Ann. Inst. Statist. Math.
Vol. 50, No. 2, 295–310 (1998)

f-DISSIMILARITY OF SEVERAL DISTRIBUTIONS IN TESTING STATISTICAL HYPOTHESES

K. ZOGRAFOS

Department of Mathematics, Probability-Statistics & Operational Research Unit, University of Ioannina, 45110 Ioannina, Greece

(Received July 15, 1996; revised June 30, 1997)

Abstract. Various problems in statistics have been treated by the decision rule, based on the concept of distance between distributions. The aim of this paper is to give an approach for testing statistical hypotheses, using a general class of dissimilarity measures among $k \geq 2$ distributions. The test statistics are obtained by the replacement, in the expression of the dissimilarity measure, of the unknown parameters by their maximum likelihood estimators. The asymptotic distributions of the resulting test statistics are investigated and the results are applied to multinomial and multivariate normal populations.

Key words and phrases: f-dissimilarity, f-dissimilarity statistic, asymptotic distribution, multinomial distribution, multivariate normal distribution, testing statistical hypotheses.

1. Introduction

The concept of the distance between probability distributions is of fundamental importance in statistics. A lot of methods have been developed using distance-based decision rules, in various areas in statistics like hypotheses testing, nonparametric statistic, discriminant analysis, multiple regression models. On the other hand, statistical information theory provides us with a number of measures for discriminating among distributions. A divergence or dissimilarity measure between distributions can be considered as a measure of the "distance" among the respective distributions from the point of view that, the smaller the divergence between two distributions is, the harder it is to separate, to discriminate among them.

In the sequel, we focus our interest on the use of divergence or dissimilarity type "distances" for testing statistical hypotheses. In this context, a measure of affinity of several distributions has been used by Matusita (1966, 1967), as a decision rule for testing hypotheses about the parameters of several multivariate normal distributions. Since Matusita's pioneer work, a lot of papers have been appeared in the literature, where divergence or dissimilarity measures are used for the construction and the study of multinomial goodness of fit and independence tests. We refer, among others, to Cressie and Read (1984), Nayak (1985), Gil (1989), Zografos *et al.* (1990), Zografos (1993, 1994), Morales *et al.* (1994), Menendez *et al.* (1995) and references therein. Recently, Salicru *et al.* (1994) and Pardo *et al.* (1995) extented the above ideas and methods to the space of a parametric family of distributions, for testing hypotheses about the parameters of one or two populations.

The approach for testing statistical hypotheses, using a divergence or dissimilarity type "distance", can be summarized as follows: Let $\{P_\theta : \theta \in \Theta\}$ be a parametric family of distribution functions and $I(\cdot,\cdot)$ a divergence type "distance" on $\{P_\theta : \theta \in \Theta\}$. If $\theta_0 \in \Theta$ is known, then, $I(P_\theta, P_{\theta_0})$ is a measure of dissimilarity of P_θ and P_{θ_0} and therefore a measure of the closeness of θ and θ_0, for $\theta \subset \Theta$. If $\hat{\theta}$ is an estimator of θ, then, $I(P_{\hat{\theta}}, P_{\theta_0})$ can be used as a test statistic for testing the hypothesis $\theta = \theta_0$. In a similar manner, for $\hat{\theta}_i$ an estimator of $\theta_i \in \Theta$, $i = 1, 2$, $I(P_{\hat{\theta}_1}, P_{\hat{\theta}_2})$ can be used as a test statistic for testing homogeneity, i.e. $\theta_1 = \theta_2$, of two populations P_{θ_1} and P_{θ_2}. The investigation of the asymptotic distribution of the above statistics is therefore the main task of the method.

The aim of this paper is to extend the above method, concerning two populations to the case of several populations. In the following section, the *f-dissimilarity*, a general class of dissimilarity measures between k ($k \geq 2$) distributions due to Gyorfi and Nemetz (1978), is introduced in a parametric family of distribution functions. In Section 3, the asymptotic behaviour of the estimated *f-dissimilarity* is investigated, under a variety of assumptions concerning the relationships of the parameters of the distributions. The results of this section are applied, in the final Section 4, to multinomial and multivariate normal populations.

2. Preliminary concepts

Let (X, A, P_θ) be a probability space for $\theta \in \Theta$, where Θ is an open subset of R^M. Consider the generalized probability density function $f_\theta(x) = dP_\theta/d\mu$ relative to a σ-finite measure μ. Suppose that the following regularity conditions (cf. Serfling (1980), p. 144) are satisfied.

($R1$) The support of f_θ does not depend on $\theta \in \Theta$ and the derivatives $\frac{\partial f_\theta(x)}{\partial \theta_i}$, $\frac{\partial^2 f_\theta(x)}{\partial \theta_i \partial \theta_j}$, $\frac{\partial^3 f_\theta(x)}{\partial \theta_i \partial \theta_j \partial \theta_k}$, $i, j, k = 1, \ldots, M$, exist for all x.

($R2$) For each $\theta_0 \in \Theta$, there exist functions $g(x)$, $h(x)$ and $H(x)$, such that, for θ in a neighborhood $N(\theta_0)$, the relations $|\frac{\partial f_\theta(x)}{\partial \theta_i}| \leq g(x)$, $|\frac{\partial^2 f_\theta(x)}{\partial \theta_i \partial \theta_j}| \leq h(x)$, $|\frac{\partial^3 f_\theta(x)}{\partial \theta_i \partial \theta_j \partial \theta_k}| \leq H(x)$ hold for all x and $\int g(x)dx < \infty$, $\int h(x)dx < \infty$, $E_\theta(H(X)) < \infty$, for $\theta \in N(\theta_0)$.

($R3$) For $\theta \in \Theta$, the $M \times M$ Fisher's information matrix $\|E_\theta\{\frac{\partial \log f_\theta(x)}{\partial \theta_i} \cdot \frac{\partial \log f_\theta(x)}{\partial \theta_j}\}\|$, $i, j = 1, \ldots, M$, is positive definite with finite elements.

For $\theta_i \in \Theta$, let the generalized densities $f_{\theta_i}(x) = dP_{\theta_i}/d\mu$, $i = 1, \ldots, \nu$. The *f-dissimilarity* of $f_{\theta_1}, f_{\theta_2}, \ldots, f_{\theta_\nu}$ (cf. Gyorfi and Nemetz (1978)) is defined by

(2.1) $$D_f = D_f(\theta_1, \theta_2, \ldots, \theta_\nu) = \int f(f_{\theta_1}(x), f_{\theta_2}(x), \ldots, f_{\theta_\nu}(x))d\mu,$$

where f is a continuous, convex, homogeneous function, defined on the set $S = \{(s_1, \ldots, s_\nu) : 0 \le s_i < \infty, i = 1, \ldots, \nu\}$. D_f measures the dissimilarity ("distance") of f_{θ_i}, $i = 1, \ldots, \nu$, and it is intuitively intended as a measure which reflects the difference between the ν populations. If $f(s_1, \ldots, s_\nu) = -\prod_{i=1}^{\nu} s_i^{1/\nu}$ and $f(s_1, \ldots, s_\nu) = -\prod_{i=1}^{\nu} s_i^{\alpha_i}$, $\alpha_i \ge 0$, with $\sum_{i=1}^{\nu} \alpha_i = 1$, then *$f$-dissimilarity* is the negative of Matusita's (1967) affinity of ν populations and the negative of Toussaint's (1974) affinity, respectively. *f-dissimilarity* leads also to the *ϕ-divergence*, introduced by Csiszar (1967), if $f(s_1, s_2) = s_2\phi(s_1/s_2)$, with $s_1, s_2 \in [0, \infty)$ and ϕ a continuous, convex function defined on $[0, \infty)$, satisfying some regularity conditions (cf. Csiszar (1967)). Special choices of the convex function ϕ lead to the Kullback-Leibler directed divergence, or Renyi's order α information, or Cressie and Read's power divergence, etc.

In order to estimate D_f, motivated by Salicru *et al.* (1994), we suppose that the parameter $\theta_1^t = (\theta_{11}, \ldots, \theta_{1M})$ of the first population is completely unknown, while the parameters $\theta_m^t = (\theta_{m1}, \ldots, \theta_{mk}, \theta_{m(k+1)}, \ldots, \theta_{mM_0}, \theta^*_{m(M_0+1)}, \ldots, \theta^*_{mM})$, $m = 2, \ldots, \nu$, of the remaining populations are partially known, where the superscript t denotes the transpose of a vector or a matrix. In particular we assume that $\theta_{mi} = \theta_{1i}$, if $i \in I_1 = \{1, \ldots, k\}$, θ_{mi} is unknown when $i \in I_2 = \{k+1, \ldots, M_0\}$ and θ_{mi} is known and equal to θ^*_{mi}, for $i \in I_3 = \{M_0 + 1, \ldots, M\}$ and $m = 2, \ldots, \nu$. In this context, the joint parameter space is an open subset of $R^{M+(\nu-1)(M_0-k)}$ with elements $\gamma^t = (\theta_{11}, \ldots, \theta_{1M}, \theta_{2(k+1)}, \ldots, \theta_{2M_0}, \ldots, \theta_{\nu(k+1)}, \ldots, \theta_{\nu M_0})$.

The partitioning setup for the parameters, considered above, is motivated by the desire to construct statistical tests of homogeneity of ν independent, r-variate normal populations $N_r(\mu_i, V_i)$, with $V_i = \|u_{mj}^{(i)}\|_{r\times r}$, $m, j = 1, \ldots, r$, being a positive definite unknown dispersion matrix of order r, for $i = 1, \ldots, \nu$. In this context, consider for example, the null hypothesis,

$$H_0 : \mu_1 = \mu_2^* = \cdots = \mu_\nu^* \quad \text{and} \quad u_{mm}^{(1)} = \cdots = u_{mm}^{(\nu)}, \qquad m = 1, \ldots, r,$$

given that $u_{mj}^{(1)} = \cdots = u_{mj}^{(\nu)}$, for $m < j$, $m, j = 1, \ldots, r$ and μ_i^* are predicted values for μ_i, $i = 2, \ldots, \nu$. This hypothesis, can be written in the form, $H_0 : \theta_1 = \theta_2 = \cdots = \theta_\nu$, with $\theta_1^t = ((u_{mj}^{(1)}, m < j, m, j = 1, \ldots, r), (u_{mm}^{(1)}, m = 1, \ldots, r), \mu_1^t)$ and $\theta_i^t = ((u_{mj}^{(1)}, m < j, m, j = 1, \ldots, r), (u_{mm}^{(i)}, m = 1, \ldots, r), (\mu_i^*)^t)$, for $i = 2, \ldots, \nu$. In this case $k = \frac{r(r-1)}{2}$, $M_0 = k + r$ and $M = \frac{r^2+3r}{2}$. In a similar manner, for the hypothesis of complete homogeneity,

$$H_0 : \mu_1 = \mu_2 = \cdots = \mu_\nu \quad \text{and} \quad V_1 = V_2 = \cdots = V_\nu,$$

we have $\theta_i = (\mu_i, V_i)$, $i = 1, \ldots, \nu$ and $k = 0$, $M_0 = M = \frac{r^2+3r}{2}$.

Consider independent random samples $x_1^{(m)}, \ldots, x_{n_m}^{(m)}$, of sizes n_m, from the populations f_{θ_m}, $m = 1, \ldots, \nu$. Let $\hat\theta_{1i}$ and $\hat\theta_{mj}$ be the estimators of θ_{1i} and θ_{mj}, $i = 1, \ldots, M$, $j = k+1, \ldots, M_0$, $m = 2, \ldots, \nu$, respectively, maximizing the likelihood function

$$\log L(\gamma) = \sum_{i=1}^{n_1} \log f_{\theta_1}(x_i^{(1)}) + \sum_{m=2}^{\nu} \sum_{i=1}^{n_m} \log f_{\theta_m}(x_i^{(m)}). \tag{2.2}$$

Denote by $\hat{\gamma}^t = (\hat{\theta}_{11}, \ldots, \hat{\theta}_{1M}, \hat{\theta}_{2(k+1)}, \ldots, \hat{\theta}_{2M_0}, \ldots, \hat{\theta}_{\nu(k+1)}, \ldots, \hat{\theta}_{\nu M_0})$ the maximum likelihood estimator of γ and $n = n_1 + n_2 + \cdots + n_\nu$.

The sample estimator of D_f is obtained from (2.1), if we replace the unknown parameters θ_m, $m = 1, \ldots, \nu$, by their maximum likelihood estimators $\hat{\theta}_1^t = (\hat{\theta}_{11}, \ldots, \hat{\theta}_{1M})$ and $\hat{\theta}_m^t = (\hat{\theta}_{11}, \ldots, \hat{\theta}_{1k}, \hat{\theta}_{m(k+1)}, \ldots, \hat{\theta}_{mM_0}, \theta^*_{m(M_0+1)}, \ldots, \theta^*_{mM})$, for $m = 2, \ldots, \nu$. This estimator is the *f-dissimilarity* between $f_{\hat{\theta}_1}, \ldots, f_{\hat{\theta}_\nu}$ defined by

$$\hat{D}_f = D_f(\hat{\theta}_1, \hat{\theta}_2, \ldots, \hat{\theta}_\nu) = \int f(f_{\hat{\theta}_1}(x), f_{\hat{\theta}_2}(x), \ldots, f_{\hat{\theta}_\nu}(x))d\mu. \tag{2.3}$$

The estimated *f-dissimilarity*, measures the distance between $\hat{\theta}_i$, $i = 1, \ldots, \nu$, in the sense that, the smaller the $\hat{D}_f$ is, the harder it is to separate to discriminate among $\hat{\theta}_i$, $i = 1, \ldots, \nu$ and can therefore be used as a test statistic for testing homogeneity.

In the sequel, we assume that the convex function f admits continuous first and second order derivatives on S. Denote by $f_i^{(1)}(u)$ the first order partial derivative of $f(s)$ with respect to s_i, $i = 1, \ldots, \nu$, at the point u and by $f_{ij}^{(2)}(u)$ the second order partial derivative of $f(s)$ with respect to s_i and s_j, $i, j = 1, \ldots, \nu$, at the point u. Let $H(u)$ be the $\nu \times \nu$ non-negative definite Hessian matrix of the convex function f at the point u with elements $f_{ij}^{(2)}(u)$, for $i, j = 1, \ldots, \nu$. Let also $H = H(1)$.

3. Main results

Consider the vector $W = (W_1^t, W_2^t, \ldots, W_\nu^t)^t$. W_1 is the $M \times 1$ vector with elements

$$\begin{aligned} W_{1i} &= \frac{\partial}{\partial \theta_{1i}} D_f \\ &= \begin{cases} \displaystyle\sum_{m=1}^{\nu} \int \frac{\partial f_{\theta_m}(x)}{\partial \theta_{1i}} f_m^{(1)}(f_{\theta_1}(x), \ldots, f_{\theta_\nu}(x))d\mu, & 1 \le i \le k \\ \displaystyle\int \frac{\partial f_{\theta_1}(x)}{\partial \theta_{1i}} f_1^{(1)}(f_{\theta_1}(x), \ldots, f_{\theta_\nu}(x))d\mu, & k+1 \le i \le M \end{cases}, \end{aligned} \tag{3.1}$$

and W_m, $m = 2, \ldots, \nu$, the $(M_0 - k) \times 1$ vector with elements

$$\begin{aligned} W_{mi} &= \frac{\partial}{\partial \theta_{mi}} D_f \\ &= \int \frac{\partial f_{\theta m}(x)}{\partial \theta_{mi}} f_m^{(1)}(f_{\theta_1}(x), \ldots, f_{\theta_\nu}(x))d\mu, \qquad k+1 \le i \le M_0. \end{aligned} \tag{3.2}$$

Let Σ be the Fisher information matrix associated to the parameter γ. In view of (2.2) and regularity condition $(R3)$, Σ is a symmetric, positive definite block matrix of order $M + (\nu - 1)(M_0 - k)$ defined by

$$\Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}, \tag{3.3}$$

with

$$\Sigma_{11} = \sum_{m=1}^{\nu} \frac{n_m}{n} \, {}^{1,k}_{1,k}I^F(\theta_m),$$

$$\Sigma_{12} = \left(\frac{n_1}{n} \, {}^{1,k}_{k+1,M}I^F(\theta_1), \frac{n_2}{n} \, {}^{1,k}_{k+1,M_0}I^F(\theta_2), \ldots, \frac{n_\nu}{n} \, {}^{1,k}_{k+1,M_0}I^F(\theta_\nu) \right),$$

$$\Sigma_{21} = \Sigma_{12}^t,$$

$$\Sigma_{22} = \operatorname{diag}\left(\frac{n_1}{n} \, {}^{k+1,M}_{k+1,M}I^F(\theta_1), \frac{n_2}{n} \, {}^{k+1,M_0}_{k+1,M_0}I^F(\theta_2), \ldots, \frac{n_\nu}{n} \, {}^{k+1,M_0}_{k+1,M_0}I^F(\theta_\nu) \right).$$

Σ_{11} is a $k \times k$, Σ_{12} a $k \times [(M-k) + (\nu - 1)(M_0 - k)]$ block matrix, while Σ_{22} is a block diagonal matrix of order $(M-k) + (\nu-1)(M_0-k)$. ${}^{i,j}_{r,s}I^F(\theta_m)$ is the $(j-i+1) \times (r-s+1)$ submatrix of $I^F(\theta_m)$, which have the $i, i+1, \ldots, j$ rows and the $r, r+1, \ldots, s$ columns of $I^F(\theta_m)$ and $I^F(\theta_m)$ is the Fisher information matrix associated to θ_m for $m = 1, \ldots, \nu$.

Under the assumption $\theta_1 = \theta_2 = \cdots = \theta_\nu$, let $I^F(\theta)$ be the Fisher information matrix, with θ the common value of θ_i, $i = 1, \ldots, \nu$. In this case, consider the square block matrix A of order $(M + M_0 - 2k)$, defined by

$$A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}, \tag{3.4}$$

with

$$A_{11} = \begin{bmatrix} {}^{k+1,M_0}_{k+1,M_0}I^F(\theta) & {}^{k+1,M_0}_{M_0+1,M}I^F(\theta) \\ {}^{M_0+1,M}_{k+1,M_0}I^F(\theta) & {}^{M_0+1,M}_{M_0+1,M}I^F(\theta) \end{bmatrix}, \qquad A_{12} = \begin{bmatrix} {}^{k+1,M_0}_{k+1,M_0}I^F(\theta) \\ {}^{M_0+1,M}_{k+1,M_0}I^F(\theta) \end{bmatrix},$$

$$A_{21} = A_{12}^t \quad \text{and} \quad A_{22} = {}^{k+1,M_0}_{k+1,M_0}I^F(\theta).$$

Consider also the following partition of the Hessian matrix H of the convex function f at the point 1,

$$H = \begin{bmatrix} H_{11} & H_{12} \\ H_{21} & H_{22} \end{bmatrix}, \quad \text{with} \quad H_{11} = {}^{1,1}_{1,1}H, \quad H_{12} = {}^{1,1}_{2,\nu}H, \tag{3.5}$$

$$H_{22} = {}^{2,\nu}_{2,\nu}H$$

and

$$H_{21} = H_{12}^t.$$

In view of (3.4) and (3.5), define the square matrix of order $(M-k) + (\nu - 1) \times (M_0 - k)$,

$$B = \begin{bmatrix} A_{11} \otimes H_{11} & A_{12} \otimes H_{12} \\ A_{21} \otimes H_{21} & A_{22} \otimes H_{22} \end{bmatrix}, \tag{3.6}$$

with $\otimes$ the direct product of the respective matrices.

After the notation introduced above, we state the main result concerning the asymptotic distribution of the statistic $\hat{D}_f$. The proof of the theorem is given in the Appendix.

THEOREM 3.1. *Suppose that the regularity conditions* $(R1)$–$(R3)$ *are satisfied. If* $\frac{n_i}{n} \to \lambda_i > 0$, *as* $n_i \to \infty$, *for* $i = 1, \ldots, \nu$ *and* $n = \sum_{i=1}^{\nu} n_i$, *then,*

a)

$$\sqrt{n}(\hat{D}_f - D_f) \underset{n_i \to \infty}{\xrightarrow{L}} N(0, W^t \Sigma_*^{-1} W), \qquad \textit{provided that} \quad W^t \Sigma_*^{-1} W > 0,$$

where W *is given by* (3.1), (3.2) *and* Σ_* *is obtained from* Σ, *given by* (3.3), *if we replace* $\frac{n_i}{n}$ *by* λ_i, $i = 1, \ldots, \nu$.

b) *If* $\theta_1 = \theta_2 = \cdots = \theta_\nu$, *then*

$$2n(\hat{D}_f - f(1)) \underset{n_i \to \infty}{\xrightarrow{L}} \sum_{i=1}^{k^*} \beta_i X_{1,i}^2,$$

where $X_{1,i}^2$ *are independent random variables, each having a chi-square distribution with* 1 *degree of freedom,* $k^* = \min\{(\nu - 1)(M - k), (M - k) + (\nu - 1)(M_0 - k)\}$, β_i, $i = 1, \ldots, k^*$, *are the non-zero eigenvalues of the matrix* $B\Sigma_\beta^{-1}$, *with* B *given by* (3.6) *and* Σ_β *is obtained from* Σ_{22}, *given by* (3.3), *if we replace* $\frac{n_i}{n}$ *by* λ_i, $i = 1, \ldots, \nu$.

The estimation procedure for the *f-dissimilarity*, considered above, is invariant under reparametrizations. Indeed, since the maximum likelihood estimators are used to estimate the parameters of the respective populations, the *f-dissimilarity* statistic, given by (2.3), is parametrization invariant. The *f-dissimilarity* measure, is also invariant under one-to-one reparametrizations, done individually on each of the ν populations. For example, consider, reparametrizations of the form $\gamma_l \mapsto \psi_l$, $l = 1, \ldots, \nu$, with $\gamma = (\gamma_1^t, \gamma_2^t, \ldots, \gamma_\nu^t)^t$ the joint parameter, where γ_1 is M-dimensional and γ_l is $(M_0 - k)$-dimensional for each $l = 2, \ldots, \nu$. In this case, it is easily seen that the *f-dissimilarity*, given by (2.1), remains invariant. This is also true for the variance of the estimated *f-dissimilarity*, obtained in Theorem 3.1(a). To see this, let $I^F(\gamma)$ and $I^F(\psi)$ be the Fisher information matrices associated to the parameters γ and ψ respectively, with $\psi = (\psi_1^t, \psi_2^t, \ldots, \psi_\nu^t)^t$. Denote by Σ_ψ the matrix obtained from $I^F(\psi)$, if we replace $\frac{n_i}{n}$ by λ_i, $i = 1, \ldots, \nu$, and let $C = \|\frac{\partial \gamma_i}{\partial \psi_j}\|$, $i, j = 1, \ldots, \nu$, be the square, of order $M + (\nu - 1)(M_0 - k)$, block diagonal Jacobian matrix of the transformation $\gamma \mapsto \psi$. In virtue of (3.1) and (3.2) we have

$$W_\psi \stackrel{\triangle}{=} \nabla_\psi D_f(\psi_1, \ldots, \psi_\nu) = C^t \nabla_\gamma D_f(\gamma_1, \ldots, \gamma_\nu) = C^t W.$$

Using the reparametrization formula (cf. Papaioannou and Kempthorne (1971), p. 30),

$$I^F(\psi) = C^t I^F(\gamma) C = C I^F(\gamma) C^t,$$

we obtain,

$$\Sigma_\psi = C^t \Sigma_* C = C \Sigma_* C^t.$$

In this context, let σ^2_ψ be the asymptotic variance of the f-*dissimilarity* statistic $\hat{D}_f$, under reparametrizations of the form $\gamma \mapsto \psi$. In view of part (a) of Theorem 3.1 we have,

$$\sigma^2_\psi = W^t_\psi \Sigma^{-1}_\psi W_\psi = W^t C (C^{-1} \Sigma^{-1}_* (C^t)^{-1}) C^t W = W^t \Sigma^{-1}_* W.$$

Therefore, the f-*dissimilarity* statistic $\hat{D}_f$ and its asymptotic distribution, remains invariant under reparametrizations, done individually on each of the populations considered.

A special case of Theorem 3.1 appears when $M_0 = M$. In this case the parameter of the m-population $\theta^t_m = (\theta_{m1}, \ldots, \theta_{mk}, \theta_{m(k+1)}, \ldots, \theta_{mM})$, $m = 1, \ldots, \nu$, is unknown and we suppose that $\theta_{1i} = \theta_{2i} = \cdots = \theta_{\nu i}$, for $i = 1, \ldots, k$. In the following corollary, this case is treated as well as the case $M_0 = M$ and $k = 0$.

COROLLARY 3.1. *Suppose that the regularity conditions (R1)–(R3) are satisfied. If* $\frac{n_i}{n} \to \lambda_i > 0$, *as* $n_i \to \infty$, *for* $i = 1, \ldots, \nu$, *then*

a)

$$\sqrt{n}(\hat{D}_f - D_f) \underset{n_i \to \infty}{\overset{L}{\to}} N(0, \sigma^2), \qquad \textit{provided that} \quad \sigma^2 > 0.$$

i) *If* $M_0 = M$, *then* $\sigma^2 = W^t \Sigma^{-1}_* W$, *with* W *and* Σ_* *are obtained from part (a) of Theorem 3.1 taking* $M_0 = M$.

ii) *If* $M_0 = M$ *and* $k = 0$, *then* $\sigma^2 = \sum_{i=1}^{\nu} \frac{1}{\lambda_i} W_i^t I^F(\theta_i)^{-1} W_i$, *where* W_i *is the* $M \times 1$ *vector with elements*

$$W_{ij} = \int \frac{\partial f_{\theta_i}(x)}{\partial \theta_{ij}} f_i^{(1)}(f_{\theta_1}(x), \ldots, f_{\theta_\nu}(x)) d\mu, \qquad i = 1, \ldots, \nu, \quad j = 1, \ldots, M$$

and $I^F(\theta_i)$ *is the Fisher information matrix associated to* θ_i, $i = 1, \ldots, \nu$.

b) *Under the hypothesis that* $\theta_1 = \theta_2 = \cdots = \theta_\nu$, *if* $M_0 = M$, *we have*

$$2n(\hat{D}_f - f(1)) \underset{n_i \to \infty}{\overset{L}{\to}} \sum_{i=1}^{\nu-1} \beta_i X^2_{M-k,i},$$

where $X^2_{r,i}$ *are independent random variables, each having a chi-square distribution with* r *degrees of freedom and* β_i *are the non-zero eigenvalues of the matrix* HD_λ, *with* H *the Hessian matrix of the convex function* f *at the point* 1 *and* D_λ *the diagonal matrix with elements* λ_i^{-1}, $i = 1, \ldots, \nu$.

PROOF. a) Direct application of part (a) of Theorem 3.1.

b) In view of Theorem 3.1(b), for $M_0 = M$, we have that

$$2n(\hat{D}_f - f(1)) \underset{n_i \to \infty}{\overset{L}{\to}} \sum_{i=1}^{(\nu-1)(M-k)} \beta_i X^2_{1,i},$$

with β_i the non-zero eigenvalues of $B\Sigma_\beta^{-1}$, where B is given by (3.6) and Σ_β is obtained from Σ_{22} given by (3.3), if we replace $\frac{n_i}{n}$ by λ_i, $i = 1, \ldots, \nu$. If $M_0 = M$, after considerable algebra, we obtain that

$$\det(B\Sigma_\beta^{-1} - \omega I) \propto [\det(HD_\lambda - \omega I)]^{M-k},$$

with H and D_λ as defined above and I the identity matrix of order ν. Because of $\operatorname{rank}(HD_\lambda) \leq \nu - 1$, the non-zero eigenvalues of $B\Sigma_\beta^{-1}$ are β_i, $i = 1, \ldots, \nu - 1$, with multiplicity $M - k$.

Consider the case $M_0 = k$. The parameter $\theta_1^t = (\theta_{11}, \ldots, \theta_{1M_0}, \theta_{1(M_0+1)}, \ldots, \theta_{1M})$ of the first population is completely unknown, while the parameter of the m-population becomes $\theta_m^t = (\theta_{11}, \ldots, \theta_{1M_0}, \theta^*_{m(M_0+1)}, \ldots, \theta^*_{mM})$, with θ^*_{mj} known, for $m = 2, \ldots, \nu$ and $j = M_0 + 1, \ldots, M$. In this case we consider a random sample of size n from the first population. From (3.1), (3.2) we have that $W^t = (W_1, \ldots, W_{M_0}, W_{M_0+1}, \ldots, W_M)$, with

$$W_i = \begin{cases} \displaystyle\sum_{m=1}^{\nu} \int \frac{\partial f_{\theta_m}(x)}{\partial \theta_{mi}} f_m^{(1)}(f_{\theta_1}(x), \ldots, f_{\theta_\nu}(x)) d\mu, & 1 \leq i \leq M_0 \\ \displaystyle\int \frac{\partial f_{\theta_1}(x)}{\partial \theta_{1i}} f_1^{(1)}(f_{\theta_1}(x), \ldots, f_{\theta_\nu}(x)) d\mu, & M_0 + 1 \leq i \leq M \end{cases}, \tag{3.7}$$

while, in this case, Fisher's information matrix is $I^F(\theta_1)$.

In the following theorem, the asymptotic distribution of $\hat{D}_f$ is investigated, when $M_0 = k$ and $M_0 = k = 0$. The proof of the theorem can be obtained by a similar argument to that of Theorem 3.1, and it is outlined in the Appendix.

THEOREM 3.2. *Suppose that the regularity conditions (R1)–(R3) are satisfied.*

a)

$$\sqrt{n}(\hat{D}_f - D_f) \underset{n\to\infty}{\overset{L}{\to}} N(0, \sigma^2), \qquad \textit{provided that} \quad \sigma^2 = W^t I^F(\theta_1)^{-1} W > 0.$$

i) *If $M_0 = k$, then W is given by (3.7).*
ii) *If $M_0 = k = 0$, then $W^t = (W_1, \ldots, W_M)$ with*

$$W_i = \int \frac{\partial f_{\theta_1}(x)}{\partial \theta_{1i}} f_1^{(1)}(f_{\theta_1}(x), \ldots, f_{\theta_\nu}(x)) d\mu, \qquad \textit{for} \quad i = 1, \ldots, M.$$

b) *Under the hypothesis that $\theta_1 = \theta_2 = \cdots = \theta_\nu$, if $M_0 = k$, we have,*

$$\{2n(\hat{D}_f - f(1))/f_{11}^{(2)}(1)\} \underset{n\to\infty}{\overset{L}{\to}} X^2_{M-M_0},$$

provided that $f_{11}^{(2)}(1) \neq 0$.

Remark 3.1. a) The above results, concerning the asymptotic distribution of *f-dissimilarity* statistic, can be used in various settings to testing statistical hypotheses. In particular, part (a) of theorems and corollaries can be used to test:

i) $H_0 : D_f = D_{f,0}$, i.e., *f-dissimilarity* is of certain magnitude $D_{f,0}$,

ii) $H_0 : D_{f,1} = D_{f,2} = \cdots = D_{f,l}$, i.e., *f-dissimilarities* of $l \geq 2$ groups of populations are equal.

Statistical tests, for testing these hypotheses, can be obtained by using the results of this section, in a similar manner as in Section 4 in Zografos *et al.* (1990) and Section 3 of Salicru *et al.* (1994).

Applications of the above theorems, for special choices of the convex function f, lead to the asymptotic distributions of Matusita's and Toussaint's affinities, as well as Csiszar's and other divergence measures, mentioned in Section 2.

b) In the case of equal sample sizes, the asymptotic distributions of part (b) of the theorems and corollaries above can be simplified for Matusita's affinity, as we will see in the Subsection 4.2 below.

4. Applications to multinomial and multivariate normal populations

4.1 *Multinomial populations*

There is an extensive literature, mentioned in introduction, dealing with the estimation of various divergence measures in a multinomial distribution context. The estimators of these measures have been used for the construction and the study of optimality properties of multinomial goodness of fit and independence tests. The population and sampling set-ups have been simple or stratified. In this subsection, a unification of the existing results and an extension to the case of $k > 2$ multinomial populations are given as application of the results of the previous section.

Consider the probability space (X, A, P_θ), $\theta \in \Theta$, with $X = \{x_1, \ldots, x_M\}$ and $\Theta = \{(p_1, \ldots, p_{M-1}) : p_i > 0, \sum_{i=1}^{M-1} p_i = 1 - p_M\}$. For μ a counting measure on X and $\theta_m \in \Theta$, $m = 1, \ldots, \nu$, let $f_{\theta_m}(x_i) = p_{mi}$, for $i = 1, \ldots, M$ and $m = 1, \ldots, \nu$. In this context, if $P_m^t = (p_{m1}, p_{m2}, \ldots, p_{mM})$, $m = 1, \ldots, \nu$, the *f-dissimilarity* of $P_1, \ldots, P_\nu$ is defined by

$$D_f = D_f(P_1, \ldots, P_\nu) = \sum_{j=1}^{M} f(p_{1j}, p_{2j}, \ldots, p_{\nu j}). \tag{4.1}$$

Consider ν independent random samples, of sizes n_m, $m = 1, \ldots, \nu$ and $n = \sum_{i=1}^{\nu} n_i$. Let $\hat{p}_{mj}$ be the relative frequency of the value x_j from the m-population for $m = 1, \ldots, \nu$ and $j = 1, \ldots, M$. In this context, let $\hat{P}_m^t = (\hat{p}_{m1}, \hat{p}_{m2}, \ldots, \hat{p}_{mM})$. The estimator of D_f is defined by

$$\hat{D}_f = D_f(\hat{P}_1, \ldots, \hat{P}_\nu) = \sum_{j=1}^{M} f(\hat{p}_{1j}, \hat{p}_{2j}, \ldots, \hat{p}_{\nu j}). \tag{4.2}$$

The asymptotic distribution of $\hat{D}_f$ is established in the following corollaries.

COROLLARY 4.1. *As* $n_i \to \infty$, *with* $\frac{n_i}{n} \to \lambda_i > 0$, *for* $i = 1, \ldots, \nu$, *we have*, a)

$$\sqrt{n}(\hat{D}_f - D_f) \underset{n_i\to\infty}{\overset{L}{\to}} N(0,\sigma^2), \qquad \textit{provided that} \quad \sigma^2 > 0,$$

with

$$(4.3)\quad \sigma^2 = \sum_{m=1}^{\nu}\sum_{i=1}^{M} \frac{p_{mi}(1-p_{mi})}{\lambda_m}(f_m^{(1)}(U_i))^2 - \sum_{m=1}^{\nu}\sum_{\substack{i,j=1\\ i\neq j}}^{M} \frac{p_{mi}p_{mj}}{\lambda_m} f_m^{(1)}(U_i) f_m^{(1)}(U_j),$$

and $U_j^t = (p_{1j},\ldots,p_{\nu j})$, $j = 1,\ldots,M$.

b) *If* $P_1 = P_2 = \cdots = P_\nu$, *then*

$$2n(\hat{D}_f - f(1)) \underset{n_i\to\infty}{\overset{L}{\to}} \sum_{i=1}^{\nu-1} \beta_i X^2_{M-1,i},$$

with β_i *the non-zero eigenvalues of* HD_λ.

PROOF. a) Based on Corollary 3.1(a(ii)), we have the asymptotic normality of $\sqrt{n}(\hat{D}_f - D_f)$ with zero mean and variance

$$(4.4)\qquad \sigma^2 = \sum_{m=1}^{\nu} \frac{1}{\lambda_m} W_m^t I^F(\theta_m)^{-1} W_m,$$

where $W_m^t = (W_{m1},\ldots,W_{mM})$ and

$$(4.5)\qquad W_{ij} = \int \frac{\partial f_{\theta_i}(x)}{\partial\theta_{ij}} f_i^{(1)}(f_{\theta_1}(x),\ldots,f_{\theta_\nu}(x))d\mu = f_i^{(1)}(p_{1j},\ldots,p_{\nu j}),$$

for $i = 1,\ldots,\nu$ and $j = 1,\ldots,M$. Furthermore, it is easily seen that

$$(4.6)\qquad I^F(\theta_m)^{-1} = \|p_{mi}(\delta_{ij} - p_{mj})\|, \qquad i,j = 1,\ldots,M-1,$$

and (4.3) is obtained from (4.4), (4.5) and (4.6).

b) Follows immediately from Corollary 3.1(b), taking $k = 0$.

Consider now the case, when we observe a random sample of size n from the first multinomial population P_1, while the remaining P_m, $m = 2,\ldots,\nu$, populations are completely known. We have the following consequence of Theorem 3.2.

COROLLARY 4.2. a)

$$\sqrt{n}(\hat{D}_f - D_f) \underset{n\to\infty}{\overset{L}{\to}} N(0,\sigma^2), \qquad \textit{provided that} \quad \sigma^2 > 0,$$

with

$$\sigma^2 = \sum_{i=1}^{M} p_{1i}(1-p_{1i})(f_1^{(1)}(U_i))^2 - \sum_{\substack{i,j=1\\ i\neq j}}^{M} p_{1i}p_{1j} f_1^{(1)}(U_i) f_1^{(1)}(U_j),$$

and $U_j^t = (p_{1j},\ldots,p_{\nu j})$, $j = 1,\ldots,M$.

b) *If* $P_1 = P_2 = \cdots = P_\nu$, *then*

$$\{2n(\hat{D}_f - f(1))/f_{11}^{(2)}(1)\} \underset{n\to\infty}{\overset{L}{\to}} X^2_{M-1},$$

provided that $f_{11}^{(2)}(1) \neq 0$.

4.2 *Multivariate normal populations*

Consider ν independent r-variate normal populations $N_r(\mu_m, V_m)$, with mean μ_m and positive definite variance-covariance matrix V_m, for $m = 1, \ldots, \nu$. In this section we focus our interest on the use of the negative of Matusita's affinity, obtained from (2.1) for $f(s_1, \ldots, s_\nu) = -\prod_{i=1}^{\nu} s_i^{1/\nu}$. Straightforward calculations (cf. Matusita (1967)), lead that the negative of Matusita's affinity of the normal populations $N_r(\mu_m, V_m)$, $m = 1, \ldots, \nu$, is given by

$$\begin{aligned}(4.7)\qquad \rho &= D_f((\mu_1, V_1), \ldots, (\mu_\nu, V_\nu)) \\ &= -\frac{\prod_{m=1}^{\nu} |V_m^{-1}|^{1/2\nu}}{\left|\frac{1}{\nu}\sum_{m=1}^{\nu} V_m^{-1}\right|^{1/2}} \\ &\quad\cdot \exp\left\{\frac{1}{2\nu}\left[\left\langle \sum_{m=1}^{\nu} V_m^{-1}\mu_m, \left(\sum_{m=1}^{\nu} V_m^{-1}\right)^{-1} \sum_{m=1}^{\nu} V_m^{-1}\mu_m\right\rangle \right.\right. \\ &\qquad\qquad \left.\left. - \sum_{m=1}^{\nu} \langle V_m^{-1}\mu_m, \mu_m\rangle \right]\right\},\end{aligned}$$

with $\langle \cdot, \cdot \rangle$ the inner product of the vectors involved.

Let $X_1^{(m)}, X_2^{(m)}, \ldots, X_{n_m}^{(m)}$ a random sample of size n_m, from the m-population and $\bar{X}^{(m)}$, S_m the sample mean and covariance matrix respectively, defined by

$$(4.8)\qquad \begin{aligned}\bar{X}^{(m)} &= \frac{1}{n_m}\sum_{j=1}^{n_m} X_j^{(m)}, \\ S_m &= \frac{1}{n_m - 1}\sum_{j=1}^{n_m}(X_j^{(m)} - \bar{X}^{(m)})(X_j^{(m)} - \bar{X}^{(m)})^t,\end{aligned}$$

for $m = 1, \ldots, \nu$, with $n_m > r$. Let also $n = n_1 + \cdots + n_\nu$.

An estimator of ρ can be obtained from (4.7), if population parameters μ_m and V_m are replaced by their estimators $\bar{X}^{(m)}$ and S_m, $m = 1, \ldots, \nu$, respectively. The statistic obtained, can be used for testing homogeneity.

Consider the statistical hypothesis,

$$H_0 : \mu_1 = \mu_2 = \cdots = \mu_\nu \quad \text{given that} \quad V_1 = V_2 = \cdots = V_\nu = V.$$

The m.l.e. of μ_m and V are respectively $\bar{X}^{(m)}$ and $S = [\sum_{i=1}^{\nu}(n_i - 1)S_i]/n$.

Under H_0, from (4.7), we have

$$\rho = -\exp\left\{\frac{1}{2\nu}\left[\left\langle V^{-1}\sum_{m=1}^{\nu}\mu_m, \frac{1}{\nu}\sum_{m=1}^{\nu}\mu_m\right\rangle - \sum_{m=1}^{\nu}\langle V^{-1}\mu_m, \mu_m\rangle\right]\right\}$$

and its sample estimator is

$$\hat{\rho} = -\exp\left\{\frac{1}{2\nu}\left[\left\langle S^{-1}\sum_{m=1}^{\nu}\bar{X}^{(m)}, \frac{1}{\nu}\sum_{m=1}^{\nu}\bar{X}^{(m)}\right\rangle - \sum_{m=1}^{\nu}\langle S^{-1}\bar{X}^{(m)}, \bar{X}^{(m)}\rangle\right]\right\}.$$

An application of Corollary 3.1(b), for $f(s_1, \ldots, s_\nu) = -\prod_{i=1}^{\nu} s_i^{1/\nu}$ gives that, as $n_i \to \infty$, with $\frac{n_i}{n} \to \lambda_i > 0$, $i = 1, \ldots, \nu$,

$$2n(\hat{\rho} + 1) \xrightarrow{L} \sum_{i=1}^{\nu-1} \beta_i X_{r,i}^2, \qquad \text{under the hypothesis } H_0, \tag{4.9}$$

where β_i are the non-zero eigenvalues of HD_λ, with H the Hessian matrix of the convex function $f(s_1, \ldots, s_\nu) = -\prod_{i=1}^{\nu} s_i^{1/\nu}$ at the point 1 and D_λ the diagonal matrix with elements λ_i^{-1}, $i = 1, \ldots, \nu$. We obtain that

$$HD_\lambda = \left\| \frac{\lambda_j^{-1}}{\nu} \left(\delta_{ij} - \frac{1}{\nu} \right) \right\|, \qquad i, j = 1, \ldots, \nu. \tag{4.10}$$

When the sample sizes are equal, (4.10) remains true, if we replace λ_j by λ, the common limit of the ratios n_i/n, $i = 1, \ldots, \nu$. In this case, after a little algebra, we can see that the non-zero eigenvalues of HD_λ are 1 with multiplicity $\nu - 1$ and from (4.9), the asymptotic distribution of $2n(\hat{\rho} + 1)$ is $X_{(\nu-1)r}^2$, under the hypothesis H_0. Therefore, for large n, H_0 is rejected at a level α, if $2n(\hat{\rho} + 1) \geq X_{(\nu-1)r,\alpha}^2$.

The negative of Matusita's affinity ρ can also be used to produce an asymptotic statistical test for testing complete homogeneity of ν independent normal populations, i.e., the hypothesis

$$H_0 : \mu_1 = \mu_2 = \cdots = \mu_\nu \quad \text{and} \quad V_1 = V_2 = \cdots = V_\nu.$$

In this case, for $\hat{V}_m = (n_m - 1)S_m/n_m$, $m = 1, \ldots, \nu$, the estimator of ρ is given by

$$\begin{aligned}
\hat{\rho} &= D_f((\bar{X}^{(1)}, \hat{V}_1), \ldots, (\bar{X}^{(\nu)}, \hat{V}_\nu)) \\
&= -\frac{\prod_{m=1}^{\nu} |\hat{V}_m^{-1}|^{1/2\nu}}{\left| \frac{1}{\nu} \sum_{m=1}^{\nu} \hat{V}_m^{-1} \right|^{1/2}} \\
&\quad \cdot \exp\left\{ \frac{1}{2\nu} \left[\left\langle \sum_{m=1}^{\nu} \hat{V}_m^{-1} \bar{X}^{(m)}, \left(\sum_{m=1}^{\nu} \hat{V}_m^{-1} \right)^{-1} \sum_{m=1}^{\nu} \hat{V}_m^{-1} \bar{X}^{(m)} \right\rangle \right.\right. \\
&\qquad \left.\left. - \sum_{m=1}^{\nu} \langle \hat{V}_m^{-1} \bar{X}^{(m)}, \bar{X}^{(m)} \rangle \right] \right\},
\end{aligned}$$

where $\bar{X}^{(m)}$ and S_m are given by (4.8). Under the hypothesis H_0, from Corollary 3.1(b), if we take $k = 0$, the test statistic $2n(\hat{\rho} + 1)$ has the asymptotic distribution $\sum_{i=1}^{\nu-1} \beta_i X_{u,i}^2$, with $u = (r^2 + 3r)/2$ and β_i the non-zero eigenvalues of HD_λ given by (4.10). When the sample sizes are equal, the asymptotic distribution of the above statistic is the $X_{[(\nu-1)(r^2+3r)]/2}^2$, under the assumption of complete homogeneity.

Acknowledgements

The author wish to thank the referees for their helpful comments and suggestions which improved the presentation of the paper.

Appendix

PROOF OF THEOREM 3.1. Following the notation introduced in Sections 2 and 3, consider the function

$$\varphi(\gamma) = D_f(\theta_1, \theta_2, \ldots, \theta_\nu) = \int f(f_{\theta_1}(x), f_{\theta_2}(x), \ldots, f_{\theta_\nu}(x))d\mu.$$

a) A Taylor series expansion of $\varphi(\hat{\gamma})$ around γ gives,

$$\begin{aligned}\varphi(\hat{\gamma}) &= \varphi(\gamma) + \sum_{i=1}^{M}(\hat{\gamma}_i - \gamma_i)\frac{\partial\varphi(\gamma)}{\partial\gamma_i} + \sum_{i=M+1}^{M+(\nu-1)(M_0-k)}(\hat{\gamma}_i - \gamma_i)\frac{\partial\varphi(\gamma)}{\partial\gamma_i} + R_n \\ &= \varphi(\gamma) + \sum_{i=1}^{k}\left(\sum_{m=1}^{\nu}\frac{n_m}{n}\hat{\theta}_{mi} - \theta_{1i}\right)\frac{\partial\varphi}{\partial\theta_{1i}} + \sum_{i=k+1}^{M}(\hat{\theta}_{1i} - \theta_{1i})\frac{\partial\varphi}{\partial\theta_{1i}} \\ &\quad + \sum_{m=2}^{\nu}\sum_{i=k+1}^{M_0}(\hat{\theta}_{mi} - \theta_{mi})\frac{\partial\varphi}{\partial\theta_{mi}} + R_n,\end{aligned}$$

with $R_n = \epsilon_n\|\hat{\gamma} - \gamma\|$ and $\epsilon_n \to 0$, in probability as $n_i \to \infty$ for $i = 1, \ldots, \nu$. After a little algebra, we have

$$\hat{D}_f = D_f + \sum_{i=1}^{M}(\hat{\theta}_{1i} - \theta_{1i})\frac{\partial D_f}{\partial\theta_{1i}} + \sum_{m=2}^{\nu}\sum_{i=k+1}^{M_0}(\hat{\theta}_{mi} - \theta_{mi})\frac{\partial D_f}{\partial\theta_{mi}} + R_n,$$

which leads to

$$\hat{D}_f = D_f + W^t(\hat{\gamma} - \gamma) + R_n, \tag{A.1}$$

with W defined by (3.1) and (3.2). For the m.l.e. $\hat{\gamma}$ of γ it holds that: as $n_i \to \infty$, with $\frac{n_i}{n} \to \lambda_i > 0$, for $i = 1, \ldots, \nu$, we have

$$\sqrt{n}(\hat{\gamma} - \gamma) \xrightarrow{L} N(0, \Sigma_*^{-1}), \tag{A.2}$$

where Σ_* is obtained from Σ, given by (3.3), if we replace $\frac{n_i}{n}$ by λ_i, $i = 1, \ldots, \nu$. In view of (A.2), $\sqrt{n}\|\hat{\gamma} - \gamma\|$ has an asymptotic distribution as $n_i \to \infty$, $i = 1, \ldots, \nu$ and therefore $\sqrt{n}R_n = \sqrt{n}\epsilon_n\|\hat{\gamma} - \gamma\| \to 0$, in probability as $n_i \to \infty$ for $i = 1, \ldots, \nu$. Relations (A.1) and (A.2) complete the proof of part (a) of the theorem.

b) If $\theta_1 = \theta_2 = \cdots = \theta_\nu$, then $W^t\Sigma_*^{-1}W = 0$. Taking again a Taylor series expansion of $\varphi(\hat{\gamma})$ around γ, we have

$$\varphi(\hat{\gamma}) = \varphi(\gamma) + \sum_{i=1}^{k^*}(\hat{\gamma}_i - \gamma_i)\frac{\partial\varphi(\gamma)}{\partial\gamma_i} + \frac{1}{2}\sum_{i,j=1}^{k^*}(\hat{\gamma}_i - \gamma_i)(\hat{\gamma}_j - \gamma_j)\frac{\partial^2\varphi(\gamma)}{\partial\gamma_i\partial\gamma_j} + R_n,$$

with $k^* = M + (\nu - 1)(M_0 - k)$ and $R_n = \epsilon_n \|\hat{\gamma} - \gamma\|^2$, $\epsilon_n \to 0$, in probability as $n_i \to \infty$ for $i = 1, \dots, \nu$. Because of homogeneity of f on S, we have that: $\sum_{i=1}^{\nu} s_i f_i^{(1)}(s) = f(s)$, for $s \in S$. Using this relation, we can easily see that $\frac{\partial \varphi(\gamma)}{\partial \gamma_i} = 0$, $i = 1, \dots, M + (\nu - 1)(M_0 - k)$, under the assumption that $\theta_1 = \theta_2 = \dots = \theta_\nu$.

Therefore

$$2(\varphi(\hat{\gamma}) - \varphi(\gamma)) = \sum_{i,j=1}^{k^*} (\hat{\gamma}_i - \gamma_i)(\hat{\gamma}_j - \gamma_j) \frac{\partial^2 \varphi(\gamma)}{\partial \gamma_i \partial \gamma_j} + 2R_n. \tag{A.3}$$

Homogeneity of f, entails also that,

$$\sum_{m=1}^{\nu} f_m^{(1)}(t, \dots, t) = f(1), \qquad \sum_{m=1}^{\nu} f_{mi}^{(2)}(t, \dots, t) = 0, \qquad f_i^{(1)}(ts) = f_i^{(1)}(s)$$

and $f_{ij}^{(2)}(ts) = t^{-1} f_{ij}^{(2)}(s)$, for $s \in S$, $t > 0$ and $i, j = 1, \dots, \nu$.

Under the assumption $\theta_1 = \theta_2 = \dots = \theta_\nu$, using relation (A.3) and homogeneity of f, we obtain that

$$\begin{aligned} 2(\hat{D}_f - f(1)) = & \sum_{i,j=k+1}^{M} (\hat{\theta}_{1i} - \theta_{1i})(\hat{\theta}_{1j} - \theta_{1j}) f_{11}^{(2)}(1) \alpha_{ij} \\ & + 2 \sum_{m=2}^{\nu} \sum_{i=k+1}^{M} \sum_{j=k+1}^{M_0} (\hat{\theta}_{1i} - \theta_{1i})(\hat{\theta}_{mj} - \theta_{mj}) f_{m1}^{(2)}(1) \alpha_{ij} \\ & + \sum_{i,j=k+1}^{M_0} \sum_{m,l=2}^{\nu} (\hat{\theta}_{mi} - \theta_{mi})(\hat{\theta}_{lj} - \theta_{lj}) f_{ml}^{(2)}(1) \alpha_{ij} + 2R_n, \end{aligned}$$

with $\alpha_{ij} = \int \frac{\partial f_{\theta_1}}{\partial \theta_{1i}} \frac{\partial f_{\theta_1}}{\partial \theta_{1j}} \frac{1}{f_{\theta_1}} d\mu$.

The above relation can be written,

$$2(\hat{D}_f - f(1)) = (\hat{\beta} - \beta)^t B (\hat{\beta} - \beta) + 2R_n, \tag{A.4}$$

for $\beta^t = (\theta_{1(k+1)}, \dots, \theta_{1M_0}, \theta_{1(M_0+1)}, \dots, \theta_{1M}, \theta_{2(k+1)}, \dots, \theta_{2M_0}, \dots, \theta_{\nu(k+1)}, \dots, \theta_{\nu M_0})$, $\hat{\beta}$ the m.l.e. of β and B the matrix defined by (3.6).

As $n_i \to \infty$, with $\frac{n_i}{n} \to \lambda_i > 0$, for $i = 1, \dots, \nu$, we have

$$\sqrt{n}(\hat{\beta} - \beta) \xrightarrow{L} N(0, \Sigma_\beta^{-1}), \tag{A.5}$$

where Σ_β is obtained from Σ_{22}, given by (3.3), if we replace $\frac{n_i}{n}$ by λ_i, $i = 1, \dots, \nu$.

From Salicru *et al.* ((1994), p. 379), we have that $\text{rank}(A) \le M - k$, while $\text{rank}(H) \le \nu - 1$, because of $\sum_{i=1}^{\nu} f_{ij}^{(2)}(1) = 0$, $j = 1, \dots, \nu$.

Therefore

$$\text{rank}(B) = \text{rank}(A \otimes H) = \text{rank}(A)\,\text{rank}(H) \le (\nu - 1)(M - k)$$

and

$$\begin{aligned}\operatorname{rank}(B\Sigma_\beta^{-1}) &\le \min\{\operatorname{rank}(B), \operatorname{rank}(\Sigma_\beta^{-1})\}\\ &\le \min\{(\nu-1)(M-k), (M-k)+(\nu-1)(M_0-k)\}.\end{aligned}$$

This relation, with (A.4), (A.5) and the convergence $2nR_n \to 0$, in probability as $n_i \to \infty$, $i = 1, \dots, \nu$, completes the proof of the theorem.

PROOF OF THEOREM 3.2. The proof of part (a) is omitted. It can be obtained by a similar argument to that of the proof of Theorem 3.1(a). For part (b), following the steps of the proof of Theorem 3.1(b), we have

$$2(\hat{D}_f - f(1))/f_{11}^{(2)}(1) = \sum_{i,j=M_0+1}^{M} (\hat{\theta}_{1i} - \theta_{1i})(\hat{\theta}_{1j} - \theta_{1j})\alpha_{ij} + 2R_n,$$

with $\alpha_{ij} = \int \frac{\partial f_{\theta_1}}{\partial \theta_{1i}} \frac{\partial f_{\theta_1}}{\partial \theta_{1j}} \frac{1}{f_{\theta_1}} d\mu$ and $2nR_n \to 0$, in probability as $n \to \infty$. This relation can be written

$$2(\hat{D}_f - f(1))/f_{11}^{(2)}(1) = (\hat{\beta} - \beta)^t B(\hat{\beta} - \beta) + 2R_n, \tag{A.6}$$

with $\beta^t = (\theta_{1(M_0+1)}, \dots, \theta_{1M})$, $\hat{\beta}$ the m.l.e. of β and $B = {}^{M_0+1,M}_{M_0+1,M} I^F(\theta_1)$. Furthermore, as $n \to \infty$, we have

$$\sqrt{n}(\hat{\beta} - \beta) \xrightarrow{L} N(0, B^{-1}),$$

which in view of (A.6) completes the proof of the theorem.

REFERENCES

Cressie, N. and Read, T. R. C. (1984). Multinomial goodness of fit tests, *J. Roy. Statist. Soc. Ser. B*, **46**, 440–464.

Csiszar, I. (1967). Information type measures of difference of probability distributions and indirect observations, *Studia Sci. Math. Hungar.*, **2**, 299–318.

Gil, M. A. (1989). A note in stratification and gain in precision in estimating diversity from large samples, *Comm. Statist. Theory Methods*, **18**, 1521–1526.

Gyorfi, L. and Nemetz, T. (1978). f-dissimilarity: a generalization of the affinity of several distributions, *Ann. Inst. Statist. Math.*, **30**, 105–113.

Matusita, K. (1966). A distance and related statistics in multivariate analysis, *Multivariate Analysis* (ed. P. R. Krishnaiah), 187–200, Academic Press, New York.

Matusita, K. (1967). On the notion of affinity of several distributions and some of its applications, *Ann. Inst. Statist. Math.*, **19**, 181–192.

Menendez, M. L., Morales, D., Pardo, L. and Salicru, M. (1995). Asymptotic behaviour and statistical applications of divergence measures in multinomial populations: a unified study, *Statistical Papers*, **36**, 1–29.

Morales, D., Pardo, L., Salicru, M. and Menendez, M. L. (1994). Asymptotic properties of divergence statistics in a stratified random sampling and its applications to test statistical hypotheses, *J. Statist. Plann. Inference*, **38**, 201–222.

Nayak, T. K. (1985). On diversity measures based on entropy functions, *Comm. Statist. Theory Methods*, **14**, 203–215.

Papaioannou, P. C. and Kempthorne, O. (1971). On statistical information theory and related measures of information, ARL. Tech. Report 71-0059, Aerospace Research Laboratories, Wright-Patterson A. F. B., Ohio.

Pardo, L., Salicru, M., Menendez, M. L. and Morales, D. (1995). Divergence measures based on entropy functions and statistical inference, *Sankhyā Ser. B*, **57**, 315–337.

Salicru, M., Morales, D., Menendez, M. L. and Pardo, L. (1994). On the applications of divergence type measures in testing statistical hypotheses, *J. Multivariate Anal.*, **51**, 372–391.

Serfling, R. J. (1980). *Approximation theorems of mathematical statistics*, Wiley, New York.

Toussaint, G. T. (1974). Some properties of Matusita's measure of affinity of several distributions, *Ann. Inst. Statist. Math.*, **26**, 389–394.

Zografos, K. (1993). Asymptotic properties of Φ-divergence statistic and its applications in contingency tables, *International Journal of Mathematical and Statistical Sciences*, **2**, 5–21.

Zografos, K. (1994). Asymptotic distributions of estimated f-dissimilarity between populations in stratified random sampling, *Statist. Probab. Lett.*, **21**, 147–151.

Zografos, K., Ferentinos, K. and Papaioannou, T. (1990). Φ-divergence statistics: sampling properties and multinomial goodness of fit and divergence tests, *Comm. Statist. Theory Methods*, **19**, 1785–1802.