*Sankhyā : The Indian Journal of Statistics*
1995, Volume 57, Series B, Pt. 3, pp. 365-376

# COMPARISONS OF THE BLENDED WEIGHT HELLINGER DISTANCE BASED GOODNESS-OF-FIT TEST STATISTICS*

*By* DONG WAN SHIN[1]
*University of Suwon*
AYANENDRANATH BASU
*University of Texas at Austin*
and
SAHADEB SARKAR
*Oklahoma State University*

*SUMMARY.* The class of goodness-of-fit tests based on the blended weight Hellinger distance (BWHD) is a rich subfamily of the family of disparity tests introduced by Basu & Sarkar (1994a). In small samples, for most members of the BWHD family, the limiting chi-square null distribution can produce significance levels that are very different from the desired nominal levels. In this paper we derive three alternative approximations of their exact distributions, leading to more accurate significance levels. Numerical results are presented for the symmetric null hypothesis for different multinomial sample sizes with various cell numbers. Exact power comparisons under specific alternatives to the symmetric null hypothesis show that the well-known Pearson's chi-square have smaller power than some other members of the BWHD family.

## 1. Introduction

Pearson's chi-square (Pearson, 1900) is the most commonly used test statistic for testing goodness of fit. There are other alternative tests like the log likelihood ratio, the Neyman modified chi-square, the Freeman-Tukey statistic and the modified likelihood ratio. Cressie and Read (1984) and Read and Cressie (1988) introduced the class of tests known as the family of power divergence statistics

Paper received. July 1994; revised February 1995.

*AMS* (1991) *subject classification.* Primary 62F03, 62F04, 62F05; secondary 62H15.

*Key words and phrases.* Goodness-of-fit, chi-squared statistic, blended weight Hellinger distance, Hellinger distance, large sparse multinomial, likelihood ratio test.

* The research of the first author was supported by Korea Science and Engineering Foundation, the research of the second author was supported in part by the URI summer research grant, University of Texas at Austin and the research of the third author was supported by a Grant from the College of Arts and Sciences Fund at Oklahoma State University.

[1] Present affiliation : Department of Statistics, Ewha Womans University, Seoul, Republic of Korea..

$\{I^\lambda : \lambda \in \Re\}$ which contain all the above statistics as members. This family is shown to be a subclass of a more general class of goodness-of-fit test statistics by Basu and Sarkar (1994a), hereafter referred to as B&S. These test statistics, called the disparity tests, use the minimum disparity parameter estimators (Lindsay 1994, Basu and Sarkar 1994b, 1994c) when the null hypothesis of interest is composite. B&S have shown that, like the power divergence statistics, there exists another very rich subfamily of disparity tests called the blended weight Hellinger distance family $\{BWHD_\alpha, \alpha \in \Re\}$.

Let $\mathbf{X} = (X_1, \ldots, X_k)$ denote the random vector of frequencies having a multinomial distribution with sample size $n$, number of categories $k$ and the probability vector $\boldsymbol{\pi} = (\pi_1, \ldots, \pi_k)$ with $\Sigma_{i=1}^k \pi_i = 1$. Consider the simple null hypothesis

$$\mathbf{H}_0 : \boldsymbol{\pi} = \boldsymbol{\pi}_0 \qquad \ldots(1.1)$$

where $\boldsymbol{\pi}_0 = (\pi_{01}, \ldots, \pi_{0k})$ with $\pi_{0i} > 0$ for each $i$ and $\Sigma_{i=1}^k \pi_{0i} = 1$. Let $\mathbf{p} = (p_1, p_2, \ldots, p_k) = (X_1/n, \ldots, X_k/n)$. Then, the $BWHD_\alpha$ tests for (1.1) are defined by

$$2nBWHD_\alpha(\mathbf{p}, \boldsymbol{\pi}_0) = n \sum_{i=1}^{k} \left\{ \frac{p_i - \pi_{0i}}{\alpha(p_i)^{1/2} + (1-\alpha)(\pi_{0i})^{1/2}} \right\}^2 . \qquad \ldots(1.2)$$

B&S have shown that under the null hypothesis (1.1), $2nBWHD_\alpha$ has an asymptotic $\chi^2_{k-1}$ distribution. The $BWHD_0$, $BWHD_1$ and $BWHD_{1/2}$ tests correspond to the Pearson's chi-square, the Neyman's chi-square and the Freeman-Tukey chi-square tests respectively. The power divergence statistic $I^{2/3}$ of Cressie and Read (1984) provides an excellent alternative to the standard Pearson's chi-square and the log likelihood ratio statistic. B&S have shown that the $BWHD_{1/9}$ statistic also does the same.

The small sample properties of the $BWHD_\alpha$ test statistics are examined in this paper for testing the simple null hypothesis (1.1). In Section 2 we have derived three approximations of the exact null distributions of the $BWHD_\alpha$ tests that depend on the index parameter $\alpha$. We examine in Section 3 the applicability of these approximating distributions by measuring the inaccuracy of the critical regions that result from these approximations when used in small samples. The inaccuracy in using the $\chi^2$ approximation in different multinomial distributions for the Pearson's chi-square and log likelihood ratio tests was studied by Yarnold (1970), Odoroff (1970) and Larntz (1978). Read (1984a) has discussed similar small sample properties of the power divergence goodness-of-fit statistics. Finally, in Section 4, we present some exact power comparisons for various $BWHD_\alpha$ tests.

As in Read (1984a), all our numerical results for the $BWHD_\alpha$ tests are obtained under the equiprobable null hypothesis, also called the symmetric hypothesis, defined by

$$\mathbf{H}_0 : \boldsymbol{\pi} = \boldsymbol{\pi}_0^* = (1/k, 1/k, \ldots, 1/k) \qquad \ldots(1.3)$$

Recommendations are made on the use of the approximations to obtain critical regions for different members of the $BWHD_\alpha$ family of tests for different values of sample sizes and category numbers. We have also computed exact powers of the $BWHD_\alpha$ tests for various $\alpha$ values under several specific alternatives to the symmetric null hypothesis. The exact power calculations show that the power of the commonly used Pearson's chi-square test can be improved by choosing other members from the $BWHD_\alpha$ family.

## 2. APPROXIMATIONS OF THE EXACT NULL DISTRIBUTIONS

In this section we consider the limiting $\chi^2$ and three other approximations of the exact distribution $F_E$ of the $BWHD_\alpha$ tests under the null hypothesis (1.1). Let $F_{\chi^2_\nu}(.)$ denote the $\chi^2$ distribution function with $\nu$ degrees of freedom. Theorem 3.1 of B&S implies that if $k$ is fixed, for each value of the family parameter $\alpha$,

$$Pr\{2nBWHD_\alpha < t\} = F_E(t) = F_{\chi^2_{k-1}}(t) + o(1) \text{ as } n \to \infty \qquad \ldots(2.1)$$

for all $t$. Because the limiting chi-square distribution $F_{\chi^2_{k-1}}(t)$ does not depend on $\alpha$, it can be used to compute the rejection regions for the $BWHD_\alpha$ tests. The asymptotic result (2.1) does not, however, give any insight into how the rate of convergence to $F_{\chi^2_{k-1}}$ varies with $\alpha$.

The first approximation is based on first and second moment corrections of the $\chi^2_{k-1}$ limit. It is given by

$$F_C(t) = F_{\chi^2_{k-1}}(d_\alpha^{-1/2}[t - c_\alpha]) \qquad \ldots(2.2)$$

where

$$c_\alpha = (k-1)[1 - d_\alpha^{1/2}] + n^{-1}a_\alpha, \; d_\alpha = 1 + [n(2(k-1))]^{-1}b_\alpha$$

with

$$a_\alpha = -\alpha(S - 3k + 2) + (3/4)(3\alpha^2 + \alpha)(S + 1 - 2k),$$

$$b_\alpha = (2 - 2k - k^2 + S) + \alpha^2(30 - 54k - 9k^2 + 33S) + \alpha(-18 + 24k + 6k^2 - 12S),$$

and

$$S = \Sigma_{i=1}^k \pi_{0i}^{-1}. \qquad \ldots(2.3)$$

The $c_\alpha$ and $d_\alpha$ terms define the asymptotic means and variances of the $BWHD_\alpha$ tests up to the order $o(n^{-1})$. We have derived (2.2) using equation (5.1) of B&S. Expectations of the terms involved in equation (5.1) of B&S are given in Appendix A11 of Read and Cressie (1988). The percentiles of $F_C$ are easier to compute than the the other two approximations $F_S$ and $F_N$ considered below, error for $F_C$ in approximating the right tail of $F_E$ is generally the lowest.

Following Read (1984b) we obtain the second closer approximation of $F_E(t)$ by extracting the $\alpha$-dependent second order components from the $o(1)$ term in (2.1) :

$$\begin{aligned} F_S(t) \; = \; & F_{\chi^2_{k-1}}(t) + \frac{1}{24n}\Big\{2(1-S)F_{\chi^2_{k-1}}(t) + \\ & [3(3S-k^2-2k) + \alpha(-27S+18k^2+18k-9) \\ & +\alpha^2(18S-27k^2+9)]F_{\chi^2_{k+1}}(t) + [-6(2S-k^2-2k+1) \\ & +\alpha(57S-36k^2-54k+33) + \alpha^2(-63S+54k^2+54k-45)]F_{\chi^2_{k+3}}(t) \\ & + [(-6\alpha+9\alpha^2+1)(5S-3k^2-6k+4)]F_{\chi^2_{k+5}}(t)\Big\} + \\ & \left[N_\alpha(t) - n^{(k-1)/2}V_\alpha(t)\right]\left[e^{-t/2}(2\pi n)^{-(k-1)/2}Q^{-1/2}\right] \end{aligned} \qquad \ldots(2.4)$$

where $S$ is defined in (2.3), $N_\alpha(t)$ = number of multinomial $\mathbf{X}$ vectors such that $2nBWHD(\frac{\mathbf{X}}{n};\boldsymbol{\pi}_0) < t$,

$$V_\alpha(t) = \left(\frac{(\pi t)^{(k-1)/2}}{\Gamma\{(k+1)/2\}}\right)Q^{1/2}\left\{1 + \frac{t[\alpha(-9S+18k-9)+\alpha^2(18S-27k^2+9)]}{24n(k+1)}\right\},$$

and $Q = \Pi_{i=1}^k \pi_{0i}$. In the Appendix we give the derivation of $F_S(t)$ for any general disparity test $2n\rho_G$ where

$$\rho_G(\mathbf{p},\boldsymbol{\pi}_0) = \sum_{i=1}^{k} G\Big(\frac{p_i}{\pi_{0i}} - 1\Big)\pi_{0i}, \qquad \ldots(2.5)$$

and $G$ is a thrice differentiable function with $G^{(3)}$, the third derivative of $G$, continuous at 0 and $G^{(3)}(0)$ finite. The approximation for the $BWHD_\alpha$ family given in expression (2.4) is then obtained as a special case when one uses $G(\delta)$ given by

$$G(\delta) = 2^{-1}\left\{\frac{\delta}{[\alpha(\delta+1)^{1/2} + (1-\alpha)]}\right\}^2. \qquad \ldots(2.6)$$

The derivation in the Appendix also show that $F_S$ is not a probability distribution function and because it contains $N_\alpha(t)$ it is not continuous. However, for all practical purposes, $F_S$ can be treated as a distribution function. Computation of $F_S$ is complicated like that of $F_E$ but it provides the best overall approximation over the entire range of $F_E$.

The third approximation of $F_E(t)$ is obtained under the assumption that $k \to \infty$ as $n \to \infty$ such that $k^{-1}n \to a$ for $0 < a < \infty$ fixed. In this case the blended weight Hellinger distance tests have a limiting normal distribution. The normal limit $F_N(t)$ is derived under the specific null hypothesis (1.3) by applying Theorem 2.4 of Cressie and Read (1984) due to Holst (1972) with

$$f_i(x) = \left\{\frac{[(\frac{kx}{n}) - 1]}{\alpha(\frac{kx}{n})^{1/2} + (1-\alpha)}\right\}^2.$$

It is given by

$$F_E(t) = F_N(t) + o(1) \text{ as } n \to \infty, \qquad \dots (2.7)$$

where $F_N(t) = Pr\{N(0,1) < \sigma_n^{-1}(t - \mu_n)\}$, $N(0,1)$ denotes a standard normal random variable,

$$\mu_n = nE\left\{\frac{(Y/a) - 1}{\alpha(Y/a)^{1/2} + (1-\alpha)}\right\}^2,$$

$$\sigma_n^2 = a^2 k \left\{Var\left(\left[\frac{(Y/a) - 1}{\alpha(Y/a)^{1/2} + (1-\alpha)}\right]^2\right) - aCov^2\left(Y/a, \left[\frac{(Y/a) - 1}{\alpha(Y/a)^{1/2} + (1-\alpha)}\right]^2\right)\right\},$$

and $Y$ is a Poisson(a) random variable. Using Monte carlo techniques, Koehler and Larntz (1980) have investigated applicability of normal approximations for the Pearson's chi-square and the log likehood ratio tests when sample sizes and cell numbers are moderate.

## 3. Small sample comparison of the approximations

Comparison of the approximation errors associated with the four approximations $F_{\chi^2_{k-1}}, F_C, F_S$ and $F_N$ in small samples are done using two criteria. The first criterion is different from that considered by Read (1984a, Section 2.2) for $I^\lambda$ statistics. Our first method measures more directly the approximation error in estimating the right tail of the true distribution $F_E$. The small sample computations presented in this section were done for all combinations of $(n, k)$ for $n = 10, 15, 20$ and $k = 2, 3, 5$. We summarize these results in Section 4.3, although for brevity we graphically present the findings only for the case $n = 20$ and $k = 5$.

3.1. *The significance levels obtained with $F_E, F_{\chi^2_{k-1}}, F_C, F_S$ and $F_N$.* We compute $100(1-\gamma)$-th percentiles of $F_E, F_{\chi^2_{k-1}}, F_C, F_S$ and $F_N$ for $\gamma = 0.10$ and $0.01$. For a fixed $i, i = E, \chi^2_{k-1}, C, S, N$, let $t_{\gamma,i}$ denote the corresponding percentile of the $i$-th distribution function $F_i$, defined by

$$t_{\gamma,i} = \min\{t : Pr(U \le t) \ge 1 - \gamma\}, \qquad \dots (3.1)$$

where $U$ is a random variable with the distribution function $F_i$. These are illustrated in Figures 1 and 2 for $\alpha \in [-1, 1], n = 20$ and $k = 5$. The method described in Section 2.1 of Read (1984a) is used to compute $F_E$. Computation of $F_S$ is also done by considering all possible multinomial vectors as in the case of $F_E$. The $t_{\gamma,C}$ and $t_{\gamma,N}$ percentile points are easily computed as

$$t_{\gamma,C} = c_\alpha + d_\alpha^{1/2} t_{\gamma,\chi^2_{k-1}} \text{ and } t_{\gamma,N} = \mu_n + \sigma_n z_\gamma,$$

were $z_\gamma$ is the $100(1-\gamma)$-th percentile of the $N(0,1)$ distribution. In Figures 1 and 2 the percentiles of $F_{\chi^2_{k-1}}, F_E, F_C, F_S$ and $F_N$ are denoted by CV-CHI, CV-E, CV-C, CV-S and CV-N respectively. The unbroken lines represent the 90% and 99% $\chi_4^2$ critical values 7.78 and 13.28 respectively. The figures reveal that the percentile points of $F_S$ and $F_N$ approximate those of $F_E$ very well for all the values $\alpha$ considered here at the 10% level. At the 1% level also these approximations are much better than the $\chi^2$ approximation. The normal approximation $F_N$ is poor at the 10% level, but is better at the 1% level.

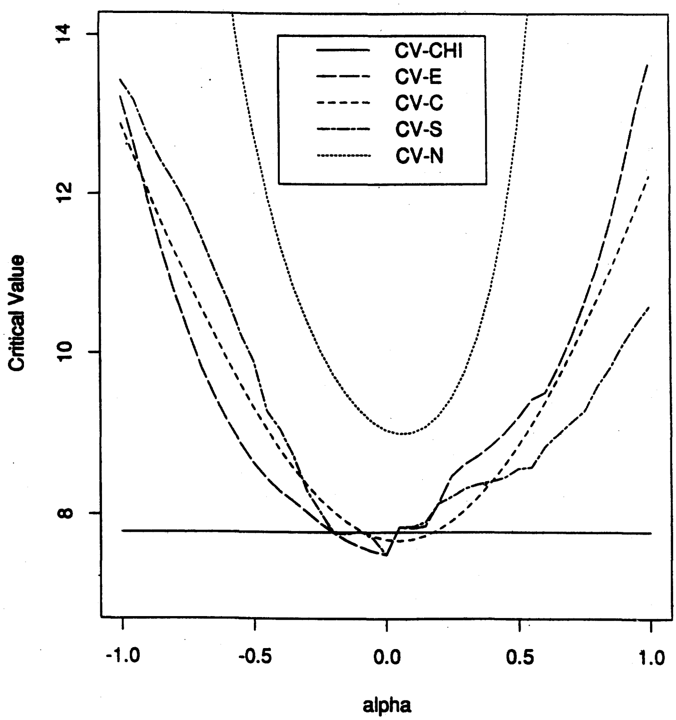


Figure 1. True and approximate critical values for the equiprobable null hypothesis at the 10% nominal level ($n = 20, k = 5$). In the graph CV-CHI, CV-E, CV-C, CV-S and CV-N denote the critical values corresponding to $F_{\chi^2_{k-1}}, F_E, F_C, F_S$ and $F_N$ respectively

3.2 *Maximum approximation error*. The choice of nominal level $\gamma$ determines the results under the first method of error measurement. For a fixed $(n, k, \alpha)$ combination a second criterion used by Read (1984a, Section 2.3) measures the worst error made across an entire approximating distribution in estimating the exact distribution $F_E$. It is called the maximum approximation error, and is defined by

$$M_i = \max_{\mathbf{x}} \mid F_E(2nBWHD_\alpha(\frac{\mathbf{x}}{n}, \pi_0^*)) - F_i(2nBWHD_\alpha(\frac{\mathbf{x}}{n}, \pi_0^*)) \mid \quad \ldots (3.2)$$

for a fixed $\alpha$ and $i = \chi^2_{k-1}, C, S, N$, where $BWHD_\alpha(\cdot, \cdot), \pi_0^*$ are defined in (1.2), (1.3) respectively and $\mathbf{x}$ represents the observed value of the multinomial random vector **X**. The sign associated with the maximum difference $M_i$ is also recorded and the results are shown in Figure 3 for $\alpha$ in $[-1.0, 1.0]$ and for $n = 20$ and $k = 5$. For $\alpha$ away from zero the errors tend to increase more rapidly than for $\alpha$ close to zero. In Figure 3, the $M_i$, $i = \chi^2_{k-1}, C, S, N$, are denoted by MCHI, MC, MS and MN respectively. It appears that $F_S$ is the best approximation according to this criterion, $F_C$ being the close second.

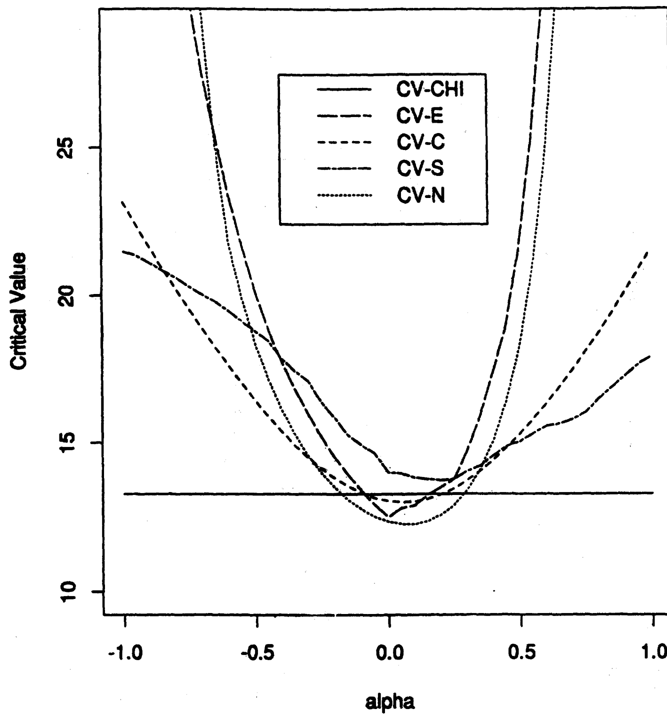


Figure 2. True and approximate critical values for the equiprobable null hypothesis at the 1% nominal level ($n = 20, k = 5$). In the graph CV-CHI, CV-E, CV-C, CV-S and CV-N denote the critical values corresponding to $F_{\chi^2_{k-1}}$, $F_E, F_C, F_S$ and $F_N$ respectively

## 4. EXACT POWER COMPARISONS

In the last section we discussed how one can obtain very good approximations of the exact critical regions for members of the $BWHD_\alpha$ statistics. In this section we present small sample powers of the $BWHD_\alpha$ tests for testing (1.3) against

$$\mathbf{H}_1 : \pi_i = \begin{cases} \{1 - \delta/(k-1)\}/k & i = 1, 2, \ldots, (k-1), \\ (1+\delta)/k & i = k, \end{cases}$$

where $-1 \le \delta \le k - 1$ is fixed. We have computed exact powers for three alternative hypotheses defined by $\delta = 1.5$ (alt 1), 0.5 (alt 2) and $-0.9$ (alt 3).

Read (1984a) considered these alternative hypotheses and he discussed their importance and interpretation. Because the attainable discrete levels for the

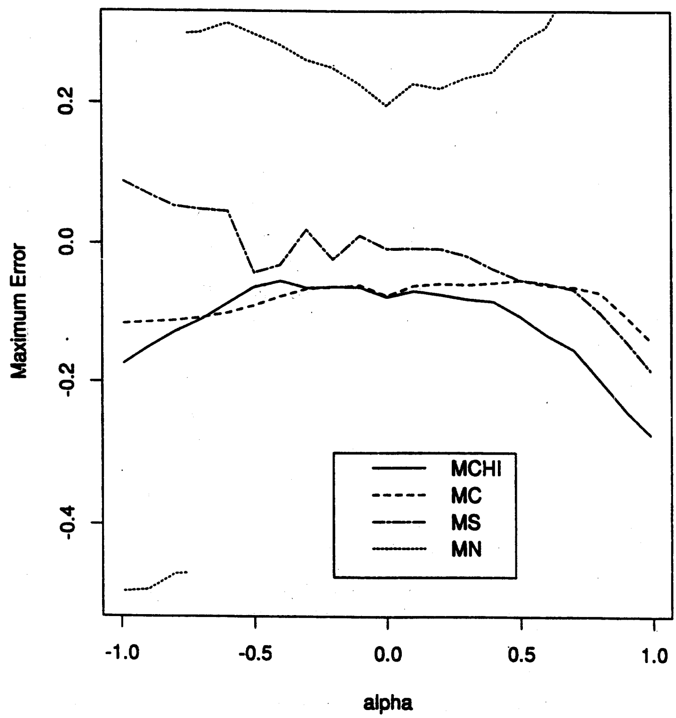


Figure 3. Maximum approximation errors for the equiprobable null hypothesis ($n = 20, k = 5$). In the graph MCHI, MC, MS and MN denote the maximum aproximation errors for $F_{\chi^2_{k-1}}$, $F_C$, $F_S$ and $F_N$ respectively

exact $BWHD_\alpha$ tests vary with $\alpha$, to make the power functions of $BWHD_\alpha$ tests comparable across $\alpha$ values we have considered the randomized tests of same size 0.05. Power values for $n = 20$ and $k = 5$ and for different $\alpha$ values are presented in Table 1.

Table 1. EXACT POWER FUNCTION FOR THE $BWHD_\alpha$ RANDOMIZED SIZE .05 TESTS ($n = 20, k = 5$)

| | $\delta$ | | |
|---|---|---|---|
| $\alpha$ | 1.5 | 0.5 | -0.9 |
| 1.00 | 0.2574 | 0.0785 | 0.5893 |
| 0.70 | 0.2574 | 0.0785 | 0.5893 |
| 0.50 | 0.3361 | 0.0800 | 0.5875 |
| 0.30 | 0.6088 | 0.1066 | 0.4589 |
| 0.10 | 0.6815 | 0.1190 | 0.3214 |
| 0.00 | 0.6997 | 0.1228 | 0.2720 |
| -0.10 | 0.7124 | 0.1250 | 0.2310 |
| -0.30 | 0.7306 | 0.1278 | 0.1895 |
| -0.50 | 0.7430 | 0.1291 | 0.1606 |
| -0.70 | 0.7488 | 0.1296 | 0.1491 |
| -1.00 | 0.7498 | 0.1295 | 0.1451 |

From Table 1 we see that the power increases as $\alpha$ decreases for hypotheses alt 1 and alt 2, whereas for hypothesis alt 3 the power increases as $\alpha$ increases. This means the power of the Pearson's chi-square ($BWHD_0$) statistic can be improved upon by considering another suitable member from the $BWHD_\alpha$ for all the three alternatives considered. Note that the direction of increase in power for the $\mathrm{BWHD}_\alpha$ tests over $\alpha$ is the opposite of that for the power divergence tests $I^\lambda$ over $\lambda$ (Read 1984a, Table 1).

## Appendix

### DERIVATION OF $F_S(t)$ FOR A GENERAL DISPARITY TEST $2n\rho_G$

Assume that (1.1) is true. We also assume that the fourth derivative of $G$ exists; this assumption is used in expanding $2n\rho_G(\mathbf{p}, \boldsymbol{\pi}_0)$ in a fourth order Taylor series. Let $W_j = n^{1/2}(p_j - \pi_{0j})$ for $j = 1, 2, \ldots, k$ and let $r = k - 1$. Then, the normalized vector $\mathbf{W} = (W_1, \ldots, W_r)$ takes values in the lattice

$$L = \left\{ \mathbf{w} = (w_1, \ldots, w_r) : \mathbf{w} = n^{1/2}(n^{-1}\mathbf{m} - \tilde{\boldsymbol{\pi}}_0) \text{ and } \mathbf{m} \in M \right\}$$

where $\tilde{\boldsymbol{\pi}}_0 = (\pi_{01}, \ldots, \pi_{0r})$ and $M = \{\mathbf{m} = (m_1, \ldots, m_r) : m_j, j = 1, \ldots, r$ are nonnegative integers satisfying $\Sigma_{j=1}^{r} m_j \le n\}$. Using a general asymptotic probability result for lattice random variables of Yarnold (1972), Read (1984b) derived the asymptotic expansion of the limiting distribution of the $I^\lambda$ statistic under the null hypothesis (1.1). Read's result for $I^\lambda$ statistics contains that of Yarnold (1972) for the Pearson's chi-square ($I^1$) and that of Siotani and Fujikoshi (1980) for the log likelihood ratio ($I^0$) as special cases. We generalize

Read's result to the more general class of disparity test statistics $2n\rho_G$. We use Read's (1984b) method in exploiting Theorem 2 of Yarnold (1972), which gives a useful expression for the probability of lattice random variables belonging to an extended convex set B. Definition of an extended convex set is given in Read (1984b, Definition 2.1). Let

$$B_G(t) = \{\mathbf{w} = (w_1, \ldots, w_r) : 2n\rho_G(n^{-1}(\mathbf{m}, m_k); \pi_0) < t\} \qquad \ldots (A.1)$$

where

$$w_k = -\Sigma_{j=1}^r w_j, \mathbf{m} = n^{1/2}\mathbf{w} + n\tilde{\boldsymbol{\pi}}_0, m_k = n^{1/2} w_k + n\tilde{\pi}_{0k}.$$

Expanding $2n\rho_G(\mathbf{p}, \boldsymbol{\pi}_0)$ (as a function of $p_i$ around $\pi_{0i}$) in a fourth order Taylor series we get the following.

Theorem A.1. *The asymptotic expansion for the distribution function $F_E(t)$ of the $2n\rho_G(\mathbf{p}, \boldsymbol{\pi}_0)$ is given by*

$$F_E(t) = J_1^G + J_2^G + J_3^G + O(n^{-3/2}), \qquad \ldots (A.2)$$

*where $J_1^G, J_2^G$ and $J_3^G$ are defined by $J_1, J_2$, and $J_3$ respectively in Theorem 2.1 of Read (1984b) with $B = B_G(t)$ defined in (A.1). Furthermore,*

$$\begin{aligned} J_1^G(t) \;=\; & F_{\chi^2_{k-1}}(t) + \frac{1}{24n}\Big\{2(1-S)F_{\chi^2_{k-1}}(t) + \\ & [3g_1 + 6g_2 + 6d_3g_1 + d_3^2g_4 - 3d_4g_3]F_{\chi^2_{k+1}}(t) + \\ & [-6g_1 - 6g_3 - 6d_3g_1 - 2d_3g_4 - 2d_3^2g_4 + 3d_4g_3]F_{\chi^2_{k+3}}(t) + \\ & [(2d_3 + d_3^2 + 1)g_4]F_{\chi^2_{k+5}}(t)\Big\} \end{aligned} \qquad \ldots (A.3)$$

*where*

$$d_3 = G^{(3)}(0), \; d_4 = G^{(4)}(0), \qquad \ldots (A.4)$$

$$g_1 = (S - k^2), g_2 = (S - k), g_3 = (S - 2k + 1), g_4 = (5S - 3k^2 - 6k + 4)$$

*and $S$ is as defined in (2.3). An approximation of $J_2^G$ to the first order is given by*

$$\widehat{J_2^G} = \Big\{N_G(t) - n^{(k-1)/2}V_G(t)\Big\}\Big\{e^{-t/2}(2\pi n)^{-(k-1)/2}Q^{-1/2}\Big\}$$

*where $N_G(t)$ = number of multinomial $\mathbf{X}$ vectors such that $2n\rho_G(\frac{\mathbf{X}}{n}, \boldsymbol{\pi}_0) <$ $t$, and*

$$V_G(t) = \left(\frac{(\pi t)^{(k-1)/2}}{\Gamma\{(k+1)/2\}}\right) Q^{1/2}\left\{1 + \frac{t(d_3^2g_4 - 3d_4g_3)}{24n(k+1)}\right\},$$

*with $Q = \Pi_{i=1}^k \pi_{0i}$.* ▱

Note that if the distribution of $2n\rho_G(\mathbf{p}, \boldsymbol{\pi}_0)$ was continuous, the term $J_1^G$ could have been obtained by the multivariate Edgeworth approximation. The discontinuous nature of the $2n\rho_G(\mathbf{p}, \boldsymbol{\pi}_0)$ statistic is accounted for by the term $J_2^G$. Theorem 2.1 of Read (1984b) implies that $J_3^G = O(n^{-1})$. since the members of the family of $2n\rho_G(\mathbf{p}, \boldsymbol{\pi}_0)$ tests are asymptotically equivalent (B&S, Theorem 3.1) we have $n(J_3^G - J_3^{PCS}) = o(1)$ as $n \to \infty$ where $J_3^{PCS}$ denotes the corresponding term for the Pearson chi-square distribution expansion. All the $G$-dependent terms in $J_3^G$ are, therefore, $O(n^{-3/2})$. In light of the expansion in (A.2), $J_3^G$ can be viewed as independent of $G$. Because the evaluation of $J_3^G$ is complex (see e.g. Yarnold (1972) for $J_3^0$), as was done by Read (1984b) in the case of power divergence goodness-of-fit statistics, we may ignore the term $J_3^G$ in (4). As a closer approximation that $F_{\chi^2_{k-1}}(t)$ of $F_E(t)$ up to the order $n^{-1}$, we may use

$$F_S(t) = J_1^G + \widehat{J}_2^G. \qquad \ldots (A.5)$$

In equation (2.4) we presented the simplified form of $F_S(t)$ defined in (A.5) above for the $BWHD_\alpha$ subfamily of disparity tests with $d_3 = -3\alpha, d_4 = 3\alpha(1+3\alpha)$.

## References


Basu, A. and Sarkar, S. (1994a). On disparity based goodness-of-fit tests for multinomial models. *Stat. and Probab. Lett.* **19**, 307-312.

— — — (1994b). Minimum disparity estimation in the errors-in-variables model. *Statistics and Probability Letters* **20**, 69-73.

— — — (1994c). The trade-off between robustness and efficiency and the effect of model smoothing in minimum disparity inference. *Journal of Statistical Computation and Simulation* **50**, 173-185.

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

Holst, L. (1972). Asymptotic normality and efficiency for certain goodness-of-fit tests. *Biometrika* **59**, 137-145.

Koehler, K.J. and Larntz, K. (1980). An empirical investigation of goodness-of-fit statistics for sparse multinomials. *J. Amer. Statist. Assoc.* **75**, 336-344.

Larntz, K. (1978). Small sample comparisons of exact levels of chi-squared goodness-of-fit statistics. *J. Amer. Staist. Assoc.* **73**, 253-263.

Lindsay, B.G. (1994). Efficiency versus robustness : The case for minimum Hellinger distance and related methods. *Ann. Statist.* **22**, 1081-1114.

Neyman, J. (1949). Contribution to the theory of the $\chi^2$ test. *Proc. 1st Berkley Symp. Math. Statist. Prob.*, 239-273.

Odoroff, C.L. (1970). A comparison of minimum logit chi-square estimation and maximum likelihood estimation in 2 × 2 × 2 and 3 × 2 × 2 contigency tables : tests for interaction. *J. Amer. Statist. Assoc.*, **65**, 1617-1631.

Pearson, K. (1900). 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 supposed to have arisen from random sampling. *Philosophy Magazine Series* **50**, 157-172.

READ, T.R.C. (1984a). Small sample comparisons for the power divergence goodness-of-fit statistics. *J. Amer. Statist. Assoc.* **79**, 929-935.

—— (1984b). Closer asymptotic approximations for the distributions of the power divergence goodness-of-fit statistics. *Ann. Inst. Statist. Math.* **A, 36**, 59-69.

—— and CRESSIE, N. (1988). *Goodness-of-fit statistics for discrete multivariate data.* New York : Springer-Verlag.

SIOTANI, M. and FUJIKOSHI, Y. (1980). Asymptotic approximations for the distributions of multinomial goodness-of-fit tests, *Technical Report* **14**, Statistical Research Group, Horishima University.

YARNOLD, J.K. (1970). The minimum expectation in $X^2$ goodness of fit tests and the accuracy of approximations for the null distribution. *J. Amer. Statist. Assoc.* **65**, 864-886.

—— (1972). Asymptotic approximations for the probability that a sum of lattice random vectors lies in a convex set. *Ann. Math. Statist.* **43**, 1566-1580.

DEPARTMENT OF APPLIED STATISTICS
UNIVERSITY OF SUWON
REPUBLIC OF KOREA

DEPARTMENT OF MATHEMATICS
UNIVERSITY OF TEXAS AT AUSTIN
AUSTIN, TX 78712
USA

DEPARTMENT OF STATISTICS
OKLAHOMA STATE UNIVERSITY
STILLWATER, OK 74078 USA