JOURNAL OF MULTIVARIATE ANALYSIS 51, 372–391 (1994)

On the Applications of Divergence Type Measures in Testing Statistical Hypotheses*

M. SALICRÚ

Departamento de Estadística, Avd. Diagonal 645, Universidad de Barcelona, 08028 Barcelona, Spain

D. MORALES

Departamento de Estadística e I.O., Facultad de Matemáticas, Universidad Complutense de Madrid, 28040 Madrid, Spain

M. L. MENÉNDEZ

Departamento de Matemática Aplicada, E.T.S. de Arquitectura, Universidad Politécnica de Madrid, 28040 Madrid, Spain

AND

L. PARDO

Departamento de Estadística e I.O., Facultad de Matemáticas, Universidad Complutense de Madrid, 28040 Madrid, Spain

The fundamentals of information theory and also their applications to testing statistical hypotheses have been known and available for some time. There is currently a new and heterogeneous development of statistical procedures, based on information measures, scattered through the literature. In this paper a unification is attained by consistent application of the concepts and properties of information theory. Our aim is to examine a wide range of divergence type measures and their applications to statistical inferences, with special emphasis on multinomial and multivariate normal distributions, The "maximum likelihood" and the "minimum discrepancy" principles are combined here in order to derive new approaches to the discrimination between two groups or populations. To study the asymptotic properties of divergence statistics, we propose a unified expression, called $(\underline{h}, \underline{\phi})$-divergence, which includes as particular cases most divergences. Under different

Received January 12, 1993; revised May 6, 1994.

AMS 1991 subject classifications: 62B10, 62E20.

Key words and phrases: $(\underline{h}, \underline{\phi})$-divergence, $(\underline{h}, \underline{\phi})$-divergence statistics, asymptotic distributions, multivariate normal distribution, multinomial distribution, testing statistical hypotheses.

* The research in this paper was supported in part by DGICYT Grants PB91-0387 and PB91-0155. Their financial support is gratefully acknowledged.

0047-259X/94 $6.00

assumptions it is shown that the asymptotic distributions of the $(\underline{h}, \phi)$-divergences are either normal or chi square. From the previous results a wide range of statistical hypotheses about the parameters of one or two populations can be tested. To help clarify the discussion and provide a simple illustration examples are given.

1. Introduction

Divergence measures play an important role in statistical theory, especially in large sample theories of estimation and testing. The underlying reason is that they are indices of statistical distance between probability distributions P and Q; the smaller these indices are the harder it is to discriminate between P and Q. Many divergence measures have been proposed since the publication of the paper of Kullback and Leibler (1951). Renyi (1961) gave the first generalization of Kullback–Leibler divergence, Jeffreys (1946) defined the J-divergences, Burbea and Rao (1982a, b) introduced the R-divergences, Sharma and Mittal (1977) the (r, s)-divergences, Csiszar (1967) the ϕ-divergences, Taneja (1989) the generalized J-divergences and the generalized R-divergences, and so on. In order to conduct a unified study of their statistical properties, here we propose a generalized divergence, called $(\underline{h}, \phi)$-divergence, which includes as particular cases the above mentioned divergence measures.

Let $(\mathfrak{X}, \beta_{\mathfrak{X}}, P_\theta)_{\theta \in \Theta}$ be a probability space, where Θ is an open subset of $\mathbb{R}^M$. We shall assume that there exists a generalized probability density function (p.d.f) $f_\theta(x)$ for the distribution P_θ with respect to a σ-finite measure μ. In this context, one obtains

$$H(\theta_1) \leqslant H(\theta_1 \parallel \theta_2), \tag{1.1}$$

where

$$H(\theta_1) = -\int_{\mathfrak{X}} f_{\theta_1}(x) \log f_{\theta_1}(x)\, d\mu(x) \tag{1.2}$$

is the Shannon entropy (Shannon, 1948), and

$$H(\theta_1 \parallel \theta_2) = -\int_{\mathfrak{X}} f_{\theta_1}(x) \log f_{\theta_2}(x)\, d\mu(x)$$

is the Kerridge inaccuracy (Kerridge, 1961). Inequality (1.1) is known as the Shannon–Gibbs inequality. The difference

$$D(\theta_1, \theta_2) = H(\theta_1 \parallel \theta_2) - H(\theta_1) = \int_{\mathfrak{X}} f_{\theta_1}(x) \log \frac{f_{\theta_1}(x)}{f_{\theta_2}(x)}\, d\mu(x) \tag{1.3}$$

is known as the Kullback–Leibler divergence (Kullback and Leibler, 1951). While Shannon entropy either quantifies the information that a random variable, X with p.d.f. $f_\theta(x)$, gives or measures the variability of X among the individual within the population, the Kullback–Leibler divergence is used in order to express in a quantitative way analogies and differences between two populations by means of their respective p.d.f. $f_{\theta_1}(x)$ and $f_{\theta_2}(x)$.

Csiszar (1967) generalized the Kullback–Leibler divergence as

$$D_\varphi(\theta_1, \theta_2) = \int_{\mathcal{X}} f_{\theta_2}(x)\, \varphi\left(\frac{f_{\theta_1}(x)}{f_{\theta_2}(x)}\right) d\mu(x),$$

where φ is a real valued convex function on $(0, \infty)$ being strictly convex in some point x, $0 < x < \infty$. Important φ-divergences are: Kullback and Leibler with $\varphi(x) = x \log x$, variational or statistical with $\varphi(x) = |x - 1|$, χ^2-divergence or Kagan with $\varphi(x) = (1 - x)^2$, Matusita with $\varphi(x) = (1 - x^a)^{1/a}$ $0 < a \leqslant 1$, Balakrishman and Sanghvi with $\varphi(x) = (x-1)^2/(x+1)$, Havrda and Charvat with $\varphi(x) = (x - x^s)/(1 - s)$. Further examples can be found in Vajda (1989).

Of course many other divergence measures not enumerated above can be found in the literature of information theory and sometimes one feels like asking for the following questions: Is there any reason for such a variety? Have all these measures been introduced having in mind a real problem? Due to historical type arguments, it is understandable to find works on different divergence measures where similar results are obtained and/or the same tools are employed. For these reasons, we have tried to give a very general functional, which can be used to conduct global studies instead of measure-to-measure individualized studies. The final purpose is to save time and work. So in this paper we propose a unified expression, called $(\underline{h}, \underline{\phi})$-divergence, as follows

$$D_{\underline{\phi}}^{\underline{h}}(\theta_1, \theta_2) = \int_A h_\alpha \left\{ \int_{\mathcal{X}} f_{\theta_2}(x)\, \phi_\alpha(f_{\theta_1}(x)/f_{\theta_2}(x))\, d\mu(x) - \phi_\alpha(1) \right\} d\eta(\alpha),$$

where $\underline{h} = (h_\alpha)_{\alpha \in A}$, $\underline{\phi} = (\phi_\alpha)_{\alpha \in A}$, ϕ_α and h_α are real valued C^2 functions with $h_\alpha(0) = 0$, and η is σ-finite measure on the measurable space (A, β). It is assumed that, for every $\alpha \in A$, either ϕ_α is convex and h_α increasing or ϕ_α is concave and h_α decreasing. Furthermore, we suppose that $\lim_{x \to 0^+} \phi_\alpha(x)$ and $\lim_{x \to 0^+} (\phi_\alpha(x)/\alpha)$ exist, in the extended sense, for all α, and $\underline{h}$ is a family of functions in order that the integral makes sense; i.e.,

$$\int_A h_\alpha\, d\eta = \int_A h_\alpha^+\, d\eta - \int_A h_\alpha^-\, d\eta,$$

provided at least one of the integrals is not ∞, where $h_\alpha^+ = \max\{h_\alpha, 0\}$ and $h_\alpha^- = \max\{-h_\alpha, 0\}$. In Morales *et al.* (1994) a table is given with the functions h_α and ϕ_α corresponding to many well known divergence measures.

In what follows, the following regularity assumptions hold:

(i) The set $A = \{x \in \mathfrak{X}/f(x, \theta) > 0\}$ does not depend on θ and for all $x \in A$, $\theta \in \Theta$

$$\frac{\partial f(x, \theta)}{\partial \theta_i}, \frac{\partial^2 f(x, \theta)}{\partial \theta_i \, \partial \theta_j}, \frac{\partial^3 f(x, \theta)}{\partial \theta_i \, \partial \theta_j \, \partial \theta_k}, \qquad i, j, k = 1, ..., M$$

exist and are finite.

(ii) There exist real valued functions $F(x)$ and $H(x)$ such that

$$\left|\frac{\partial f(x, \theta)}{\partial \theta_i}\right| < F(x), \qquad \left|\frac{\partial^2 f(x, \theta)}{\partial \theta_i \, \partial \theta_j}\right| < F(x), \qquad \left|\frac{\partial^3 f(x, \theta)}{\partial \theta_i \, \partial \theta_j \, \partial \theta_k}\right| < H(x),$$

where F is finitely integrable and $E[H(X)] < M$, with M independent of θ.

(iii) $$\left(E\left\{\frac{\partial \log f(X, \theta)}{\partial \theta_i} \frac{\partial \log f(X, \theta)}{\partial \theta_j}\right\}\right)_{i,j=1,...,M}$$

is finite and positive definite.

Kupperman (1957) established that the statistic based on the Kullback–Leibler measure of information

$$2nD(\hat{\theta}, \theta) = 2n \int_{\mathfrak{X}} f_{\hat{\theta}}(x) \log \frac{f_{\hat{\theta}}(x)}{f_\theta(x)} \, d\mu(x),$$

where $\hat{\theta} = (\hat{\theta}_1, ..., \hat{\theta}_M)$ is the maximum likelihood estimator of $\theta = (\theta_1, ..., \theta_M)$, is asymptotically chi-square distributed with M degrees of freedom (χ_M^2). Based on this result, the null hypothesis $H_0: \theta = \theta_0$ can be tested. If we now consider a K-variate normal distribution, i.e., $\theta = (\mu_i, \sigma_{ii}, \sigma_{ij}; i = 1, ..., K, j = 1, ..., K, j > i)$ with dimension $(K^2 + 3K)/2$, then we can test $H_0: (\mu, \Sigma) = (\mu_0, \Sigma_0)$, where $\mu = \mu_1, ..., \mu_K)'$ is the mean vector and $\Sigma = (\sigma_{ij})$ is the variance–covariance matrix.

For the two sample case, Kupperman (1957) established that if $\theta_1 = \theta_2$, then

$$\frac{2mn}{m+n} D(\hat{\theta}_1, \hat{\theta}_2) \xrightarrow[n \to \infty]{L} \chi_M^2,$$

where $\hat{\theta}_1 = (\hat{\theta}_{11}, ..., \hat{\theta}_{1M})$ and $\hat{\theta}_2 = (\hat{\theta}_{21}, ..., \hat{\theta}_{2M})$ are the maximum likelihood estimators of $\theta_1 = (\theta_{11}, ..., \theta_{1M})$ and $\theta_2 = (\theta_{21}, ..., \theta_{2M})$ based on independent samples of sizes n and m, respectively. Based on this result, the null hypothesis $H_0: \theta_1 = \theta_2$ can be tested. If we now consider a K-variate normal case, then we can test $H_0: (\mu_1, \Sigma_1) = (\mu_2, \Sigma_2)$, i.e., the complete homogeneity test. Clearly, there are some other possibilities not covered by the above two tests. For the one sample problem, consider $H_0: \Sigma = \Sigma_0$ (when μ is unknown) or $H_0: \mu = \mu_0$ (when Σ is unknown). For the two samples problem, consider $H_0: \mu_1 = \mu_2$ (when Σ_1 and Σ_2 are equal but unknown), $H_0: \Sigma_1 = \Sigma_2$ (when μ_1 and μ_2 are equal but unknown), $H_0: \mu_1 = \mu_2^*$ and $\Sigma_1 = \Sigma_2$ (when μ_2^* is a predicted value of μ_2), and $H_0: \mu_1 = \mu_2^*$ and $\sigma_{1ii} = \sigma_{2ii}$ (when σ_{1ij} and σ_{2ij} are equal but unknown for every $i < j$).

In this paper, we give an answer to the above tests on the basis of the $(\underline{h}, \phi)$-divergence functional. So, on one side we deal with new problems and on the other side we give a general procedure which could be used with almost every divergence measure. More concretely, we suppose that

$$\theta_1 = (\theta_{11}, ..., \theta_{1M})$$

is unknown and that

$$\theta_2 = (\theta_{21}, ..., \theta_{2k}, \theta_{2(k+1)}, ..., \theta_{2M_0}, \theta^*_{(M_0+1)}, ..., \theta^*_M)$$

is partially known. We assume that $\theta_{2i} = \theta_{1i}$ if $i \in I_1 = \{1, 2, ..., k\}$, θ_{2i} is unknown and different from θ_{1i} when $i \in I_2 = \{k+1, ..., M_0\}$ and θ_{2i} is known and equal to θ_i^* when $i \in I_3 = \{M_0 + 1, ..., M\}$.

Therefore the joint parameter space Γ is an open subset of $\mathbb{R}^{M+M_0-k}$, however, if we add the hypothesis $\theta_1 = \theta_2$, the joint parameter subspace Γ_0 is an open subset of $\mathbb{R}^{M_0}$. Moreover the elements of the parameter space $\gamma = (\gamma_1, ..., \gamma_{M+M_0-k}) \in \Gamma$ are as follows:

$$\gamma_i = \begin{cases} \theta_{1i} & \text{if } \ 1 \leqslant i \leqslant M \\ \theta_{2(i-M+k)} & \text{if } \ M+1 \leqslant i \leqslant M+M_0-k; \end{cases}$$

i.e., $\gamma = (\theta_{11}, ..., \theta_{1M}, \theta_{2(k+1)}, ..., \theta_{2M_0})$.

From each population independent random samples of sizes n and m respectively are drawn. Let $\hat{\theta}_{1i}$, $\hat{\theta}_{2j}$, $i = 1, ..., M$, $j = k+1, ..., M_0$, be the estimators which maximize the logarithm of the joint likelihood function

$$\log L(\gamma) = \sum_{i=1}^{n} \log f_{\theta_1}(x_i) + \sum_{i=1}^{m} \log f_{\theta_2}(y_i),$$

and let $\hat{\gamma} = (\hat{\gamma}_1, ..., \hat{\gamma}_{M+M_0-k})$ be the maximum likelihood estimator of γ; i.e.,

$$\hat{\gamma}_i = \begin{cases} \hat{\theta}_{1i} & \text{if } \ 1 \leqslant i \leqslant M \\ \hat{\theta}_{2(i-M+k)} & \text{if } \ M+1 \leqslant i \leqslant M+M_0-k. \end{cases}$$

In this paper the asymptotic distribution of $D_{\phi}^{h}(\hat{\theta}_1, \hat{\theta}_2)$ is obtained, where $\hat{\theta}_1 = (\hat{\theta}_{11}, ..., \hat{\theta}_{1M})'$, $\hat{\theta}_2 = (\hat{\theta}_{21}, ..., \hat{\theta}_{2k}, \hat{\theta}_{2(k+1)}, ..., \hat{\theta}_{2M_0}, \theta^*_{(M_0+1)}, ..., \theta^*_M)'$, $\hat{\theta}_{2i} = \hat{\gamma}_i$ if $1 \leqslant i \leqslant k$ and $\hat{\theta}_{2i} = \theta_i^*$ if $M_0 + 1 \leqslant i \leqslant M$. We also obtain the asymptotic distributions of $D_{\phi}^{h}(\hat{\theta}_1, \hat{\theta}_2)$ in some particular but very important cases. Applications of these results in testing statistical hypotheses are presented. Examples of multinomial and multivariate normal distributions are given.

2. Asymptotic Distribution of $D_{\phi}^{h}(\hat{\theta}_1, \hat{\theta}_2)$

Consider the function

$$H_\alpha(\gamma) = h_\alpha \left\{ \int_{\mathfrak{X}} f_{\theta_2}(x)\, \phi_\alpha(f_{\theta_1}(x)/f_{\theta_2}(x))\, d\mu(x) - \phi_\alpha(1) \right\}.$$

A Taylor's expansion of $H_\alpha(\hat{\gamma})$ around γ yields

$$D_{\phi}^{h}(\hat{\theta}_1, \hat{\theta}_2) = D_{\phi}^{h}(\theta_1, \theta_2) + T'(\hat{\gamma} - \gamma) + R_{nm},$$
$$\text{where} \quad T' = (t_1, ..., t_{M+M_0-k}),$$

with

$$t_i = \int_A \Bigg(h'_\alpha \left\{ \int_{\mathfrak{X}} f_{\theta_2}(x)\, \phi_\alpha(f_{\theta_1}(x)/f_{\theta_2}(x))\, d\mu - \phi_\alpha(1) \right\}$$
$$\times \int_{\mathfrak{X}} \left(\frac{\partial f_{\theta_2}(x)}{\partial \theta_{1i}} \phi_\alpha(f_{\theta_1}(x)/f_{\theta_2}(x)) + \frac{\partial f_{\theta_1}(x)}{\partial \theta_{1i}} \phi'_\alpha(f_{\theta_1}(x)/f_{\theta_2}(x)) \right.$$
$$\left. - \phi'_\alpha(f_{\theta_1}(x)/f_{\theta_2}(x)) \frac{\partial f_{\theta_2}(x)}{\partial \theta_{1i}} \frac{f_{\theta_1}(x)}{f_{\theta_2}(x)} \right) d\mu \Bigg)\, d\eta \qquad \text{if} \quad 1 \leqslant i \leqslant k,$$

$$t_i = \int_A \Bigg(h'_\alpha \left\{ \int_{\mathfrak{X}} f_{\theta_2}(x)\, \phi_\alpha(f_{\theta_1}(x)/f_{\theta_2}(x))\, d\mu - \phi_\alpha(1) \right\}$$
$$\times \int_{\mathfrak{X}} \left(\frac{\partial f_{\theta_1}(x)}{\partial \theta_{1i}} \phi'_\alpha(f_{\theta_1}(x)/f_{\theta_2}(x))\, d\mu \right) \Bigg)\, d\eta \qquad \text{if} \quad k+1 \leqslant i \leqslant M,$$

$$t_i = \int_A \Bigg(h'_\alpha \left\{ \int_{\mathfrak{X}} f_{\theta_2}(x)\, \phi_\alpha(f_{\theta_1}(x)/f_{\theta_2}(x))\, d\mu - \phi_\alpha(1) \right\}$$
$$\times \int_{\mathfrak{X}} \left(\frac{\partial f_{\theta_2}(x)}{\partial \theta_{2,i-M+k}} \phi_\alpha(f_{\theta_1}(x)/f_{\theta_2}(x)) - \phi'_\alpha(f_{\theta_1}(x)/f_{\theta_2}(x)) \right.$$
$$\left. \times \frac{f_{\theta_2}(x)}{\partial \theta_{2,i-M+k}} \frac{f_{\theta_1}(x)}{f_{\theta_2}(x)} \right) d\mu \Bigg)\, d\eta \qquad \text{if} \quad M+1 \leqslant i \leqslant M+M_0-k.$$

If $m/(m+n) \xrightarrow[m,n\to\infty]{} \lambda \in (0,1)$, then

$$\left(\frac{mn}{m+n}\right)^{1/2} (\hat{\gamma}-\gamma) \xrightarrow[m,n\to\infty]{L} N(0, \Sigma(\theta_1,\theta_2)^{-1}),$$

where

$$\Sigma(\theta_1,\theta_2) = \begin{pmatrix} \frac{1}{\lambda}\, {}^{1,k}_{1,k}I_F(\theta_1) + \frac{1}{1-\lambda}\, {}^{1,k}_{1,k}I_F(\theta_2) & \frac{1}{\lambda}\, {}^{1,k}_{k+1,M}I_F(\theta_1) & \frac{1}{1-\lambda}\, {}^{1,k}_{k+1,M_0}I_F(\theta_2) \\ \frac{1}{\lambda}\, {}^{k+1,M}_{1,k}I_F(\theta_1) & \frac{1}{\lambda}\, {}^{k+1,M}_{k+1,M}I_F(\theta_1) & 0 \\ \frac{1}{1-\lambda}\, {}^{k+1,M_0}_{1,k}I_F(\theta_2) & 0 & \frac{1}{1-\lambda}\, {}^{k+1,M_0}_{k+1,M_0}I_F(\theta_2) \end{pmatrix},$$

${}^{i,j}_{r,s}I_F(\theta)$ is the $(j-i+1)\times(r-s+1)$ submatrix of $I_F(\theta)$ which have the rows i, $i+1, ..., j$ and the columns $r, r+1, ..., s$ and $I_F(\theta)$ is the Fisher information matrix associated to θ. Therefore

$$\left(\frac{mn}{m+n}\right)^{1/2} (D^h_\phi(\hat{\theta}_1,\hat{\theta}_2) - D^h_\phi(\theta_1,\theta_2)) \xrightarrow[m,n\to\infty]{L} N(0, T'\Sigma(\theta_1,\theta_2)^{-1}\, T),$$

provided $T'\Sigma(\theta_1,\theta_2)^{-1}\, T>0$, because $(mn/(m+n))^{1/2}\, R_{nm}$ converges in probability to zero.

If $\theta_1=\theta_2$, then $T'\Sigma(\theta_1,\theta_2)^{-1}\, T=0$. Therefore using again a Taylor's expansion of $H_\alpha(\hat{\gamma})$ we get

$$\begin{aligned} \frac{2D^h_\phi(\hat{\theta}_1,\hat{\theta}_2)}{\int_\Lambda h'_\alpha(0)\,\phi''_\alpha(1)\,d\eta} &= \sum_{i,j=k+1}^{M} a_{ij}(\hat{\theta}_{1i}-\theta_{1i})(\hat{\theta}_{1j}-\theta_{1j}) \\ &\quad -2\sum_{i=k+1}^{M}\sum_{j=k+1}^{M_0} a_{ij}(\hat{\theta}_{1i}-\theta_{1i})(\hat{\theta}_{2j}-\theta_{2j}) \\ &\quad +\sum_{i,j=k+1}^{M_0} a_{ij}(\hat{\theta}_{2i}-\theta_{2i})(\hat{\theta}_{2j}-\theta_{2j}) + R_{nm} \\ &= (\hat{\beta}-\beta)'\, A(\hat{\beta}-\beta) + R_{nm}, \end{aligned}$$

where

$$\hat{\beta} = (\hat{\theta}_{1(k+1)}, ..., \hat{\theta}_{1M}, \hat{\theta}_{2(k+1)}, ..., \hat{\theta}_{2M_0})$$

and

$$\beta = (\theta_{1(k+1)}, ..., \theta_{1M}, \theta_{2(k+1)}, ..., \theta_{2M_0})$$

are vectors of dimension $M+M_0-2k$ and A is a partitioned matrix of dimension $(M+M_0-2k)\times(M+M_0-2k)$, which is given by

$$A=\begin{pmatrix} {}^{k+1,M_0}_{k+1,M_0}I_F(\theta) & {}^{k+1,M_0}_{M_0+1,M}I_F(\theta) & -({}^{k+1,M_0}_{k+1,M_0}I_F(\theta)) \\ {}^{M_0+1,M}_{k+1,M_0}I_F(\theta) & {}^{M_0+1,M}_{M_0+1,M}I_F(\theta) & -({}^{M_0+1,M}_{k+1,M_0}I_F(\theta)) \\ -({}^{k+1,M_0}_{k+1,M_0}I_F(\theta)) & -({}^{k+1,M_0}_{M_0+1,M}I_F(\theta)) & {}^{k+1,M_0}_{k+1,M_0}I_F(\theta) \end{pmatrix}.$$

If we assume that $m/(m+n)\xrightarrow[m,n\to\infty]{}\lambda\in(0,1)$ then

$$\left(\frac{nm}{n+m}\right)^{1/2}(\hat{\beta}-\beta)\xrightarrow[n,m\to\infty]{L}N(0,\Sigma_\beta),$$

where

$$\Sigma_\beta=\begin{pmatrix} \lambda\,{}^{k+1,M}_{k+1,M}I_F(\theta_2)^{-1} & 0 \\ 0 & (1-\lambda)\,{}^{k+1,M_0}_{k+1,M_0}I_F(\theta_2)^{-1} \end{pmatrix}.$$

Therefore

$$\frac{2nm}{m+n}\frac{D^h_\phi(\hat{\theta}_1,\hat{\theta}_2)}{\int_A h'_x(0)\,\phi''_x(1)\,d\eta}\xrightarrow[n,m\to\infty]{L}\sum_{i=1}^{M+M_0-2k}\beta_i\chi_1^2,$$

where the β_is are the eigenvalues of the matrix $A\Sigma_\beta$ and the χ_1^2s are independent.

Let us observe that in the matrix A, the last M_0-k columns (rows) are exactly the first M_0-k first columns (rows) with opposite sign. Therefore the number of linearly independent columns (rows) is less than or equal to $M-k$, hence the rank of A is less than or equal to $M-k$. Moreover

$$r(A\Sigma_\beta)\leqslant\min(r(A),r(\Sigma_\beta)))\leqslant\min\{M-k,M+M_0-2k\}=M-k,$$

since $k\leqslant M_0$. For this reason, the maximum number of non-null eigenvalues of $A\Sigma_\beta$ is $M-k$. So we get the following result:

THEOREM 1. *Assume the regularity conditions* (i)–(iii) *hold.*

(a) *If* $m/(m+n)\xrightarrow[m,n\to\infty]{}\lambda\in(0,1)$ *and* $T'\Sigma(\theta_1,\theta_2)^{-1}T>0$, *then*

$$\left(\frac{mn}{m+n}\right)^{1/2}\left(D^h_{\underline{\phi}}(\hat{\theta}_1,\hat{\theta}_2)-D^h_{\underline{\phi}}(\theta_1,\theta_2)\right)\xrightarrow[n,m\to\infty]{L}N(0,T'\Sigma(\theta_1,\theta_2)^{-1}T).$$

(b) *If* $\theta_1=\theta_2$, *then*

$$\frac{2nm}{m+n}\frac{D^h_{\underline{\phi}}(\hat{\theta}_1,\hat{\theta}_2)}{\int_A h'_x(0)\,\phi''_x(1)\,d\eta}\xrightarrow[n,m\to\infty]{L}\sum_{i=1}^{M-k}\beta_i\chi_1^2,$$

where the χ_1^2*s are independent and the* β_i*s are the non-null eigenvalues of the matrix* $A\Sigma_\beta$, *provided that* $\int_A h'_\alpha(0)\,\phi''_\alpha(1)\,d\eta \neq 0$.

To prove Theorem 1a (1b), we suppose that H_α has continuous second (third) partial derivatives verifying regularity conditions (such as dominated convergence theorem) which enable us to interchange the symbols $\lim_{n,m\to\infty}$ and $\int_A$. Note that, as A is finite, this assumption holds for every divergence measure appearing in Section 1.

An important particular case of this theorem appears when $M_0 = M$, i.e., when the parameter of the second distribution is

$$\theta_2 = (\theta_{21}, ..., \theta_{2k}, \theta_{2(k+1)}, ..., \theta_{2M}),$$

where $\theta_{21}, ..., \theta_{2k}$ are unknown and equal to $\theta_{11}, ..., \theta_{1k}$, while $\theta_{2(k+1)}, ..., \theta_{2M}$ are unknown and different, in general, from the parameter in the first distribution.

COROLLARY 1. *Under the regularity conditions* (i)–(iii), *if the k first components of the parameters* θ_1 *and* θ_2 *are equal, then*

(a) *If* $m/(m+n) \xrightarrow[m,n\to\infty]{} \lambda \in (0,1)$ *and* $T'\Sigma(\theta_1,\theta_2)^{-1}\,T > 0$, *then*

$$\left(\frac{mn}{m+n}\right)^{1/2} (D^h_\phi(\hat\theta_1,\hat\theta_2) - D^h_\phi(\theta_1,\theta_2)) \xrightarrow[n,m\to\infty]{L} N(0, T'\Sigma(\theta_1,\theta_2)^{-1}\,T),$$

where $T = (t_1, ..., t_{2M-k})^t$ *is given in Theorem* 1.

(b) *If* $m/(m+n) \xrightarrow[m,n\to\infty]{} \lambda \in (0,1)$, $\int_A h'_\alpha(0)\,\phi''_\alpha(1)\,d\eta \neq 0$ *and* $\theta_1 = \theta_2$, *then*

$$\frac{2nm}{m+n}\,\frac{D^h_\phi(\hat\theta_1,\hat\theta_2)}{\int_A h'_\alpha(0)\,\phi''_\alpha(1)\,d\eta} \xrightarrow[n,m\to\infty]{L} \chi^2_{M-k}.$$

Proof. (a) Immediate, taking $M_0 = M$ in Theorem 1.

(b) In this case, it is easy to check that the matrix $A\Sigma_\beta$ is idempotent and its trace is $M-k$; therefore, $(\hat\beta-\beta)^t A(\hat\beta-\beta)$ has a chi-square distribution with $M-k$ degrees of freedom.

Another interesting case appears when $M_0 = M$ and $k = 0$; then we have the following result:

COROLLARY 2. *Let* $\hat\theta_1$ *and* $\hat\theta_2$ *be the maximum likelihood estimators of* θ_1 *and* θ_2, *and suppose that* (i–iii) *hold; then*

(a) *If* $m/(m+n) \xrightarrow[m,n\to\infty]{} \lambda \in (0,1)$ *and* $\lambda T' I_F(\theta_1)^{-1} T + (1-\lambda) S' I_F(\theta_2)^{-1} S > 0$, *then*

$$\left(\frac{mn}{m+n}\right)^{1/2} (D_\phi^h(\hat\theta_1, \hat\theta_2) - D_\phi^h(\theta_1, \theta_2))$$

$$\xrightarrow[n,m\to\infty]{L} N(0, \lambda T' I_F(\theta_1)^{-1} T + (1-\lambda) S' I_F(\theta_2)^{-1} S),$$

where $T = (t_1, ..., t_M)'$ *and* $S = (s_1, ..., s_M)'$ *with*

$$t_i = \int_\Lambda \left(h'_\alpha \left\{ \int_{\mathfrak{X}} f_{\theta_2}(x)\, \phi_\alpha(f_{\theta_1}(x)/f_{\theta_2}(x))\, d\mu - \phi_\alpha(1) \right\} \times \int_{\mathfrak{X}} \left(\frac{\partial f_{\theta_1}(x)}{\partial \theta_{1i}} \phi'_\alpha(f_{\theta_1}(x)/f_{\theta_2}(x))\, d\mu \right) \right) d\eta$$

and

$$s_i = \int_\Lambda \left(h'_\alpha \left\{ \int_{\mathfrak{X}} f_{\theta_2}(x)\, \phi_\alpha(f_{\theta_1}(x)/f_{\theta_2}(x))\, d\mu - \phi_\alpha(1) \right\} \times \int_{\mathfrak{X}} \left(\frac{\partial f_{\theta_2}(x)}{\partial \theta_{2i}} \phi_\alpha(f_{\theta_1}(x)/f_{\theta_2}(x)) - \phi'_\alpha(f_{\theta_1}(x)/f_{\theta_2}(x)) \frac{\partial f_{\theta_2}(x)}{\partial \theta_{2i}} \frac{f_{\theta_1}(x)}{f_\theta(x)} \right) d\mu \right) d\eta.$$

(b) *If* $m/(m+n) \xrightarrow[m,n\to\infty]{} \lambda \in (0,1)$, $\int_\Lambda h'_\alpha(0)\, \phi''_\alpha(1)\, d\eta \neq 0$ *and* $\theta_1 = \theta_2$, *then*

$$\frac{2mn}{m+n} \frac{D_\phi^h(\hat\theta_1, \hat\theta_2)}{\int_\Lambda h'_\alpha(0)\, \phi''_\alpha(1)\, d\eta} \xrightarrow[n,m\to\infty]{L} \chi^2_M.$$

Proof. Taking $k=0$ in Corollary 1, the result follows.

Another important case appears when $M_0 = k$. In this case we only have to observe a sample of size n in the first population and a new theorem must be derived in a similar way to Theorem 1. Also note that the first M_0 components of θ_2 coincide with the corresponding components in θ_1, while the last $M - M_0$ components are known. Let us define $\theta = \theta_1$ and $\theta^* = \theta_2$; we state without proof the following result.

THEOREM 2. *Let* $\hat\theta = (\hat\theta_1, ..., \hat\theta_M)$ *and* $\hat\theta^* = (\hat\theta_1, ..., \hat\theta_{M_0}, \theta^*_{M_0+1}, ..., \theta^*_M)$ *be the maximum likelihood estimators of* $\theta = (\theta_1, ..., \theta_M)$ *and* $\theta^* = (\theta_1, ..., \theta_{M_0}, \theta^*_{M_0+1}, ..., \theta^*_M)$, *respectively, based on a random sample of size* n *from* $f_\theta(x)$. *Assume that the regularity conditions* (i)–(iii) *hold; then*

(a) $n^{1/2}(D^h_\phi(\hat\theta, \hat\theta^*) - D^h_\phi(\theta, \theta^*)) \xrightarrow[n\to\infty]{L} N(0, T' I_F(\theta)^{-1} T)$, *when* $T' I_F(\theta)^{-1} T > 0$, *where* $T = (t_1, ..., t_M)'$, *with*

$$t_i = \int_A \left(h'_\alpha \left\{ \int_{\mathcal{X}} f_{\theta^*}(x)\, \phi_\alpha(f_\theta(x)/f_{\theta^*}(x))\, d\mu - \phi_\alpha(1) \right\} \right.$$

$$\times \int_{\mathcal{X}} \left(\frac{\partial f_{\theta^*}(x)}{\partial \theta_i} \phi_\alpha(f_\theta(x)/f_{\theta^*}(x)) + \frac{\partial f_\theta(x)}{\partial \theta_i} \phi'_\alpha(f_\theta(x)/f_{\theta^*}(x)) \right.$$

$$\left. \left. - \phi'_\alpha(f_\theta(x)/f_{\theta^*}(x)) \frac{\partial f_\theta(x)}{\partial \theta_i} \frac{f_\theta(x)}{f_{\theta^*}(x)} \right) d\mu \right) d\eta \qquad \text{if} \quad 1 \leqslant i \leqslant M_0,$$

$$t_i = \int_A \left(h'_\alpha \left\{ \int_{\mathcal{X}} f_{\theta^*}(x)\, \phi_\alpha(f_\theta(x)/f_{\theta^*}(x))\, d\mu - \phi_\alpha(1) \right\} \right.$$

$$\left. \times \int_{\mathcal{X}} \left(\frac{\partial f_\theta(x)}{\partial \theta_i} \phi'_\alpha(f_\theta(x)/f_{\theta^*}(x))\, d\mu \right) \right) d\eta \qquad \text{if} \quad M_0 + 1 \leqslant i \leqslant M.$$

(b) *If* $\theta = \theta^*$ *and* $\int_A h'_\alpha(0)\, \phi''_\alpha(1)\, d\eta \neq 0$, *then*

$$\frac{2n\, D^h_\phi(\hat\theta, \hat\theta^*)}{\int_A h'_\alpha(0)\, \phi''_\alpha(1)\, d\eta} \xrightarrow[n\to\infty]{L} \chi^2_{M-M_0}.$$

Another important case appears when $M_0 = k = 0$. In this case we only have to observe a sample of size n from the first population. Also note that θ_2 is completely known. Let us define $\theta = \theta_1$ and $\theta_0 = \theta_2$, so we have the following result:

COROLLARY 3. *Assume that the regularity conditions* (i)–(iii) *hold. If* $\hat\theta$ *is the maximum likelihood estimator of* θ *and* θ_0 *is known*; *then*

(a) $n^{1/2}(D^h_\phi(\hat\theta, \theta_0) - D^h_\phi(\theta, \theta_0)) \xrightarrow[n\to\infty]{L} N(0, T' I_F(\theta_1)^{-1} T)$, *where* $T = (t_1, ..., t_M)'$ *is defined in Corollary* 2a *and* $T' I_F(\theta_1)^{-1} T > 0$.

(b) *If* $\theta = \theta_0$ *and* $\int_A h'_\alpha(0)\, \phi''_\alpha(1)\, d\eta \neq 0$, *then*

$$\frac{2n\, D^h_\phi(\hat\theta, \theta_0)}{\int_A h'_\alpha(0)\, \phi''_\alpha(1)\, d\eta} \xrightarrow[n\to\infty]{L} \chi^2_M,$$

Proof. Taking $M_0 = 0$ in Theorem 2, the result is immediate.

3. STATISTICAL APPLICATIONS

The previous results giving the asymptotic distribution of the D^h_ϕ-divergence statistics in random sampling can be used in various settings to construct confidence intervals and to test statistical hypotheses based on one or more samples. First, we give some examples with one sample:

(1) To test the hypothesis that the divergence between θ and θ_0, a predicted value of θ available beforehand to the experimenter, is of a certain magnitude D_0, i.e., $H_0: D_{\phi}^{h}(\theta, \theta_0) = D_0$, we can use the statistic

$$Z_1 = \frac{n^{1/2}(D_{\phi}^{h}(\hat{\theta}, \theta_0) - D_0)}{\hat{\sigma}},$$

which has approximately a standard normal distribution under H_0 for sufficiently large n, and $\hat{\sigma}$ is obtained from Corollary 3a by replacing θ by its maximum likelihood estimator $\hat{\theta}$ in $(T^t I_F^{-1}(\theta) T)^{1/2}$.

(2) To test the hypothesis $H_0: \theta = \theta_0$, where θ_0 is a given value of the parameter, we consider the statistic T_1 given in Corollary 3b, whose asymptotic probability distribution function under H_0 is a chi-square with M degrees of freedom. If H_0 is true, then T_1 will be small. Thus a large value of T_1 indicates data less compatible with the null hypothesis. Hence for large n, when $T_1 = t$, one must reject H_0 at a level α if $P(\chi_M^2 > t) \leqslant \alpha$.

(3) Let $\theta = (\theta_1, ..., \theta_M)$ be the unknown true value of the parameter and let $\theta^* = (\theta_1, ..., \theta_{M_0}, \theta^*_{M_0+1}, ..., \theta^*_M)$ be a value of θ having the same first M_0 components as θ and with its last $M - M_0$ components being prespecified values of $\theta_{M_0+1}, ..., \theta_M$. To test the hypothesis that the divergence between θ and θ^* is of a certain magnitude D_0; i.e., $H_0: D_{\phi}^{h}(\theta, \theta^*) = D_0$, we can use the statistic

$$Z_2 = \frac{n^{1/2}(D_{\phi}^{h}(\hat{\theta}, \hat{\theta}^*) - D_0}{\hat{\sigma}},$$

which has approximately a standard normal distribution under H_0 for sufficiently large n, and $\hat{\sigma}$ is obtained from Theorem 2a by replacing θ by its maximum likelihood estimator $\hat{\theta}$.

(4) Suppose the conditions given in (3) hold. To test the hypothesis $H_0: \theta = \theta^*$, we use the statistic T_2 given in Theorem 2b. Hence for large n, when $T_2 = t$, one must reject H_0 at a level α if $P(\chi_{M-M_0}^2 > t) \leqslant \alpha$.

We shall now show some examples of two sample tests.

(5) To test the hypothesis that the divergence between θ_1 and θ_2 is of a certain magnitude D_0, i.e., $H_0: D_{\phi}^{h}(\theta_1, \theta_2) = D_0$, we can use the statistic

$$Z_3 = \frac{\left(\frac{mn}{m+n}\right)^{1/2} (D_{\phi}^{h}(\hat{\theta}_1, \hat{\theta}_2) - D_0)}{\hat{\sigma}},$$

which has approximately a standard normal distribution under H_0 for sufficiently large n and m, and $\hat{\sigma}$ is obtained from Corollary 2a by replacing θ_1 and θ_2 by their maximum likelihood estimators $\hat{\theta}_1$ and $\hat{\theta}_2$ in $(\lambda T' I_F^{-1}(\theta_1) T + (1-\lambda) S' I_F^{-1}(\theta_2) S)^{1/2}$.

(6) To test the hypothesis $H_0: \theta_1 = \theta_2$, we consider the statistic T_3 given in Corollary 2b, whose asymptotic probability distribution function under H_0 is a chi-square with M degrees of freedom. For large n and m, when $T_3 = t$, we reject H_0 at a level α if $P(\chi^2_M > t) \leqslant \alpha$.

(7) Let $\theta_{11} = \theta_{21}, ..., \theta_{1k} = \theta_{2k}$ be unknown but equal components of θ_1 and θ_2, respectively. To test the hypothesis that the divergence between θ_1 and θ_2 is of a certain magnitude, i.e., $H_0: D^h_\phi(\theta_1, \theta_2) = D_0$, we can use the statistic

$$Z_4 = \frac{\left(\dfrac{mn}{m+n}\right)^{1/2} (D^h_\phi(\hat{\theta}_1, \hat{\theta}_2) - D_0)}{\hat{\sigma}},$$

where $\hat{\sigma}$ is obtained from Corollary 1a by replacing θ_1 and θ_2 by $\hat{\theta}_1$ and $\hat{\theta}_2$, respectively, in $(T' \Sigma(\theta_1, \theta_2)^{-1} T)^{1/2}$.

(8) Suppose the conditions given in (7) hold. To test the hypothesis $H_0: \theta_1 = \theta_2$, we use the statistic T_4 given in Corollary 1b. Hence for large n and m, when $T_4 = t$, one must reject H_0 at a level α if $P(\chi^2_{M-k} > t) \leqslant \alpha$.

(9) Let $\theta_{11} = \theta_{21}, ..., \theta_{1k} = \theta_{2k}$ be unknown but equal components of θ_1 and θ_2, respectively. Let $\theta^*_{2,M_0+1}, ..., \theta^*_{2,M}$ be predicted values of $\theta_{2,M_0+1}, ..., \theta_{2,M}$, respectively, available beforehand to the experimenter. To test the hypothesis that the divergence between θ_1 and θ_2 is of a certain magnitude, i.e., $H_0: D^h_\phi(\theta_1, \theta_2) = D_0$, we can use the statistic

$$Z_5 = \frac{\left(\dfrac{mn}{m+n}\right)^{1/2} (D^h_\phi(\hat{\theta}_1, \hat{\theta}_2) - D_0)}{\hat{\sigma}},$$

where $\hat{\sigma}$ is obtained from Theorem 1a by replacing θ_1 and θ_2 by $\hat{\theta}_1$ and $\hat{\theta}_2$, respectively, in $(T' \Sigma(\theta_1, \theta_2)^{-1} T)^{1/2}$.

(10) Suppose the conditions given in (9) hold. To test the hypothesis $H_0: \theta_1 = \theta_2$, we can use the statistic T_5 given in Theorem 1b. Hence for large n, when $T_5 = t$, one must reject H_0 at a level α if $P(\sum_{i=1}^{M-M_0} \beta_i \chi^2_1 > t) \leqslant \alpha$, where the χ^2_1s are independent and the $\hat{\beta}_i$s are the eigenvalues of the matrix obtained by replacing $\theta = \theta_1 = \theta_2$ by $\hat{\theta}$ (e.g., the maximum likelihood estimator based on the joint sample of size $n+m$) in $A\Sigma_\beta$.

Remark 1. In (2), (4), (6), and (8), the degrees of freedom of the asymptotic chi-square distributions are all equal to $\dim(\Theta) - \dim(\Theta_0)$.

Remark 2. If for given α, $0 \leqslant \alpha \leqslant 1$, a nonrandomized Neyman–Pearson test of a simple hypothesis against a simple alternative exists, then there exists an equivalent D_ϕ^h-divergence test statistic. To prove this statement, consider the following estimator of $\theta \in \Theta = \{\theta_0, \theta_1\}$:

$$\hat{\theta} = \begin{cases} \theta_1 & \text{if } f(x_1, ..., x_n/\theta_1) \geqslant \lambda f(x_1, ..., x_n/\theta_0) \\ \theta_0 & \text{if } f(x_1, ..., x_n/\theta_1) < \lambda f(x_1, ..., x_n/\theta_0) \end{cases}.$$

We reject the null hypothesis if $D_\phi^h(\hat{\theta}, \theta_0) > c$ for some proper $c > 0$, but

$$D_\phi^h(\hat{\theta}, \theta_0) = \begin{cases} D_\phi^h(\hat{\theta}, \theta_0) & \text{if } f(x_1, ..., x_n/\theta_1) \geqslant \lambda f(x_1, ..., x_n/\theta_0) \\ 0 & \text{if } f(x_1, ..., x_n/\theta_1) < \lambda f(x_1, ..., x_n/\theta_0) \end{cases}.$$

So the decision rule is

$$\varphi(x_1, ..., x_n) = \begin{cases} 1 & \text{if } f(x_1, ..., x_n/\theta_1) \geqslant \lambda f(x_1, ..., x_n/\theta_0) \\ 0 & \text{if } f(x_1, ..., x_n/\theta_0) < \lambda f(x_1, ..., x_n/\theta_0) \end{cases},$$

where $\lambda > 0$ is chosen to verify that $\alpha = E_{\theta_0}[\varphi(X_1, ..., X_n)]$.

Remark 3. If the D_ϕ^h-divergence test statistic is constructed on the basis of maximum likelihood estimators, as it happens in Section 2, then, for testing $\theta \in \Theta_0$ against $\theta \in \Theta_1$, this statistic is a function of every sufficient statistic for θ.

To illustrate the above results we consider the following example

EXAMPLE 1. (Seal 1964, p. 106). Measures of cranial length (x_1) and cranial breadth (x_2) on a sample of 35 mature female frogs led to the following statistics

$$\bar{x}_1 = \begin{pmatrix} 22.860 \\ 24.397 \end{pmatrix} \quad \text{and} \quad S_1 = \begin{pmatrix} 17.178 & 19.710 \\ 19.710 & 23.710 \end{pmatrix},$$

where S_1 is the maximum likelihood estimator of the variance–covariance matrix Σ_1. Similar measurements on 14 male frogs led to the statistics

$$\bar{x}_2 = \begin{pmatrix} 21.821 \\ 22.843 \end{pmatrix} \quad \text{and} \quad S_2 = \begin{pmatrix} 17.159 & 17.731 \\ 17.731 & 19.273 \end{pmatrix},$$

where we have assumed that both populations are bivariate normally distributed.

We will use the r order and s degree divergence, $D_r^s(\theta_1, \theta_2)$, with $r=0.5$ and $s=2$, to test the following hypotheses:

(a) $H_0: \mu_1 = \mu_2$ given that $\Sigma_1 = \Sigma_2$.

(b) $H_0: \mu_1 = \mu_2$ and $\Sigma_1 = \Sigma_2$ (test of complete homogeneity).

(a) In this case the maximum likelihood estimator of μ_i under the alternative hypothesis is $\bar{x}_i$, the ith sample mean $(i=1, 2)$, and the maximum likelihood estimator of the common variance–covariance matrix is

$$S = \frac{nS_1 + mS_2}{n+m} = \begin{pmatrix} 17.173 & 19.145 \\ 19.145 & 22.442 \end{pmatrix}$$

and

$$S^{-1} = \begin{pmatrix} 1.18958 & -1.01482 \\ -1.01482 & 0.91028 \end{pmatrix}.$$

The expression of the (r, s)-divergence, obtained from $(\underline{h}, \phi)$-divergence for $\Lambda = \{1\}$, $\eta(1)=1$, $h_1(x) = (1/(s-1))[(x+1)^{(s-1)/(r-1)} - 1]$ and $\phi_1(x) = x^r$, $r, s > 0$, $r, s \neq 1$, between two K-variate normal distributions, see Pardo *et al.* (1992), is

$$D_r^s((\mu_1, \Sigma_1), (\mu_2, \Sigma_2))$$
$$= (s-1)^{-1}\left(\exp\left\{\frac{r(s-1)}{2}(\mu_1-\mu_2)'\,[r\Sigma_2 + (1-r)\,\Sigma_1]^{-1}(\mu_1-\mu_2)\right\}\right.$$
$$\left.\cdot\frac{|r\Sigma_2 + (1-r)\,\Sigma_1|^{(1-s)/2(r-1)}}{|\Sigma_1|^{(s-1)/2}\,|\Sigma_2|^{(1-s)r/2(r-1)}} - 1\right), \qquad s \neq 1, \quad r \neq 1, \quad r > 0.$$

As $\int_\Lambda h'_\alpha(0)\,\phi''_\alpha(1)\,d\eta = r$, we have to evaluate the test statistic given in Corollary 1b, i.e.,

$$T_4 = \frac{2nm}{r(m+n)} D_r^s((\bar{x}_1, S), (\bar{x}_2, S))$$
$$= 40\left(\exp\left\{0.25(1.039, 1.554)\begin{pmatrix} 1.18958 & -1.01482 \\ -1.01482 & 0.91028 \end{pmatrix}\begin{pmatrix} 1.039 \\ 1.554 \end{pmatrix}\right\} - 1\right)$$
$$= 2.107.$$

Since $\chi^2_{2,0.05} = 5.94$, we conclude that μ_1 is equal to μ_2 when $\Sigma_1 = \Sigma_2$. Also oberve that from the proposed test a new test with exact distribution can be derived as follows. $T_4 > c_1$ iff $(nm/(n+m))D^2 = (nm/(n+m))(\bar{x}_1 - \bar{x}_2)'\,S_u^{-1}(\bar{x}_1 - \bar{x}_2) > c_2$, where $S_u = ((n+m)/(n+m-2))S$

and $(nm/(n+m))\,D^2$ is the Hotelling's two-sample T^2 statistic. Furthermore, if $\mu_1=\mu_2$ and $\Sigma_1=\Sigma_2$, then $(nm/(n+m))\,D^2 \overset{d}{=} T^2(K, n+m-2)$. Using the relationship between T^2 and F distributions, we may also deduce that, under the stated conditions,

$$T'_4=\frac{nm(n+m-K-1)}{(n+m)(n+m-2)K}D^2 \overset{d}{=} F_{K,n+m-K-1}.$$

One must reject the null hypothesis when $T'_4>F_{K,n+m-K-1,\alpha}$. Finally, we have $T'_4=0.964$ and $F_{2,46,0.05}=3.20$, so we again conclude that μ_1 is equal to μ_2 when $\Sigma_1=\Sigma_2$.

(b) Now we have that $\theta_1=(\mu_1,\Sigma_1)$ and that $\theta_2=(\mu_2,\Sigma_2)$. The statistic that we have to evaluate is given in Corollary 2b, i.e.,

$$\begin{aligned}T_3&=\frac{2nm}{r(m+n)}D_r^s((\bar{x}_1,S_1),(\bar{x}_2,S_2))\\&=40\left(\exp\{0.25\cdot 0.06970\}\frac{18.51970}{(18.80628)^{1/2}\,(16.31705)^{1/2}}-1\right)\\&=3.03174.\end{aligned}$$

Since $\chi^2_{50,05}=11.07$, we conclude that $\mu_1=\mu_2$ and $\Sigma_1=\Sigma_2$.

4. Applications to Multinomial Populations

Recently, many works have appeared in the scientific literature treating statistical problems in a multinomial distribution context by means of divergence type measures. In Zografos *et al.* (1990) the sampling properties of estimated φ-divergences are studied. Approximate means and variances are derived and asymptotic distributions are obtained. Furthermore, tests of goodness of fit of observed frequencies to expected ones and tests of equality of divergences based on two or more multinomial samples are given. Zografos (1992) and Pardo *et al.* (1992) studied a test of independence based on φ-divergences measures. Also related with the stratified sampling setup is a recent paper by Zografos (1991). With a different divergence measure, Morales *et al.* (1993) have studied the same problem in a stratified random sampling with proportional allocation and independence among strata. Other interesting works in this line can be seen in Menéndez *et al.* (1992) and Salicrú *et al.* (1993).

In this section we shall study the asymptotic behaviour of the $(\underline{h}, \underline{\phi})$-divergences in multinomial populations. More concretely, consider $(\mathfrak{X}, \beta_{\mathfrak{X}}, P_\theta)_{\theta\in\Theta}$ with $\Theta=\{\theta=(p_1, ..., p_{M-1})/p_i>0,\ \sum_{i=1}^{M-1} p_i=1-p_M\}$

and $\mathfrak{X} = \{x_1, ..., x_M\}$. Consider the parameters $\theta_1 = (p_1, ..., p_{M-1})$ and $\theta_2 = (q_1, ..., q_{M-1})$, where $p_i \geq 0$, $q_i \geq 0$ $(i = 1, 2, ..., M)$ and $\sum_{i=1}^{M} p_i = \sum_{i=1}^{M} q_i = 1$ (Note that $(M-1)$ is now the parameter dimension and not M). Observe that for every value x_i of X

$$f_{\theta_1}(x_i) = p_i \qquad \text{and} \qquad f_{\theta_2}(x_i) = q_i,$$

where we have supposed that μ is a counting measure giving mass one to each of the values x_i of $\mathfrak{X}$. Then, $D_{\underline{\phi}}^{h}(\theta_1, \theta_2)$ can be written in the following way

$$D_{\underline{\phi}}^{h}(P, Q) = D_{\underline{\phi}}^{h}(\theta_1, \theta_2) = \int_{A} h_\alpha\left(\sum_{i=1}^{M} q_i \phi_\alpha\left(\frac{p_i}{q_i}\right) - \phi_\alpha(1)\right) d\eta.$$

Estimation of multinomial population $(h, \underline{\phi})$-divergences can be made in two ways: estimating both distributions involved in the argument or estimating one distribution and considering the other as given. In the first case we have an index of similarity or dissimilarity of P and Q. In the last case we have the relative information or directed divergence between the sample and a given probability model. In this section we shall only consider the previous two cases as particular cases of Corollaries 2 and 3. From these results it is possible to construct tests of goodness of fit and homogeneity. Furthermore, it is not difficult to check that

$$I_F(\theta_1)^{-1} = (p_i(\delta_{ij} - p_j))_{i,j=1,...,M-1}$$

and

$$I_F(\theta_2)^{-1} = (q_i(\delta_{ij} - q_j))_{i,j=1,...,M-1}.$$

After some straightforward calculus, all the results in Section 2 can be obtained. Here we only present the particularized versions of Corollaries 2 and 3:

COROLLARY 4. *Consider the analogue estimate,* $D_{\underline{\phi}}^{h}(\hat{P}, \hat{Q})$, *obtained by replacing* p_i*s and* q_i*s by the observed relative frequencies* $\hat{p}_i$ *and* $\hat{q}_i$ $(i = 1, ..., M)$, *respectively. Suppose that* $(\hat{p}_1, ..., \hat{p}_M)$ *and* $(\hat{q}_1, ..., \hat{q}_M)$ *are based on independent samples of sizes* n *and* m, *respectively. If* $(m/(m+n)) \xrightarrow[n,m \to \infty]{} \lambda \in (0, 1)$, *then*

(a) $(mn/(m+n))^{1/2} (D_{\underline{\phi}}^{h}(\hat{P}, \hat{Q}) - D_{\underline{\phi}}^{h}(P, Q)) \xrightarrow[n,m \to \infty]{L} N(0, \sigma_1^2)$, where $\sigma_1^2 = \lambda \sigma_P^2 + (1-\lambda)\sigma_Q^2 > 0$,

$$\sigma_P^2 = \sum_{i=1}^{M} p_i \left(\int_{A} h'_\alpha\left(\sum_{i=1}^{M} q_i \phi_\alpha\left(\frac{p_i}{q_i}\right) - \phi_\alpha(1)\right) \phi'_\alpha\left(\frac{p_i}{q_i}\right) d\eta\right)^2 - \left(\sum_{i=1}^{M} p_i \int_{A} h'_\alpha\left(\sum_{i=1}^{M} q_i \phi_\alpha\left(\frac{p_i}{q_i}\right) - \phi_\alpha(1)\right) \phi'_\alpha\left(\frac{p_i}{q_i}\right) d\eta\right)^2$$

and

$$\sigma_Q^2 = \sum_{i=1}^{M} q_i \left(\int_A h'_\alpha \left(\sum_{i=1}^{M} q_i \phi_\alpha \left(\frac{p_i}{q_i} \right) - \phi_\alpha(1) \right) \left(\phi_\alpha \left(\frac{p_i}{q_i} \right) - \frac{p_i}{q_i} \phi'_\alpha \left(\frac{p_i}{q_i} \right) \right) d\eta \right)^2$$
$$- \left(\sum_{i=1}^{M} q_i \int_A h'_\alpha \left(\sum_{i=1}^{M} q_i \phi_\alpha \left(\frac{p_i}{q_i} \right) - \phi_\alpha(1) \right) \left(\phi_\alpha \left(\frac{p_i}{q_i} \right) - \frac{p_i}{q_i} \phi'_\alpha \left(\frac{p_i}{q_i} \right) \right) d\eta \right)^2.$$

(b) *If* $P = Q$ *and* $\int_A h'_\alpha(0)\, \phi''_\alpha(1)\, d\eta \neq 0$, *then*

$$\frac{2mn}{n+m} \frac{D^h_\phi(\hat{P}, \hat{Q})}{\int_A h'_\alpha(0)\, \phi''_\alpha(1)\, d\eta} \xrightarrow[n,m \to \infty]{L} \chi^2_{M-1}.$$

COROLLARY 5. *If Q is known, then*

(a) $n^{1/2}(D^h_\phi(\hat{P}, Q) - D^h_\phi(P, Q)) \xrightarrow[n,m \to \infty]{L} N(0, \sigma_P^2)$,

where σ_P^2 *has been defined in Corollary* 4.

(b) *If Q is known,* $\int_A h'_\alpha(0)\, \phi''_\alpha(1)\, d\eta \neq 0$ *and* $P = Q$, *then*

$$2n \frac{D^h_\phi(\hat{P}, Q)}{\int_A h'_\alpha(0)\, \phi''_\alpha(1)\, d\eta} \xrightarrow[n \to \infty]{L} \chi^2_{M-1}.$$

Based on Corollaries 4 and 5 we can construct the following tests:

(1) Test for a predicted value of population divergence, i.e.,

$$H_0\colon D^h_\phi(P, Q) = D_0.$$

See Section 3(1) if Q is known and Section 3(5) if Q is unknown.

(2) Test for a common predicted value of r population divergences, i.e.,

$$H_0\colon D^h_\phi(P_1, Q_1) = \cdots = D^h_\phi(P_r, Q_r) = D_0.$$

Q_i could be known or unknown, $i = 1, ..., r$.

(3) Test for equality of r population divergences, i.e.,

$$H_0\colon D^h_\phi(P_1, Q_1) = \cdots = D^h_\phi(\mathrm{Pr}, Q_r).$$

Q_i could be known or unknown, $i = 1, ..., r$.

(4) Test for goodness of fit, i.e.,

$$H_0\colon P = P^*.$$

P^* is known. See Section 3(2).

(5) Test for homogeneity of m populations with a known distribution Q; i.e.,

$$H_0: P_1 = P_2 = \cdots = P_m = Q.$$

(6) Test for homogeneity of two parallel samples, i.e.,

$$H_0: P = Q.$$

See Section 3(6).

EXAMPLE 2 (Rohatgi, 1984, p. 624). Suppose we have the following record of 340 fatal automobile accidents, per hour, on a long holiday weekend in the United States.

No. of fatal accidents per hour	$\leqslant 1$	2	3	4	5	6	7	$\geqslant 8$
No. of hours	5	8	10	11	11	9	8	10

The number X of fatal auto accidents per hour is clearly a discrete random variable. We feel that the Poisson distribution with $\lambda = 5$ provides an adequate representation for the distribution of X. The following table compares the observed and the postulated relative frequencies:

X	$\leqslant 1$	2	3	4	5	6	7	$\geqslant 8$
$\hat{p}_i$	.07	.11	.14	.15	.15	.13	.11	.14
q_i	.04	.085	.14	.175	.176	.146	.105	.133

We will use the r order and s degree divergence, $D_r^s(P, Q)$, with $r = 0.5$ and $s = 2$, to test the above hypotheses. So, according to Corollary 5(b), we have to evaluate the following test statistic:

$$T_2 = \frac{2n}{r} D_r^s(\theta_1, \theta_2) = \frac{2n}{r}(s-1)^{-1}\left\{\left[\sum_{i=1}^{M} \hat{p}_i^r q_i^{1-r}\right]^{s-1/r-1} - 1\right\} = 2.45433.$$

Since $\chi^2_{7,0.05} = 14.07$, we conclude that the Poisson model with $\lambda = 5$ does provide a reasonable fit for the data.

REFERENCES

BURBEA, J., AND RAO, C. R. (1982a). Entropy differential metric, distance and divergence measures in probability spaces: An unified approach. *J. Multivariate Anal.* **12** 575–596.

BURBEA, J., AND RAO, C. R. (1982b). On the convexity of some divergence measures based on entropy functions. *IEEE Trans. Inform. Theory* **IT-28** 489–495.

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

JEFFREYS, H. (1946). An invariant form of the prior probability in estimation problems. *Proc. Roy. Soc. Ser. A* **186** 453–561.

KERRIDGE, D. F. (1961). Inaccuracy and inference. *J. Roy. Stat. Soc. Ser. B* **23**(1), 184–194.

KULLBACK, S., AND LEIBLER, A. (1951). On the information and sufficiency. *Ann. Math. Statist.* **27** 986–1005.

KUPPERMAN, M. (1957). *Further Applications of information Theory to Multivariate Analysis and Statistical Inference*, Ph.D. dissertation.

MENÉNDEZ, M. L., MORALES, D., PARDO, L., AND SALICRÚ, M. (1992). Some applications of (r, s)-directed divergences. *Utilitas Math.* **42** 115–127.

MORALES, D., PARDO, L., SALICRÚ, M., AND MENÉNDEZ, M. L. (1993). The λ-divergence and the λ-mutual information: Estimation in the stratified sampling. *J. Comput. Appl. Math.* **47** 1–10.

MORALES, D., PARDO, L., SALICRÚ, M., AND MENÉNDEZ, 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** 201–221.

PARDO, L., PARDO, J. A., MENÉNDEZ, M. L., AND ESTEBAN, D. (1992). *The Convexity with respect to Gaussian Distribution of (r, s)-Directed Divergences*, Tech. Rep. Departamento de Estadística e I.O. Universidad Complutense de Madrid.

RENYI, A. (1961). On measures of entropy and information. In *Proceedings 4th Berkeley Symp. Math. Statist. and Prob.*, Vol. 1, pp. 547–561.

ROHATGI, V. K. (1984). *Statistical Inference*. J. Wiley, New York.

SALICRÚ, M., MENÉNDEZ, M. L., MORALES, D., AND PARDO, L. (1993). Asymptotic properties of (r, s)-directed divergence in a stratified sampling. *Appl. Math. Comput.* **55** 131–152.

SEAL, H. L. (1964). *Multivariate Statistical Analysis for Biologists*. Methuen, London.

SHANNON, C. (1948). A mathematical theory of communications. *Bell. System. Tech. J.* **27** 379–423.

SHARMA, B. D., AND MITTAL, D. P. (1977). New nonadditive measures of entropy for discrete probability distributions. *J. Math. Sci.* **10** 28–40.

TANEJA, I. J. (1989). On generalized information measures and their applications. *Adv. Electron. Electron Phys.* **76** 327–413.

VAJDA, I. (1989). *Theory of Statistical Inference and Information*. Kluwer, Dordrecht.

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**(5), 1785–1802.

ZOGRAFOS, K. (1991). *Asymptotic Distributions of Estimated f-Dissimilarity between Populations in a Stratified Random Sampling*. Tech. Rep. 191, Math. Dept., Univ. of Ioannina.

ZOGRAFOS, K. (1992). Asymptotic properties of φ-divergence statistic and its application in contingency tables. Tech. Rep. 185, Math. Dept., Univ. of Ioannina; *Internat. J. Math. Statist. Sci.*