Journal of Multivariate Analysis **69**, 65–87 (1999)
Article ID jmva.1998.1799, available online at http://www.idealibrary.com on IDEAL®

On Burbea–Rao Divergence Based Goodness-of-Fit Tests for Multinomial Models*

M. C. Pardo

Complutense University of Madrid, Madrid, Spain

Received May 30, 1997; revised April 14, 1998

This paper investigates a new family of statistics based on Burbea–Rao divergence for testing goodness-of-fit. Under the simple and composite null hypotheses the asymptotic distribution of these tests is shown to be chi-squared. For composite hypothesis, the unspecified parameters are estimated by maximum likelihood as well as minimum Burbea–Rao divergence. © 1999 Academic Press

AMS 1991 subject classifications: 62B10; 62E20.

Key words and phrases: goodness-of-fit; minimum R_ϕ-divergence estimate; Pitman efficiency; power function.

INTRODUCTION

Many tests of goodness-of-fit can be reduced to testing a hypothesis about the parameter $\pi = (\pi_1, ..., \pi_M)^t$ from a multinomial random variable $X = (X_1, ..., X_M)^t$ of parameters $(n; \pi_1, ..., \pi_M)^t$. This is possible when we group the model into M mutual classes $A_1, A_2, ..., A_M$ with corresponding probabilities $\pi_1, \pi_2, ..., \pi_M$. In this case, the general goodness-of-fit problem reduces to testing a hypothesis H_0 which describes the generally unknown class probability vector $\pi = (\pi_1, ..., \pi_M)^t$, i.e.,

$$H_0 : \pi = \pi_0 = (\pi_{01}, ..., \pi_{0M})^t \in T, \tag{1}$$

where $T \subset \Delta_M = \{(p_1, ..., p_M)^t / \sum_{i=1}^M p_i = 1,\ p_i \geqslant 0,\ i = 1, ..., M\}$ is the null model space of probability vectors. The null hypothesis (e.g., a simple hypothesis) may completely specify π, in which case T is simply a one-point set. Otherwise the null hypothesis is composite, specifying π as a function of a smaller number of unknown parameters (i.e., π lies in the subset T of Δ_M) which needs to be estimated from the experimental data.

To test the general null hypothesis (1), multinomial goodness-of-fit tests measure the discrepancy between the observed proportions X/n and the hypothesized proportions π_0. If the discrepancy is "too large" the null

* This work was supported by Grant DGES PB96-0635 and PR156/97-7159.

0047-259X/99 $30.00

hypothesis is rejected. The key is the choice of a good statistic to measure the discrepancy between X/n and π_0.

The most commonly used statistic is Pearson's X^2 (Pearson [27]),

$$X^2 = \sum_{i=1}^{M} \frac{(X_i - n\hat{\pi}_i)^2}{n\hat{\pi}_i},$$

which is asymptotically distributed as a chi-squared with $M-1$ degrees of freedom when $\hat{\pi}_i = \pi_{0i}$, $i = 1, ..., M$. In the case where the $\hat{\pi}_i$'s depend on parameters that need to be estimated, Pearson argued that using the chi-squared distribution with $M-1$ degrees of freedom would still be adequate for practical decisions. This case was finally settled by Fisher [13], who gave the first derivation of the correct degrees of freedom, namely, $M - M_0 - 1$ when M_0 parameters are estimated efficiently from the data.

Cressie and Read [9] and Read and Cressie [30] proposed a generalized statistic which they called the power divergence statistic or power divergence family of statistics and which is defined as

$$2nI^{\lambda}(X/n, \hat{\pi}) = \frac{2}{\lambda(\lambda+1)} \sum_{i=1}^{M} X_i \left(\left(\frac{X_i}{n\hat{\pi}_i} \right)^{\lambda} - 1 \right), \qquad -\infty < \lambda < \infty, \tag{2}$$

where $\hat{\pi}$ is a BAN estimator of π_0 under the null hypothesis (1). Here λ is a parameter indexing which member of the family is to be used for goodness-of-fit testing. It can easily be seen that Pearson's X^2 ($\lambda = 1$), the loglikelihood ratio statistic ($\lambda = 0$), the Freeman–Tukey statistic ($\lambda = -1/2$), the modified loglikelihood ratio statistic ($\lambda = -1$), and the Neyman modified X^2 ($\lambda = -2$) are all special cases. Furthermore, this generalization provides a family of competitors to the well-known statistics which might prove to be superior in some situations. These authors show that under the same regularity conditions each member of the power divergence family (2) follows the same asymptotic distribution (a $\chi^2_{M-M_0-1}$).

But there is a more general family of statistics that contains the above family, namely

$$T_{\varphi}(X/n, \hat{\pi}) = \frac{2n}{\varphi''(1)} D_{\varphi}(X/n, \hat{\pi}),$$

where $D_{\varphi}(X/n, \hat{\pi})$ is the Csiszár φ-divergence [10] between the observed proportions and the hypothesized proportions, defined by

$$D_{\varphi}(X/n, \hat{\pi}) = \sum_{i=1}^{M} \hat{\pi}_i \varphi \left(\frac{X_i}{n\hat{\pi}_i} \right),$$

for every convex function $\varphi : [0, \infty) \to R \cup \{\infty\}$, where $0\varphi(0/0) = 0$ and $0\varphi(p/0) = \lim_{u\to\infty} \varphi(u)/u$. (See Csiszár [10] and Ali and Silvey [1] or the book of Liese and Vajda [21] devoted to the φ-divergences.) For $\varphi_\lambda(x) = (\lambda(\lambda+1))^{-1}(x^{\lambda+1} - x)$, $\lambda \neq 0, -1$, we have the power divergence family of statistics. Under the simple hypothesis, Zografos *et al.* [35] established that $T_\varphi(X/n, \hat{\pi})$ is asymptotically distributed as chi-squared with $M-1$ degrees of freedom. Morales *et al.* [23], under the composite hypothesis, proved that

$$T_\varphi(X/n, \hat{\pi}) \xrightarrow[n\to\infty]{L} \chi^2_{M-M_0-1},$$

where

$$\hat{\pi} = \arg\min D_\varphi(X/n, \pi)$$

and M_0 is the number of unspecified parameters of π.

In any case we can observe that implicitly or explicitly the goodness-of-fit tests are based on distances, dissimilarities, or simply divergences. For this reason, we can use measures of divergences different from the φ-divergences. For example, there is an important family of divergences, the Jensen differences of R_ϕ-divergences introduced and studied by Rao [28], Burbea and Rao [4, 5], and Burbea [6], which can be used for this proposal. This family is defined, for two probability distributions π^1 and π^2, as

$$R_\phi(\pi^1, \pi^2) = H_\phi\left(\frac{\pi^1 + \pi^2}{2}\right) - \frac{1}{2}\{H_\phi(\pi^1) + H_\phi(\pi^2)\},$$

where $H_\phi(\pi) = \sum_{i=1}^M \phi(\pi_i)$ is the ϕ-entropy, $\phi : (0, \infty) \to R$ being a continuous concave function and $\phi(0) = \lim_{t\downarrow 0} \phi(t) \in (-\infty, \infty]$. Some interesting properties of the ϕ-entropies can be seen in Vajda and Vasek [34]. The convexity of the R_ϕ-divergence is obtained if the function $\phi(x)$ is concave and $\phi''(x)^{-1}$ is convex.

An important family of R_ϕ-divergences is obtained if we consider the entropies of degree α due to Havrda and Charvát [16], with

$$\phi_\alpha(x) = \begin{cases} (1-\alpha)^{-1}(x^\alpha - x), & \alpha \neq 1 \\ -x\log x, & \alpha = 1. \end{cases}$$

In this case the R_{ϕ_α}-divergence is convex if and only if $\alpha \in [1, 2]$, for $M > 2$, and if only if $\alpha \in [1, 2]$ or $\alpha \in [3, 11/3]$, for $M = 2$. Rao [28] used the family of ϕ_α-entropies in genetic diversity between populations. In the particular case of $\alpha = 2$ we obtain the Gini–Simpson index. This measure of entropy was introduced by Gini [14] and by Simpson [32] in biometry and, its properties have been studied by various authors. Note that if we consider the Gini–Simpson index, then the associated R_ϕ-divergence is proportional to the square of the Euclidean distance,

$$R_{\phi_2}(\pi^1, \pi^2) = \tfrac{1}{4} \sum_{i=1}^{M} (\pi_i^1 - \pi_i^2)^2.$$

Another important family of R_ϕ-divergences is obtained if we consider the Bose–Einstein entropy (introduced by Burbea and Rao [4]) or the Fermi–Dirac entropy (Kapur [20]), among others.

On the basis of the R_ϕ-divergence one can introduce a new statistic for the goodness-of-fit problem which will be denoted by $R_\phi(X/n, \hat{\pi})$. For $\phi = \phi_\alpha$ and $\hat{\pi}_i = \pi_{0i} = 1/M$, $i = 1, ..., M$, we have

$$T_{\varphi_\alpha}(X/n, \pi_0) = \frac{-8Mn}{\phi_\alpha''(1/M)} R_{\phi_\alpha}(X/n, \pi_0),$$

where

$$\varphi_\alpha(x) = \begin{cases} 4\,\dfrac{-2(\frac{1}{2}x + \frac{1}{2})^\alpha + x^\alpha + 1}{\alpha(\alpha - 1)}, & \alpha \neq 1 \\ 4\left(x \ln \dfrac{2x}{x+1} + \ln \dfrac{2}{x+1}\right), & \alpha = 1. \end{cases}$$

In the following we put

$$S_\phi(X/n, \pi_0) = \frac{-8Mn}{\phi''(1/M)} R_\phi(X/n, \pi_0).$$

It is interesting to observe that for $\hat{\pi}_i = \pi_{0i} = 1/M$, $i = 1, ..., M$,

$$T_{\varphi_2}(X/n, \pi_0) = S_{\phi_2}(X/n, \pi_0) = \sum_{i=1}^{M} \frac{(X_i - n/M)^2}{n/M}$$

is the classical Pearson statistic.

A question is whether for any ϕ associated to a R_ϕ-divergence there exists a function φ associated to a φ-divergence such that $T_\varphi(X/n, \pi_0) = S_\phi(X/n, \pi_0)$. The answer is negative. An interesting counter example is $\phi(x) = -1/x - x$, where $T_\varphi(X/n, \pi_0) = S_\phi(X/n, \pi_0)$ for $\varphi(x) = 2(x-1)^2/x(x+1)$, which is not convex in the domain $x > 0$. Therefore in this example we cannot apply the result given by Zografos *et al.* [35] and Morales *et* *al.* [23]. So the family of statistics $S_\phi(X/n, \pi_0)$ provides new possibilities although there are many important statistics included in the family $T_\varphi(X/n, \pi_0)$ but not in the family $S_\phi(X/n, \pi_0)$. In this sense we can refer to the result established in Pardo and Vajda [25], according to which the condition

$$\frac{1}{t}\phi\left(t\frac{u+v}{2}\right)-\frac{\phi(tu)+\phi(tv)}{2t}=\phi\left(\frac{u+v}{2}\right)-\frac{\phi(u)+\phi(v)}{2},$$

valid for all positive t, u, v, implies the identity

$$D_\varphi(X/n, \pi_0) = R_\phi(X/n, \pi_0)$$

for

$$\varphi(x)=\phi\left(\frac{x+1}{2}\right)-\frac{\phi(x)+\phi(1)}{2}.$$

For example, the function $\phi(x) = -x \log x$ satisfies the above condition.

In Section 2 we obtain that the asymptotic distribution of the statistic $S_\phi(X/n, \pi_0)$ under the equiprobable null hypothesis is X^2_{M-1}. Furthermore, the members of the family S_{ϕ_α} are compared by means of two criteria, using the methodology introduced by Cressie and Read [9]. The statistic corresponding to ϕ_α with $\alpha = 13/7$ emerges as a good competitor of the classical Pearson statistic. Note that the power divergence statistic corresponding T_{φ_λ} to $\lambda = 2/3$ is also a good competitor of the Pearson statistic (Cressie and Read [9]).

In Section 3 we obtain the asymptotic distribution of the statistic $8nR_\phi(X/n, \hat{\pi})$ under composite hypotheses as well as the asymptotic power of the goodness-of-fit test based on the statistic S_{ϕ_α}. Here it is possible to estimate π by maximum likelihood method but we propose to use the minimum R_ϕ-divergenc estimate ([26]) defined as

$$\hat{\pi} = \arg\min R_\phi(X/n, \pi).$$

2. SIMPLE NULL HYPOTHESIS. OPTIMALITY

Pardo *et al.* [24] proved that if $\phi : (0, \infty) \to R$ is a twice continuously differentiable concave function, under the simple null hypothesis

$$8nR_\phi(X/n, \pi_0) \xrightarrow[n\to\infty]{L} \sum_{i=1}^{r} \beta_i Z_i^2,$$

where $r = \text{rank}(\Sigma_{\pi_0} D(\pi_0)\, \Sigma_{\pi_0})$, $\beta_1, ..., \beta_r$ are the nonzero eigenvalues of $D(\pi_0)\, \Sigma_{\pi_0}$, and Z_i, $i = 1, ..., r$, are independent random variables with standard normal distribution. In this case

$$\Sigma_{\pi_0} = \text{diag}(\pi_0) - \pi_0 \pi_0^t, \qquad \text{and} \qquad D(\pi_0) = \text{diag}(-\phi''(\pi_{01}), ..., -\phi''(\pi_{0M})).$$

From this result we have that the corresponding γ size goodness-of-fit test would reject the null hypothesis if $8nR_\phi(X/n, \pi_0) > t_\gamma$ where t_γ verifies that $P(\sum_{i=1}^{r} \beta_i Z_i^2 > t_\gamma) = \gamma$.

One decision to take on this type of tests is the choice of the lengths of the class intervals in which the range of the variable is divided. In this paper we choose equiprobable intervals:

$$H_0 : \pi_i = \pi_{0i} = 1/M, \qquad \forall i = 1, ..., M. \tag{3}$$

Several reasons that justify this choice. On one hand, Cohen and Sackrowitz [8] proved that for hypothesis (3) a rejected region of the form $\sum_{i=1}^{M} h_i(x_i) > c$, where c is a positive constant, h_i, $i = 1, ..., M$, are convex functions, and $x_i \geqslant 0$, $i = 1, ..., M$, is unbiased. In our case if we choose ϕ so that R_ϕ is convex, the tests proposed are unbiased for equal cell probabilities. On the other hand, Bednarski and Ledwina [2] established that for every fixed number of observations, every continuous and reflexive function $h : \Delta_M \times \Delta_M \to R^+$, and every $0 < c < \sup\{c/P(h(p, x) \geqslant c) < 1, p \in \Delta_M\}$, there is $q \in \Delta_M$ such that the test with rejected region $h(q, x) > c$ is biased for testing $H_0 : p = q$. Here the statistic proposed is a continuous function in $\Delta_M \times \Delta_M - \{(0, 0)\}$ and thus it is not unbiased in general for the unequal cell probabilities case. Finally, as we will prove in Theorem 1, the statistical asymptotic distribution under hypothesis (3) is chi-squared with $M - 1$ degrees of freedom independent of the ϕ function.

THEOREM 1. *Let $\phi : (0, \infty) \to R$ be a twice continuously differentiable concave function and let the second derivative $\phi''(1/M)$ be negative. Under hypothesis* (3),

$$S_\phi(X/n, \pi_0) = -\frac{M}{\phi''(1/M)}\, 8nR_\phi(X/n, \pi_0) \xrightarrow[n \to \infty]{L} \chi^2_{M-1}.$$

Proof. Under the null hypothesis (3), the $D(\pi_0)\, \Sigma_{\pi_0}$ matrix of the above result can be written as $-\phi''(1/M)\, A/M$ where $A = I - 1/M(\mathbf{1})_{i, j, ..., M}$, I being the identity matrix.

Now, as the eigenvalues of A are 0 with multiplicity 1 and 1 with multiplicity $M - 1$; then, the eigenvalues of $-\phi''(1/M)\, A/M$ are 0 with multiplicity 1 and $-\phi''(1/M)/M$ with multiplicity $M - 1$.

Consequently,

$$8nR_{\phi}(X/n, \pi_0) \xrightarrow[n\to\infty]{L} -\phi''(1/M)\frac{1}{M}\chi^2_{M-1}.$$

Thus the desired result holds. ▮

By the above theorem we have that

$$P(S_{\phi}(X/n, \pi_0) > \chi^2_{M-1,\gamma} \mid H_0) \xrightarrow[n\to\infty]{} \gamma,$$

where $\chi^2_{M-1,\gamma}$ verifies that $P(\chi^2_{M-1} > \chi^2_{M-1,\gamma}) = \gamma$. So, for large n sample size and M fixed, the corresponding γ size goodness-of-fit test would reject the null hypothesis (3) if $S_{\phi}(X/n, \pi_0) > \chi^2_{M-1,\gamma}$.

2.1. *Pitman Efficiency*

So far we have compared the large-sample distributions of the R_{ϕ}-divergence family when the hypothesized model (3) is true. An acceptance of this model based on the S_{ϕ} test statistic is accompanied by some possibility of having made an incorrect decision. The question we must ask now is the following: How efficient is the test statistic at distinguishing the hypothesized null model from some alternative “true” model for the sampled population?

The power function of the S_{ϕ} family of statistics quantifies the chance of accepting the alternative model when it is true. It is used to compare the members of the family as the closer the function is to 1, the better the test.

Under fixed alternatives the power function of the family S_{ϕ} converges to 1 as $n \to \infty$. However, it is possible for the alternative probability vector to converge to the null vector as $n \to \infty$ at such a rate that the limiting power is less than one (and greater than the test size γ). We call this limiting value the asymptotic efficiency of the test.

To produce some less trivial asymptotic powers that are not all equal to 1, Cochran [7] describes using a set of local alternative hypotheses,

$$H_{1,n} : \pi = \pi_0 + n^{-1/2}c, \tag{4}$$

where $c = (c_1, ..., c_M)^t$ with $\sum_{i=1}^{M} c_i = 0$, which converges to (3).

The power function of the family of statistics $S_{\phi}(X/n, \pi_0)$ is then defined as

$$\beta_{\phi}^{(n)}(\pi_0 + n^{-1/2}c) = P(S_{\phi}(X/n, \pi_0) > \chi^2_{M-1,\gamma} \mid H_{1,n}).$$

Thus the Pitman asymptotic relative efficiency for comparing two tests based on, say, S_{ϕ_1} and S_{ϕ_2} is defined to be the ratio of their respective efficiencies, i.e., e_{ϕ_1}/e_{ϕ_2}, where

$$e_\phi = \lim_{n \to \infty} \beta_\phi^{(n)}(\pi_0 + n^{-1/2}c). \tag{5}$$

In order to evaluate e_ϕ, it is necessary to obtain the asymptotic distribution of S_ϕ under $H_{1,n}$. It is not difficult to establish, under the same conditions of Theorem 1, that $S_\phi(X/n, \pi_0)$ with $\pi_{0i} = 1/M$, $\forall i = 1, ..., M$ is asymptotically distributed, under $H_{1,n}$, as a non-central chi-squared with $M-1$ degrees of freedom and non-centrality parameter $\delta = M \sum_{i=1}^{M} c_i^2$. From this result it follows that Eq. (5) can be rewritten independent of ϕ. That is,

$$e_\phi = P(\chi^2_{M-1}(\delta) > \chi^2_{M-1,\gamma})$$

and the Pitman asymptotic relative efficiency for any two members S_{ϕ_1} and S_{ϕ_2} is one. This implies that no discrimination between family members is possible using Pitman asymptotic relative efficiency.

In any case the above relevant alternatives help to solve the important problem of finding the sample size required to obtain a fixed power at a fixed alternative for a given level of significance. To solve this problem we need only use the tables of the non-central chi-squared in a convenient way.

2.2. *Comparison of Family Members via Moments*

It is also possible to compare arbitrary goodness-of-fit tests ($S_\phi(X/n, \pi_0)$, $\chi^2_{M-1,\gamma}$), $0 < \gamma < 1$, from the point of view of how the first three moments of the test statistic $S_\phi(X/n, \pi_0)$ match the first three moments of the limiting chi-squared random variable χ^2_{M-1}. The method used here is similar to the method given by Read and Cressie [30]. The proximity is interpreted as a coincidence between moments. The Taylor series expansion of $S_\phi(X/n, \pi_0)$ around $\pi_{0i} = 1/M$, $\forall i = 1, ..., M$ is considered. Thus the statistic only depends on powers of a multinomial variable. So the moments of $S_\phi(X/n, \pi_0)$ are obtained using the moments of a multinomial variable. For the moment $\mu_\beta(S_\phi(X/n, \pi_0)) = E[(S_\phi(X/n, \pi_0))^\beta]$ of given order β with $\pi_{0i} = 1/M$, $i = 1, ..., M$, the asymptotic expansion is given by

$$\mu_\beta(S_\phi(X/n, \pi_0)) = m_{\beta,0}(\phi) + \frac{m_{\beta,1}(\phi)}{n} + o(n^{-1}),$$

where the parameters $m_{\beta,i}(\phi)$, $i=0, 1$, are given by

$$m_{1,0}(\phi)=M-1,$$

$$m_{2,0}(\phi)=M^2-1,$$

$$m_{3,0}(\phi)=M^3+3M^2-M-3,$$

$$m_{1,1}(\phi)=\frac{\phi'''(1/M)}{2\phi''(1/M)}\left(\frac{2}{M}-3+M\right)+\frac{7\phi^{IV}(1/M)}{16\phi''(1/M)}\left(\frac{1}{M^2}-\frac{2}{M}+1\right),$$

$$m_{2,1}(\phi)=-2M+2+\left(\frac{10}{M}-13+2M+M^2\right)\frac{\phi'''(1/M)}{\phi''(1/M)}$$
$$+\frac{1}{4}\left(\frac{\phi'''(1/M)}{\phi''(1/M)}\right)^2\left(\frac{12}{M^2}-\frac{18}{M}+6\right)+\frac{7\phi^{IV}(1/M)}{8\phi''(1/M)}\left(\frac{3}{M^2}-\frac{5}{M}+1+M\right),$$

and

$$m_{3,1}(\phi)=26-24M-2M^2+\frac{\phi'''(1/M)}{2\phi''(1/M)}\left(\frac{210}{M}-243+3M+27M^2+3M^3\right)$$
$$+\frac{7\phi^{IV}(1/M)}{16\phi''(1/M)}\left(\frac{45}{M^2}-\frac{66}{M}+18M+3M^2\right)$$
$$+\frac{1}{4}\left(\frac{\phi'''(1/M)}{\phi''(1/M)}\right)^2\left(\frac{180}{M^2}-\frac{234}{M}+36+18M\right).$$

According to this criterion, for optimal values of ϕ we take the set R_β of roots of the equations $m_{\beta,1}(\phi)=0$, $\beta=1, 2, 3$. Thus we have proved the following result:

THEOREM 2. *The moment $\mu_\beta(S_\phi(X/n, \pi_0))$ matches $\mu_\beta(\chi^2_{M-1})$ up to terms of order n^{-1} if and only if ϕ satisfies the equations $m_{\beta,1}(\phi)=0$, $\beta=1, 2, 3$.*

The equations system given in the above theorem is not easy to solve in general, but if we consider functions ϕ depending on a parameter, for instance $\phi_\alpha(x)=(x^\alpha-x)/(1-\alpha)$ and $M\to\infty$ we have that the roots, irrespective of β, are the solution of the equation $7\alpha^2-27\alpha+26=0$. That is to say we get $\alpha=2$ and $\alpha=13/7$. We must not forget that with this method Read and Cressie [30] got $\lambda=1$ (the chi-squared statistic) and $\lambda=2/3$. Therefore, these authors propose the statistic corresponding to $\lambda=2/3$ as a good alternative to the chi-squared statistic. Observe that in our case $S_{\phi_2}(X/n, \pi_0)$ is the chi-squared statistic and it is reasonable to think that $S_{\phi_{13/7}}(X/n, \pi_0)$ is also a good competitor to the classical chi-squared statistic.

TABLE I

Roots of $m_{\beta,1}(\phi)=0$, $\beta=1,2,3$ $(\alpha_1<\alpha_2)$

	M	2	3	4	5	10	20	40	50	100	200	500
$m_{1,1}(\phi)$	α_1	3.0	2.42	2.23	2.14	2.0	2.0	2.0	2.0	2.0	2.0	2.0
	α_2	2.0	2.0	2.0	2.0	2.98	1.91	1.88	1.88	1.86	1.86	1.85
$m_{2,1}(\phi)$	α_1	3.34	2.52	2.31	2.21	2.07	2.02	2.0	2.0	2.0	2.0	2.0
	α_2	1.65	1.68	1.7	1.71	1.76	1.8	1.83	1.83	1.84	1.85	1.85
$m_{3,1}(\phi)$	α_1	3.69	2.62	2.37	2.27	2.10	2.04	2.01	2.01	2.0	2.0	2.0
	α_2	1.3	1.41	1.47	1.51	1.62	1.72	1.78	1.79	1.82	1.84	1.85

The above result is for large M. For smaller M, Table I shows the roots of the equations $m_{\beta,1}(\phi)=0$, $\beta=1,2,3$ for $M<\infty$ fixed. For $M>20$ all roots are very near to the limiting roots $\alpha=2$ and $\alpha=13/7$. Therefore for $M>20$ choosing $\alpha=13/7$ or 2 results in the fastest convergence of the first three moments to those of a chi-squared random variable with $M-1$ degrees of freedom. For $M\leqslant 20$, it would be reasonable to choose $\alpha\in[1.5,2]$.

2.3. *Exact Power Comparisons*

In the above subsection an optimality criterion to choose the best members of the R_ϕ-divergence family of statistics is given. Now, the most important criterion for comparing tests, namely the power function, is for finite samples often mathematically intractable. However, for given specific choices of sample size, class size, equiprobable null hypothesis, and specified alternative hypothesis,

$$H_1:\pi_i=\begin{cases}\dfrac{M-1-\delta}{M(M-1)} & \text{if} \quad i=1,\ldots,M-1\\[2ex] \dfrac{1+\delta}{M} & \text{if} \quad i=M,\end{cases}$$

where $-1\leqslant\delta\leqslant M-1$ is fixed, it is possible to calculate the exact power on a computer. Read and Cressie [30] first proposed this alternative, which results from the Mth probability being perturbed by δ/M, while the rest are adjusted so that they still sum to one.

First, it is necessary to choose a test size γ and calculate the asociated critical region. If we rely on the chi-squared approximation studied then it is clear that the magnitude of the approximation errors in calculating a level γ test will depend on α parameter value. Any such consistent over- or under-estimate of the true size of the test will confound the power comparisons with α dependent approximation errors. So we calculate the critical region of the exact size γ test. However due to the discrete nature of the critical regions for exact tests based on $S_{\phi_\alpha}(X/n, \pi_0)$, the attainable levels for these test statistics will also be discrete. Therefore it is unlikely that for any given γ, we would be able to match sizes for every α. To overcome this problem the appropriate randomized test of size γ has been used for each α.

We tabulated a subset of possible α values against the power of S_{ϕ_α} for a given δ (see Table II).

Five δ values have been chosen to illustrate the alternative model, $\delta = -0.9$, -0.5, 0.5, 1, and 1.5, for $M = 5$, $n = 10$ and 20, and test size $\gamma = 0.05$.

For alternatives $\delta < 0$ the power decreases with α. For alternatives $\delta > 0$ the reverse occurs except for $n = 10$, where the power stops increasing for $\alpha > 2$. For fixed n, M, and α we can see that the power increases with increasing $|\delta|$. Furthermore, the effect of increasing n from 10 to 20 is to increase the power throughout.

TABLE II

Exact Power Functions for the Randomized Size 0.05 Test of the Symmetric Hypothesis

	$(n = 10, M = 5)$				
	δ				
α	-0.9	-0.5	0.5	1.0	1.5
0.3	0.1619	0.0797	0.0721	0.1414	0.2618
0.5	0.1521	0.0760	0.0733	0.1423	0.2622
0.7	0.1538	0.0765	0.0741	0.1440	0.2639
1.0	0.1582	0.0791	0.0741	0.1552	0.3039
13/7	0.1305	0.0758	0.0846	0.2078	0.4186
2.0	0.1305	0.0756	0.0872	0.2144	0.4276
2.5	0.1246	0.0735	0.0851	0.2076	0.4177
5.0	0.1199	0.0759	0.0816	0.2027	0.4131
	$(n = 20, M = 5)$				
0.3	0.5952	0.1289	0.0783	0.1470	0.2573
0.5	0.5952	0.1289	0.0783	0.1470	0.2577
0.7	0.5689	0.1244	0.0797	0.1566	0.3063
1.0	0.5627	0.1281	0.0886	0.2204	0.4705
13/7	0.2839	0.1081	0.1218	0.3677	0.6950
2.0	0.2684	0.1063	0.1229	0.3725	0.7007
2.5	0.1958	0.0965	0.1255	0.3897	0.7236
5.0	0.1432	0.0860	0.1294	0.4091	0.7484

If we are interested in choosing a test with reasonable power against this family of alternatives, Table II indicates that one should choose $\alpha \in [13/7, 2.5]$. (Although not reported, powers for other values of n and M were calculated and the same conclusion is obtained.)

3. COMPOSITE NULL HYPOTHESIS

In this section, we consider the composite hypothesis

$$H_0 : \pi \in T \subset \Delta_M, \tag{6}$$

where $T = \{Q(\theta), \theta \in \Theta\}$ being $Q(\theta) = (q_1(\theta), ..., q_M(\theta))^t$ and $\theta = (\theta_1, ..., \theta_{M_0})^t \in \Theta \subseteq R^{M_0}$ the unspecified parameters vector.

This goodness-of-fit test requires estimation of the unspecified parameters, i.e., choice of one value $Q(\hat{\theta}) \in T$ that is "most consistent" with the observed proportions x/n. The best-known method to choose $Q(\hat{\theta})$ consists of estimating θ by maximum likelihood, but another sensible way to estimate π_0 is to choose the $Q(\hat{\theta}) \in T$ that is closest to x/n with respect to the measure $R_\phi(\hat{P}, Q(\theta))$. This leads to the minimum R_ϕ-divergence estimate (Pardo [26]), defined as a $\hat{\theta}_\phi \in \Theta$ that verifies

$$R_\phi(\hat{P}, Q(\hat{\theta}_\phi)) = \inf_{\theta \in \Theta} R_\phi(\hat{P}, Q(\theta)).$$

Therefore, it is necessary to obtain the asymptotic distribution of $R_\phi(\hat{P}, Q(\hat{\theta}))$ under H_0, where $\hat{P} = (\hat{p}_1, ..., \hat{p}_M)^t$ is the relative frequencies vector and $Q(\hat{\theta}) = (q_1(\hat{\theta}), ..., q_M(\hat{\theta}))^t$, $\hat{\theta}$ being the maximum likelihood or minimum R_ϕ-divergence estimator.

Through the paper we abbreviate $(-\phi''(q_1(\theta)), ..., -\phi''(q_M(\theta)))$ by $-\phi''(Q(\theta))$. Before the asymptotic distribution of $R_\phi(\hat{P}, Q(\hat{\theta}))$ is calculated, the following general result is established.

LEMMA 3. *Let $\hat{P} = (\hat{p}_1, ..., \hat{p}_M)^t$ and $\hat{Q} = (\hat{q}_1, ..., \hat{q}_M)^t$ be c_n-consistent estimates of the unknown distribution $\pi = Q(\theta^0)$ for some $c_n \uparrow \infty$. If $Q(\theta^0)$ satisfies $\pi_i = q_i(\theta^0) > 0$ for $i = 1, ..., M$ then, for all R_ϕ-divergence with ϕ concave and twice continuously differentiable on* $(0, \infty)$,

$$c_n^2 R_\phi(\hat{P}, \hat{Q}) \approx \tfrac{1}{8} c_n^2 (\hat{P} - \hat{Q})^t \, D(\theta^0)(\hat{P} - \hat{Q}),$$

where

$$D(\theta^0) = \operatorname{diag}(-\phi''(Q(\theta^0))).$$

Proof. Consider the random vector $\hat{W}=(\hat{w}_1, ..., \hat{w}_{2M})^t=(\hat{P}, \hat{Q})$ and the vector variable $W=(w_1, ..., w_{2M})^t=(P, Q)$, where $P=(p_1, ..., p_M)^t$ and $Q=(q_1, ..., q_M)^t$ are probability distributions pertaining to Δ_M. Further, define

$$\psi(W)=R_\phi(P, Q) \qquad \text{and} \qquad W^0=(Q(\theta^0), Q(\theta^0)).$$

By the mean value theorem

$$\psi(\hat{W})=\psi(W^0)+(\hat{W}-W^0)^t\, a(W^0)+\tfrac{1}{2}(\hat{W}-W^0)^t\, K(W^*)(\hat{W}-W^0),$$

where the $2M$-vector function $a(W)=(a_j(W))_{j=1, ..., 2M}$ is defined by

$$a_j(W)=\frac{\partial\psi(W)}{\partial w_j},$$

the $2M\times 2M$-matrix function $K(W)=(k_{jr}(W))_{j, r=1, ..., 2M}$ is defined by

$$k_{jr}(W)=\frac{\partial^2\psi(W)}{\partial w_j\,\partial w_r},$$

and W^* is a random vector satisfying the condition

$$\|W^*-W^0\|\leqslant\|\hat{W}-W^0\|. \tag{7}$$

Further, the continuity of ϕ'' implies that all functions $k_{jr}(W)$ are continuous in W. Therefore the consistency of $\hat{P}$ and $\hat{Q}$ together with (7) implies that the matrix $K(W^*)$ tends elementwise to $K(W^0)$ in probability.

Letting $K(W^0)=\frac{1}{4}K$ with

$$K=\begin{pmatrix} D(\theta^0) & -D(\theta^0) \\ -D(\theta^0) & D(\theta^0) \end{pmatrix},$$

it follows that

$$\begin{aligned}
&(\hat{W}-W^0)^t\, K(\hat{W}-W^0)\\
&\quad=(\hat{P}-Q(\theta^0))^t\, D(\theta^0)(\hat{P}-Q(\theta^0))-2(\hat{P}-Q(\theta^0))^t\, D(\theta^0)\\
&\qquad\times(\hat{Q}-Q(\theta^0))+(\hat{Q}-Q(\theta^0))^t\, D(\theta^0)(\hat{Q}-Q(\theta^0))\\
&\quad=(\hat{P}-\hat{Q})^t\, D(\theta^0)(\hat{P}-\hat{Q}).
\end{aligned}$$

Finally, taking into account the identity

$$\psi(W^0)=0 \qquad \text{and} \qquad a(W^0)=(0)_{i=1, ..., 2M}$$

we obtain that, for every random variable X and every sequence c_n

$$|c_n^2 R_\phi(\hat{P}, \hat{Q}) - X| \leqslant |\tfrac{1}{2} c_n^2 (\hat{W} - W^0)^t K(W^0)(\hat{W} - W^0) - X|$$
$$+ |\tfrac{1}{2} c_n^2 (\hat{W} - W^0)^t (K(W^*) - K(W^0))(\hat{W} - W^0)|.$$

The first term equals

$$|\tfrac{1}{8} c_n^2 (\hat{P} - \hat{Q})^t D(\theta^0)(P - \hat{Q}) - X|,$$

so that it suffices to prove that the second term tends in probability to zero.

The second term is upper-bounded by

$$\frac{(c_n \|\hat{P} - Q(\theta^0)\|)^2 + (c_n \|\hat{Q} - Q(\theta^0)\|)^2}{2} \max_{j, r} |k_{jr}(W^*) - k_{jr}(W^0)|.$$

In this bound the c_n-consistency of $\hat{P}$ and $\hat{Q}$ implies

$$\frac{(c_n \|\hat{P} - Q(\theta^0)\|)^2 + (c_n \|\hat{Q} - Q(\theta^0)\|)^2}{2} \leqslant O_p(1).$$

Hence the elementwise convergence of $K(W^*)$ to $K(W^0)$ established previously implies that $\max |k_{jr}(W^*) - k_{jr}(W^0)|$ tends to 0 in probability. Thus the desired convergence to zero holds. ∎

The following theorem obtains the asymptotic distribution of $R_\phi(\hat{P}, Q(\theta))$ under H_0, when θ is estimated by minimum R_ϕ-divergence. We restrict ourselves to unknown parameters θ^0 satisfying the regularity conditions introduced by Birch [3].

1. θ^0 is an interior point of Θ.

2. $\pi_i = q_i(\theta^0) > 0$ for $i = 1, ..., M$. Thus $\pi = (\pi_1, ..., \pi_M)^t$ is an interior point of T.

3. The mapping $Q : \Theta \to \Delta_M$ is totally differentiable at θ^0 so that the partial derivatives of q_i with respect to each θ_j exist at θ^0 and $Q(\theta)$ has a linear approximation at θ^0 given by

$$q_i(\theta) = q_i(\theta^0) + \sum_{j=1}^{M} (\theta_j - \theta_j^0) \frac{\partial q_i(\theta^0)}{\partial \theta_j} + o(\|\theta - \theta^0\|) \qquad \text{as} \quad \theta \to \theta^0.$$

4. The Jacobian matrix

$$\left(\frac{\partial Q(\theta)}{\partial \theta}\right)_{\theta=\theta^0}=\left(\frac{\partial q_i(\theta^0)}{\partial \theta_j}\right)_{\substack{i=1,\ldots,M\\ j=1,\ldots,M_0}}$$

is of full rank.

5. The inverse mapping $Q^{-1}: T \to \Theta$ is continuous at $Q(\theta^0)=\pi$.

Condition 1 ensure the existence of an estimator belonging to Θ. Condition 2 merely ensures that there really are T cells in the multinomial and not fewer. Conditions 3 and 4 combine to ensure that the model really does have M_0 parameters and not fewer. Condition 5 ensures the consistency of the estimates of θ^0.

THEOREM 4. *Let $\phi^*:(0,\infty)\to R$ be a twice continuously differentiable concave function. Let $\hat{P}$ be the relative frequencies vector, $Q:\Theta\to\Delta_M$ a function with continuous second partial derivatives in a neighborhood of θ^0, and $\hat{Q}_\phi = Q(\hat{\theta}_\phi)$; then under the Birch regularity conditions [3] we have that*

$$8nR_{\phi^*}(\hat{P},\hat{Q}_\phi)\xrightarrow[n\to\infty]{L}\sum_{i=1}^{r}\beta_i Z_i^2,$$

where $r=\operatorname{rank}(\Sigma_1 D(\theta^0)\,\Sigma_1)$, the Z_i are independent random variables with standard normal distribution and the β_i are the eigenvalues of the matrix $D(\theta^0)\,\Sigma_1$, where

$$D(\theta^0)=\operatorname{diag}(-\phi^{*\prime\prime}(Q(\theta^0)))$$

and

$$\Sigma_1=(I-J(\theta^0)\,B(\theta^0))\,\Sigma_{Q(\theta_0)}(I-J(\theta^0)\,B(\theta^0))^{\mathrm{t}}$$

with

$$B(\theta^0)=(A(\theta^0)^{\mathrm{t}}\,A(\theta^0))^{-1}\,A(\theta^0)^{\mathrm{t}}\operatorname{diag}(\sqrt{-\phi''(Q(\theta^0))}),$$

$$J(\theta^0)=\left(\frac{\partial q_j(\theta^0)}{\partial \theta_r}\right)_{\substack{j=1,\ldots,M\\ r=1,\ldots,M_0}}, \qquad A(\theta^0)=\operatorname{diag}(\sqrt{-\phi''(Q(\theta^0))})\,J(\theta^0)$$

and

$$\Sigma_{Q(\theta^0)}=\operatorname{diag}(Q(\theta^0))-Q(\theta^0)\,Q(\theta^0)^{\mathrm{t}}.$$

Proof. By Lemma 3, for $\hat{P}$ and $\hat{Q}_\phi$ $\sqrt{n}$-consistent estimates, we have that

$$8nR_\phi * (\hat{P}, \hat{Q}_\phi) \approx n(\hat{P}-\hat{Q}_\phi)^{t} D(\theta^0)(\hat{P}-\hat{Q}_\phi).$$

By Theorems 1 and 2(b) of Pardo [26] we know that

$$\sqrt{n}(\hat{Q}_\phi - Q(\theta^0)) \approx \sqrt{n}\, J(\theta^0)\, B(\theta^0)(\hat{P} - Q(\theta^0)),$$

so

$$\begin{aligned}\sqrt{n}\,(\hat{P}-\hat{Q}_\phi) &= \sqrt{n}\,(\hat{P} - Q(\theta^0)) + \sqrt{n}(Q(\theta^0) - \hat{Q}_\phi)\\ &\approx \sqrt{n}(I - J(\theta^0)\, B(\theta^0))(\hat{P} - Q(\theta^0)).\end{aligned}$$

Consequently,

$$\sqrt{n}(\hat{P}-\hat{Q}_\phi) \xrightarrow[n\to\infty]{L} N(0, \Sigma_1),$$

where

$$\Sigma_1 = (I - J(\theta^0)\, B(\theta^0))\, \Sigma_{Q(\theta^0)}(I - J(\theta^0)\, B(\theta^0))^{t},$$

so by Corollary 2.1 of Dik and Gunst [11], $8nR_{\phi*}(\hat{P}, Q(\hat{\theta}_\phi))$ is asymptotically distributed as $\sum_{i=1}^{r} \beta_i Z_i^2$, where $r = \text{rank}(\Sigma_1 D(\theta^0)\, \Sigma_1)$, $\beta_1, ..., \beta_r$ are the nonzero eigenvalues of $D(\theta^0)\, \Sigma_1$, and Z_i, $i = 1, ..., r$, are independent standard normal random variables. ∎

Remark 1. By the above theorem for large n and significance level γ the proposed test rejects the null hypothesis if $8nR_{\phi*}(\hat{P}, Q(\hat{\theta}_\phi)) > t_\gamma$, where t_γ satisfies the condition

$$\sup_{\theta^0 \in \Theta} P\left(\sum_{i=1}^{r} \beta_i(\theta^0)\, Z_i^2 > t_\gamma\right) \leqslant \gamma.$$

Note that we write $\beta_i(\theta^0)$ instead of β_i because it depends on the true value of θ. From a practical point of view we have two ways to carry out the test:

(i) Given θ^0 fixed we can find the value $t_\gamma(\theta^0)$ verifying

$$P\left(\sum_{i=1}^{r} \beta_i(\theta^0)\, Z_i^2 > t_\gamma(\theta^0)\right) \leqslant \gamma$$

and then we can define $t_\gamma = \sup_{\theta^0 \in \Theta} t_\gamma(\theta^0)$.

(ii) Given a value of the statistic we can calculate for each θ

$$P(\theta) = P\left(\sum_{i=1}^{r} \beta_i(\theta)\, Z_i^2 > 8nR_{\phi^*}(\hat{P}, Q(\hat{\theta}_\phi))\right)$$

and if $\sup_{\theta^0 \in \Theta} P(\theta^0) < \gamma$ then we have evidence to reject the null hypothesis.

We ought to have to calculate a probability of a linear combination of chi-squared distributions for each θ. One may feel a little worried but after reading the papers of Rao and Scott [29] and Modarres *et al.* [22] that feeling disappears. They give some ideas on how to overcome this situation. In fact, a variety of problems in statistical inference and applied probability require either percentiles or probabilities from the distribution of a combination of chi-squares per se; see Jensen and Solomon [18].

Corollary 1 of Rao and Scott [29] propose to consider the statistic

$$(\beta^*(\theta^0))^{-1}\, 8nR_{\phi^*}(\hat{P}, Q(\hat{\theta}_\phi)) \leqslant \sum_{i=1}^{r} Z_i^2 \simeq \chi_r^2,$$

where $\beta^*(\theta^0) = \max\{\beta_1(\theta^0), ..., \beta_r(\theta^0)\}$. In this case if we consider the statistic

$$a(\theta^*)\, 8nR_{\phi^*}(\hat{P}, Q(\hat{\theta}_\phi))$$

with $a(\theta^*) = \inf_{\theta^0 \in \Theta} \beta^*(\theta^0)^{-1}$, we have $\forall \theta^0 \in \Theta$

$$P(8na(\theta^*)\, R_{\phi^*}(\hat{P}, Q(\hat{\theta}_\phi)) > \chi^2_{r,\gamma}) \leqslant P\left(\frac{8n}{\beta^*(\theta^0)}\, R_{\phi^*}(\hat{P}, Q(\hat{\theta}_\phi)) > \chi^2_{r,\gamma}\right) = \gamma$$

and then we would reject the null hypothesis with a significance level γ if

$$8na(\theta^*)\, R_{\phi^*}(\hat{P}, Q(\hat{\theta}_\phi)) > \chi^2_{r,\gamma}$$

or if

$$P(\chi_r^2 > 8na(\theta^*)\, R_{\phi^*}(\hat{P}, Q(\hat{\theta}_\phi))) < \gamma.$$

It is clear that in this case we get an asymptotically conservative test.

Another simple approach to the asymptotic distribution of the statistic $8nR_{\phi^*}(\hat{P}, Q(\hat{\theta}_\phi))$ is the modified statistic

$$(\bar{\beta}(\theta^0))^{-1}\, 8nR_{\phi^*}(\hat{P}, Q(\hat{\theta}_\phi)) \leqslant \sum_{i=1}^{r} Z_i^2 \simeq \chi_r^2$$

with $\bar{\beta}(\theta^0)=\sum_{i=1}^{r}\beta_i(\theta^0)/r$ as a χ_r^2 random variable under H_0. We can observe that

$$E[(\bar{\beta}(\theta^0))^{-1}\,8nR_{\phi*}(\hat{P},Q(\hat{\theta}_\phi))]=r=E[\chi_r^2],$$

$$V[(\bar{\beta}(\theta^0))^{-1}\,8nR_{\phi*}(\hat{P},Q(\hat{\theta}_\phi))]=2r+2\sum_{i=1}^{r}\frac{(\beta_i(\theta^0)-\bar{\beta}(\theta^0))^2}{\bar{\beta}(\theta^0)^2}>V[\chi_r^2]$$

Note that, if we denote by $\Lambda=\mathrm{diag}(\beta_1(\theta^0),\ldots,\beta_r(\theta^0))$ we get

$$E\left[\sum_{i=1}^{r}\beta_i(\theta^0)\,Z_i^2\right]=\sum_{i=1}^{r}\beta_i(\theta^0)=\mathrm{trace}\,\Lambda=\mathrm{trace}(D(\theta^0)\,\Sigma_1)$$
$$=-\sum_{i=1}^{r}\frac{s_{ii}(\theta^0)}{\phi''(q_i(\theta^0))},$$

where $s_{ii}(\theta^0)$ are the diagonal elements of the matrix Σ_1. Then we can calculate $\bar{\beta}(\theta^0)$ by

$$\bar{\beta}(\theta^0)=-\frac{1}{r}\sum_{i=1}^{r}\frac{s_{ii}(\theta^0)}{\phi''(q_i(\theta^0))}.$$

Now in a similar way to the previous approximation we would reject the null hypothesis if $8nb(\theta^*)\,R_{\phi*}(\hat{P},Q(\hat{\theta}_\phi))>\chi_{r,\gamma}^2$ with $b(\theta^*)=\inf_{\theta^0\in\Theta}\bar{\beta}(\theta^0)^{-1}$.

Satterthwaite [31] presented an approximation based on the statistic $c8nR_{\phi*}(\hat{P},Q(\hat{\theta}_\phi))+d$ such that c and d are chosen to get that its expectation and variance coincide with the expectation and variance of a χ_r^2. Jensen and Solomon [18] presented a normal approximation and employed a Wilson–Hilferty type scheme to accelerate the rate of convergence to normality. Imhof [17] considered a nonstatistical approximation based directly on the numerical inversion of the characteristic function. Apart from the above approximations it is possible to consider tables of the cumulative distribution of $\sum_{i=1}^{k}a_iZ_i^2$ in the case of small k (see Solomon [33], Johnson and Kotz [19], Eckler [12], and Gupta [15]).

Note that the proposed tests are asymptotically consistent. Since

$$R_\phi(\hat{P},Q(\hat{\theta}_\phi))\xrightarrow[n\to\infty]{P}R_\phi(\pi,Q_\phi)>0\qquad\text{as}\quad n\to\infty,$$

it follows that

$$P(8nR_\phi(\hat{P},Q(\hat{\theta}_\phi))>t_\gamma)=P(R_\phi(\hat{P},Q(\hat{\theta}_\phi))>t_\gamma/8n)\to 1\qquad\text{as}\quad n\to\infty.$$

This holds true for $R_{\phi*}(\hat{P},Q(\hat{\theta}_\phi))$.

If some alternative $\pi \neq Q(\theta^0)$ is the true value of the parameter, then the probability of rejecting H_0, using the above test with a fixed α, tends to 1 as $n \to \infty$. It is important, in the same way that in the simple hypothesis, to obtain an approximation to the power by considering a sequence of alternatives hypothesis instead of a fixed alternative. In this sense, we now investigate the following alternatives hypotheses

$$H_{1,n} : Q = Q(\theta^0) + n^{-1/2} d,$$

d being a fixed vector in R^M with $d = (d_1, ..., d_M)^t$ and $\sum_{i=1}^{M} d_i = 0$. We know that

$$\sqrt{n}(\hat{P} - Q(\theta^0)) = \sqrt{n}(\hat{P} - Q) + d \quad \text{and} \quad \sqrt{n}(\hat{P} - \hat{Q}_\phi) \xrightarrow[n \to \infty]{L} N(K(\theta^0)\, d, \Sigma_1)$$

with $K(\theta^0) = J(\theta^0)\, B(\theta^0)$. Therefore, applying Corollary 2.1 of Dik and Gunst [11], we have the following result.

THEOREM 5. *Let $\phi^* : (0, \infty) \to R$ be a twice continuously differentiable concave function. Let $\hat{P}$ be the relative frequencies vector, $Q : \Theta \to \Delta_M$ a function with continuous second partial derivatives in a neighborhood of θ^0 and $\hat{Q}_\phi = Q(\hat{\theta}_\phi)$, $r = \operatorname{rank}(\Sigma_1 D(\theta^0)\, \Sigma_1)$, $r \geqslant 1$ and $\beta_1, ..., \beta_r$ the positive eigenvalues of $D(\theta^0)\, \Sigma_1$. Then*

$$8nR_{\phi^*}(\hat{P}, \hat{Q}_\phi) - (d^t K(\theta^0)^t\, \Sigma_1 d + \xi) = \sum_{i=1}^{r} \beta_i (Z_i + w_i)^2,$$

where

$$w = \Lambda^{-1} R^t (\Sigma_1^{1/2})^t\, S^t \Sigma_1 K(\theta^0)\, d, \qquad \xi = d^t K(\theta^0)^t\, K(\theta^0)\, d - w^t \Lambda w,$$

where $\Lambda = \operatorname{diag}(\beta_1, ..., \beta_r)$ and R is the matrix of corresponding orthonormal eigenvectors.

The comments in Remark 1 can be used in this case to find an appropriate approach to the above linear combination of noncentral chi-square distribution.

COROLLARY 6. *Let $\hat{P}$ be the relative frequencies vector and $\hat{Q}_1 = Q(\hat{\theta}_1)$, where $\hat{\theta}_1$ is the minimum R-divergence ($\phi(x) = -x \ln x$), then under Birch regularity conditions [3] and assuming that $Q : \Theta \to \Delta_M$ is a function with continuous second partial derivatives in a neighborhood of θ^0,*

$$8nR(\hat{P}, \hat{Q}_1) \xrightarrow[n \to \infty]{L} \chi^2_{M - M_0 - 1}.$$

Proof. By Theorem 4,

$$8nR(\hat{P}, \hat{Q}_1) \xrightarrow[n\to\infty]{L} \sum_{i=1}^{r} \beta_i Z_i^2,$$

where the β_i are the eigenvalues of the matrix

$$T = \mathrm{diag}(Q(\theta^0)^{-1/2})(\Sigma_{Q(\theta^0)} - J(\theta^0)\, B(\theta^0)\, \Sigma_{Q(\theta^0)} - \Sigma_{Q(\theta^0)} B(\theta^0)^{\mathrm{t}}\, J(\theta^0)^{\mathrm{t}}$$
$$+ J(\theta^0)\, B(\theta^0)\, \Sigma_{Q(\theta^0)} B(\theta^0)^{\mathrm{t}}\, J(\theta^0)^{\mathrm{t}})\, \mathrm{diag}(Q(\theta^0)^{-1/2}).$$

It is easy to obtain that

$$T = I - (Q(\theta^0)^{1/2})(Q(\theta^0)^{1/2})^{\mathrm{t}} - A(\theta^0)(A(\theta^0)^{\mathrm{t}}\, A(\theta^0))^{-1}\, A(\theta^0)^{\mathrm{t}}.$$

This matrix is idempotent so it only has eigenvalues 0 and 1, with the number of unitary eigenvalues as

$$\mathrm{trace}(T) = \mathrm{trace}(I) - \mathrm{trace}((Q(\theta^0)^{1/2})(Q(\theta^0)^{1/2})^{\mathrm{t}})$$
$$- \mathrm{trace}(A(\theta^0)(A(\theta^0)^{\mathrm{t}}\, A(\theta^0))^{-1}\, A(\theta^0)^{\mathrm{t}}) = M - 1 - M_0.$$

Thus the desired result holds. ∎

Remark 2. Under the assumptions of Corollary 6 if we consider the alternative hypotheses $H_{1,n}: Q = Q(\theta^0) + n^{-1/2} d$ using Corollary 2.3 in Dik and Gunst we have

$$8nR(\hat{P}, \hat{Q}_1) = \sum_{i=1}^{r} (Z_i + w_i)^2,$$

since in this case $\beta_1 = \cdots = \beta_r = 0$. That is to say, the asymptotic distribution of the statistic $8nR_{\phi*}(\hat{P}, \hat{Q}_\phi)$ is a noncentral chi-square distribution with $M - 1 - M_0$ degrees of freedom and noncentrality parameter

$$\lambda = d^{\mathrm{t}}\, \mathrm{diag}(Q(\theta^0)^{-1/2})(I - A(\theta^0)(A(\theta^0)^{\mathrm{t}}\, A(\theta^0))^{-1}\, A(\theta^0)^{\mathrm{t}})\, \mathrm{diag}(Q(\theta^0)^{-1/2})\, d.$$

The following theorem obtains the asymptotic distribution of $R_\phi(\hat{P}, Q(\hat{\theta}_{\mathrm{MLE}}))$ under H_0 when θ is estimated by maximum likelihood from the discrete model.

THEOREM 7. *Let $\phi: (0, \infty) \to R$ be a twice continuously differentiable concave function. Let $\hat{P}$ the relative frequencies vector and $\hat{Q}_{\mathrm{MLE}} = Q(\hat{\theta})$ where $\hat{\theta}_{\mathrm{MLE}}$ is the maximum likelihood estimate, then under Birch regularity* *conditions* [3] *and assuming that* $Q : \Theta \to \Delta_M$ *is a function with continuous second partial derivatives in a neighborhood of* θ^0,

$$8nR_\phi(\hat{P}, \hat{Q}_{\text{MLE}}) \xrightarrow[n\to\infty]{L} \sum_{i=1}^{r} \beta_i Z_i^2,$$

where the Z_i^2 *are independent and the* β_i *are the eigenvalues of the matrix* $D(\theta^0)\, \Sigma_2$, *where*

$$D(\theta^0) = \text{diag}(-\phi''(Q(\theta^0)))$$

and

$$\begin{aligned}\Sigma_2 = {} & (I - J(\theta^0)\, I(\theta^0))^{-1}\, J(\theta^0)\, \text{diag}(Q(\theta^0)^{-1}))\, \Sigma_{Q(\theta^0)} \\ & \times (I - J(\theta^0)\, I(\theta^0))^{-1}\, J(\theta^0)\, \text{diag}(Q(\theta^0)^{-1}))^{\text{t}}\end{aligned}$$

where $I(\theta^0)$ *is the Fisher Information matrix of the discrete model and* $\Sigma_{Q(\theta^0)} = \text{diag}(Q(\theta^0)) - Q(\theta^0)\, Q(\theta^0)^{\text{t}}$.

Proof. By Lemma 3, with $\hat{P}$ and $\hat{Q}_{\text{MLE}}$ $\sqrt{n}$-consistent estimates, we have that

$$8nR_\phi(\hat{P}, \hat{Q}_{\text{MLE}}) \approx n(\hat{P} - \hat{Q}_{\text{MLE}})^{\text{t}}\, \text{diag}(-\phi''(Q(\theta^0)))(\hat{P} - \hat{Q}_{\text{MLE}}).$$

Furthermore, from Lemma 2 of Morales *et al.* [23] it follows that

$$\sqrt{n}(\hat{P} - \hat{Q}_{\text{MLE}}) \xrightarrow[n\to\infty]{L} N(0, \Sigma_2),$$

where

$$\begin{aligned}\Sigma_2 = {} & (I - J(\theta^0)\, I(\theta^0))^{-1}\, J(\theta^0)\, \text{diag}(Q(\theta^0)^{-1}))\, \Sigma_{Q(\theta_0)} \\ & \times (I - J(\theta^0)\, I(\theta^0))^{-1}\, J(\theta^0)\, \text{diag}(Q(\theta^0)^{-1}))^{\text{t}}.\end{aligned}$$

Therefore, $8nR_\phi(\hat{P}, \hat{Q}_{\text{MLE}})$ is asymptotically distributed as $\sum_{i=1}^{r} \beta_i Z_i^2$, where the Z_i^2 are independent and the β_i are the eigenvalues of the matrix $D(\theta^0)\, \Sigma_2$. ∎

COROLLARY 8. *Under Birch regularity conditions* [3] *and assuming that* $Q : \Theta \to \Delta_M$ *is a function with continuous second partial derivatives in a neighborhood of* θ^0,

$$8nR(\hat{P}, \hat{Q}_{\text{MLE}}) \xrightarrow[n\to\infty]{L} \chi^2_{M - M_0 - 1}.$$

Proof. The result follows from Theorem 7. ∎

ACKNOWLEDGMENT

The author thanks the referees for some helpful comments which have improved the paper.

REFERENCES

1. M. S. Ali and D. Silvey, A general class of coefficients of divergence of one distribution from another, *J. Roy. Statist. Soc. Ser. B* **28** (1966), 131–140.
2. T. Bednarski and T. Ledwina, A note on a biasedness of tests of fit, *Math. Oper. Statist. Ser. Statist.* **9** (1978), 191–193.
3. M. W. Birch, A new proof of the Pearson–Fisher theorem, *Ann. of Math. Statist.* **35** (1964), 817–824.
4. J. Burbea and C. R. Rao, On the convexity of some divergence measures based on Entropy functions, *IEEE Trans. Inform. Theory* **28** (1982a), 489–495.
5. J. Burbea and C. R. Rao, On the convexity of higher order Jensen differences based on entropy function, *IEEE Trans. Inform. Theory* **28** (1982b), 961–963.
6. J. Burbea, *J*-divergences and related topics, *Encyclopedia Statist. Sci.* **44** (1983), 290–296.
7. W. G. Cochran, The χ^2 test of goodness of fit, *Ann. of Math. Statist.* **23** (1952), 315–345.
8. A. Cohen and H. B. Sackrowitz, Unbiasedness of the chi-square, likelihood ratio, and other goodness of fit tests for the equal cell case, *Ann. Statist.* **3** (1975), 959–964.
9. N. Cressie and T. R. C. Read, Multinomial goodness of fit test, *J. Roy. Statist. Soc. Ser. B* **46** (1984), 440–464.
10. I. Csiszár, Eine Informationtheoretische Ungleichung und ihre Anwendung auf den Beweis der Ergodizität von Markhoffschen Ketten, *Publ. Math. Inst. Hung. Acad. Sci. Ser. A* **8** (1963), 85–108.
11. J. J. Dik and M. C. M. Gunst, The distribution of general quadratic forms in normal variables, *Statist. Neerlandica* **39** (1985), 14–26.
12. A. R. Eckler, A survey of coverage problems associated with point and area targets, *Technometrics* **11** (1969), 561–589.
13. R. A. Fisher, The conditions under which χ^2 measures the discrepancy between observation and hypothesis, *J. Roy. Statist. Soc.* **87** (1924), 442–450.
14. C. Gini, Variabilitá e mutabilitá, Studi Economico–Giuridici della Facolta di Giurisprudenza dell Universitá di Cagliari, a III, parte II, 1912.
15. S. S. Gupta, Bibliography on the multivariate normal integrals and related topics, *Ann. Math. Statist.* **34** (1963), 829–838.
16. M. E. Havrda and F. Charvát, Quantification method of classification processes: Concept of structural α-entropy, *Kybernetika* **3** (1975), 30–35.
17. J. P. Imhof, Computing the distribution of quadratic forms in normal variables, *Biometrika* **48** (1961), 419–426.
18. D. R. Jensen and H. Solomon, A Gaussian approximation to the distribution of a definite quadratic form, *J. Amer. Statist. Assoc.* **67**, No. 340 (1972), 898–902.
19. N. L. Johnson and S. Kotz, Tables of distributions of positive definite quadratic forms in central normal variables, *Sankhyā Ser. B* **30** (1968), 303–314.
20. J. N. Kapur, Measures of uncertainty, mathematical programming and physics, *J. Ind. Soc. Agricultural Statist.* **24** (1972), 47–66.
21. F. Liese and I. Vajda, "Convex Statistical Distances," Teubner, Leipzig, 1987.
22. R. Modarres and R. W. Jernigan, Testing the equality of correlation matrices, *Comm. Statist. Theory Methods* **21**, No. 8 (1992), 2107–2125.

23. D. Morales, L. Pardo, and I. Vajda, Asymptotic divergence of estimates of discrete distributions, *J. Statist. Plan. Inference* **48** (1995), 347–369.
24. L. Pardo, D. Morales, M. Salicrú, and M. L. Menéndez, R_ϕ^h-divergence statistics in applied categorical data analysis with stratified sampling, *Utilitas Math.* **44** (1993), 145–164.
25. M. C. Pardo and I. Vajda, About distances of discrete distributions satisfying the data process theorem of information theory, *Trans. IEEE Inform. Theory* **43**, No. 4 (1997), 1288–1293.
26. M. C. Pardo, Asymptotic behaviour of an estimator based on Rao's divergence, *Kybernetika* **33**, No. 5 (1997), 489–504.
27. K. Pearson, On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling, *Philos. Mag.* **50** (1900), 157–172.
28. C. R. Rao, Diversity and dissimilarity coefficients: an unified approach, *J. Theoret. Pop. Biol.* **21** (1982), 24–43.
29. 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. Stat. Assoc.* **76** (1981), 221–230.
30. T. R. C. Read and N. Cressie, "Goodness of Fit Statistics for Discrete Multivariate Data," Springer-Verlag, New York, 1988.
31. F. E. Satterthwaite, An approximate distribution of estimates of variance components, *Biometrics* **2** (1946), 110–114.
32. E. H. Simpson, Measurement of diversity, *Nature* **163** (1949), 688.
33. H. Solomon, Distribution of quadratic forms—tables and applications, Technical Report 45, Applied mathematics and statistics laboratories, Stanford University, Stanford, CA, 1960.
34. I. Vajda and K. Vasek, Majorization, concave entropies, and comparison of experiments, *Prob. Control Inform. Theory* **14** (1985), 105–115.
35. K. Zografos, K. Ferentinos, and T. Papaioannou, φ-divergence statistics: Sampling properties and multinomial goodness of fit divergence tests, *Comm. Statist. Theory Methods* **19**, No. 5 (1990), 1785–1802.