Kybernetika

Elena Landaburu; Leandro Pardo

Goodness of fit tests with weights in the classes based on (h, ϕ)-divergences

Kybernetika, Vol. 36 (2000), No. 5, [589]--602

Persistent URL: `http://dml.cz/dmlcz/135373`

KYBERNETIKA — VOLUME 36 (2000), NUMBER 5, PAGES 589-602

GOODNESS OF FIT TESTS WITH WEIGHTS IN THE CLASSES BASED ON (h, ϕ)–DIVERGENCES[1]

ELENA LANDABURU AND LEANDRO PARDO

The aim of the paper is to present a test of goodness of fit with weigths in the classes based on weighted (h, ϕ)-divergences. This family of divergences generalizes in some sense the previous weighted divergences studied by Frank et al [5] and Kapur [11]. The weighted (h, ϕ)-divergence between an empirical distribution and a fixed distribution is here investigated for large simple random samples, and the asymptotic distributions are shown to be either normal or equal to the distribution of a linear combination of independent chi-square variables. Some approximations to the linear combination of independent chi-square variables are presented.

1. INTRODUCTION

Several coefficients have been suggested in the statistical literature to reflect the fact that some probability distributions are "closer together" than others and, consequently, that it may be easier to distinguish between the distributions of one pair than between those of another. While these coefficients, called divergence measures, have not been introduced for exactly the some purpose, they have the common property of increasing as the two distributions involved "more apart".

Let $(\mathcal{X}, \beta_{\mathcal{X}}, P)_{P\in\Delta_M}$ be a statistical space, where $\mathcal{X} = \{x_1, \dots, x_M\}$, $\Delta_M = \left\{P = (p_1, \dots, p_M)^t : p_i > 0 \text{ and } \sum_{i=1}^M p_i = 1\right\}$ and $\beta_{\mathcal{X}}$ is the σ-field of all the subsets of $\mathcal{X}$. For any P, $Q \in \Delta_M$ the most important family of divergences was given by Csiszár [1] and defined by the following expression

$$D_\phi(P, Q) = \sum_{i=1}^{M} q_i \phi\left(\frac{p_i}{qi}\right) \tag{1}$$

where ϕ is a continuous convex function, $\phi : [0, \infty) \to R^+ \cup \{\infty\}$, $0\phi(0/0) = 0$ and $0\phi(p/0) = \lim_{u\to\infty} \frac{\phi(u)}{u}$. For the properties of ϕ-divergences we refer Liese and Vajda [13] or Vajda [21].

However, there are some important measures of divergence that cannot be obtained as particular cases of ϕ-divergence. For this reason Menéndez et al [15]

[1]This work was supported by Grant DGES PB96-0635.

introduced an extended expression, called (h, ϕ)-divergence, which is defined by

$$D_{\phi}^{h}(P,Q)=\sum_{a=1}^{A}\eta_a h_a\left(\sum_{i=1}^{M}q_i\phi_a\left(\frac{p_i}{q_i}\right)\right) \tag{2}$$

where $h = (h_a)_{a=1,\ldots,A}$, $\phi = (\phi_a)_{a=1,\ldots,A}$, and for $a = 1,\ldots,A$, ϕ_a satisfy the condition of the Csiszár's divergence definition, h_a are nondecreasing and continuous functions on $\left[0, \phi_a(0) + \lim_{u\to\infty} \frac{\phi_a(u)}{u}\right]$, i. e., on the range of the function $D_{\phi_a}(P,Q)$ (cf. Theorem 9.1 in Vajda [21]) with $h_a(0) = 0$, $a = 1,\ldots,A$ and η_a are positive numbers.

The expressions (1) and (2) depend only on the probabilities of the events and do not take into account the effectiveness of the events under consideration. But there exist many fields dealing with random events where it is necessary to take into account both these probabilities and some qualitative characteristics of events. A criterion for a qualitative differentiation of the possible events of a given experiment is represented by the relevance, the significance, or the utility of the information they carry, with respect to a qualitative characteristic. The occurrence of an event thus removes a double uncertainly: the quantitative one, related to the probability with it occurs, and the qualitative one, related to a given qualitative characteristic. An interesting motivation about the necessity of introducing weights is given, for instance in Guiaşu [7] and Longo [14]. In this paper, these considerations are taken into account, so that we shall suppose that this is done by means of some qualitative weights which are nonnegative, finite, real numbers, as the usual weights in physics or as the utilities in decision theory.

If in the expression (1) we consider $\phi(x) = x\log x$ or $\phi^*(x) = x\log x - x + 1$ we obtain the Kullback–Leibler divergence. This divergence measure was generalized by Taneja [20] who introduced a weighted version by the formula

$$D_U(P,Q)=\sum_{i=1}^{M}u_ip_i\log\left(\frac{p_i}{q_i}\right) \tag{3}$$

where the numbers u_i are positive weights or utilities reflecting some consequences attached to the element $x_i \in \mathcal{X}$, $i = 1,\ldots,M$. Taneja also investigated an estimator of the weighted Kullback–Leibler divergence obtained by replacing p_i, $i = 1,\ldots,M$, by the relative frequencies $\hat{p}_i$, $i = 1,\ldots,M$, in a simple random sample.

The expression (3) has two problems: the first one appears when $u_i = u \neq 1$, $i = 1,\ldots,M$, because in that case the expression (3) does not reduce to the Kullback–Leibler divergence. The second one is more serious: It would be important to solve the problem of goodness of fit with weights that $D_U(P,Q) \geq 0$ and the equality holds when $P = Q$, but it is possible in (3) that $D_U(P,Q) < 0$. For instance, if we consider the probability distributions $P = (p, 1-p)^t$ and $Q = (1-p, p)^t$ we have

$$D_U(P,Q)=\log\left(\frac{1-p}{p}\right)\left(u_2(1-p)-u_1p\right).$$

It is clear if $1-p > p$ and $u_2/u_1 < p/(1-p)$ or $1-p < p$ and $u_2/u_1 > p/(1-p)$ then $D_U(P,Q) < 0$.

The first problem can be solved multiplying the expression (3) by $E_P(U)^{-1} = \left(\sum_{i=1}^M u_i p_i\right)^{-1}$ or by $E_Q(U)^{-1} = \left(\sum_{i=1}^M u_i q_i\right)^{-1}$. The second problem can be solved considering the function

$$\phi^*(x) = x \log x - x + 1$$

instead of $\phi(x) = x \log x$. Then to solve both the previous problems we can define the weighted Kullback–Leibler divergence by

$$WD(P,Q) = \frac{1}{E_P(U)} \sum_{i=1}^{M} u_i \left(p_i \log \frac{p_i}{q_i} + q_i \right) - 1. \tag{4}$$

It is important that for $\phi(x) = x \log x$ used instead of $\phi^*(x) = x \log x - x + 1$ we do not get the weighted Kullback–Leibler divergence given in (4).

It is clear that $WD(P,Q)$ coincides with the Kullback–Leibler divergence if $u_i = u$, $i = 1, \ldots, M$, and $WD(P,Q) \geq 0$ with equality when $P = Q$. The expression (2) was generalized for the ϕ-divergences, solving the first problem, by Frank et al [5]. The second problem was considered by Kapur [11] in relation with the Kullback–Leibler divergence.

In this paper we present a unified expression for solving the two problems simultaneously. Instead of considering the unified expression for (1) we shall consider the weighted version of the expression (2) in the following way

$$WD_\phi^h(P,Q) = \sum_{a=1}^{A} \eta_a h_a \left(\sum_{i=1}^{M} \frac{u_i q_i}{E_P(U)} \phi_a \left(\frac{p_i}{q_i} \right) \right) \tag{5}$$

where $h = (h_a)_{a=1,\ldots,A}$, $\phi = (\phi_a)_{a=1,\ldots,A}$, ϕ_a and h_a are real valued C^2 functions, and for $a = 1, \ldots, A$, ϕ_a satisfies the condition of the Csiszár's divergence definition with $\phi_a(1) = \phi'_\alpha(1) = 0$, $a = 1, \ldots, A$, h_a are nondecreasing and continuous functions with $h_a(0) = 0$, $a = 1, \ldots, A$, η_a, $a = 1, \ldots, A$, are positive numbers and u_i, $i = 1, \ldots, M$, positive weights. The nonnegativity of the measure (5) as well as the equality to zero when $P = Q$ hold because ϕ_a, $a = 1, \ldots, A$, are assume to be convex twice differentiable functions, vanishing at $x = 1$ together with their first derivatives ϕ'_a , $a = 1, \ldots, A$.

In Section 2 we consider the nonparametric estimator $\hat{P} = (\hat{p}_1, \ldots, \hat{p}_M)^t$, based on a random sample of size n, $X_1, \ldots, X_M$, and defined by relative frequencies $\hat{p}_j = N_j/n$, with $N_j = \sum_{i=1}^n I_{\{x_j\}}(X_i)$ and the asymptotic distribution of the statistic $WD_\phi^h\left(\hat{P}, P_0\right)$ is obtained under the null hypothesis $P = P_0$. On the basis of this asymptotic distribution a test of goodness of fit with weights is introduced. In Section 3 we assume $M = 2$, binomial case, and we present a ramification of the results obtained in Section 2.

2. TEST FOR GOODNESS OF FIT WITH WEIGHTS

Suppose we are sampling from a distribution $F_X(x)$. Divide the range of the distribution into M mutually exclusive and exhaustive classes, say $A_1, \ldots, A_M$. Each class has a probability of containing the random variable X, $P(X \in A_i) = p_i$, $i = 1, \ldots, M$, and a nonnegative weights w_i, $i = 1, \ldots, M$, directly proportional to their importance, and each sample value x falls into exactly one of the intervals. Let $X_1, \ldots, X_n$ be a size n random sample from $F_X(x)$ and let $(N_1, \ldots, N_M)$ be the respective observed number of sample values falling in the classes $A_1, \ldots, A_M$. Then the vector $(N_1, \ldots, N_M)$ has a multinomial distribution with parameters $(n; p_1, \ldots, p_M)$. Now, we want to test the hypothesis $H_0 : F_X(x) = F_0(x)$. Firstly, we compute $p_{i0} = P(X \in A_i)$, $i = 1, \ldots, M$, under H_0. If H_0 is true, then $P = P_0$ and intuitively it is expected $n\hat{p}_i \approx np_{i0}$ in which case $WD_\phi^h\left(\hat{P}, P_0\right)$ is small. Thus larger $WD_\phi^h\left(\hat{P}, P_0\right)$ indicates data less compatible with the claimed null distribution and we must reject the null hypothesis iff

$$WD_\phi^h\left(\hat{P}, P_0\right) > c$$

where c must be chosen for getting a level α test. In some situations it will be possible to get the exact distribution of the statistic $WD_\phi^h\left(\hat{P}, P_0\right)$ and then the value c. But in general this is not possible and we must use the asymptotic distribution of the statistic $WD_\phi^h\left(\hat{P}, P_0\right)$. In the following theorem we present this asymptotic distribution.

Theorem 1. Consider the weighted (h, ϕ)-divergence $WD_\phi^h(P, P_0)$ and its estimator $WD_\phi^h\left(\hat{P}, P_0\right)$. Under the null hypothesis $H_0 : P = P_0$ and assuming $\phi_a''(1) > 0$, $h_a'(0) > 0$, $a = 1, \ldots, A$, we have

$$2nWD_\phi^h\left(\hat{P}, P_0\right) \xrightarrow[n\to\infty]{L} \sum_{i=1}^{r} \beta_i Z_i^2,$$

where $Z_i, i = 1, \ldots, r$, are iid normal variables with mean zero and variance 1, β_i, $i = 1, \ldots, r$,are the non null eigenvalues of the matrix $A\Sigma_{P_0}$,and $r = rank\left(\Sigma_{P_0} A \Sigma_{P_0}\right)$, being

$$\Sigma_{P_0} = diag(P_0) - P_0 P_0^t$$

and

$$A = \left(\frac{\partial^2 WD_\phi^h(P, P_0)}{\partial p_i \partial p_j}\right)_{i,j=1,\ldots,M} = (a_{ij})_{i,j=1,\ldots,M}$$

where

$$a_{ij} = \begin{cases} 0 & \text{if } i \neq j \\ l\frac{u_i}{p_{i0}} & \text{if } i = j \end{cases}$$

and $l = \sum_{a=1}^{A} \eta_a h_a'(0) \phi_a''(1) \left(\sum_{i=1}^{M} u_i p_{i0}\right)^{-1}$.

Proof. The second-order expansion of $WD_\phi^h(P, P_0)$ about P_0 gives

$$WD_\phi^h(P, P_0) = \frac{1}{2}(P - P_0)^t A (P - P_0) + o\left(\|P - P_0\|^2\right).$$

But $\sqrt{n}\left(\hat{P}-P_0\right) \xrightarrow[n\to\infty]{L} N\left(0,\Sigma_{P_0}\right)$, then $\left\|\hat{P}-P_0\right\|^2 = O_p\left(n^{-1}\right)$ and

$$o\left(\left\|\hat{P}-P_0\right\|^2\right) = o_p\left(n^{-1}\right).$$

Therefore the random variables $2nWD_\phi^h\left(\hat{P},P_0\right)$ and

$$\sqrt{n}\left(\hat{P}-P_0\right)^t A\sqrt{n}\left(\hat{P}-P_0\right)$$

have the same asymptotic distribution. Now by Corollary 2.1 in Dik and Gunst [2] the result follows. □

Corollary 1. If we consider the weighted ϕ-divergences, $A=\{1\}$, $\eta_1=1$, $h(x)=x$, $\phi_1(x)=\phi(x)$, then under the null hypothesis $H_0: P=P_0$ and assuming $\phi''(1)>0$,

$$\frac{2nD_\phi\left(\hat{P},P_0\right)}{\phi''(1)} \xrightarrow[n\to\infty]{L} \sum_{i=1}^{r}\beta_i^* Z_i^2$$

where β_i^* are the eigenvalues of the matrix $A^*\Sigma_{P_0}$, with $A^*=\left(a_{ij}^*\right)_{i,j=1,\dots,M}$, and

$$a_{ij}^* = \begin{cases} 0 & \text{if } i\neq j \\ \frac{u_i}{p_{i0}E_{P_0}(U)} & \text{if } i=j \end{cases}.$$

Corollary 2. If we consider the (h,ϕ)-divergences satisfying the assumptions of Theorem 1 and $u_i=u$, $i=1,\dots,M$, then under the null hypothesis $H_0: P=P_0$,

$$\frac{2nD_\phi^h\left(\hat{P},P_0\right)}{\sum_{a=1}^{A}\eta_a h_a'(0)\,\phi_a''(1)} \xrightarrow[n\to\infty]{L} \sum_{i=1}^{r}\beta_i^{**}Z_i^2$$

where β_i^{**} are the eigenvalues of the matrix $A^{**}\Sigma_{P_0}$, $A^{**}=diag\left(P_0^{-1}\right)$. Since $\Sigma_{P_0}A^{**}\Sigma_{P_0}A^{**}\Sigma_{P_0}=\Sigma_{P_0}A^{**}\Sigma_{P_0}$ the random variable $\sum_{i=1}^{r}\beta_i^{**}Z_i^2$ is a chi-square distribution with $M-1=trace(A^{**}\Sigma_{P_0})$ degrees of freedom. This result was obtained in Menéndez et al [15].

If we consider the ϕ-divergences, $u_i=u$, $i=1,\dots,M$, $A=\{1\}$, $\eta_1=1$, $h(x)=x$, $\phi_1(x)=\phi(x)$, we have

$$\frac{2n}{\phi''(1)}D_\phi^h\left(\hat{P},P_0\right) \xrightarrow[n\to\infty]{L} \mathcal{X}_{M-1}^2.$$

This result was obtined for the first time in Zografos et al [22].

Using Theorem 2.1 we can test

$$H_0 : P = P_0 \text{ against } H_1 : P \neq P_0$$

as follows

$$\varphi(\hat{p}_1, \cdots, \hat{p}_M) = \begin{cases} 1 & \text{if } T_{\phi,h,W} > t_\alpha \\ 0 & \text{otherwise} \end{cases}$$

where

$$T_{\phi,h,W} = 2nWD_\phi^h\left(\hat{P}, P_0\right),$$

t_α is the critical value of $T_{\phi,h,W}$ verifying

$$P\left(\sum_{i=1}^{r} \beta_i Z_i^2 > t_\alpha\right) = \alpha, \tag{6}$$

and the β_i's are given in Theorem 1.

Remark 1. In order to apply the theorem above, we have to calculate a probability of a linear combination of chi-squared distributions and one can feel a little worried, but after reading the papers of Rao and Scott [18] and Modarres et al [17] that feeling disappears. They give some ideas to overcome this situation. In fact, a variety of problems in statistical inference and applied probability requires either percentiles or probabilities from the distribution of a combination of chi-squared distributed random variables (cf. Jensen and Solomon [9]).

Corollary 1 of Rao and Scott [18] proposes to consider the statistic

$$S_{\phi,h,W}^1 = \frac{2n}{\beta_*} WD_\phi^h\left(\hat{P}, P_0\right) \leq \sum_{i=1}^{r} Z_i^2$$

where $\beta_* = \max\{\beta_1, \ldots, \beta_r\}$. We know that the asymptotic distribution of the random variable $\sum_{i=1}^r Z_i^2$ is a chi-square distribution with r degrees of freedom. Then if we assume that the statistic $S_{\phi,h,W}^1$ is asymptotically distributed as a chi-square distribution with r degrees of freedom, we must reject the null hypothesis $H_0 : P = P_0$, with a significance level α, if

$$S_{\phi,h,W}^1 \geq \mathcal{X}_{r,\alpha}^2.$$

We can observe that this test is a more conservative than the previous one given by $T_{\phi,h,W}$.

Another approach to the asymptotic distribution of the statistic is the modified statistic

$$S_{\phi,h,W}^2 = \frac{2n}{\bar{\beta}} WD_\phi^h\left(\hat{P}, P_0\right) \leq \sum_{i=1}^{r} Z_i^2$$

where $\bar{\beta} = \frac{1}{r}\sum_{i=1}^r \beta_i$. This test is more conservative too. In this case we can observe that

$$E\left[S_{\phi,h,W}^2\right] = r = E\left[\mathcal{X}_r^2\right]$$

and

$$\mathrm{Var}\left(S^2_{\phi,h,W}\right) = \frac{2\sum\limits_{i=1}^{r}\beta_i}{\bar{\beta}^2} = 2r + \sum_{i=1}^{r}\frac{\left(\beta_i - \bar{\beta}\right)^2}{\bar{\beta}^2} > \mathrm{Var}\left(\mathcal{X}_r^2\right).$$

If we denote by $\Lambda = diag\,(\beta_1, \dots, \beta_r)$, we get

$$E\left[\sum_{i=1}^{r}\beta_i Z_i^2\right] = \sum_{i=1}^{r}\beta_i = trace(\Lambda) = trace\left(A\Sigma_{P_0}\right) = l\sum_{i=1}^{r}u_i\left(1 - p_{i0}\right).$$

Then $\bar{\beta}$ is given by

$$\bar{\beta} = \frac{l}{r}\left(\sum_{i=1}^{r}u_i\left(1 - p_{i0}\right)\right);$$

where l is given in Theorem 1.

Finally, we can also consider the statistic

$$S^3_{\phi,h,W} = \frac{2n}{\bar{\beta}\left(1+\lambda\right)}WD^h_\phi\left(\hat{P}, P_0\right)$$

with $\lambda = \sum\limits_{i=1}^{r}\frac{\left(\beta_i - \bar{\beta}\right)^2}{r\bar{\beta}^2}$. This statistic verifies

$$E\left[S^3_{\phi,h,W}\right] = \frac{r}{1+\lambda} \text{ and } \mathrm{Var}\left(S^3_{\phi,h,W}\right) = \frac{2r}{1+\lambda}.$$

Then the approximate asymptotic distribution of the statistic $S^3_{\phi,h,W}$ is a chi-square distribution with $v = \frac{r}{1+\lambda}$ degrees of freedom.

Apart from the above approximations it is possible to consider tables of the cumulartive distribution $\sum_{i=1}^{r} a_i Z_i^2$ in the case of small r (see Solomon [19], Johnson and Kotz [10], Eckler [3] and Gupta [8]).

Theorem 2. The test of goodness of fit considered in Theorem 1 is consistent in the Fraser sense, i. e. for every alternative hypothesis $P^* \neq P_0$,

$$\lim_{n\to\infty}\beta_{n,\phi,h,W}\left(P^*\right) = 1.$$

Proof. Since

$$WD^h_\phi\left(\hat{P}, P_0\right) \underset{n\to\infty}{\overset{P}{\longrightarrow}} WD^h_\phi\left(P^*, P_0\right) > 0$$

under the alternative hypothesis $P = P^*$, it follows that

$$P\left(2nWD^h_\phi\left(\hat{P}, P_0\right) > t_\alpha\right) = P\left(WD^h_\phi\left(\hat{P}, P_0\right) > \frac{t_\alpha}{2n}\right) \to 1$$

as $n \to \infty$.

Then the power function, $\beta_{n,\phi,h,W}(P^*)$, verifies

$$\beta_{n,\phi,h,W}(P^*) = P\left(2nWD_\phi^h\left(\hat{P},P_0\right) > t_\alpha\right) = P\left(WD_\phi^h\left(\hat{P},P_0\right) > \frac{t_\alpha}{2n}\right) \underset{n\to\infty}{\longrightarrow} 1.$$

□

Now we will present a theorem providing an approximation for the power function of the test based on the statistic $T_{\phi,h,W}$.

Theorem 3. If h, ϕ satisfy the assumptions of Theorem 1 then, under the alternative hypothesis $P^* \neq P_0$,

$$n^{1/2}\left(WD_\phi^h\left(\hat{P},P_0\right) - WD_\phi^h(P^*,P_0)\right) \underset{n\to\infty}{\overset{L}{\longrightarrow}} N\left(0,\sigma^2_{P^*,P_0}\right),$$

where

$$\sigma^2_{P^*,P_0} = T^t\Sigma_{P^*}T = \sum_{i=1}^M t_i^2p_i^* - \left(\sum_{i=1}^M t_ip_i^*\right)^2 > 0,$$

and $\Sigma_{P^*} = diag(P^*) - P^*(P^*)^t$. The vector T is defined by $T = (t_1,\dots,t_M)^t$, with

$$t_i = \sum_{a=1}^{\Lambda}\left\{\eta_a h_a'\left(\frac{\sum_{i=1}^M u_ip_i^0}{\sum_{i=1}^M u_ip_i^*}\phi_a\left(\frac{p_i^*}{p_i^0}\right)\right)\left[\frac{u_i}{\left(\sum_{i=1}^M u_ip_i^*\right)^2}\left(\phi_a'\left(\frac{p_i^*}{p_i^0}\right)\sum_{i=1}^M u_ip_i^* - \sum_{i=1}^M u_ip_i^0\phi_a\left(\frac{p_i^*}{p_i^0}\right)\right)\right]\right\}.$$

Proof. A first order Taylor expansion of $WD_\phi^h(P,P_0)$ gives

$$n^{1/2}\left(WD_\phi^h(P,P_0) - WD_\phi^h(P^*,P_0)\right) = n^{1/2}T^t(P-P^*)^t + n^{1/2}o(\|P-P^*\|),$$

where $T^t = \nabla WD_\phi^h(P_0,P^*)$ with $\nabla = \left(\frac{\partial}{\partial p_1},\dots,\frac{\partial}{\partial p_M}\right)$ and the result follows in view of the relation $n^{1/2}o\left(\left\|\hat{P}-P^*\right\|\right) = o_p(1)$ and applying the fact that

$$\sqrt{n}\left(\hat{P}-P^*\right) \underset{n\to\infty}{\overset{L}{\longrightarrow}} N(0,\Sigma_{P^*}).$$

□

Remark 2. On the basis of Theorem 3, the power function at $P^* \neq P_0$ when testing $H_0 : P = P_0$ with weights, is given by the formula

$$\beta_{n,\phi,h,W}\left(P^*\right) = 1 - \Phi_n\left(\frac{t_\alpha - 2nWD_\phi^h\left(P^*, P_0\right)}{2n^{1/2}\sigma_{P^*,P_0}}\right) \tag{7}$$

for a sequence of distribution functions $\Phi_n(x)$ tending uniformly to the standard normal distribution function $\Phi(x)$, t_α is the critical value of $T_{\phi,h,W}$, given in (6) and σ_{P^*,P_0} is given in Theorem 2.

It is clear in (7) that

$$\lim_{n\to\infty} \beta_{n,\phi,h,W}\left(P^*\right) = 1$$

and we get the result given in Theorem 2.4.

Under fixed alternatives the power function of the family $T_{\phi,h,W}$ converges to 1 as $n \to \infty$, Theorem 2. However, it is possible to get another approximation if we consider alternative hypotheses that converge to the null vector as $n \to \infty$. We consider the contiguous alternatives hypotheses

$$H_{1,n} : P^{(n)} = P_0 + n^{-1/2}C$$

where $C = (c_1, \ldots, c_M)^t$ with $\sum_{i=1}^M c_i = 0$, which converge to the null hypothesis $H_0 : P = P_0$. In this case the power function of the family of statistics $T_{\phi,h,W}$ is given by

$$\beta_{n,\phi,h,W}\left(P^{(n)}\right) = P\left(T_{\phi,h,W} > t_\alpha / H_{1,n}\right).$$

Now the problem is to obtain the asymptotic distribution of the statistic $T_{\phi,h,W}$ under the hypothesis $H_{1,n}$. This asymptotic distribution is given in the following theorem:

Theorem 4. If h, ϕ satisfy the assumptions of Theorem 1 then, under the alternative hypotheses

$$H_{1,n} : P^{(n)} = P_0 + n^{-1/2}C$$

where $C = (c_1, \ldots, c_M)^t$ with $\sum\limits_{i=1}^{M} c_i = 0$, we have

$$T_{\phi,h,W} - \xi \xrightarrow[n\to\infty]{L} \sum_{i=1}^{r} \beta_i \left(Z_i + \omega_i\right)^2$$

where $r = rank\left(\Sigma_{P_0} A \Sigma_{P_0}\right)$,$\beta_1, \ldots, \beta_r$ are the nonnegative eigenvalues of the matrix $A\Sigma_{P_0}$, A is the matrix given in Theorem 1, Z_i, $i = 1, \ldots, r$, are independent random variables with standard normal distribution, $\omega = \Lambda^{-1}R^tS^tAC$, $\xi = C^tAC - \omega^t\Lambda\omega$ where $\Lambda = diag\left(\beta_1, \ldots, \beta_r\right)$, S is an arbitrarily chosen root of $\Sigma_{P_0}$and R is the matrix of eigenvalues of S^tAS.

Proof. Since,

$$\sqrt{n}\left(\hat{P}-P_0\right)=\sqrt{n}\left(\hat{P}-P^{(n)}\right)+C,$$

then under $H_{1,n}$,

$$\sqrt{n}\left(\hat{P}-P_0\right)\xrightarrow[n\to\infty]{L}N\left(C,\Sigma_{P_0}\right).$$

But the statistic $T_{\phi,h,W}$ has the same asymptotic distribution as the quadratic form

$$\sqrt{n}\left(\hat{P}-P_0\right)^t A\sqrt{n}\left(\hat{P}-P_0\right).$$

Therefore the result follows by Corollary 2.2 in Dik and Gunst [2]. □

Remark 3. On the basis of Theorem 2 the power function at $P^{(n)}$ when testing $H_0 : P = P_0$ with weights is given by the formula

$$\beta_{n,\phi,h,W}\left(P^{(n)}\right)=1-G_n\left(t_\alpha\right),$$

for a sequence of distribution functions $G_n(x)$ tending to the distribution function associated to the random variable $\sum_{i=1}^{r}\beta_i\left(Z_i+\omega_i\right)^2+\xi$ and t_α is the critical value of $T_{\phi,h,W}$ given in (6).

If we consider the ϕ-divergences, $u_i = u$, $i = 1,\ldots,M$, $A = \{1\}$, $\eta_1 = 1$, $h(x) = x$, $\phi_1(x) = \phi(x)$, we have on the basis of Theorem 4 that, under contiguous alternatives $H_{1,n}$, the statistic $T_{\phi,h,W}$ converges in law to the non-central chi-squared variable with with $M-1$ degrees of freedom and non-centrality parameter δ given by

$$\delta=\sum_{i=1}^{M}\frac{c_i^2}{p_{i0}}.$$

This result was obtained for the first time in Menéndez et al [16]. The same result is obtained if we consider (h,ϕ)-divergences.

If instead of considering the statistic $T_{\phi,h,W}$ we consider the statistics

a) $S^1_{\phi,h,W}=\left(\beta^*\right)^{-1}2nWD^h_\phi\left(\hat{P},P_0\right)$ where $\beta^*=\max\{\beta_1,\ldots,\beta_r\}$,

b) $S^2_{\phi,h,W}=\left(\bar{\beta}\right)^{-1}2nWD^h_\phi\left(\hat{P},P_0\right)$ where $\bar{\beta}=\frac{1}{r}\sum_{i=1}^{r}\beta_i$,

and

c) $S^3_{\phi,h,W}=\left(\bar{\beta}\right)^{-1}(1+\lambda)\,2nWD^h_\phi\left(\hat{P},P_0\right)$ where $\lambda=\sum_{i=1}^{r}\frac{\left(\beta_i-\bar{\beta}\right)^2}{r\bar{\beta}^2}$,

that are asymptotically distributed as a chi-square distribution with r degrees of freedom, then we have

$$\beta_{n,\phi,h,W}\left(P^{(n)}\right)=1-F_n\left(\mathcal{X}^2_{s,\alpha}-\xi\right)$$

for a sequence of distribution functions $F_n(x)$ tending uniformly to the distribution function, F, of a non-central random variable chi-squared with s degrees of freedom

($s = r$ if we consider the statistics $S^1_{\phi,h,W}$ or $S^2_{\phi,h,W}$ and $s = r/\left(1+\lambda^2\right)$ for the statistic $S^3_{\phi,h,W}$) and non-centrality parameter $\delta = \sum_{i=1}^r w_i^2$. This result follows from the fact that the random variable $\sum_{i=1}^r (Z_i + w_i)^2$ is a non-central chi-squared random variable with r degrees of freedom and non-centrality parameter $\delta = \sum_{i=1}^r w_i^2$.

3. TESTS FROM BINOMIAL DATA WITH WEIGHTS

In this section we consider the particularization of the results obtained in Section 2 to binomial data. In this case the null hypothesis is

$$H_0 : p = p_0$$

where p is the probability of having outcome 1 for the binary observation and $q = 1 - p$ is the probability of having outcome 0. We denote by u_1 and u_2 the weights of outcome 1 and 0 respectively.

Theorem 5. If h, ϕ satisfy the assumptions of Theorem 1 then, under the hypothesis of Theorem 1 and with binomial data, we have

$$T^1_{\phi,h,W} = \frac{2n}{L} W D^h_\phi\left(\hat{P}, P_0\right) \xrightarrow[n\to\infty]{L} \mathcal{X}^2_1$$

where $L = l^{-1}\left(u_1\left(1-p_0\right) + u_2 p_0\right)$ and l is given in Theorem 1.

Proof. In this case we have

$$T^1_{\phi,h,W} = \sqrt{n}\left(\hat{P} - P_0\right) B \left(\hat{P} - P_0\right) + o_P(1)$$

with

$$B = \left(u_1\left(1-p_0\right) + u_2 p_0\right)^{-1} \begin{pmatrix} u_1 p_0^{-1} & 0 \\ 0 & u_2\left(1-p_0\right)^{-1} \end{pmatrix}.$$

Then

$$T^1_{\phi,h,W} = X^t X + o_P(1)$$

where

$$X \equiv \sqrt{n} B^{1/2} \left(\hat{P} - P_0\right).$$

It is clear that the asymptotic distribution of the random vector X is normal with vector mean 0 and variance-covariance matrix given by

$$\Sigma^* = \begin{pmatrix} u_1\left(1-p_0\right) & -\sqrt{u_1 u_2 p_0 \left(1-p_0\right)} \\ -\sqrt{u_1 u_2 p_0\left(1-p_0\right)} & u_2 p_0 \end{pmatrix}. \tag{8}$$

It is easy to establish that Σ^* is a projection of *rank* 1, i.e., $\Sigma^*\Sigma^* = \Sigma^*$ and *rank* $(\Sigma^*) = 1$. Then

$$T^1_{\phi,h,W} \xrightarrow[n\to\infty]{L} \mathcal{X}^2_1. \qquad \square$$

In relation with the power function for $P^* = (p^*, 1-p^*)^t \neq P_0 = (p_0, 1-p_0)^t$ we have the same result obtained in Remark 2 but now with $M = 2$.

If we consider non local alternatives, that is to say, alternative hypotheses of the form

$$H_{1,n} : P^{(n)} = P_0 + \frac{C^*}{\sqrt{n}} \tag{9}$$

where $P_0 = (p_0, 1-p_0)^t$ and $C^* = (c, -c)^t$, we have the following result.

Theorem 6. If h, ϕ satisfy the assumptions of Theorem 1 then, under the local alternatives $H_{1,n}$ given in (9), and with binomial data,

$$T^1_{\phi,h,W} \xrightarrow[n\to\infty]{L} \mathcal{X}_1^2(\delta)$$

where $\mathcal{X}_1^2(\delta)$ is the non-central chi-square distribution with 1 degree of freedom and non-centrality parameter

$$\delta = \frac{c^2}{p_0(1-p_0)}.$$

Proof. Under the alternative hypotheses $H_{1,n}$, given in (9), we have

$$\sqrt{n}\left(\hat{P} - P_0\right) \xrightarrow[n\to\infty]{L} N\left(C^*, \Sigma_{P_0}\right)$$

where

$$\Sigma_{P_0} = \begin{pmatrix} p_0(1-p_0) & -p_0(1-p_0) \\ -p_0(1-p_0) & p_0(1-p_0) \end{pmatrix}$$

because

$$\begin{aligned} \sqrt{n}\left(\hat{P} - P_0\right) &= \sqrt{n}\left(\hat{P} - P^{(n)}\right) + \sqrt{n}\left(P^{(n)} - P_0\right) \\ &= \sqrt{n}\left(\hat{P} - P^{(n)}\right) + C^*. \end{aligned}$$

On the other hand we know that

$$T^1_{\phi,h,W} = X^t X + o_P(1)$$

and

$$X \xrightarrow[n\to\infty]{L} N\left(B^{1/2}C^*, \Sigma^*\right)$$

where Σ^* is given in given in (8).

We know that Σ^* is a projection of *rank* 1 if we establish that $\Sigma^* B^{1/2} C^* = B^{1/2} C^*$. Thus we can apply Lemma of pp. 63 in Ferguson [4] to obtain

$$T^1_{\phi,h,W} \xrightarrow[n\to\infty]{L} \mathcal{X}_1^2(\delta)$$

with $\delta = \left(C^* B^{1/2}\right)^t \left(C^* B^{1/2}\right)$.

It is clear that

$$B^{1/2}\left(C^*\right)^t = \left(c\sqrt{u_1 p_0^{-1}\left(u_1\left(1-p_0\right)+u_2 p_0\right)^{-1}}, -c\sqrt{u_2\left(1-p_0\right)^{-1}\left(u_1\left(1-p_0\right)+u_2 p\right)^{-1}}\right)$$

and also $\Sigma^* B^{1/2} C^* = B^{1/2} C^*$. Thus we have established the desired result with

$$\delta = B^{1/2}\left(C^*\right)^t C^* B^{1/2} = \frac{c^2}{p_0\left(1-p_0\right)}.$$

□

ACKNOWLEDGEMENTS

We thank the referee for a careful examination of the paper and suggestions leading to its improvement.

(Received June 15, 1999.)

REFERENCES

[1] I. Csiszár: Eine Informationstheoretische Ungleichung und ihre Anwendung auf den Beweis der Ergodizität von Markoffschen Ketten. Publications of the Mathematical Institute of Hungarian Academy of Sciences Ser A. *8* (1963), 85–108.

[2] J. J. Dik and M. C. M. de Gunst: The distribution of general quadratic forms in normal variables. Statistica Neerlandica *39* (1985), 14–26.

[3] A. R. Eckler: A survey of coverage problems associated with point and area targets. Technometrics *11* (1969), 561–589.

[4] T. S. Ferguson: A Course in Large Sample Theory. Chapman & Hall, London 1996.

[5] O. Frank, M. L. Menéndez and L. Pardo: Asymptotic distributions of weighted divergence between discrete distributions. Comm. Statist. – Theory Methods *27* (1998), 4, 867–885.

[6] D. A. S. Fraser: Non-parametrics Methods in Statistics. Wiley, New York 1957.

[7] S. Guiaşu: Grouping data by using the weighted entropy. J. Statist. Plann. Inference *15* (1986), 63–69.

[8] S. S. Gupta: Bibliography on the multivariate normal integrals and related topics. Ann. Math. Statist. *34* (1963), 829–838.

[9] D. R. Jensen and H. Solomon: A Gaussian approximation to the distribution of a definite quadratic form. J. Amer. Statist. Assoc. *67* (1972), 340, 898–902.

[10] N. L. Johnson and S. Kotz: Tables of distributions of positive definite quadratic forms in central normal variables. Sankhya, Ser. B *30* (1968), 303–314.

[11] J. N. Kapur: Measures of Information and their Applications. Wiley, New York 1994.

[12] S. Kotz, N. M. Johnson and D. W. Boid: Series representation of quadratic forms in normal variables I. Central case. Annals Math. Statist. *38* (1967), 823–837.

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

[14] G. Longo: Quantitative and Qualitative Measure of Information. Springer, New York 1970.

[15] M. L. Menéndez, D. Morales, L. Pardo and M. Salicrú: Asymptotic behaviour and statistical applications of divergence measures in multinomial populations: A unified study. Statistical Papers *36* (1995), 1–29.
[16] M. L. Menéndez, D. Morales, L. Pardo and I. Vajda: Approximations to powers of φ-disparity goodness of fit tests. Submitted.
[17] R. Modarres and R. W. Jernigan: Testing the equality of correlation matrices. Comm. Statist. – Theory Methods *21* (1992), 2107–2125.
[18] J. N. K. Rao and A. J. Scott: The analysis of categorical data from complex sample surveys: Chi-squared tests for goodness of fit and independence in two-way tables. J. Amer. Statist. Assoc. *76* (1981), 221–230.
[19] H. Solomon: Distribution of Quadratic Forms – Tables and Applications. Applied Mathematics and Statistics Laboratories, Technical Report 45, Stanford University, Stanford, Calif. 1960.
[20] C. T. Taneja: On the mean and the variance of estimates of Kullback information and relative useful information measures. Apl. Mat. *30* (1985), 166–175.
[21] I. Vajda: Theory of Statistical Inference and Information. Kluwer Academic Publishers, Dordrecht 1989.
[22] K. Zografos, K. Ferentinos and Papaioannou: φ-divergence statistics: sampling properties and multinomial goodness of fit and divergence tests. Comm. Statist. – Theory Methods *19* (1990), 5, 1785-1802.

Elena Landaburu and Prof. Dr. Leandro Pardo, Department of Statistics and Operations Research, Complutense University of Madrid, 28040- Madrid. Spain.
e-mail: leandro_pardo@mat.ucm.es