SINJINI MITRA (*) – SRABASHI BASU (**)
AYANENDRANATH BASU (***)

# Exact minimum disparity inference in complex multinomial models

CONTENTS: 1. Introduction and motivation. — 2. Disparity based inference and the empty cell penalty. — 3. Numerical studies. — 4. Concluding remarks. References. Summary. Riassunto. Key words.

## 1. INTRODUCTION AND MOTIVATION

Consider a random variable $X$ having a $k$-cell multinomial distribution with parameters $n$ and $\boldsymbol{p} = (p_1, \ldots , p_k)$, where $\boldsymbol{p}$ is a function of $q(< k-1)$ parameters. Our goal is to develop a class of estimates of $\boldsymbol{p}$, which may act as reasonable alternatives to ordinary maximum likelihood estimates, by minimizing suitable 'disparity' measures. A disparity is a nonnegative measure of discrepancy — with a particular structure — between two densities which assumes its minimum value zero only when the densities are identical. For a detailed theoretical discussion see Lindsay (1994), Basu and Lindsay (1994) and Basu and Basu (1998). All 'minimum disparity estimator's are asymptotically first order efficient under the model. Several of them have considerable robustness property under moderate contaminations. However many of the more robust estimators can be substantially poor under the model (in terms of efficiency) compared to the maximum likelihood estimator when the sample size is small (e.g. see Simpson 1987, Park, Basu and Basu 1995, Basu, Basu and Chaudhury 1997).

(*) Department of Statistics, University of Washington - Seattle, WA 98195–4322, U.S.A., E-mail: sinjini@stat.washington.edu
(**) Theoretical Statistics & Mathematics Unit, Indian Statistical Institute - Calcutta, 700 035, India, E-mail: srabashi@isical.ac.in
(***) Applied Statistics Unit, Indian Statistical Institute, Calcutta 700 035, India, E-mail: ayanbasu@isical.ac.in

The asymptotic behavior of the minimum disparity estimators, both at the model and under deviations from it, have been studied in some detail by several authors including those mentioned in the previous paragraph. Procedures based on the Hellinger distance and the Cressie-Read subfamily of disparities (Cressie and Read 1984) have received particular attention (e.g. Beran 1977, Tamura and Boos 1986, Simpson 1987, 1989a, 1989b). While the asymptotic efficiency and the robustness of these procedures are now well established, comprehensive theoretical results about the cause of their comparatively poor behavior in small samples is still unavailable. Several authors including Harris and Basu (1994), Basu, Harris and Basu (1996) and Basu, Basu and Chaudhury (1997) have empirically observed the following: this lack of small sample efficiency can be partially corrected by an empty cell penalty which does not alter their asymptotic distributions or compromise their robustness properties. Basu and Basu (1998) have considered the small sample properties of some of the more robust Cressie-Read type methods in the multinomial model. However, they have only considered the simplest case where the multinomial probabilities are the functions of a single parameter. In the current paper we present the results of a study for the more complex two-parameter problem under some natural multinomial models. Among other things this allows us to demonstrate the performance of the penalized disparity test statistics for a complex null hypothesis in a natural way where one parameter is left unspecified by the null hypothesis.

The emphasis of the present paper is on efficiency — more precisely on small sample efficiency. We make it clear at the outset that it is not our aim to develop just another robust procedure. The robustness of the procedures considered here are already well established. What we do is exhibit that the small sample performance of these well known robust procedures can be improved, often substantially, by a simple empty cell penalty.

All the computations presented here are *exact*; the relevant quantities are calculated by enumerating all possible samples and determining their probabilities under the true distribution. This demonstrates, at least in these limited settings, the empty cell penalties lead to actual improvements in the performance of the methods. Such exact computations have also been considered by Read (1984), Cressie and Read (1984), Basu and Sarkar (1994), and Basu and Basu (1995), albeit under different circumstances.

## 2. Disparity based inference and the empty cell penalty

Let $f_{\boldsymbol{\Theta}}(x)$ be a parametric density defined on the set $\{1, 2, 3, \ldots, k\}$, $\boldsymbol{\Theta} \in \boldsymbol{\Gamma}$. Let $X_1, \ldots, X_n$ be a random sample from the distribution of $f_{\boldsymbol{\Theta}}(x)$ and $d(x), x = 1, \ldots, k$ be the observed proportion of the value $x$ among the $n$ sample observations. Cressie and Read (1984) defined a family of disparities between $\boldsymbol{d} = (d(1), \ldots, d(k))$ and $\boldsymbol{f_\Theta} = (f_{\boldsymbol{\Theta}}(1), \ldots, f_{\boldsymbol{\Theta}}(k))$ as a function of a single parameter $\lambda \in R$ as

$$I^{\lambda}(\boldsymbol{d}, \boldsymbol{f_\Theta}) = \frac{1}{\lambda(\lambda+1)} \sum_{x=1}^{k} d(x) \left[ \left( \frac{d(x)}{f_{\boldsymbol{\Theta}}(x)} \right)^{\lambda} - 1 \right].$$

Harris and Basu (1996) have considered the Cressie-Read disparity in the form

$$\begin{aligned} I_*^{\lambda}(\boldsymbol{d}, \boldsymbol{f_\Theta}) &= \sum_{x=1}^{k} \left[ \frac{d(x) \left[ \left( \frac{d(x)}{f_{\boldsymbol{\Theta}}(x)} \right)^{\lambda} - 1 \right]}{\lambda(\lambda+1)} + \frac{(f_{\boldsymbol{\Theta}}(x) - d(x))}{\lambda+1} \right], \ \lambda > -1 \\ &= \sum_{x: d(x) \neq 0} \left\{ \frac{d(x)}{\lambda(\lambda+1)} \left[ \left( \frac{d(x)}{f_{\boldsymbol{\Theta}}(x)} \right)^{\lambda} - 1 \right] + \frac{(f_{\boldsymbol{\Theta}}(x) - d(x))}{\lambda+1} \right\} \\ &\quad + \frac{1}{\lambda+1} \sum_{x: d(x) = 0} f_{\boldsymbol{\Theta}}(x) \end{aligned} \tag{2.1}$$

which makes each term in the summand non-negative. For $\lambda \leq -1$ the disparity is not defined if there are one or more empty cells. For $\lambda = 0$ the divergence is undefined, and $I_*^0(\boldsymbol{d}, \boldsymbol{f_\Theta})$ has to be defined as the limit of $I_*^{\lambda}(\boldsymbol{d}, \boldsymbol{f_\Theta})$ as $\lambda \to 0$. The minimizer of $I_*^0$ is the maximum likelihood estimator of $\boldsymbol{\Theta}$. We will call $I_*^0(\boldsymbol{d}, \boldsymbol{f_\Theta})$ the likelihood disparity. Also note that $\lambda = -0.5$ corresponds to the (twice, squared) Hellinger distance. The weight applied to the empty cells by the disparity $I_*^{\lambda}$ is $1/(\lambda+1)$, as seen from (2.1).

To counter the problem of poor small sample efficiency among some of the more robust minimum disparity estimators within the Cressie-Read family (e.g. estimators corresponding to $-0.5 \geq \lambda > -1$), one can alternatively consider the *penalized* family of disparities by simply manipulating the weight applied to the empty cells. The penalized family is defined as

$$P_{\omega}^{\lambda}(\boldsymbol{d}, \boldsymbol{f_\Theta}) = \sum_{x:d(x)\neq 0} \left\{ \frac{d(x)}{\lambda(\lambda+1)} \left[ \left( \frac{d(x)}{f_\Theta(x)} \right)^{\lambda} - 1 \right] + \frac{(f_\Theta(x) - d(x))}{\lambda+1} \right\} + \omega \sum_{x:d(x)=0} f_\Theta(x)\,. \tag{2.2}$$

The above is obtained from (2.1) by applying a penalty weight $\omega$ for the empty cells instead of its natural weight $1/(\lambda+1)$. If $\omega = 1$ the penalized disparities put the same weight on the empty cells as $I_*^0(\boldsymbol{d}, f_{\boldsymbol{\Theta}})$ would have put on them. The penalty scheme $\omega = 1/2$ puts the same weight as Pearson's chi-square ($\lambda = 1$) does on the empty cells. Note that the difference between $I_*^{\lambda}$ and $P_{\omega}^{\lambda}$ is only in the way they treat the empty cells. For both of them, the nonempty cells get equal treatment. The penalty scheme in (2.2) has been extensively studied in this paper. We have restricted the penalty weight $\omega$ between 0 and 1. For a negative penalty the disparity may not remain nonnegative. For $\omega > 1$ the efficiency of the estimators appear to be inferior compared to those for which $\omega \leq 1$; neither does it seem intuitively justified to increase the weights of the empty cells too much. As the total probability of the empty cells asymptotically go to zero, this penalty does not affect the asymptotic distribution of the estimators. The minimum disparity estimators and the penalized minimum disparity estimators are obtained by minimizing $I_*^{\lambda}$ and $P_{\omega}^{\lambda}$ respectively.

Next we look at the hypothesis testing problem using ordinary and penalized disparities. Consider the simple null hypothesis $H_0$ : $\boldsymbol{\Theta} = \boldsymbol{\Theta}_0$, and define the disparity test statistic $T^{\lambda} = 2n[I_*^{\lambda}(\boldsymbol{d}, \boldsymbol{f}_{\boldsymbol{\Theta}_0}) - I_*^{\lambda}(\boldsymbol{d}, \boldsymbol{f}_{\widehat{\boldsymbol{\Theta}}})]$, where $\widehat{\boldsymbol{\Theta}}$ represents the minimizer of $I_*^{\lambda}$. The $T^{\lambda}$ statistics are asymptotically distributed as $\chi^2(q)$ under the null for $\lambda > -1$ (see Lindsay 1994). For small samples, the chi-square approximation under the null hypothesis, however, can be quite inaccurate, with the observed levels being considerably inflated compared to the nominal levels; consequently, the confidence intervals obtained by inverting the test statistic also have true confidence coefficients lower than the nominal ones (see, for example, Simpson 1989a, Table 3).

An alternative test statistic can be based on the penalized disparities. Define the penalized family of test statistics

$$T^{\lambda}_{p,\omega} = 2n[P^{\lambda}_{\omega}(\boldsymbol{d}, \boldsymbol{f}_{\boldsymbol{\Theta}_0}) - P^{\lambda}_{\omega}(\boldsymbol{d}, \boldsymbol{f}_{\widehat{\boldsymbol{\Theta}}})]\,,$$

where $\widehat{\boldsymbol{\Theta}}$ represents the minimizer of $P^{\lambda}_{\omega}$. As they differ only in the empty cells, the families $T^{\lambda}$ and $T^{\lambda}_{p,\omega}$ have the same asymptotic distribution under the null hypothesis.

The testing procedures described above extend directly to the multidimensional case when the null hypothesis is composite. Define the hypothesis of interest to be $H_0 : \boldsymbol{\Theta} \in \boldsymbol{\Gamma}_0$, and assume that the null hypothesis imposes $r$ independent restrictions on the parameter space. The test statistics $T^{\lambda}$ and $T^{\lambda}_{p,\omega}$ now have the same form as above, but with $\boldsymbol{f}_{\boldsymbol{\Theta}_0}$ replaced by $\boldsymbol{f}_{\widehat{\boldsymbol{\Theta}}_0}$, $\widehat{\Theta}_0$ being the corresponding estimate of $\boldsymbol{\Theta}$ under the null. The asymptotic distribution of the disparity test statistics (here $T^{\lambda}$ and $T^{\lambda}_{p,\omega}$) under composite $H_0$ are $\chi^2(r)$ and has been established by Sarkar and Basu (1995). Their proof essentially follows the arguments of Serfling's (1980, Section 4.4.4) proof of the asymptotic null distribution of the likelihood ratio statistic when the null is composite. The true level of the ordinary disparity test is now defined as

$$\sup_{\boldsymbol{\Theta}\in\boldsymbol{\Gamma}_0} \Pr[T^{\lambda} \geq \chi^2_{\gamma}] \tag{2.3}$$

and the same for the penalized disparity test is defined as

$$\sup_{\boldsymbol{\Theta}\in\boldsymbol{\Gamma}_0} \Pr[T^{\lambda}_{p,\omega} \geq \chi^2_{\gamma}] \tag{2.4}$$

at nominal level $\gamma$.

In the following section we present several exact computations for disparity based methods in the multinomial model where the model probabilities are functions of two unknown parameters.

## 3. Numerical studies

A random sample of $n$ observations on $k$ categories with probabilities $p_1, \ldots, p_k$ generates a multinomial observation $X$ with parameters $n$ and $\boldsymbol{p} = (p_1, \ldots, p_k)$. For the rest of the paper we will write $\widehat{\boldsymbol{p}}$

for $\boldsymbol{d}$, the vector of observed proportions, and $\boldsymbol{p}_{\boldsymbol{\Theta}}$ for the probability function $\boldsymbol{f}_{\boldsymbol{\Theta}}$.

For illustrative purposes we have chosen $k = 4$. The probability vector $\boldsymbol{p} = (p_1, p_2, p_3, p_4)$ is a known function of a 2-dimensional parameter vector $\boldsymbol{\Theta}$. To obtain the exact probability distribution of $\widehat{\boldsymbol{\Theta}}$, the vector of estimators, all possible sample combinations in the sample space $D = \{\boldsymbol{x} = (x_1, x_2, x_3, x_4) | x_i \geq 0, i = 1, \ldots, 4, \sum_{i=1}^{4} x_i = n\}$ are enumerated; the distinct values of $\widehat{\boldsymbol{\Theta}}(\boldsymbol{x})$ and their exact probabilities can then be calculated using the multinomial probability function under any given true value of $\boldsymbol{\Theta}$.

Several values of $n$ have been used in our study subject to the restriction that the sample space is not too large to be completely enumerated. Two different structures on the multinomial cell probabilities are considered. The first cell probability structure is derived from the human blood group distribution (Rao 1973). Every human being may be classified into one of four blood groups O, A, B and AB. The inheritance of these is controlled by one of three genes O, A and B, of which O is recessive to A and B. If $\pi$ and $\eta$ are gene frequencies of A and B, and frequency of O is given by $\rho = 1 - \pi - \eta$ then expected probabilities of the four groups in random mating are given by

$$\begin{aligned} \Pr(O) &= \rho^2 \\ \Pr(A) &= \pi^2 + 2\pi\rho \\ \Pr(B) &= \eta^2 + 2\eta\rho\,, \quad \text{and} \\ \Pr(AB) &= 2\pi\eta\,. \end{aligned}$$

The model generated by $\boldsymbol{\Theta} = (\pi, \eta)$ will be called the Rao$(\pi, \eta)$ model. For illustrative purpose we have taken $\boldsymbol{\Theta} = (\pi, \eta) = (0.5, 0.3)$ as the true value in this paper.

Alternatively, assume that the cell probabilities are generated by a logistic$(\alpha, \beta)$ type distribution. In particular, this functional form is indicated when cumulative logit model gives a good fit to ordinal categorical response data. The cell probabilities are

$$\begin{aligned} p_1 &= 1/\{1 + \exp(\alpha)\} \\ p_2 &= \{\exp(\alpha)(1 - \exp(\beta))/(1 + \exp(\alpha))(1 + \exp(\alpha + \beta)) \\ p_3 &= \{\exp(\alpha+\beta)(1-\exp(\beta))/(1+\exp(\alpha+\beta))(1+\exp(\alpha+2\beta))\,, \text{ and}\,, \\ p_4 &= 1 - p_1 - p_2 - p_3 \end{aligned}$$

As a function of $\alpha$ and $\beta$, we will call this the logit($\alpha$, $\beta$) model. For illustrative purpose we have taken $\boldsymbol{\Theta} = (\alpha, \beta) = (2.0, -1.5)$ in this paper.

One objective in this study is to compare performance of different penalty schemes for small to moderate sample sizes. Three distinct values of $\omega$ have been considered, $\omega = 1.0, 0.5$ and $0.0$. We have compared the performance of the penalized minimum disparity method for different values of $\omega$, as well as against the ordinary minimum disparity method. The sample sizes considered are $n = 20, 25, 30$ and $40$. A larger sample becomes computationally infeasible. The values considered for λ are 1, 0, – 0.5, –0.6, –0.7, –0.8 and –0.9. The procedures derived from the last five cases have strong robustness properties and thus any improvement in their small sample efficiency is of considerable practical interest. In particular, for $\lambda = -0.5$ the disparity is equivalent to the Hellinger distance. For $\lambda \leq -1$, the disparities are not defined when one or more cells are empty. The disparities corresponding to $\lambda = 1$ and to $\lambda = 0$ are the Pearson's chi-square disparity and the likelihood disparity respectively. Although they are commonly used divergences, the corresponding minimum disparity estimates are also known for their lack of robustness. For the purpose of comparison we note that the natural weights attached to the empty cells by the seven disparities are 1/2, 1, 2, 2.5, 10/3, 5 and 10 for $\lambda = 1, 0, -0.5, -0.6, -0.7, -0.8, -0.9$ respectively. (The numerical computations presented in this paper are done on a *Digital Alpha Unix Station 255* running *Fortran 90* in the Theoretical Statistics and Mathematics Unit of the Indian Statistical Institute, Calcutta.)

For each sample point $\boldsymbol{x} = (x_1, x_2, x_3, x_4)$, and for each value of $n$ and λ considered, we calculate estimates of the unknown parameters by minimizing the disparity $I_*^\lambda$ and the penalized disparity $P_\omega^\lambda$. Let the estimates be denoted by $(\widehat{\pi}_I, \widehat{\eta}_I)$ and $(\widehat{\pi}_{P_\omega}, \widehat{\eta}_{P_\omega})$ respectively for Rao's model and $(\widehat{\alpha}_I, \widehat{\beta}_I)$ and $(\widehat{\alpha}_{P_\omega}, \widehat{\beta}_{P_\omega})$ respectively for the logit model. (The estimators are functions of λ also, but as the value of λ will be clear from the context, a further subscript has been avoided in the estimators to reduce notational complications). For each estimator $\widehat{\boldsymbol{\Theta}}(\boldsymbol{x})$ and for each ($n$, λ) combination, we compute the *exact* mean square error (MSE) of $\widehat{\theta}_i$, the $i$-th component of $\widehat{\boldsymbol{\Theta}}$ under the true value $\boldsymbol{\Theta}$ as $\sum(\widehat{\theta}_i(\boldsymbol{x})) - \theta_i)^2 P_{\boldsymbol{\Theta}}(\boldsymbol{x})$, where the sum is over the sample space $D$, and $P_{\boldsymbol{\Theta}}(\boldsymbol{x})$ is the probability of the sample $\boldsymbol{x}$ under the cell probability vector $\boldsymbol{p}$ generated by the true parameters $\boldsymbol{\Theta} = (\theta_1, \theta_2)$.

The results comparing the performances of $\widehat{\mathbf{\Theta}}_I$ and $\widehat{\mathbf{\Theta}}_{P_\omega}$ for different values of $n$, $\lambda$ and $\omega$ are presented in Tables 1 and 2, where the true multinomial cell frequencies are generated by Rao(0.5, 0.3) and the logit(2.0, −1.5) distributions respectively. Several observations may be made from these tables. For both ordinary and penalized cases the disparity based on Pearson's $\chi^2$ ($\lambda = 1$) is doing the best, i.e. the corresponding estimator has the smallest MSE for all the four parameters; however the maximum likelihood estimator ($\lambda = 0$) is only marginally worse. As expected MSE is smaller for larger sample sizes. The performance of the penalty is clearly remarkable, especially for large negative values of $\lambda$. While the MSEs corresponding to the ordinary minimum disparity estimators with large negative values of $\lambda$ are very high compared to likelihood disparity and the Pearson's chi-square, the corresponding MSEs for the penalized robust minimum disparity estimators are extremely competitive with the cases $\lambda = 1$ and $\lambda = 0$, especially for $\omega = 0$. It appears that the penalty weight $\omega = 1$ is doing the worse among the three. Note that we must not expect the penalties to cause any dramatic improvement in case of Pearson's chi-square or the likelihood disparity. In fact, for $\omega = 1.0$ the MSEs corresponding to $\lambda = 1.0$ are greater in magnitude than those obtained using the ordinary disparity.

Next we look at the performances of the statistics $T^\lambda$ and $T^\lambda_{p,\omega}$ in testing the null hypothesis $H_0 : \mathbf{\Theta} = \mathbf{\Theta}_0$ under the model i.e. when the probability vector is actually generated by the parameter $\mathbf{\Theta}_0$. Here we have considered two cases: for Rao's model we have used the simple null hypothesis

$$H_0 : (\pi, \eta) = (0.5, 0.3)\,,$$

while for the logit model we have considered a composite null

$$H_0 : \beta = -1.5$$

with $\alpha$ unknown. In the case of the simple null hypothesis the test statistics follow a $\chi^2$ distribution with 2 degrees of freedom. Having determined the nominal critical values based on the degrees of freedom of the $\chi^2$, we have computed the exact probabilities of the test statistics to exceed the nominal critical points for 10% and 1% level of significance for the Rao's model. The results are given in Tables 3 and 4. Once again, the effect of the penalty is very clearly visible. A test which cannot hold its level even approximately under small samples when the data are coming from the model is of little

TABLE 1: *Exact mean square errors for the parameters of Rao's model for human blood group when estimates are obtained through ordinary and penalized minimum disparity methods for three penalty schemes.*

| | | Ordinary Disparity | | Penalized Disparity | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | | | $\omega = 1.0$ | | $\omega = 0.5$ | | $\omega = 0.0$ | |
| $\lambda$ | $n$ | MSE($\widehat{\pi}$) | MSE($\widehat{\eta}$) | MSE($\widehat{\pi}$) | MSE($\widehat{\eta}$) | MSE($\widehat{\pi}$) | MSE($\widehat{\eta}$) | MSE($\widehat{\pi}$) | MSE($\widehat{\eta}$) |
| 1.0 | 20 | 0.008809 | 0.006017 | 0.009321 | 0.006362 | 0.008809 | 0.006017 | 0.009029 | 0.005870 |
| | 25 | 0.007054 | 0.004773 | 0.007453 | 0.005062 | 0.007054 | 0.004773 | 0.007116 | 0.004672 |
| | 30 | 0.005861 | 0.004034 | 0.006210 | 0.004236 | 0.005861 | 0.004034 | 0.005856 | 0.003938 |
| | 40 | 0.004317 | 0.003012 | 0.004556 | 0.003125 | 0.004317 | 0.003012 | 0.004258 | 0.002936 |
| 0.0 | 20 | 0.009661 | 0.006537 | 0.009661 | 0.006537 | 0.009011 | 0.006187 | 0.009148 | 0.006053 |
| | 25 | 0.007647 | 0.005223 | 0.007647 | 0.005223 | 0.007220 | 0.004963 | 0.007240 | 0.004851 |
| | 30 | 0.006380 | 0.004322 | 0.006380 | 0.004322 | 0.005962 | 0.004136 | 0.005933 | 0.004050 |
| | 40 | 0.004690 | 0.003210 | 0.004690 | 0.003210 | 0.004414 | 0.003091 | 0.004334 | 0.003033 |
| −0.5 | 20 | 0.011303 | 0.007298 | 0.009998 | 0.006780 | 0.009240 | 0.006371 | 0.009314 | 0.006234 |
| | 25 | 0.008939 | 0.005715 | 0.007926 | 0.005367 | 0.007414 | 0.005115 | 0.007412 | 0.004977 |
| | 30 | 0.007339 | 0.004691 | 0.006573 | 0.004431 | 0.006108 | 0.004244 | 0.006066 | 0.004140 |
| | 40 | 0.005344 | 0.003433 | 0.004842 | 0.003267 | 0.004522 | 0.003145 | 0.004437 | 0.003093 |
| −0.6 | 20 | 0.012063 | 0.007527 | 0.010124 | 0.006837 | 0.009324 | 0.006430 | 0.009391 | 0.006274 |
| | 25 | 0.009444 | 0.005881 | 0.007996 | 0.005421 | 0.007452 | 0.005138 | 0.007454 | 0.005005 |
| | 30 | 0.007769 | 0.004822 | 0.006621 | 0.004458 | 0.006151 | 0.004272 | 0.006093 | 0.004164 |
| | 40 | 0.005583 | 0.003507 | 0.004875 | 0.003281 | 0.004547 | 0.003156 | 0.004464 | 0.003104 |
| −0.7 | 20 | 0.013089 | 0.007889 | 0.010234 | 0.006906 | 0.009395 | 0.006485 | 0.009460 | 0.006331 |
| | 25 | 0.010161 | 0.006151 | 0.008132 | 0.005470 | 0.007564 | 0.005173 | 0.007549 | 0.005039 |
| | 30 | 0.008272 | 0.005014 | 0.006671 | 0.004486 | 0.006200 | 0.004302 | 0.006137 | 0.004195 |
| | 40 | 0.005930 | 0.003594 | 0.004910 | 0.003294 | 0.004574 | 0.003166 | 0.004491 | 0.003115 |
| −0.8 | 20 | 0.014633 | 0.008405 | 0.010414 | 0.006967 | 0.009552 | 0.006582 | 0.009560 | 0.006387 |
| | 25 | 0.011273 | 0.006446 | 0.008219 | 0.005513 | 0.007637 | 0.005237 | 0.007613 | 0.005095 |
| | 30 | 0.009072 | 0.005200 | 0.006774 | 0.004512 | 0.006287 | 0.004326 | 0.006217 | 0.004222 |
| | 40 | 0.006431 | 0.003717 | 0.004956 | 0.003315 | 0.004619 | 0.003182 | 0.004531 | 0.003134 |
| −0.9 | 20 | 0.017064 | 0.009113 | 0.010578 | 0.007034 | 0.009691 | 0.006621 | 0.009670 | 0.006426 |
| | 25 | 0.013067 | 0.006884 | 0.008396 | 0.005563 | 0.007783 | 0.005277 | 0.007741 | 0.005133 |
| | 30 | 0.010420 | 0.005460 | 0.006911 | 0.004547 | 0.006408 | 0.004358 | 0.006320 | 0.004250 |
| | 40 | 0.007235 | 0.003885 | 0.005026 | 0.003338 | 0.004677 | 0.003208 | 0.004587 | 0.003156 |

practical value. The penalty has made our tests approximately correct level $\gamma$ tests in these cases even in the small sample sizes that we have considered.

Table 2: *Exact mean square errors for the parameters of logit model when estimates are obtained through ordinary and penalized minimum disparity methods for three penalty schemes.*

| | | Ordinary Disparity | | Penalized Disparity | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | | | $\omega = 1.0$ | | $\omega = 0.5$ | | $\omega = 0.0$ | |
| $\lambda$ | $n$ | MSE($\widehat{\alpha}$) | MSE($\widehat{\beta}$) | MSE($\widehat{\alpha}$) | MSE($\widehat{\beta}$) | MSE($\widehat{\alpha}$) | MSE($\widehat{\beta}$) | MSE($\widehat{\alpha}$) | MSE($\widehat{\beta}$) |
| 1.0 | 20 | 0.432207 | 0.166008 | 0.479121 | 0.180245 | 0.432207 | 0.166008 | 0.385877 | 0.151382 |
| | 25 | 0.332799 | 0.125934 | 0.357322 | 0.133171 | 0.332799 | 0.125934 | 0.307295 | 0.118632 |
| | 30 | 0.277126 | 0.103839 | 0.291152 | 0.107766 | 0.277126 | 0.103839 | 0.264043 | 0.100124 |
| | 40 | 0.203599 | 0.075887 | 0.207306 | 0.076906 | 0.203599 | 0.075887 | 0.199717 | 0.074853 |
| 0.0 | 20 | 0.526697 | 0.193995 | 0.526697 | 0.193995 | 0.474182 | 0.178994 | 0.425203 | 0.163927 |
| | 25 | 0.390121 | 0.141742 | 0.390121 | 0.141742 | 0.362653 | 0.134361 | 0.336789 | 0.127045 |
| | 30 | 0.313422 | 0.114188 | 0.313422 | 0.114188 | 0.298949 | 0.110264 | 0.285110 | 0.106530 |
| | 40 | 0.224089 | 0.081701 | 0.224089 | 0.081701 | 0.219898 | 0.080608 | 0.215846 | 0.079578 |
| −0.5 | 20 | 0.760069 | 0.256820 | 0.597274 | 0.213806 | 0.520190 | 0.193166 | 0.469048 | 0.177604 |
| | 25 | 0.494823 | 0.170379 | 0.428384 | 0.153529 | 0.395437 | 0.144895 | 0.368483 | 0.137473 |
| | 30 | 0.368251 | 0.128441 | 0.337405 | 0.120693 | 0.321221 | 0.116532 | 0.306998 | 0.112728 |
| | 40 | 0.244027 | 0.086649 | 0.236458 | 0.084783 | 0.232033 | 0.083671 | 0.227925 | 0.082647 |
| −0.6 | 20 | 0.854234 | 0.284012 | 0.633313 | 0.223968 | 0.540290 | 0.199630 | 0.485965 | 0.183237 |
| | 25 | 0.535591 | 0.181213 | 0.443371 | 0.157510 | 0.405555 | 0.147913 | 0.377767 | 0.140313 |
| | 30 | 0.388144 | 0.133819 | 0.345627 | 0.123029 | 0.327913 | 0.118607 | 0.313374 | 0.114741 |
| | 40 | 0.251240 | 0.088455 | 0.240115 | 0.085723 | 0.235634 | 0.084620 | 0.231516 | 0.083594 |
| −0.7 | 20 | 0.960986 | 0.320254 | 0.678920 | 0.235761 | 0.569690 | 0.207632 | 0.506491 | 0.189175 |
| | 25 | 0.590080 | 0.197391 | 0.464559 | 0.163022 | 0.422658 | 0.152497 | 0.393448 | 0.144587 |
| | 30 | 0.415304 | 0.140629 | 0.354684 | 0.125189 | 0.336331 | 0.120655 | 0.321221 | 0.116673 |
| | 40 | 0.259698 | 0.090520 | 0.243516 | 0.086538 | 0.238871 | 0.085411 | 0.234741 | 0.084383 |
| −0.8 | 20 | 1.122466 | 0.377317 | 0.719108 | 0.247407 | 0.605472 | 0.217978 | 0.533224 | 0.197345 |
| | 25 | 0.673978 | 0.224570 | 0.484006 | 0.168299 | 0.441321 | 0.157589 | 0.409618 | 0.149073 |
| | 30 | 0.460369 | 0.154075 | 0.365543 | 0.128104 | 0.346293 | 0.123425 | 0.330829 | 0.119359 |
| | 40 | 0.273727 | 0.094052 | 0.248038 | 0.087688 | 0.243165 | 0.086523 | 0.239016 | 0.085495 |
| −0.9 | 20 | 1.465113 | 0.492359 | 0.767246 | 0.262684 | 0.647345 | 0.230690 | 0.566573 | 0.207798 |
| | 25 | 0.858971 | 0.282913 | 0.508514 | 0.175305 | 0.463537 | 0.163923 | 0.429650 | 0.154887 |
| | 30 | 0.558949 | 0.184190 | 0.381511 | 0.132600 | 0.361452 | 0.127713 | 0.345301 | 0.123522 |
| | 40 | 0.302115 | 0.102020 | 0.252776 | 0.089043 | 0.247890 | 0.087880 | 0.243666 | 0.086839 |

To better understand the improvement in the performance of the test statistics due to the penalty we looked at the histograms of the exact null distribution of the test statistics $T^{\lambda}$ and $T^{\lambda}_{p,\omega}$ with the $\chi^2(2)$

TABLE 3: *Exact levels of the ordinary and penalized minimum disparity test statistics with three penalty schemes for testing the simple null hypothesis* $H_0 : (\pi, \eta) = (0.5, 0.3)$ *regarding the parameters of Rao's model for human blood group at nominal level 10%.*

| | | Ordinary Disparity | Penalized Disparity | | |
|---|---|---|---|---|---|
| | | | $\omega = 1.0$ | $\omega = 0.5$ | $\omega = 0.0$ |
| $\lambda$ | $n$ | Observed Level | Observed Level | Observed Level | Observed Level |
| 1.0 | 20 | 0.106859 | 0.131426 | 0.106859 | 0.091283 |
| | 25 | 0.109729 | 0.123500 | 0.109729 | 0.097683 |
| | 30 | 0.109462 | 0.126155 | 0.109462 | 0.096979 |
| | 40 | 0.102732 | 0.122007 | 0.102732 | 0.093209 |
| 0.0 | 20 | 0.103581 | 0.103581 | 0.094846 | 0.074768 |
| | 25 | 0.107865 | 0.107865 | 0.094964 | 0.078068 |
| | 30 | 0.109503 | 0.109503 | 0.089107 | 0.084489 |
| | 40 | 0.105484 | 0.105484 | 0.087004 | 0.080219 |
| −0.5 | 20 | 0.184291 | 0.102021 | 0.093705 | 0.083126 |
| | 25 | 0.179125 | 0.108047 | 0.097883 | 0.081181 |
| | 30 | 0.170999 | 0.107752 | 0.090174 | 0.081212 |
| | 40 | 0.174686 | 0.102793 | 0.084721 | 0.076677 |
| −0.6 | 20 | 0.206793 | 0.101631 | 0.093461 | 0.082696 |
| | 25 | 0.225200 | 0.111287 | 0.100784 | 0.084361 |
| | 30 | 0.232040 | 0.108186 | 0.090761 | 0.081714 |
| | 40 | 0.213567 | 0.106011 | 0.088199 | 0.080542 |
| −0.7 | 20 | 0.325101 | 0.104054 | 0.096020 | 0.089523 |
| | 25 | 0.319481 | 0.111279 | 0.101340 | 0.086802 |
| | 30 | 0.301780 | 0.109779 | 0.091401 | 0.083273 |
| | 40 | 0.246856 | 0.107253 | 0.089474 | 0.082463 |
| −0.8 | 20 | 0.448369 | 0.108736 | 0.100939 | 0.095672 |
| | 25 | 0.406010 | 0.116436 | 0.102269 | 0.089825 |
| | 30 | 0.352143 | 0.112462 | 0.093964 | 0.086392 |
| | 40 | 0.258931 | 0.106751 | 0.088504 | 0.082521 |
| −0.9 | 20 | 0.497502 | 0.109408 | 0.100945 | 0.096204 |
| | 25 | 0.418517 | 0.116337 | 0.102466 | 0.089692 |
| | 30 | 0.356017 | 0.114281 | 0.100930 | 0.090341 |
| | 40 | 0.259349 | 0.107071 | 0.088716 | 0.084330 |

density superimposed under Rao's model. The null hypothesis considered was $H_0 : (\pi, \eta) = (0.5, 0.3)$; for the sake of illustration we took $n = 25$, $\omega = 0.5$ and nominal level $\gamma = 0.05$. In particular we

TABLE 4: *Exact levels of the ordinary and penalized minimum disparity test statistics with three penalty schemes for testing the simple null hypothesis $H_0 : (\pi, \eta) = (0.5, 0.3)$ regarding the parameters of Rao's model for human blood group at nominal level 1%.*

| | | Ordinary Disparity | Penalized Disparity | | |
|---|---|---|---|---|---|
| | | | $\omega = 1.0$ | $\omega = 0.5$ | $\omega = 0.0$ |
| $\lambda$ | $n$ | Observed Level | Observed Level | Observed Level | Observed Level |
| 1.0 | 20 | 0.016078 | 0.019543 | 0.016078 | 0.014625 |
| | 25 | 0.015275 | 0.017882 | 0.015275 | 0.014655 |
| | 30 | 0.014093 | 0.016457 | 0.014093 | 0.013316 |
| | 40 | 0.013605 | 0.016671 | 0.013605 | 0.012592 |
| 0.0 | 20 | 0.008983 | 0.008983 | 0.006568 | 0.006010 |
| | 25 | 0.009625 | 0.009625 | 0.006809 | 0.006014 |
| | 30 | 0.010490 | 0.010490 | 0.007952 | 0.006842 |
| | 40 | 0.012053 | 0.012053 | 0.008133 | 0.006917 |
| −0.5 | 20 | 0.021410 | 0.009578 | 0.006578 | 0.006917 |
| | 25 | 0.022216 | 0.013324 | 0.008845 | 0.006525 |
| | 30 | 0.024170 | 0.011684 | 0.008864 | 0.008389 |
| | 40 | 0.026769 | 0.011955 | 0.008334 | 0.007497 |
| −0.6 | 20 | 0.039244 | 0.010351 | 0.006873 | 0.007384 |
| | 25 | 0.031864 | 0.013461 | 0.009932 | 0.007484 |
| | 30 | 0.036821 | 0.011891 | 0.009238 | 0.008902 |
| | 40 | 0.046879 | 0.012300 | 0.008852 | 0.007659 |
| −0.7 | 20 | 0.053926 | 0.010023 | 0.008251 | 0.007625 |
| | 25 | 0.063053 | 0.014139 | 0.013052 | 0.008118 |
| | 30 | 0.071303 | 0.013189 | 0.009859 | 0.009528 |
| | 40 | 0.099205 | 0.012888 | 0.009627 | 0.008559 |
| −0.8 | 20 | 0.139614 | 0.010372 | 0.008600 | 0.008522 |
| | 25 | 0.188266 | 0.014806 | 0.013768 | 0.009576 |
| | 30 | 0.203893 | 0.014390 | 0.010548 | 0.010191 |
| | 40 | 0.189362 | 0.014172 | 0.011374 | 0.009895 |
| −0.9 | 20 | 0.450202 | 0.010638 | 0.008892 | 0.009887 |
| | 25 | 0.368709 | 0.015214 | 0.014314 | 0.012239 |
| | 30 | 0.303226 | 0.017710 | 0.012137 | 0.012093 |
| | 40 | 0.203662 | 0.014435 | 0.012453 | 0.010231 |

looked at the histograms of $T^{-0.9}$ and $T^{-0.9}_{p,0.5}$. Our interest is in the right hand tail area of the histograms, and how well the $\chi^2(2)$ density approximates it. In Figure 1, the poor approximation to the very long

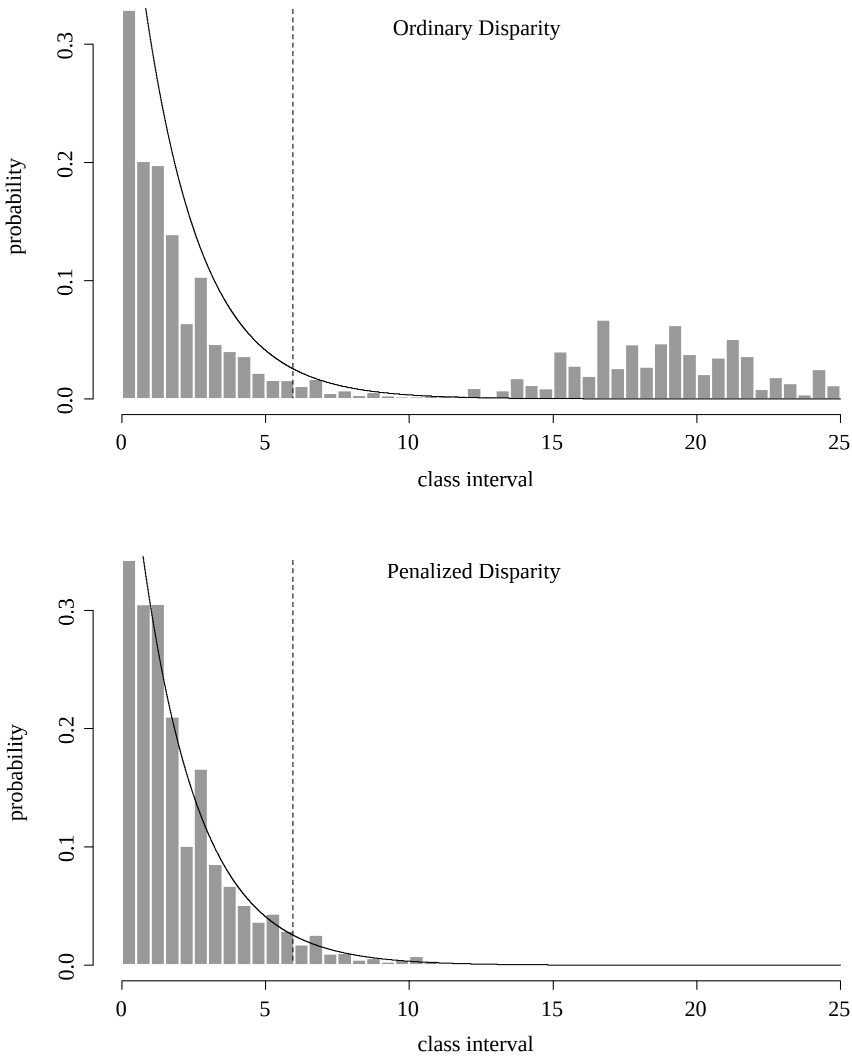


Fig. 1. Histograms of test statistics and their $\chi^2$ (2) approximations

and heavy tail of the statistic $T^{-0.9}$ provided by the $\chi^2(2)$ density is evident (the height of each bar represents the exact probability for the test statistic to lie between the respective end points). However, the right tails of the histogram of $T_{p,0.5}^{-0.9}$ around and beyond the 5% critical point is very well approximated by the overlaid density, leading to high agreement in the observed and nominal levels. Similar features were observed for $\gamma = 0.1$ and 0.01, and other values of $\lambda$ in the $[-0.5, -1)$ range, although they have not been presented here for brevity.

TABLE 5: *Exact levels of the ordinary and penalized minimum disparity test statistics with three penalty schemes for testing the composite null hypothesis $H_0 : \beta = -1.5$ regarding the parameters of logit model at nominal level 10%.*

| | | Ordinary Disparity | Penalized Disparity | | |
|---|---|---|---|---|---|
| | | | $\omega = 1.0$ | $\omega = 0.5$ | $\omega = 0.0$ |
| $\lambda$ | $n$ | Observed Level | Observed Level | Observed Level | Observed Level |
| 1.0 | 20 | 0.095901 | 0.150740 | 0.095901 | 0.076076 |
| | 25 | 0.096374 | 0.151670 | 0.096374 | 0.092667 |
| | 30 | 0.094674 | 0.156184 | 0.094674 | 0.090174 |
| | 40 | 0.096979 | 0.173511 | 0.096979 | 0.096725 |
| 0.0 | 20 | 0.142799 | 0.142799 | 0.108600 | 0.100752 |
| | 25 | 0.137898 | 0.137898 | 0.103055 | 0.098943 |
| | 30 | 0.141535 | 0.141535 | 0.101298 | 0.098609 |
| | 40 | 0.152867 | 0.152867 | 0.103420 | 0.102746 |
| −0.5 | 20 | 0.246378 | 0.137767 | 0.129488 | 0.121643 |
| | 25 | 0.250641 | 0.131497 | 0.107404 | 0.103691 |
| | 30 | 0.253139 | 0.137712 | 0.108778 | 0.106794 |
| | 40 | 0.258066 | 0.142000 | 0.111829 | 0.110989 |
| −0.6 | 20 | 0.258825 | 0.138005 | 0.129685 | 0.121883 |
| | 25 | 0.258712 | 0.127864 | 0.108037 | 0.104271 |
| | 30 | 0.256197 | 0.134621 | 0.110235 | 0.107920 |
| | 40 | 0.267819 | 0.142104 | 0.113472 | 0.112539 |
| −0.7 | 20 | 0.316689 | 0.138611 | 0.130250 | 0.122486 |
| | 25 | 0.307577 | 0.125881 | 0.110127 | 0.106360 |
| | 30 | 0.307679 | 0.132534 | 0.110968 | 0.108734 |
| | 40 | 0.328664 | 0.133389 | 0.113749 | 0.112394 |
| −0.8 | 20 | 0.422257 | 0.139242 | 0.130805 | 0.122416 |
| | 25 | 0.456079 | 0.125162 | 0.110365 | 0.106598 |
| | 30 | 0.426111 | 0.133223 | 0.115214 | 0.113748 |
| | 40 | 0.425560 | 0.131983 | 0.115363 | 0.114416 |
| −0.9 | 20 | 0.581532 | 0.140394 | 0.131923 | 0.123519 |
| | 25 | 0.612361 | 0.126689 | 0.113189 | 0.109330 |
| | 30 | 0.628707 | 0.133140 | 0.116614 | 0.115146 |
| | 40 | 0.645533 | 0.117472 | 0.116054 | 0.115323 |

Table 6: *Exact levels of the ordinary and penalized minimum disparity test statistics with three penalty schemes for testing the composite null hypothesis $H_0 : \beta = -1.5$ regarding the parameters of logit model at nominal level 5%.*

| | | Ordinary Disparity | Penalized Disparity | | |
|---|---|---|---|---|---|
| | | | $\omega = 1.0$ | $\omega = 0.5$ | $\omega = 0.0$ |
| $\lambda$ | $n$ | Observed Level | Observed Level | Observed Level | Observed Level |
| 1.0 | 20 | 0.053001 | 0.088905 | 0.053001 | 0.047738 |
| | 25 | 0.050062 | 0.076978 | 0.050062 | 0.044172 |
| | 30 | 0.047668 | 0.080257 | 0.047668 | 0.041869 |
| | 40 | 0.046802 | 0.086700 | 0.046802 | 0.046016 |
| 0.0 | 20 | 0.080175 | 0.080175 | 0.054719 | 0.049084 |
| | 25 | 0.061673 | 0.061673 | 0.050049 | 0.047691 |
| | 30 | 0.068061 | 0.068061 | 0.060122 | 0.059088 |
| | 40 | 0.073473 | 0.073473 | 0.053747 | 0.053451 |
| −0.5 | 20 | 0.155083 | 0.073200 | 0.055277 | 0.049355 |
| | 25 | 0.159944 | 0.059908 | 0.051903 | 0.049535 |
| | 30 | 0.125976 | 0.063756 | 0.062399 | 0.061308 |
| | 40 | 0.128084 | 0.065355 | 0.065160 | 0.064949 |
| −0.6 | 20 | 0.166887 | 0.073060 | 0.055417 | 0.049374 |
| | 25 | 0.170337 | 0.060815 | 0.054006 | 0.051657 |
| | 30 | 0.171345 | 0.063738 | 0.062380 | 0.061289 |
| | 40 | 0.176251 | 0.067318 | 0.067116 | 0.066911 |
| −0.7 | 20 | 0.241835 | 0.072660 | 0.054973 | 0.048942 |
| | 25 | 0.244005 | 0.062151 | 0.057057 | 0.054706 |
| | 30 | 0.251531 | 0.064642 | 0.063284 | 0.062193 |
| | 40 | 0.209547 | 0.068975 | 0.068773 | 0.068567 |
| −0.8 | 20 | 0.301509 | 0.062728 | 0.055664 | 0.049561 |
| | 25 | 0.312500 | 0.064554 | 0.061479 | 0.059127 |
| | 30 | 0.303337 | 0.066056 | 0.064648 | 0.063604 |
| | 40 | 0.322263 | 0.069256 | 0.069054 | 0.068848 |
| −0.9 | 20 | 0.461987 | 0.068229 | 0.061098 | 0.054923 |
| | 25 | 0.498336 | 0.064511 | 0.061417 | 0.058871 |
| | 30 | 0.521048 | 0.067831 | 0.066422 | 0.065357 |
| | 40 | 0.541143 | 0.070161 | 0.069960 | 0.069754 |

For the logit model we are testing a composite null hypothesis and in this case the asymptotic null distribution of the statistics $T^{\lambda}$ and $T^{\lambda}_{p,\omega}$ are both $\chi^2(1)$ distributions. Having thus calculated their

TABLE 7: *Exact levels of the ordinary and penalized minimum disparity test statistics with three penalty schemes for testing the composite null hypothesis* $H_0 : \beta = -1.5$ *regarding the parameters of logit model at nominal level 1%.*

| | | Ordinary Disparity | Penalized Disparity | | |
|---|---|---|---|---|---|
| | | | $\omega = 1.0$ | $\omega = 0.5$ | $\omega = 0.0$ |
| $\lambda$ | $n$ | Observed Level | Observed Level | Observed Level | Observed Level |
| 1.0 | 20 | 0.010420 | 0.021235 | 0.010420 | 0.006903 |
| | 25 | 0.011208 | 0.025526 | 0.011208 | 0.010538 |
| | 30 | 0.010295 | 0.025153 | 0.010295 | 0.009645 |
| | 40 | 0.010909 | 0.024718 | 0.010909 | 0.010526 |
| 0.0 | 20 | 0.014688 | 0.014688 | 0.009682 | 0.007892 |
| | 25 | 0.013370 | 0.013370 | 0.010445 | 0.009529 |
| | 30 | 0.015655 | 0.015655 | 0.009710 | 0.009261 |
| | 40 | 0.014008 | 0.014008 | 0.012442 | 0.012351 |
| −0.5 | 20 | 0.034828 | 0.014808 | 0.010485 | 0.008567 |
| | 25 | 0.040542 | 0.013448 | 0.011532 | 0.010599 |
| | 30 | 0.046836 | 0.015108 | 0.014320 | 0.013860 |
| | 40 | 0.052747 | 0.013467 | 0.013308 | 0.013216 |
| −0.6 | 20 | 0.057384 | 0.014918 | 0.010589 | 0.008671 |
| | 25 | 0.066270 | 0.013553 | 0.011690 | 0.010757 |
| | 30 | 0.069320 | 0.015341 | 0.014552 | 0.014089 |
| | 40 | 0.077393 | 0.013935 | 0.013772 | 0.013681 |
| −0.7 | 20 | 0.099741 | 0.016217 | 0.011880 | 0.009965 |
| | 25 | 0.108981 | 0.015148 | 0.013285 | 0.012351 |
| | 30 | 0.108437 | 0.016497 | 0.015708 | 0.015245 |
| | 40 | 0.090335 | 0.014448 | 0.014110 | 0.014015 |
| −0.8 | 20 | 0.156823 | 0.016470 | 0.012085 | 0.010154 |
| | 25 | 0.135369 | 0.016414 | 0.014542 | 0.013607 |
| | 30 | 0.150324 | 0.019646 | 0.018856 | 0.018384 |
| | 40 | 0.151219 | 0.014923 | 0.014484 | 0.014353 |
| −0.9 | 20 | 0.311807 | 0.016604 | 0.012214 | 0.010275 |
| | 25 | 0.311650 | 0.018557 | 0.016687 | 0.015749 |
| | 30 | 0.348043 | 0.020946 | 0.020155 | 0.019683 |
| | 40 | 0.342513 | 0.015485 | 0.015042 | 0.014908 |

asymptotic critical points, we have determined the exact levels of the tests as the maximum of the observed sizes over all the different values of the parameter $\alpha$. The results corresponding to the nominal levels

$\gamma = 0.1, 0.05$ and $0.01$ are given in Tables 5-7. For the composite null hypothesis, too, the findings are similar. The penalties again lead to major differences in the levels of the tests.

## 4. Concluding remarks

In this paper we have provided a moderate study on the effects of an empty cell penalty on some density-based minimum disparity estimators in multinomial models. These minimum disparity estimators and the corresponding parametric tests are known to have good robustness and asymptotic optimality properties, but their applicability is tempered by their observed poor performances in small samples. In this paper we have attempted to demonstrate the improved performance of these estimators and tests when a small sample penalty is applied through some exact comparisons in the multinomial model. It appears that the penalized estimators discussed do achieve good small sample efficiency in the cases that we have studied.

We have considered three different weights for the penalty, and among the cases that we have studied the penalty weight $\omega = 0$ has done well in terms of the MSE. On the other hand this penalty weight seems to slightly underestimate the nominal level in the testing problems. While it is clear that more extensive and detailed investigations have to be made before a general recommendation about an optimal value of $\omega$ can be made, it does appear that some penalty weight in the interval [0, 1] may be a reasonable thing to attempt in minimum disparity inference problems for large negative values of $\lambda$.

## REFERENCES


Basu, A. and Basu, S. (1998) Penalized minimum disparity methods in multinomial models, *Statistica Sinica,* 8, 841-860.

Basu, A., Basu, S. and Chaudhury, G. (1997) Robust minimum divergence procedures for count data models, *Sankhya B,* 59, 11-27.

Basu, A., Harris, I. R. and Basu, S. (1996) Tests of hypotheses in discrete models based on the penalized Hellinger distance, *Statistics & Probability Letters,* 27, 367-373.

Basu, A. and Lindsay, B. G. (1994) Minimum disparity estimation for continuous models, *Annals of The Institute of Statistical Mathematics,* 46, 683-705.

BASU, A. and SARKAR, S. (1994) On disparity based goodness-of-fit tests for multinomial models, *Statistics & Probability Letters*, 19, 307-312.

BASU, S. and BASU, A. (1995) Comparison of several goodness of fit tests for the kappa statistic based on exact power and coverage probability, *Statistics in Medicine*, 14, 347-356.

BERAN, R. (1997) Minimum Hellinger distance estimates for parametric models, *Ann. Statist.*, 5, 445-463.

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

HARRIS, I. R. and BASU, A. (1994) Hellinger distance as a penalized log likelihood, *Commun. Statist: Simula*, 23, 1097-1113.

HARRIS, I. R. and BASU, A. (1996) A generalized divergence measure, *Technical Report*, Computer Science Unit, Indian Statistical Institute, Calcutta 700 035, India.

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

PARK, C., BASU, A. and BASU, S. (1995) Robust minimum distance inference based on combined distances, *Communications in Statistics: Simulation and Computation*, 24, 653-673.

RAO, C. R. (1973) *Linear Statistical Inference and Its Applications*, 2nd Ed. John Wiley & Sons, New York.

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

SARKAR, S. and BASU, A. (1995) On disparity based robust tests for two discrete populations, *Sankhya B*, 57, 353-364.

SERFLING, R. (1980) *Approximation Theorems of Mathematical Statistics*, John Wiley, New York.

SIMPSON, D. G. (1987) Minimum Hellinger distance estimation for the analysis of count data, *J. Amer. Statist. Assoc.*, 82, 802-807.

SIMPSON, D. G. (1989A) Hellinger deviance test: efficiency, breakdown points, and robustness, *J. Amer. Statist. Assoc.*, 84, 107-113.

SIMPSON, D. G. (1989B) Choosing a discrepancy for minimum distance estimation: multinomial models with infinitely many cells, *Technical Report*, Department of Statistics, University of Illinois, Champaign, Illinois, U. S. A.

TAMURA, R. N. and BOOS, D. D. (1986) Minimum Hellinger distance estimation for multivariate location and covariance, *J. Amer. Statist. Assoc.*, 81, 223-229.

**Exact minimum disparity inference**
**in complex multinomial models**

SUMMARY

Estimation of the probability vector in a multinomial set-up is an important practical problem. Under moderate contaminations and model misspecifications several minimum distance estimators corresponding to the Cressie-Read family of disparities have better robustness properties than the maximum likelihood estimator. However, it has also been previously observed that when an empty cell penalty is introduced, the above mentioned estimators often show marked improvement in their small sample efficiencies. In this paper we have studied the role of different penalties in reducing the mean square errors of the estimators and in improving the chi-square approximation of the penalized test statistics under certain parametric models within the multinomial family.

**Inferenza esatta di minima disparità**
**in modelli multinomiali complessi**

RIASSUNTO

La stima del vettore delle probabilità nel contesto della multinomiale è un importante problema operativo. In caso di moderate contaminazioni ed errori di specificazione del modello, diversi stimatori di minima distanza corrispondenti alla famiglia di Cressie e Read delle disparità hanno migliori proprietà di robustezza rispetto allo stimatore di massima verosimiglianza. Tuttavia, è stato osservato che quando viene introdotta una penalità per cella vuota, i menzionati stimatori mostrano spesso un marcato miglioramento dell'efficienza nel caso di piccoli campioni. Nel presente articolo, è stato studiato il ruolo giocato da differenti penalità nella riduzione dell'errore quadratico medio degli stimatori e nel miglioramento dell'approssimazione al Chi Quadrato della statistica test penalizzata sotto alcuni modelli parametrici all'interno della famiglia multinomiale.

KEY WORDS

Exact computations; Simple and composite null hypothesis; Exact levels of test statistics; Empty cells.

[Manuscript received March 1999; final version received February 2000.]