Int. J. Math. & Stat. Sci.
Vol. 2, No. 1, June 1993, 5–22

ASYMPTOTIC PROPERTIES OF ϕ - DIVERGENCE STATISTIC AND ITS APPLICATIONS IN CONTINGENCY TABLES

K. Zografos

Department of Mathematics, University of Ioannina
Probability, Statistics and Operational Research Unit
451 10 Ioannina, Greece

Abstract

Asymptotic properties of a ϕ - divergence statistic arising in the context of a two way contingency table are investigated. It is shown that the asymptotic distribution of this statistic is either normal or a linear form in chi square variables depending on whether or not a suitable condition is satisfied. Under the assumption of independence this (asymptotic) distribution is shown to be chi square. The chi square and likelihood ratio test statistics are particular cases of the ϕ - divergence statistic considered. The Pitman and Bahadur efficiencies of tests of independence based on this statistic are obtained. Finally, tests of equality of association between two or more contingency tables are constructed.

Keywords: ϕ - divergence, ϕ - divergence statistic, asymptotic distributions, contingency tables, tests of independence, Pitman efficiency, Bahadur efficiency, tests of association in contingency tables.
1991 A.M.S. Subject Classification: Primary 62B10, Secondary 62E20.

1 Introduction

Divergence or dissimilarity measures serve as indices of the distance or discrepancy between distributions or as indices of dissimilarity between populations. A wide class of these measures, called ϕ - divergence measures, has been independently introduced by Ali and Silvey (1966) and Csiszar (1967). ϕ - divergence measures satisfy many interesting statistical properties (cf. Papaioannou (1985), Zografos et al. (1989)) and have been applied in many fields in an extensively widespread manner. They are based on a convex function ϕ and two distributions and also express the amount of information supplied by data for discrimination between these distributions. Special choices of the convex function ϕ lead to Kullback - Leibler (directed divergence), Renyi (order α information), Kagan, Cressie and Read power divergence etc.

Given a two way contingency table, denote by n_{ij} the frequency in the (i, j) cell for $i = 1, \ldots, k$, $j = 1, \ldots, m$ and suppose that the distribution of n_{ij}'s is the multinomial with parameters N and $\pi_{11}, \ldots, \pi_{km}$. Let $\pi_{i*} = \sum_{j=1}^{m} \pi_{ij}$ and $\pi_{*j} = \sum_{i=1}^{k} \pi_{ij}$ be the marginal probabilities of π_{ij} for $i = 1, \ldots, k$ and $j = 1, \ldots, m$ and $p_{ij} = n_{ij}/N$, $p_{i*} = n_{i*}/N$ and $p_{*j} = n_{*j}/N$ the sample estimators of the population proportions π_{ij}, π_{i*} and π_{*j} respectively where $n_{i*} = \sum_{j=1}^{m} n_{ij}$ and $n_{*j} = \sum_{i=1}^{k} n_{ij}$, $i = 1, \ldots, k$, $j = 1, \ldots, m$.

Consider the probability distributions $\mathbf{\Pi}^T = (\pi_{11}, \ldots, \pi_{km})$, $\mathbf{P}^T = (p_{11}, \ldots, p_{km})$, $\mathbf{\Pi}_0^T = (\pi_{1*}\pi_{*1}, \ldots, \pi_{k*}\pi_{*m})$ and $\mathbf{P}_0^T = (p_{1*}p_{*1}, \ldots, p_{k*}p_{*m})$, where boldtype letters are used to denote vectors or matrices while the superscript T denotes the transpose of a vector or a matrix.

The ϕ - divergence of $\mathbf{\Pi}$ with respect to $\mathbf{\Pi}_0$ is given by

$$(1.1) \qquad I^C(\mathbf{\Pi}, \mathbf{\Pi}_0) = \sum_{i=1}^{k} \sum_{j=1}^{m} \pi_{i*}\pi_{*j}\phi\left(\frac{\pi_{ij}}{\pi_{i*}\pi_{*j}}\right),$$

and its sample estimator, called hereafter ϕ - divergence statistic, by

$$(1.2) \qquad I^C(\mathbf{P}, \mathbf{P}_0) = \sum_{i=1}^{k} \sum_{j=1}^{m} p_{i*}p_{*j}\phi\left(\frac{p_{ij}}{p_{i*}p_{*j}}\right),$$

where ϕ is a real valued convex function on $[0, \infty)$ with $0\phi(0/0) = 0$ and $0\phi(u/0) = u\phi_\infty$ with $\phi_\infty = \lim_{u\to\infty}[\phi(u)/u]$. Thus, $I^C(\mathbf{P}, \mathbf{P}_0)$ measures, in a way, the distance between $\mathbf{P}$ and $\mathbf{P}_0$ and can serve as a test statistic for testing independence.

If $\phi(u) = u \log u$, $\phi(u) = (1-u)^2$, $\phi(u) = (1-\sqrt{u})^2$, $\phi(u) = \mid 1-u \mid^\alpha$, $\alpha \geq 1$, $\phi(u) = sgn(\alpha-1)u^\alpha$, $\alpha > 0 (\alpha \neq 1)$ and $\phi(u) = [\lambda(\lambda+1)]^{-1}[u(u^\lambda - 1)]$, $-\infty < \lambda < \infty$, ϕ - divergence yields the Kullback-Leibler, Kagan, Matusita (square), Vajda divergences, the affinity (Hellinger distance) and Cressie-Read power divergence family between $\mathbf{\Pi}$ and $\mathbf{\Pi}_0$ respectively. Renyi's measure is a logarithmic function of affinity. Kullback-Leibler's and Kagan's divergence statistics, appropriately normalized, are the well known loglikelihood ratio and Pearson statistics for testing independence in a two way contingency table respectively.

In this paper, in section 2, we investigate the asymptotic distribution of $I^C(\mathbf{P}, \mathbf{P}_0)$. Depending on whether or not a suitable condition is satisfied the limiting distribution of $I^C(\mathbf{P}, \mathbf{P}_0)$, after an appropriate normalization, is either normal or the distribution of a linear form in chi square variables. This unifies and generalizes the results of Lomnicki and Zaremba (1959) and Zvarova (1973) who examined asymptotic distributions of sample estimators of Kullback-Leibler and Renyi's information of order α respectively. Next, a normalization of $I^C(\mathbf{P}, \mathbf{P}_0)$ is shown to be asymptotically distributed as a chi square variable with $(k-1)(m-1)$ degrees of freedom, under the assumption of independence. This permits the construction of a large sample test based on the ϕ-divergence statistic. Thus the study and comparision of Pearson's and loglikelihood ratio test statistic is achieved by linking them through, the probably most general, ϕ-divergence statistic. In section 3, following work of Cressie and Read (1984), optimality criteria for test of independence based on ϕ-divergence statistics are studied. It is shown that these tests are equivalent in Pitman sense under a sequence of local alternatives while the loglikelihood ratio test ($\phi(u) = u \log u$) obtains maximal Bahadur efficiency among tests based on ϕ-divergence statistics. In section 4 the problem of statistical dependence in a two way contingency table is discussed and tests of equality of association between two or more contingency tables are constructed.

In a related work, Gil (1989) has estimated unbiasedly quadratic mutual information and considered its asymptotic distribution. Also, Zografos et al. (1990) investigated the asymptotic behaviour of a ϕ - divergence statistic based on one or two independent multinomial populations and constructed multinomial goodness of fit and divergence tests.

2. Asymptotic distribution of ϕ-divergence statistic

Before stating the main results of this section we give the notation which will be needed in the sequel. Let F be the real valued convex function on R^{km} defined for $\mathbf{Q} = (q_{11}, \ldots, q_{km})$ by

$$F(\mathbf{Q}) = \sum_{i=1}^{k} \sum_{j=1}^{m} (\sum_j q_{ij})(\sum_i q_{ij}) \phi \left(\frac{q_{ij}}{(\sum_j q_{ij})(\sum_i q_{ij})} \right).$$

Convexity of F follows from the equality $F(\Pi) = I^C(\Pi, \Pi_0)$ and the convexity of I^C in its arguments (cf. Vajda (1989), p. 271). Assuming that the convex function ϕ admits continuous first and second order derivatives, denote by $D_1(\Pi)$ the vector $D_1^T(\Pi) = (W_{11}, \ldots, W_{1m}, \ldots, W_{k1}, \ldots, W_{km})$ with $W_{ij} = \frac{\partial F(\Pi)}{\partial \pi_{ij}}$ for $i = 1, \ldots, k$, $j = 1, \ldots, m$ and by $D_2(\Pi)$ the $km \times km$ block, non negative definite Hessian matrix of the convex function F, with elements

$$D_2(\Pi) = \left[\frac{\partial^2 F(\Pi)}{\partial \pi_{iv} \partial \pi_{jl}}\right], i, j = 1, \ldots, k \text{ and } v, l = 1, \ldots, m.$$

After a little algebra we have

$$\begin{aligned} W_{ij} &= \sum_{r=1}^{m} \left[\pi_{*r}\phi_{ir} - \frac{\pi_{ir}}{\pi_{i*}}\phi_{ir}^{'}\right] \\ &+ \sum_{s=1}^{k} \left[\pi_{s*}\phi_{sj} - \frac{\pi_{sj}}{\pi_{*j}}\phi_{sj}^{'}\right] + \phi_{ij}^{'} \end{aligned} \tag{2.1}$$

for $i = 1, \ldots, k$ and $j = 1, \ldots, m$ and

$$\begin{aligned} &\frac{\partial^2 F(\Pi)}{\partial \pi_{iv} \partial \pi_{jl}} \\ &= \phi_{il} + \phi_{jv} - u_{il}\phi_{il}^{'} - u_{jv}\phi_{jv}^{'} + u_{il}^2\phi_{il}^{''} + u_{jv}^2\phi_{jv}^{''} \\ &- \delta_{ij}\left[\frac{1}{\pi_{i*}}(u_{il}\phi_{il}^{''} + u_{jv}\phi_{jv}^{''}) - \frac{1}{\pi_{i*}^2}\sum_{r=1}^{m} u_{ir}\pi_{ir}\phi_{ir}^{''}\right] \\ &- \delta_{vl}\left[\frac{1}{\pi_{*v}}(u_{il}\phi_{il}^{''} + u_{jv}\phi_{jv}^{''}) - \frac{1}{\pi_{*v}^2}\sum_{s=1}^{k} u_{sv}\pi_{sv}\phi_{sv}^{''}\right] \\ &+ \frac{\delta_{ij}\delta_{vl}}{\pi_{i*}\pi_{*v}}\phi_{il}^{''}, \quad i, j = 1, \ldots, k, \text{ and } v, l = 1, \ldots, m \end{aligned} \tag{2.2}$$

with $u_{iv} = \frac{\pi_{iv}}{\pi_{i*}\pi_{*v}}$, $\phi_{iv}^{(\mu)}$ the μ-order derivative of ϕ at u_{iv}, $\mu = 1, 2$ and δ_{iv} is Kronecker's delta for $i = 1, \ldots, k$, $v = 1, \ldots, m$.

The following theorems investigate the asymptotic distribution of $I^C(\mathbf{P}, \mathbf{P}_0)$. Depending on whether or not condition (2.6), given below, is satisfied the limiting distribution of $I^C(\mathbf{P}, \mathbf{P}_0)$, after an appropriate normalization, is either normal or the distribution of a linear form in chi square variables.

Theorem 2.1

If $\phi(u)$ is differentiable with ϕ' continuous, then

$$\sqrt{N}\left[I^C(\mathbf{P}, \mathbf{P}_0) - I^C(\mathbf{\Pi}, \mathbf{\Pi}_0)\right] \xrightarrow{L} N(0, \sigma^2),$$

with

$$\sigma^2 = \sum_{i=1}^{k}\sum_{j=1}^{m} \pi_{ij} W_{ij}^2 - \left(\sum_{i=1}^{k}\sum_{j=1}^{m} \pi_{ij} W_{ij}\right)^2, \tag{2.3}$$

provided $\sigma^2 > 0$.

Proof. The first order Taylor expansion of $F(\mathbf{P})$ around the point $\mathbf{\Pi}$ yields

$$F(\mathbf{P}) = F(\mathbf{\Pi}) + D_1^T(\mathbf{\Pi})(\mathbf{P} - \mathbf{\Pi}) + \epsilon_N \parallel \mathbf{P} - \mathbf{\Pi} \parallel$$

or in view of $F(\mathbf{P}) = I^C(\mathbf{P}, \mathbf{P}_0)$ and $F(\mathbf{\Pi}) = I^C(\mathbf{\Pi}, \mathbf{\Pi}_0)$

$$\begin{aligned}(2.4)\quad & \sqrt{N}\left[I^C(\mathbf{P}, \mathbf{P}_0) - I^C(\mathbf{\Pi}, \mathbf{\Pi}_0)\right] - \sqrt{N} D_1^T(\mathbf{\Pi})(\mathbf{P} - \mathbf{\Pi}) \\ & = \sqrt{N}\epsilon_N \parallel \mathbf{P} - \mathbf{\Pi} \parallel\end{aligned}$$

where $\epsilon_N \to 0$ in probability, as $N \to \infty$. ¿From Serfling (1980), p. 108, we have that

$$\sqrt{N}(\mathbf{P} - \mathbf{\Pi}) \xrightarrow{L} N(0, \mathbf{\Sigma})$$

where $\mathbf{\Sigma}$ is a block symmetric matrix of order km with elements

$$\sigma_{ij,vl} = \pi_{iv}(\delta_{ij}\delta_{vl} - \pi_{jl}), \quad i, j = 1, \ldots, k, \ v, l = 1, \ldots, m. \tag{2.5}$$

Therefore $\sqrt{N} \parallel \mathbf{P} - \mathbf{\Pi} \parallel$ has an asymptotic distribution as $N \to \infty$ and $\sqrt{N}\epsilon_N \parallel \mathbf{P} - \mathbf{\Pi} \parallel \to 0$ in probability as $N \to \infty$. From (2.4) we have that the random variable $\sqrt{N}[I^C(\mathbf{P}, \mathbf{P}_0) - I^C(\mathbf{\Pi}, \mathbf{\Pi}_0)]$ has the same asymptotic distribution as the linear form $\sqrt{N} D_1^T(\mathbf{\Pi})(\mathbf{P} - \mathbf{\Pi})$ which is asymptotically distributed according to a normal distribution with zero mean and variance $\sigma^2 = D_1^T(\mathbf{\Pi})\mathbf{\Sigma} D_1(\mathbf{\Pi})$ which after a little algebra leads to (2.3).

Remark

Special choices of the convex function ϕ yield analogous asymptotic results for the divergence measures mentioned in the introduction. From Cauchy-Schwarz inequality it follows that the asymptotic variance of the random variable

$$\sqrt{N}[I^C(\mathbf{P}, \mathbf{P}_0) - I^C(\mathbf{\Pi}, \mathbf{\Pi}_0)]$$

given by (2.3) vanishes if and only if

$$(2.6) \qquad \pi_{ij}(W_{ij} - c) = 0, \text{ for every } i = 1, \dots, k, \; j = 1, \dots, m$$

where c is a constant. If condition (2.6) is satisfied then the random variable considered tends to zero in probability. Whenever condition (2.6) is satisfied we have the following theorem. Condition (2.6) is satisfied in Theorem 2.3.

Theorem 2.2

If the function ϕ admits second derivative and condition (2.6) is satisfied then the asymptotic distribution of the random variable $2N[I^C(\mathbf{P}, \mathbf{P}_0) - I^C(\mathbf{\Pi}, \mathbf{\Pi}_0)]$ is the same as that of $\sum_{i=1}^{r-1} \alpha_i U_i^2$, where r is the number of positive π_{ij}, U_i are independent random variables each having a standard normal distribution and α_i, $i = 1, \dots, r-1$ are the non zero eigenvalues of the matrix $D_2(\mathbf{\Pi})\mathbf{\Sigma}$, where $D_2(\mathbf{\Pi})$ and $\mathbf{\Sigma}$ are given by (2.2) and (2.5) respectively.

Proof. By a Taylor series expansion of $F(\mathbf{P})$ around $\mathbf{\Pi}$ and in view of relations $F(\mathbf{P}) = I^C(\mathbf{P}, \mathbf{P}_0)$ and $F(\mathbf{\Pi}) = I^C(\mathbf{\Pi}, \mathbf{\Pi}_0)$ we obtain

$$(2.7) \qquad \begin{aligned} &2N[I^C(\mathbf{P}, \mathbf{P}_0) - I^C(\mathbf{\Pi}, \mathbf{\Pi}_0)] \\ &\quad - N(\mathbf{P} - \mathbf{\Pi})^T D_2(\mathbf{\Pi})(\mathbf{P} - \mathbf{\Pi}) \\ &= 2ND_1^T(\mathbf{\Pi})(\mathbf{P} - \mathbf{\Pi}) + 2N\epsilon_N \parallel \mathbf{P} - \mathbf{\Pi} \parallel^2 \end{aligned}$$

where $N\epsilon_N \parallel \mathbf{P} - \mathbf{\Pi} \parallel^2 \to 0$ in probability as $N \to \infty$. From (2.6) and the fact that $p_{ij} = 0$ with probability one whenever $\pi_{ij} = 0$ for $i = 1, \ldots, k$, $j = 1, \ldots, m$ we have that $D_1^T(\mathbf{\Pi})(\mathbf{P} - \mathbf{\Pi}) = 0$ with probability one. Therefore from (2.7) the asymptotic distribution of $2N[I^C(\mathbf{P}, \mathbf{P}_0) - I^C(\mathbf{\Pi}, \mathbf{\Pi}_0)]$ is the same as that of the quadratic form $N(\mathbf{P} - \mathbf{\Pi})^T D_2(\mathbf{\Pi})(\mathbf{P} - \mathbf{\Pi})$ as $N \to \infty$. From Corollary, p. 25 of Serfling (1980), and Corollary 2.1 of Dik and de Gunst (1985), the asymptotic normality of $\sqrt{N}(\mathbf{P} - \mathbf{\Pi})$ entails that the asymptotic distribution of $N(\mathbf{P} - \mathbf{\Pi})^T D_2(\mathbf{\Pi})(\mathbf{P} - \mathbf{\Pi})$ is given by the distribution of $\sum_{i=1}^{km} \alpha_i U_i^2$, where α_i, $i = 1, \ldots, km$, are the eigenvalues of $D_2(\mathbf{\Pi})\mathbf{\Sigma}$ and U_i, $i = 1, \ldots, km$ are independent random variables each having a standard normal distribution. Because of (2.6) let r be the number of positive π_{ij}. From (2.5) we obtain that $rank(\mathbf{\Sigma}) \leq r - 1$ and then $rank(D_2(\mathbf{\Pi})\mathbf{\Sigma}) \leq r - 1$. Therefore the nonzero eigenvalues of $D_2(\mathbf{\Pi})\mathbf{\Sigma}$ are at most $r - 1$, which complete the proof of the theorem.

The following lemma will be used to establish the asymptotic distribution of the ϕ-divergence statistic under the assumption of independence in Theorem 2.3 below.

Lemma 2.1

The eigenvalues of the $km \times km$ block matrix with elements, $(\delta_{ij} - \pi_{j*})(\delta_{vl} - \pi_{*l})$, $i, j = 1, \ldots, k$, $v, l = 1, \ldots, m$ are 1 and 0 with multiplicities $(k-1)(m-1)$ and $k + m - 1$ respectively.

The proof of the lemma can be obtained by using elementary properties of determinants.

Theorem 2.3

If the function ϕ admits second derivative and $\phi''(1) \neq 0$ then under the hypothesis $H_0 : \pi_{ij} = \pi_{i*}\pi_{*j}$, $i = 1, \ldots, k$ and $j = 1, \ldots, m$ the statistic $2N[I^C(\mathbf{P}, \mathbf{P}_0) - \phi(1)]/\phi''(1)$ is asymptotically distributed as a chi square random variable with $(k-1)(m-1)$ degrees of freedom.

Proof. Under H_0, from (2.2), we obtain that $D_2(\mathbf{\Pi}) = \phi''(1)(\mathbf{A}+\mathbf{B})$ where $\mathbf{A}$ and $\mathbf{B}$ are block symmetric matrices of order km with elements

$$\mathbf{A} = \left[2\frac{\phi(1) - \phi'(1) + \phi''(1)}{\phi''(1)}\right] \tag{2.8}$$

and

$$\mathbf{B} = \left[\frac{\delta_{ij}\delta_{vl}}{\pi_{i*}\pi_{*v}} - \frac{\delta_{ij}}{\pi_{i*}} - \frac{\delta_{vl}}{\pi_{*v}}\right] \tag{2.9}$$

for $i, j = 1, \ldots, k$ and $v, l = 1, \ldots, m$. It is easy to see that (2.6) is satisfied. Applying now the previous theorem the asymptotic distribution of the ϕ-divergence statistic above is the distribution of $\sum_{i=1}^{km} \alpha_i U_i^2$, where U_i are independent random variables with $N(0,1)$ distribution and α_i are the eigenvalues of $(\mathbf{A}+\mathbf{B})\mathbf{\Sigma}$, with $\mathbf{\Sigma}$ defined by (2.5). After a little algebra $(\mathbf{A}+\mathbf{B})\mathbf{\Sigma}$ is the matrix considered in Lemma 2.1. Therefore its eigenvalues are 1 and 0 and the distribution of $\sum_{i=1}^{km} \alpha_i U_i^2$ is the chi square with $(k-1)(m-1)$ degrees of freedom.

3. Pitman and Bahadur efficiencies of tests of independence

The asymptotic results of the previous section can be used to construct tests of independence in a two way contingency table. The test statistics and their asymptotic distribution under the assumption of independence, are given in Theorem 2.3 for appropriate choices of the convex function ϕ. For $\phi(u) = u \log u$ and $\phi(u) = (1-u)^2$ we obtain the well known loglikelihood ratio and Pearson's tests of independence.

In the following subsections we study optimality criteria for tests of independence based on ϕ-divergence statistic.

3.1 Pitman asymptotic relative efficiency (a.r.e)

For the evaluation of Pitman a.r.e between any two tests based on the ϕ-divergence statistic we need the asymptotic distribution of $2N[I^C(\mathbf{P}, \mathbf{P}_0) - \phi(1)]/\phi''(1)$ under the sequence of the local alternatives

$$(3.1) \qquad H_{a,N} : \pi_{ij} = \pi_{i\bullet}\pi_{\bullet j} + N^{-1/2}C_{ij}, \quad i = 1, \ldots, k, \; j = 1, \ldots, m$$

where the vector $\mathbf{C}^T = (C_{11}, \ldots, C_{km})$ satisfies $\sum_{i=1}^k \sum_{j=1}^m C_{ij} = 0$.

The following lemma will be used in this direction.

Lemma 3.1

Let the k-dimensional vector $\mathbf{X}$ be $N(\mu, \Sigma)$ and let $\mathbf{G}$ be a symmetric non negative definite (n.n.d) matrix of order k. Assume that $\mu \in \mathcal{M}(\mathbf{\Sigma})$, where $\mathcal{M}(\mathbf{\Sigma})$ is the linear space generated by the columns of the matrix $\mathbf{\Sigma}$. Then $\mathbf{X}^T\mathbf{G}\mathbf{X}$ has a non central chi square distribution if and only if the non zero eigenvalues of $\mathbf{G\Sigma}$ are equal to one. In this case the degrees of freedom are *trace* $(\mathbf{G\Sigma})$ and the noncentrality parameter is $\mu^T\mathbf{G}\mu$.

The proof of the lemma can be obtained by using Theorems 2.1, 3.1 and Remarks 2.2 and 2.3 of Dik and de Gunst (1985).

Theorem 3.1

If the function ϕ admits second derivative and $\phi''(1) \neq 0$ then under the hypotheses (3.1) the statistic $2N[I^C(\mathbf{P}, \mathbf{P}_0) - \phi(1)]/\phi''(1)$ is asymptotically non central chi square distributed with $(k-1)(m-1)$ degrees of freedom and non centrality parameter $\delta = \mathbf{C}^T\mathbf{B}\mathbf{C}$, where $\mathbf{C}$ and $\mathbf{B}$ are given by (3.1) and (2.9) respectively.

Proof. Following the steps of the proof of Theorem 2.2, by a Taylor series expansion of $F(\mathbf{P})$ around the point $\mathbf{\Pi}_0^T = (\pi_{1*}\pi_{*1}, \ldots, \pi_{k*}\pi_{*m})$ we obtain that the statistic considered above has the same asymptotic distribution as that of quadratic form $N(\mathbf{P}-\mathbf{\Pi}_0)^T(\mathbf{A}+\mathbf{B})(\mathbf{P}-\mathbf{\Pi}_0)$ where the matrices $\mathbf{A}$ and $\mathbf{B}$ are given by (2.8) and (2.9) respectively.

We can easily see that under (3.1)

$$\sqrt{N}(\mathbf{P}-\mathbf{\Pi}_0) \xrightarrow{L} N(C, \Sigma^*),$$

where Σ^* is a block covariance matrix of order km with elements,

$$\sigma^*_{ij,vl} = \pi_{i*}\pi_{*v}(\delta_{ij}\delta_{vl} - \pi_{j*}\pi_{*l}),\ i,j = 1,\ldots,k,\ v,l = 1,\ldots,m.$$

Also the matrix $\mathbf{A}+\mathbf{B}$ is n.n.d because the matrix $\phi''(1)(\mathbf{A}+\mathbf{B})$ is the Hessian matrix of the convex function $F(\mathbf{\Pi})$. After a little algebra $(\mathbf{A}+\mathbf{B})\mathbf{\Sigma}^*$ is the matrix considered in Lemma 2.1 and therefore its eigenvalues are 1 and 0 with multiplicities $(k-1)(m-1)$ and $k+m-1$ respectively. Obviously $\mathbf{C} \in \mathcal{M}(\mathbf{\Sigma}^*)$ because for the unity vector $\mathbf{1}$ we have $\mathbf{1}^T\mathbf{\Sigma}^* = \mathbf{0}$ and $\mathbf{1}^T\mathbf{C} = 0$. Therefore the proof of the theorem is completed as an application of Lemma 3.1.

The Pitman a.r.e between any two test statistics obtained from $2N[I^C(\mathbf{P},\mathbf{P}_0) - \phi(1)]/\phi''(1)$, for the convex functions ϕ_1 and ϕ_2, is given by the ratio of their non centrality parameters (cf. Puri and Sen (1971), p. 121) and it is therefore equal to one for any pair of ϕ_1 and ϕ_2. Thus the ϕ-divergence test statistics are equivalent in the Pitman sense for local alternatives given by (3.1).

3.2 Bahadur efficiency

Another concept of efficiency was introduced by Bahadur (cf. Bahadur (1971)). In order to evaluate the exact Bahadur efficiency between two ϕ-divergence tests, we require the exact Bahadur slope of these tests. A method for evaluating this slope is described by the following theorem, based on Theorem 7.2 of Bahadur (1971), p. 27 (see also Cressie and Read (1984), p. 447). The proof of the theorem is a generalization of Bahadur (1971), Example 8.3, p. 31 and is omitted.

Theorem 3.2

Let $T_N = \{2N[I^C(\mathbf{P},\mathbf{P}_0) - \phi(1)]/\phi''(1)\}^{1/2}$ and $\Delta = \{\mathbf{\Pi} : \pi_{ij} \geq 0, \sum_{i,j}\pi_{ij} = 1,\ i = 1,\dots,k,\ j = 1,\dots,m\ \}$. Then

a) $\lim_{N\to\infty} N^{-1/2}T_N = \{2[I^C(\mathbf{\Pi},\mathbf{\Pi}_0) - \phi(1)]/\phi''(1)\}^{1/2}$ in probability for $\pi_{ij} \neq \pi_{i*}\pi_{*j}$, $i = 1,\dots,k$, $j = 1,\dots,m$.

b) $\lim_{N\to\infty} N^{-1}\log P(T_N \geq N^{1/2}t) = -\inf_{\mathbf{v}\in A_t} I_0^C(\mathbf{v},\mathbf{v}_0)$ for each t in an open interval, where $b(\mathbf{v}) = \{2[I^C(\mathbf{v},\mathbf{v}_0) - \phi(1)]/\phi''(1)\}^{1/2}$, $A_t = \{\mathbf{v} \in \Delta : b(\mathbf{v}) \geq t\}$ and $I_0^C(\mathbf{v},\mathbf{v}_0) = \sum_{i,j} v_{ij}\log(v_{ij}/v_{i*}v_{*j})$ with $\mathbf{v}_0 = (v_{1*}v_{*1},\dots,v_{k*}v_{*m})$. This result holds under the assumption of independence.

c) The exact Bahadur slope of the test based on

$$2N[I^C(\mathbf{P},\mathbf{P}_0) - \phi(1)]/\phi''(1)$$

is given by

$$C_\phi(\mathbf{\Pi}) = \inf_{\mathbf{v}\in B} 2I_0^C(\mathbf{v},\mathbf{v}_0),\ \ \pi_{ij} \neq \pi_{i*}\pi_{*j}$$

where $B = \{\mathbf{v} \in \Delta : I^C(\mathbf{v},\mathbf{v}_0) \geq I^C(\mathbf{\Pi},\mathbf{\Pi}_0)\}$.

The exact Bahadur efficiency, between two tests obtained from ϕ-divergence statistic for two convex functions ϕ_1 and ϕ_2, is equal to the ratio of its exact Bahadur slopes. A straightforward generalization of Example, p. 448, of Cressie and Read (1984) provides that the loglikelihood ratio test ($\phi(u) = u\log u$)

obtains maximal Bahadur efficiency among all tests based on ϕ-divergence statistic for testing independence in a two way contingency table. This is to be expected in view of Theorem 10.1 in Bahadur (1971).

4. Measures of association and tests of equality of associations between contingency tables

Measures of association i.e. numerical assessments of the strength of the statistical dependence between two or more random variables play an important role in statistics. ϕ-divergence (suitably normed) of the joint distribution of two random variables with respect to the product of their marginal distributions is an appropriate measure of association between the random variables. [cf. Ali and Silvey (1965), Csiszar (1967), Zvarova (1974), Vajda (1989, Chapter 10)]. Therefore the concept of ϕ-divergence $I^C(\Pi, \Pi_0)$, defined by (1.1), can be used also to define a class of measures of association for cross classified random variables tabulated in a two way contingency table. From $I^C(\Pi, \Pi_0)$ can be obtained, for $\phi(u) = (1-u)^2$, Pearson's mean square contingency, for $\phi(u) = u \log u$, Linfoot's (1957) informational measure of dependence, for $\phi(u) = | 1 - u | /2$, Hoffding's coefficient of statistical dependence while for $\phi(u) = sgn(\alpha - 1)u^\alpha$, $\alpha > 0 (\alpha \neq 1)$, the quantity $(\alpha - 1)^{-1} \log | I^C(\Pi, \Pi_0) |$ is the informational measure of correlation of order α, introduced by Pessoa and Dial (1988).

Asymptotic results stated previously provide ϕ-divergence association measures with useful sampling properties. They can also be used to construct tests of equality of association between one or more pairs of cross classified random variables or contingency tables.

Following the notation given in the introduction denote by $n_{ij}^{(s)}$ the frequency in the (i, j) cell of the $s = 1, \ldots, r$ contingency table for $i = 1, \ldots, k_s$ and $j = 1, \ldots, m_s$. Suppose that the distribution of $n_{ij}^{(s)}$'s is the multinomial with parameters N_s and $\mathbf{\Pi}_s^T = (\pi_{11}^{(s)}, \ldots, \pi_{k_s m_s}^{(s)})$. Let $\pi_{i*}^{(s)} = \sum_{j=1}^{m_s} \pi_{ij}^{(s)}$ and $\pi_{*j}^{(s)} = \sum_{i=1}^{k_s} \pi_{ij}^{(s)}$ are the marginal probabilities of $\pi_{ij}^{(s)}$ and $p_{ij}^{(s)} = n_{ij}^{(s)}/N_s$, $p_{i*}^{(s)} = n_{i*}^{(s)}/N_s$ and $p_{*j}^{(s)} = n_{*j}^{(s)}/N_s$ the sample estimators of the population proportions $\pi_{ij}^{(s)}$, $\pi_{i*}^{(s)}$ and $\pi_{*j}^{(s)}$ respectively where $n_{i*}^{(s)} = \sum_{j=1}^{m_s} n_{ij}^{(s)}$ and $n_{*j}^{(s)} = \sum_{i=1}^{k_s} n_{ij}^{(s)}$, for $s = 1, \ldots, r$, $i = 1, \ldots, k_s$ and $j = 1, \ldots, m_s$. Let also $\mathbf{P}_s^T = (p_{11}^{(s)}, \ldots, p_{k_s m_s}^{(s)})$, $\mathbf{\Pi}_{0,s}^T = (\pi_{1*}^{(s)}\pi_{*1}^{(s)}, \ldots, \pi_{k_s*}^{(s)}\pi_{*m_s}^{(s)})$ and $\mathbf{P}_{0,s}^T = (p_{1*}^{(s)}p_{*1}^{(s)}, \ldots, p_{k_s*}^{(s)}p_{*m_s}^{(s)})$, $s = 1, \ldots, r$.

ϕ-divergence statistic $I^C(\mathbf{P}_s, \mathbf{P}_{0,s})$, $s = 1, \ldots, r$, defined as in (1.2), can be used to test the following hypotheses:

i) $H_0 : I^C(\mathbf{\Pi}_1, \mathbf{\Pi}_{0,1}) = I_0$.

ϕ-divergence association of the attributes of a contingency table is of certain magnitude I_0.

ii) $H_0 : I^C(\mathbf{\Pi}_1, \mathbf{\Pi}_{0,1}) = I^C(\mathbf{\Pi}_2, \mathbf{\Pi}_{0,2})$.

The pairs of the attributes of two contingency tables are equally associated.

iii) $H_0 : I^C(\mathbf{\Pi}_1, \mathbf{\Pi}_{0,1}) = \cdots = I^C(\mathbf{\Pi}_r, \mathbf{\Pi}_{0,r}) = I_0$.

The pairs of the attributes of r contingency tables are equally associated to a certain magnitude I_0.

To test the hypotheses i) and ii) we can use the statistics

$$\frac{\sqrt{N_1}[I^C(\mathbf{P}_1, \mathbf{P}_{0,1}) - I_0]}{\hat{\sigma}_1}$$

and

$$\frac{I^C(\mathbf{P}_1, \mathbf{P}_{0,1}) - I^C(\mathbf{P}_2, \mathbf{P}_{0,2})}{\sqrt{[(\hat{\sigma}_1^2/N_1) + (\hat{\sigma}_2^2/N_2)]}}$$

respectively with, by Theorem 2.1, $N(0,1)$ distribution under H_0. The standard deviations are obtained from (2.3) by replacing population parameters by their sample estimators.

To test hypothesis iii) we can use the statistic

$$\sum_{i=1}^{r} \frac{N_i[I^C(\mathbf{P}_i, \mathbf{P}_{0,i}) - I_0]^2}{\hat{\sigma}_i^2},$$

which, as $N_i \to \infty$, $i = 1, \ldots, r$, is distributed as χ_r^2 under H_0. The variances $\hat{\sigma}_i^2$, $i = 1, \ldots, r$ are obtained as above.

To test the hypothesis

$$H_0 : I^C(\mathbf{\Pi}_1, \mathbf{\Pi}_{0,1}) = \cdots = I^C(\mathbf{\Pi}_r, \mathbf{\Pi}_{0,r})$$

we can use the statistic

$$\sum_{i=1}^{r} \frac{N_i[I^C(\mathbf{P}_i, \mathbf{P}_{0,i}) - I_*^C]^2}{\hat{\sigma}_i^2},$$

where

$$I_*^C = \frac{\sum_{i=1}^{r}[N_i I^C(\mathbf{P}_i, \mathbf{P}_{0,i})/\hat{\sigma}_i^2]}{\sum_{i=1}^{r}(N_i/\hat{\sigma}_i^2)}.$$

As $N_i \to \infty$, $i = 1, \ldots, r$ we can easily see that the above considered test statistic is distributed as χ_{r-1}^2 under H_0.

References

[1] Ali, S.M. and Silvey, S.D. (1965) Association between random variables and the dispersion of a Radon - Nikodym derivative, *J. Roy. Statist. Soc.*, **B 27**, 100–107.

[2] Ali, S.M. and Silvey, S.D. (1966) A general class of coefficients of divergence of one distribution from another, *J. Roy. Statist. Soc.*, **B 28**, 131–142.

[3] Bahadur, R.R. (1971) *Some Limit Theorems in Statistics*, SIAM, Philadelfia.

[4] Cressie, N. and Read, T.R.C. (1984) Multinomial goodness of fit tests, *J. Roy. Statist. Soc.*, **B 46**, 440–464.
[5] Csiszar, I. (1967) Information type measures of difference of probability distributions and indirect observations, *Studia Sci. Mat. Hung.*, **2**, 299–318.
[6] Dik, J.J. and Gunst, M.C.M. (1985) The distribution of general quadratic forms in normal variables, *Statistica Neerlandica*, **39**, 14-26.
[7] Gil, M.A. (1989) A note on the choice of sample size in estimating the mutual information, *Appl. Math. Comput.*, **30**, 125–132.
[8] Linfoot, E.H. (1957) An informational measure of correlation, *Information and Control*, **1**, 85–89.
[9] Lomnicki, Z.A. and Zaremba, S.K. (1959) The asymptotic distributions of estimators of the amount of transmitted information, *Information and Control*, **2**, 260–284.
[10] Papaioannou, T. (1985) Measures of information, *Encyclopedia of Statistical Sciences, Vol. 5 (Editors: S. Kotz and N.L. Johnson)*, Wiley, New York, 391–397.
[11] Pessoa, F. and Dial, G. (1988) An informational measure of correlation of order α for contingency tables, *J. Comb. Inf. and Syst. Sci.*, **13**, 79–85.
[12] Puri, M. and Sen, P.K. (1971) *Nonparametric Methods in Multivariate Analysis*, Wiley, New York.
[13] Serfling, R.J. (1980) *Approximation Theorems of Mathematical Statistics*, Wiley, New York.
[14] Vajda, I. (1989) *Theory of Statistical Inference and Information*, Kluwer Academic Publishers, Dordrecht.
[15] Zografos, K., Ferentinos, K. and Papaioannou, T. (1989) Limiting properties of some measures of information, *Ann. Inst. Statist. Math.*, **41**, 451–460.
[16] Zografos, K., Ferentinos, K. and Papaioannou, T. (1990) ϕ - divergence statistics: Sampling properties and multinomial goodness of fit and divergence tests, *Commun. Statist.*,

Theory – Methods, **19**, 1785–1802.

[17] Zvarova, J. (1973) On asymptotic behaviour of a sample estimator of Renyi's information of order α, *Trans. Sixth Prague Conf. on Inf. Theory, Stat. Dec. Func. and Rand. Proc.*, Czech. Acad. of Sci., Prague, 919–924.

[18] Zvarova, J. (1974) On measures of statistical dependence, *Cas. Pest. Mathemat.*, **99**, 15–29.

Received by P.N. Rathie *on* March 20, 1992.

Refereed by Silvio Guiasu, Department of Mathematics, York University, North York, Ontario, Canada M3J 1P3.

Accepted on June 22, 1992 ■