1,\dots,s}}$$

is of full rank;

5. The mapping inverse to $\theta \rightarrow \mathbf{p}(\theta)$ exists and is continuous at θ_0 ;
6. The mapping $\mathbf{p}: \Theta \rightarrow \mathcal{P}$ is continuous at every point $\theta \in \Theta$.

References

1. Salicrú, M.; Morales, D.; Menéndez, M.; Pardo, L. On the Applications of Divergence Type Measures in Testing Statistical Hypotheses. *J. Multivar. Anal.* **1994**, *51*, 372–391. [\[CrossRef\]](#)
2. Contreras-Reyes, J.E.; Kahrari, F.; Cortés, D.D. On the modified skew-normal-Cauchy distribution: Properties, inference and applications. *Commun. Stat. Theory Methods* **2021**, *50*, 3615–3631. [\[CrossRef\]](#)
3. Mattheou, K.; Karagrigoriou, A. A New Family of Divergence Measures for Tests of Fit. *Aust. N. Z. J. Stat.* **2010**, *52*, 187–200. [\[CrossRef\]](#)
4. Csizsár, I. Eine Informationstheoretische Ungleichung und Ihre Anwendung auf Beweis der Ergodizität von Markoffschen Ketten. *Magyar Tud. Akad. Mat. Kut. Int. Koezl.* **1963**, *8*, 85–108.
5. Kullback, S.; Leibler, R.A. On Information and Sufficiency. *Ann. Math. Stat.* **1951**, *22*, 79–86. [\[CrossRef\]](#)
6. Cressie, N.; Read, T.R.C. Multinomial Goodness-of-Fit Tests. *J. R. Stat. Soc. Ser. B Methodol.* **1984**, *46*, 440–464. [\[CrossRef\]](#)
7. Pearson, K. On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling. *Lond. Edinb. Dublin Philos. Mag. J. Sci.* **1900**, *50*, 157–175. [\[CrossRef\]](#)
8. Basu, A.; Harris, I.R.; Hjort, N.L.; Jones, M.C. Robust and Efficient Estimation by Minimising a Density Power Divergence. *Biometrika* **1998**, *85*, 549–559. [\[CrossRef\]](#)
9. Pardo, L. *Statistical Inference Based on Divergence Measures*; Chapman and Hall/CRC: New York, NY, USA, 2006.
10. Morales, D.; Pardo, L.; Vajda, I. Asymptotic Divergence of Estimates of Discrete Distributions. *J. Stat. Plan. Inference* **1995**, *48*, 347–369. [\[CrossRef\]](#)
11. Meselidis, C.; Karagrigoriou, A. Statistical Inference for Multinomial Populations Based on a Double Index Family of Test Statistics. *J. Stat. Comput. Simul.* **2020**, *90*, 1773–1792. [\[CrossRef\]](#)
12. Pardo, J.; Pardo, L.; Zografos, K. Minimum ϕ -divergence Estimators with Constraints in Multinomial Populations. *J. Stat. Plan. Inference* **2002**, *104*, 221–237. [\[CrossRef\]](#)
13. Read, T.R.; Cressie, N.A. *Goodness-of-Fit Statistics for Discrete Multivariate Data*; Springer: New York, NY, USA, 1988.
14. Alin, A.; Kurt, S. Ordinary and Penalized Minimum Power-divergence Estimators in Two-way Contingency Tables. *Comput. Stat.* **2008**, *23*, 455–468. [\[CrossRef\]](#)
15. Toma, A. Optimal Robust M-estimators Using Divergences. *Stat. Probab. Lett.* **2009**, *79*, 1–5. [\[CrossRef\]](#)
16. Jiménez-Gamero, M.; Pino-Mejías, R.; Alba-Fernández, V.; Moreno-Rebollo, J. Minimum ϕ -divergence Estimation in Misspecified Multinomial Models. *Comput. Stat. Data Anal.* **2011**, *55*, 3365–3378. [\[CrossRef\]](#)
17. Kim, B.; Lee, S. Minimum density power divergence estimator for covariance matrix based on skew t distribution. *Stat. Methods Appl.* **2014**, *23*, 565–575. [\[CrossRef\]](#)
18. Neath, A.A.; Cavanaugh, J.E.; Weyhaupt, A.G. Model Evaluation, Discrepancy Function Estimation, and Social Choice Theory. *Comput. Stat.* **2015**, *30*, 231–249. [\[CrossRef\]](#)
19. Ghosh, A. Divergence based robust estimation of the tail index through an exponential regression model. *Stat. Methods Appl.* **2016**, *26*, 181–213. [\[CrossRef\]](#)
20. Jiménez-Gamero, M.D.; Batsidis, A. Minimum Distance Estimators for Count Data Based on the Probability Generating Function with Applications. *Metrika* **2017**, *80*, 503–545. [\[CrossRef\]](#)
21. Basu, A.; Ghosh, A.; Mandal, A.; Martin, N.; Pardo, L. Robust Wald-type tests in GLM with random design based on minimum density power divergence estimators. *Stat. Methods Appl.* **2021**, *30*, 973–1005. [\[CrossRef\]](#)
22. Birch, M.W. A New Proof of the Pearson-Fisher Theorem. *Ann. Math. Stat.* **1964**, *35*, 817–824. [\[CrossRef\]](#)
23. Krantz, S.G.; Parks, H.R. *The Implicit Function Theorem: History, Theory, and Applications*; Birkhäuser: Basel, Switzerland, 2013.
24. Ferguson, T.S. *A Course in Large Sample Theory*; Chapman and Hall: Boca Raton, FL, USA, 1996.
25. McManus, D.A. Who Invented Local Power Analysis? *Econom. Theory* **1991**, *7*, 265–268. [\[CrossRef\]](#)
26. Neyman, J. “Smooth” Test for Goodness of Fit. *Scand. Actuar. J.* **1937**, *1937*, 149–199. [\[CrossRef\]](#)
27. Pardo, J.A. An approach to multiway contingency tables based on ϕ -divergence test statistics. *J. Multivar. Anal.* **2010**, *101*, 2305–2319. [\[CrossRef\]](#)
28. R Core Team. *R: A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria, 2016.

29. Johnson, S.G. The NLOpt Nonlinear-Optimization Package. 2014. Available online: <http://ab-initio.mit.edu/nlopt> (accessed on 27 March 2022).
30. Dale, J.R. Asymptotic Normality of Goodness-of-Fit Statistics for Sparse Product Multinomials. *J. R. Stat. Soc. Ser. B Methodol.* **1986**, *48*, 48–59. [[CrossRef](#)]
31. Batsidis, A.; Martin, N.; Pardo Llorente, L.; Zografos, K. φ -Divergence Based Procedure for Parametric Change-Point Problems. *Methodol. Comput. Appl. Probab.* **2016**, *18*, 21–35. [[CrossRef](#)]
32. Lloyd, C.J. Estimating test power adjusted for size. *J. Stat. Comput. Simul.* **2005**, *75*, 921–933. [[CrossRef](#)]
33. Dubrova, Y.E.; Grant, G.; Chumak, A.A.; Stezhka, V.A.; Karakasian, A.N. Elevated Minisatellite Mutation Rate in the Post-Chernobyl Families from Ukraine. *Am. J. Hum. Genet.* **2002**, *71*, 801–809. [[CrossRef](#)]
34. Znaor, A.; Brennan, P.; Gajalakshmi, V.; Mathew, A.; Shanta, V.; Varghese, C.; Boffetta, P. Independent and combined effects of tobacco smoking, chewing and alcohol drinking on the risk of oral, pharyngeal and esophageal cancers in Indian men. *Int. J. Cancer* **2003**, *105*, 681–686. [[CrossRef](#)]
35. Merková, M. Use of Investment Controlling and its Impact into Business Performance. *Procedia Econ. Financ.* **2015**, *34*, 608–614. [[CrossRef](#)]
36. Geenens, G.; Simar, L. Nonparametric tests for conditional independence in two-way contingency tables. *J. Multivar. Anal.* **2010**, *101*, 765–788. [[CrossRef](#)]
37. Bartolucci, F.; Scaccia, L. Testing for positive association in contingency tables with fixed margins. *Comput. Stat. Data Anal.* **2004**, *47*, 195–210. [[CrossRef](#)]