Sankhyā : The Indian Journal of Statistics
2002, Volume 64, Series B, Pt. 2, pp 128-140

OUTLIER RESISTANT MINIMUM DIVERGENCE METHODS IN DISCRETE PARAMETRIC MODELS

By AYANENDRANATH BASU
Indian Statistical Institute, Kolkata

SUMMARY. Minimum Hellinger distance and related methods have been shown to simultaneously possess first order asymptotic efficiency and attractive robustness properties (Beran 1997; Simpson 1987, 1989a; Lindsay 1994). It has been noted, however, that these minimum divergence procedures are generally associated with unbounded influence functions, a property considered undesirable in traditional robust procedures. Lindsay has demonstrated the limitations of the influence function approach in this case. Following Lindsay's outlier stability approach, we show in this paper that there exists a similar outlier resistance property for the corresponding tests of hypotheses, and that this outlier resistance property leads to some useful and interesting results for the estimators and the corresponding tests of hypotheses for the generalized Hellinger divergence family (Simpson 1989b; Basu et al., 1997) in discrete models.

1. **Introduction**

The popularity of the minimum Hellinger distance and related methods in statistical inference (Beran 1977; Tamura and Boos 1986; Simpson 1987, 1989a, 1989b; Eslinger and Woodward 1991; Lindsay 1994; Basu and Lindsay 1994; Basu and Sarkar 1994a,b; Markatou et al. 1998) is mainly due to the ability of the corresponding techniques to combine the property of asymptotic efficiency with certain attractive robustness properties. While such methods require a nonparametric estimate of the true density, it is relatively simple in discrete parametric models since one can use the 'empirical' density for this estimate (Simpson 1987).

At the same time, however, robust techniques constructed through some density based minimum divergence methods such as those based on the Hellinger distance do not generally have bounded influence functions; in fact

---

Paper received July 2000; revised May 2002.
AMS (2000) *subject classification.* Primary 62F03; secondary 62F35.
Keywords and phrases. Bounded effective influence, disparity, generalized hellinger divergence; influence function; residual adjustment function.

their influence functions are the same as those of the maximum likelihood estimators – as they must be if they are to be asymptotically efficient. Many of the authors mentioned above have discussed the limitations of the influence function approach in measuring the robustness of these estimators. Beran has considered the "α-influence function" of the minimum Hellinger distance estimator, while Lindsay has looked at the outlier stability of the estimating equations for several divergences, including the Hellinger distance. Simpson has demonstrated nice breakdown properties for procedures resulting from the Hellinger distance.

Here we extend Lindsay's outlier stability approach, and show in this paper that there exists a similar outlier resistance property for the corresponding tests of hypotheses. This property leads to particularly interesting results for the estimators and the tests of hypotheses for the generalized Hellinger divergence (GHD) family (See Simpson 1989b and Basu et al. 1997). The inflation factors for the asymptotic chi-square distribution of the test statistics under a parametric model contaminated by a large outlier turn out to be simple functions of the contaminating proportion which are reasonably close to 1 for large outliers for some members of the generalized Hellinger divergence family.

The rest of the paper is organized as follows: minimum 'disparity' methods are discussed in Section 2. The outlier stability of minimum disparity estimators are discussed in Section 3 and its consequence on the generalized Hellinger divergence family investigated. Section 4 establishes and studies their outlier resistance properties in the context of parametric tests of hypotheses based on disparities with illustrations and also demonstrates that the chi- square inflation factor has a simple form in the case of the generalized Hellinger divergence.

2. Minimum Disparity Estimation in Discrete Models

Consider a discrete model with density $f_\theta(x)$. While our treatment will also include random variables with finite support, as a general case we will assume that the random variable of interest is supported on $\{0, 1, 2, \ldots\}$, $\theta \in \Omega$, the parameter space. Let $X_1, \ldots, X_n$ be a random sample from the true distribution modeled by the above family, and let $d(x) = \#(X_i = x)/n$ be the relative frequency (the 'empirical density') of the value x in the sample. We denote $d = (d(0), d(1), \ldots)$, and the vector f_θ is similarly defined.

A 'disparity' between d and f_θ corresponding to a strictly convex, thrice differentiable function G on $[-1, \infty)$ is given by

$$\rho_G(d, f_\theta) = \sum_{x=0}^{\infty} G(\delta(x)) f_\theta(x), \tag{1}$$

where $\delta(x) = (d(x) - f_\theta(x))/f_\theta(x)$, the 'Pearson residual' at x. Notice that $\rho_G \geq 0$, with equality if and only if $d \equiv f_\theta$, identically. When there is no scope for confusion we will simply write ρ for ρ_G. The Kullback-Leibler divergence, Pearson's chi-square, Neyman's modified chi-square, and the squared Hellinger distance are well known examples in the class of disparities. A famous subclass of disparities is the Cressie-Read family of power divergences (Cressie and Read 1984) given by

$$I^\lambda(d, f_\theta) = \frac{1}{\lambda(\lambda+1)} \sum_{x=0}^{\infty} d(x) \left[\left(\frac{d(x)}{f_\theta(x)} \right)^\lambda - 1 \right], \quad \lambda \in I\!R, \tag{2}$$

which generates all the standard examples of common disparities stated above as special cases.

The minimum disparity estimator $\hat{\theta}$ of θ based on the disparity in (1) is then defined by the relation

$$\rho_G(d, f_{\hat{\theta}}) = \min_{\theta \in \Omega} \rho_G(d, f_\theta) \tag{3}$$

provided such a $\hat{\theta}$ exists. We will denote the corresponding functional by $T_\rho(d) = \hat{\theta}$.

Under differentiability of the model, the minimization of the above disparity corresponds to solving an estimating equation of the form

$$-\nabla \rho(d, f_\theta) = \sum_{x=0}^{\infty} (G'(\delta(x))(\delta(x)+1) - G(\delta(x))) \nabla f_\theta(x) = 0, \tag{4}$$

where ∇ represents the gradient with respect to θ, and G' is the first derivative of G with respect to its argument (G'' will denote the corresponding second derivative). Letting $G'(\delta)(\delta+1) - G(\delta) = A(\delta)$, the estimating equation has the form

$$\sum_{x=0}^{\infty} A(\delta(x)) \nabla f_\theta(x) = 0. \tag{5}$$

Often we choose the function G to satisfy

$$G'(0) = 0, \quad G''(0) = 1. \tag{6}$$

This can be done because the disparity in (1) may be centred and rescaled to the form

$$\rho_{G^*}(d, f_\theta) = \sum G^*(\delta(x)) f_\theta(x) = \sum \left(\frac{G(\delta(x)) - G'(0)\delta(x)}{G''(0)} \right) f_\theta(x). \quad (7)$$

This does not change the estimating properties of the disparity in the sense that $\hat{\theta}$ which is the minimizer of ρ_G is also the minimizer of ρ_{G^*}; however G^* satisfies the conditions in (6).

Under (6), the function $A(\delta)$ satisfies $A(0) = 0$, and $A'(0) = 1$ (A' being the first derivative of A with respect to its argument). This function $A(\delta)$ is called the residual adjustment function (*RAF*) of the disparity ρ_G. Since the estimating equations of the different disparities differ only in the form the *RAF*, it is clear that the *RAF* plays a crucial role in determining the efficiency and robustness properties of the minimum disparity estimator. See Lindsay for more details on minimum disparity estimation.

3. Outlier Stability of Minimum Disparity Estimators

Consider a fixed model $f_\theta(x)$, the observed relative frequencies $d(x)$ obtained from a random sample generated by the unknown true distribution, a contamination proportion ϵ (which will be taken to be less than 0.5 here as well as in the rest of the paper), and let $\{\xi_j : j = 1, 2, \ldots,\}$ be a sequence of elements of the sample space. Consider the ϵ- contaminated data $d_j(x) = (1-\epsilon)d(x) + \epsilon\chi_{\xi_j}(x)$, $\chi_y(x)$ being the indicator function at y, and let $\delta_\theta^j(\cdot) = d_j(x)/f_\theta(x) - 1$ denote the Pearson residual for the ϵ-contaminated data. We will say that $\{\xi_j\}$ constitutes an outlier sequence for the model f_θ and data d if $\delta_\theta^j(\xi_j) \to \infty$ and $d(\xi_j) \to 0$ as $j \to \infty$. Lemma 9, Lindsay (1994) shows that $\{\xi_j\}$ constitutes an outlier sequence if and only if $d(\xi_j) \to 0$ and $f_\theta(\xi_j) \to 0$ as $j \to \infty$.

Let us consider the limiting behavior of the disparity measure $\rho(d, f_\theta)$ under contamination through an outlier sequence $\{\xi_j\}$. Let $d_\epsilon^*(x) = (1 - \epsilon)d(x)$. While $d_\epsilon^*(x)$ is not a density function, one can formally calculate $\rho(d_\epsilon^*, f_\theta)$. Following Lindsay, we note that

$$\rho(d_\epsilon^*, f_\theta) \to \rho(d, f_\theta) \text{ as } \epsilon \to 0 \quad (8)$$

under mild conditions of dominated convergence. And if in addition

$$\rho(d_j, f_\theta) \to \rho(d_\epsilon^*, f_\theta) \text{ as } j \to \infty, \quad (9)$$

then, for extreme outliers and small contaminating fractions ϵ, the disparity between the contaminated data d_j and f_θ is close to $\rho(d, f_\theta)$, the disparity obtained by simply deleting the outlier from the sample. Equation (8) exhibits a continuity property of the divergence measure that is closely related to the notion of qualitative robustness. Equation (9) represents a key stability property of the divergence which demonstrates its outlier rejection capability under an outlier sequence. A sufficient condition for the disparity under which the convergence in (9) holds is that $G(-1)$ is finite and $G(\delta)/\delta$ converges to zero as $\delta \to \infty$ (Lindsay 1994, Proposition 12). (Henceforth we will denote this by condition C1). Notice that the condition C1 is satisfied by the Cressie-Read family for $\lambda < 0$, and therefore this subfamily has this stability property in (9). The disparities of the generalized Hellinger divergence family (Basu et al., 1997), which is the Cressie-Read family restricted to $\lambda \in [-1, 0]$, has the form $GHD^\alpha(d, f_\theta) = \left(1 - \sum_x d^\alpha(x) f_\theta^{1-\alpha}(x)\right) / [\alpha(1-\alpha)]$, $\alpha \in [0, 1]$; this family, therefore, also satisfies condition C1 for $\alpha < 1$ (the disparities for $\alpha = 0, 1$ being defined through the corresponding limiting values at $\alpha = 0, 1$). Here $GHD^{1/2} = HD$, the (twice) squared Hellinger distance.

For $\alpha \in (0, 1)$ the minimization of the generalized Hellinger divergence $GHD^\alpha(d, f_\theta)$ is equivalent to maximizing $S^\alpha(d, f_\theta) = \sum_x d^\alpha(x) f_\theta^{1-\alpha}(x)$, and conditions (8) and (9) may be stated in terms of the convergence of the S^α's. Notice that $S^\alpha(d_\epsilon^*, f_\theta) = (1-\epsilon)^\alpha \sum_x d^\alpha(x) f_\theta^{1-\alpha}(x) = (1-\epsilon)^\alpha S^\alpha(d, f_\theta)$, so that the maximizers of $S^\alpha(d, f_\theta)$ and $S^\alpha(d_\epsilon^*, f_\theta)$ (or the minimizers of $GHD^\alpha(d, f_\theta)$ and $GHD^\alpha(d_\epsilon^*, f_\theta)$) are one and the same. Notice also that for $\alpha = 1$, one gets the likelihood disparity

$$LD(d, f_\theta) = \sum_{x=0}^{\infty} \left[d(x) log(d(x)/f_\theta(x)) + (f_\theta(x) - d(x)) \right], \qquad (10)$$

minimization of which generates the maximum likelihood estimator (MLE) of θ.

There is a corresponding outlier stability property of the estimating equations themselves. Let u_θ be the maximum likelihood score function for the model. If for some $k > 1$, $E_\theta[|u_\theta(X)|^k]$ is finite for all θ, then any disparity for which $A(\delta) = O(\delta^{(k-1)/k})$, and $A(-1)$ is finite has an outlier stability property (Lindsay, Proposition 14) in the sense that, under the above definitions of an outlier sequence, as $j \to \infty$

$$\sum A(\delta^j(x)) \nabla f_\theta(x) \to \sum A(\delta_\epsilon^*(x)) \nabla f_\theta(x)$$

where $\delta_j(x) = d_j(x)/f_\theta(x) - 1$ and $\delta_\epsilon^*(x) = d_\epsilon^*(x)/f_\theta(x) - 1$ (the possible θ

subscripts have been suppressed). As the estimating equations converge, the solutions will converge as well provided the convergence is uniform.

Lindsay, in fact, gives another set of sufficient conditions for $T_j = T_\rho(d_j)$ to converges to $\theta^* = T_\rho(d_\epsilon^*)$ as $j \to \infty$. Assume that: (C2) $\rho(d_j, f_\theta)$ and $\rho(d_\epsilon^*, f_\theta)$ are continuous in θ, with the latter having unique minimum at $T_\rho(d_\epsilon^*) = \theta^*$; (C3) the convergence in (9) is uniform in θ for any compact set Θ of parameter values containing θ^*; and (C4) for each $0 < \gamma < 1$ there exists a subset $\boldsymbol{S}$ of the sample space such that (i) $d(\boldsymbol{S}) = \sum_{x \in \boldsymbol{S}} d(x) \geq 1 - \gamma$, and (ii) $\boldsymbol{C} = \{\theta : f_\theta(\boldsymbol{S}) \geq \gamma\}$ is a compact set. Then (C1), (C2), (C3) and (C4) (assumptions 10, 17, 18, and 19 of Lindsay, 1994) imply that $T_j \to \theta^*$ as $j \to \infty$. Estimators having this property will be said to be 'outlier stable' in the sense that a large point mass contamination fails to have a serious impact on the estimtor, and in the limit can only displace it as far as θ^*.

To illustrate the effect of an outlier sequence on the estimators within the generalized Hellinger divergence family, we present a small numerical study here. A sample of size 50 was generated from the Poisson (2) distribution and the empirical density d was calculated. The largest observed value in the sample was 5. In the following, we have determined the minimum generalized Hellinger divergence estimates of θ under the Poisson (θ) model assuming the observed density to be $d_\epsilon(x) = (1 - \epsilon)d(x) + \epsilon\chi_y(x)$, where $\chi_y(x)$ is the indicator function at y, and $y = 6, 7, \ldots, 20$ (a sequence of large values starting with the smallest integer larger than the largest observed value). The value of ϵ was chosen to to be 0.19, for the simple reason that the contaminated test statistics for such a contamination (Section 4) are linked to $(1 - \epsilon)^\alpha$ times the original test statistic, and this factor becomes equal to 0.9 for the Hellinger distance ($\alpha = 0.5$). For each value of $\alpha = 0.5, 0.4, \ldots, 0.1$, the estimates of θ as functions of y are recorded. We chose these values of α as the method provided a degree of downweighting equal to greater than that of the Hellinger distance in these cases. As the patterns are similar, we graphically exhibit the results for $\alpha = 0.5, 0.3, 0.1$, by plotting the estimates of θ as functions of y in Figure 1. The solid horizontal line represents the value of the estimator corresponding to the uncontaminated data d. As expected, the estimates eventually converge to that for uncontaminated data, when the outlier is large enough. Thus a big outlier, instead of badly affecting the estimate, does not change it at all! Although we have not presented similar calculations for the maximum likelihood estimator, one can easily imagine how badly they will be affected by such an outlier sequence.

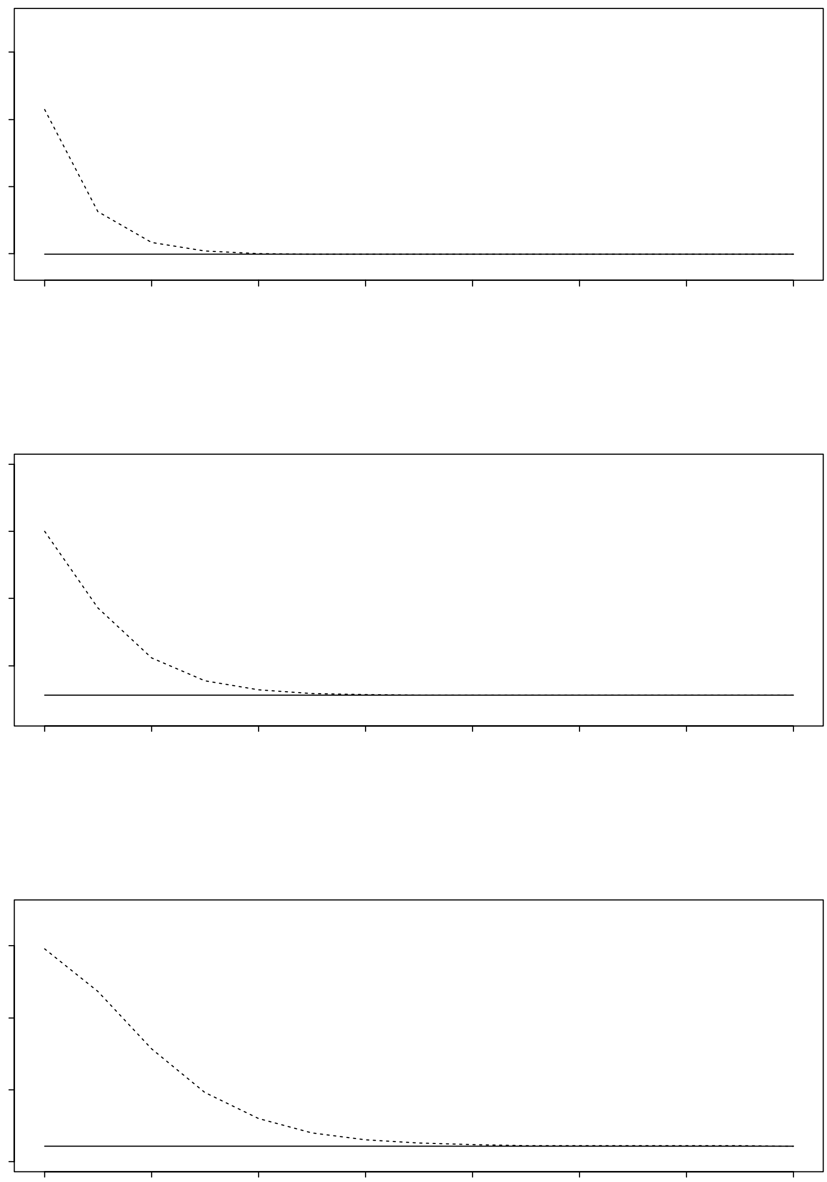

Figure 1. Behavior of the estimators of the mean under the Poisson model in the presence of an outlier sequence.

4. Hypothesis Testing Based on Disparities

4.1 *Outlier stability.* Consider the same set up as before: $f_\theta(x)$ represents a parametric model, $\theta \in \Omega$, and $d(x)$ are the relative frequencies based on a random sample of size n from the unknown true distribution. Let T_ρ be the minimum disparity functional. Consider testing the hypothesis $H_0 : \theta = \theta_0$ against a suitable alternative. The likelihood ratio test statistic (LRT), negative of twice log likelihood ratio, can be expressed as:

$$LRT(d) = 2n[LD(d, f_{\theta_0}) - LD(d, f_T)], \text{ with } T = T_{LD}(d) = MLE. \quad (11)$$

The LRT has an asymptotic $\chi^2(r)$ distribution under the null hypothesis, where r is the dimension of θ. The analogous disparity test statistics for the disparity ρ is

$$D_\rho(d) = 2n[\rho(d, f_{\theta_0}) - \rho(d, f_T)] \text{ with } T = T_\rho(d). \quad (12)$$

Consider the effect of contaminating the data d with an outlier sequence $\{\xi_j\}$ on the disparity test statistic. Let d_j, and d^*_ϵ be defined as in Section 3. Define the disparity test statistic D_ρ to be outlier stable if

$$D_\rho(d_j) \to D_\rho(d^*_\epsilon) \text{ as } j \to \infty. \quad (13)$$

Let $T_j = T_\rho(d_j)$. In the following we provide the conditions under which the disparity test statistic D_ρ is outlier stable.

THEOREM 1: *Let $\delta^j_\theta(\xi_j) = ((1-\epsilon)d(\xi_j) + \epsilon)/f_\theta(\xi_j) - 1$, where $d(\cdot)$ are the relative frequencies from a given random sample, and $\{\xi_j\}$ is a corresponding outlier sequence. Suppose that the disparity ρ satisfies conditions (C1), (C2), (C3) and (C4). Then the disparity test statistic D_ρ is outlier stable.*

PROOF. Let $\theta^* = T_\rho(d^*_\epsilon)$. Under the given conditions, $T_j \to \theta^*$ as $j \to \infty$, and by Scheffe's theorem (see, for example, Billingsley, 1986, pp. 218) $f_{T_j}(\xi_j) \to 0$ as $j \to \infty$. For a finite sample of size n, $d(\xi_j) = 0$ whenever $j > m$, for some integer $m \geq 0$ depending on the sample. Notice that for $j > m$,

$$\begin{aligned} |D_\rho(d_j) - D_\rho(d^*_\epsilon)| \;\leq\; & 2n\{|\rho(d_j, f_{\theta_0}) - \rho(d^*_\epsilon, f_{\theta_0})|\} + 2n\{|\rho(d^*_\epsilon, f_{T_j}) - \rho(d^*_\epsilon, f_{\theta^*})| \\ & + |G(-1) f_{T_j}(\xi_j)| + |G(\delta^j_{T_j}(\xi_j)) f_{T_j}(\xi_j)|\} \end{aligned}$$

From the given conditions, $|\rho(d_j, f_{\theta_0}) - \rho(d_\epsilon^*, f_{\theta_0})| \to 0$. Since condition (C2) holds and $T_j \to \theta^*$ as $j \to \infty$, $|\rho(d_\epsilon^*, f_{T_j}) - \rho(d_\epsilon^*, f_{\theta^*})| \to 0$. As $G(-1)$ is finite, $G(-1)f_{T_j}(\xi_j)$ also converges to zero. Note that

$$G(\delta_{T_j}^j(\xi_j))f_{T_j}(\xi_j) = \frac{G(\delta_{T_j}^j(\xi_j))}{\delta_{T_j}^j(\xi_j)}(d_j(\xi_j) - f_{T_j}(\xi_j)$$

and the right hand side converges to zero from the given conditions. This completes the proof. □

In particular, for $\rho = HD$ (or $GHD^{1/2}$), we get

$$\begin{aligned} D_\rho(d_j) &\to D_\rho(d_\epsilon^*) \\ &= (1-\epsilon)^{1/2} D_\rho(d) \end{aligned} \tag{15}$$

so that a single outlying value, however large, cannot arbitrarily perturb the test statistic. Here $G(\delta) = 2(\sqrt{\delta+1} - 1)^2$. For an $\alpha \in (0,1)$, $D_{GHD^\alpha}(d_j) \to (1-\epsilon)^\alpha D_{GHD^\alpha}(d)$ as $j \to \infty$.

We now present a small example of this outlier stability using the binomial (12, p) model. We generated a pseudo random sample of size 50 from the binomial (12, 0.1) distribution. Consider testing the hypothesis $H_0 : p = 0.1$ against $H_1 : p \neq 0.1$. The likelihood ratio and the Hellinger distance test statistics for the original data

$$LRT(d) = 1.23545 \quad \text{and} \quad D_{HD}(d) = 1.61255.$$

Next we chose $y = 12$ and $\epsilon = 0.19$, and calculated the LRT and the Hellinger distance test statistic for the contaminated version of the data $d_y(x) = (1-\epsilon)d(x) + \epsilon\chi_y(x)$. The values now are

$$LRT(d_y) = 124.5748 \quad \text{and} \quad D_{HD}(d_y) = 1.45119.$$

Clearly the presence of the outlier blows up the LRT, but fails to affect the HD test statistic in any major way. Notice that the latter is practically equivalent to $(1-\epsilon)^{1/2}[D_{HD}(d)] = 1.45129$, which is what we should expect from (15). The values of the test statistic $D_{GHD^\alpha}(d_y)$ for $\alpha = 0.1, 0.2, \ldots, 0.5$, and $y = 8, 9, 10, 11, 12$ are presented in Table 1. Notice how closely the statistics for $y = 12$ match with $(1-\epsilon)^\alpha\times$ [the uncontaminated statistics].

TABLE 1. OBSERVED TEST STATISTICS FOR THE CONTAMINATED BINOMIAL DISTRIBUTION. MODEL IS BINOMIAL(12, p), SAMPLE SIZE n=50.

	α					
	0.1	0.2	0.3	0.4	0.5	LRT
Without contamination	3.78409	2.44005	1.99613	1.76313	1.61255	1.23545
y=8	3.69676	2.32694	1.84874	1.56294	1.31025	50.27147
y=9	3.70459	2.33815	1.87053	1.61021	1.41639	66.23774
y=10	3.70516	2.33927	1.87352	1.61915	1.44437	84.00180
y=11	3.70518	2.33935	1.87382	1.62045	1.45024	103.47326
y=12	3.70518	2.33935	1.87384	1.62060	1.45119	124.57481
- - - - - - - - - - - -	- - - - -	- - - - -	- - - - -	- - - - -	- - - - -	- - - - - - -
$(1-\epsilon)^{\alpha}\times$ Uncontaminated Statistic	3.70519	2.33935	1.87385	1.62061	1.45129	–

4.2 *The GHD and the chi-square inflation factor.* The present section was motivated by the fact that in empirical investigations involving the chi-square inflation factor for the members of the generalized Hellinger divergence family under point mass contaminations, sometimes the observed inflation factors seemed to have a remarkably close approximation based only on α and the contamination proportion ϵ. Let $f_\theta(x)$ be the parametric model under consideration, $t(x)$ be the true density and θ be a scalar parameter. Under the set up and notations of Section 4.1, consider testing the hypothesis $H_0 : T(t) = \theta^*$ against $H_1 : T(t) \neq \theta^*$. Under the null hypothesis,

$$D_\rho(d) \to c(t)\chi_1^2 \tag{16}$$

in distribution (Lindsay, 1994, Theorem 6). Here

$$c(t) = Var_t(T'(X, t, \theta^*))\nabla^2\rho(t, f_\theta)|_{\theta=\theta^*},$$

$T'(y, t, \theta^*)$ being the influence function of the functional $T(t)$ at y evaluated under $\theta^* = T(t)$, and χ_1^2 is a χ^2 random variable with 1 degree of freedom. When the unknown t belongs to the model family, $c(t) = 1$. For the rest of this section, we concentrate on the generalized Hellinger divergence and denote $T_{GHD^\alpha}(\cdot)$ by $T_\alpha(\cdot)$, and the corresponding inflation factor in (16) by $c_\alpha(t)$. Let $u'_\theta(x)$ represent the first derivative of $u_\theta(x)$ with respect to θ.

PROPOSITION 2. *For fixed* $\theta_0 \in \Omega$, *assume that* $t(x) = (1-\epsilon)f_{\theta_0}(x) + \epsilon\chi_\xi(x)$. *For fixed* $\alpha \in (0, 1)$, *let* ξ *be such that* $u^2_{\theta_0}(\xi)f^{1-\alpha}_{\theta_0}(\xi)$ *and* $u'_{\theta_0}(\xi)f^{1-\alpha}_{\theta_0}(\xi)$

are approximately zero; then $\nabla^2\rho(t,f_\theta)|_{\theta=\theta_0}$ *is approximately equal to* $(1-\epsilon)^\alpha I(\theta_0)$ *for the generalized Hellinger divergence family.*

PROOF.

$$\begin{aligned}\nabla^2\rho(t,f_\theta) &= \nabla[-\sum t^\alpha(x) f_\theta^{1-\alpha}(x)u_\theta(x)]/\alpha\\ &= -(1-\alpha)\sum t^\alpha(x)f_\theta^{1-\alpha}(x)u_\theta^2(x)/\alpha\\ &\quad -\sum t^\alpha(x)f_\theta^{1-\alpha}(x)u_\theta'(x)/\alpha\end{aligned}$$

$$\begin{aligned}&\nabla^2\rho(t,f_\theta)|_{\theta=\theta_0}\\ &= -(1-\epsilon)^\alpha(1-\alpha)\sum u_{\theta_0}^2(x)f_{\theta_0}(x)/\alpha\\ &\quad -(1-\alpha)\Big([(1-\epsilon)f_{\theta_0}(\xi)+\epsilon]^\alpha f_{\theta_0}^{(1-\alpha)}(\xi)u_{\theta_0}^2(\xi)\\ &\quad -(1-\epsilon)^\alpha f_{\theta_0}(\xi)u_{\theta_0}^2(\xi)\Big)/\alpha-(1-\epsilon)^\alpha\sum u_{\theta_0}'(x)f_{\theta_0}(x)/\alpha\\ &\quad -\Big([(1-\epsilon)f_{\theta_0}(\xi)+\epsilon]^\alpha f_{\theta_0}^{(1-\alpha)}(\xi)u_{\theta_0}'(\xi)-(1-\epsilon)^\alpha f_{\theta_0}(\xi)u_{\theta_0}'(\xi)\Big)/\alpha\\ &= (1-\epsilon)^\alpha\left[I(\theta_0)-(1-\alpha)I(\theta_0)\right]/\alpha\\ &\quad -(1-\alpha)\Big([(1-\epsilon)f_{\theta_0}(\xi)+\epsilon]^\alpha f_{\theta_0}^{(1-\alpha)}(\xi)u_{\theta_0}^2(\xi)-(1-\epsilon)^\alpha f_{\theta_0}(\xi)u_{\theta_0}^2(\xi)\Big)/\alpha\\ &\quad -\Big([(1-\epsilon)f_{\theta_0}(\xi)+\epsilon]^\alpha f_{\theta_0}^{(1-\alpha)}(\xi)u_{\theta_0}'(\xi)-(1-\epsilon)^\alpha f_{\theta_0}(\xi)u_{\theta_0}'(\xi)\Big)/\alpha\\ &\approx (1-\epsilon)^\alpha I(\theta_0).\end{aligned}$$

under the stated assumptions. □

The influence function of the minimum GHD^α functional is given by $T_\alpha'(y)=T_\alpha'(y,t,\theta^*)=K_\alpha(y,\theta^*)/J_\alpha(\theta^*)$ where

$$\begin{aligned}K_\alpha(y,\theta^*) &= \alpha u_{\theta^*}(y)t^{\alpha-1}(y)f_{\theta^*}^{1-\alpha}(y),\\ J_\alpha(\theta^*) &= -[(1-\alpha)\sum t^\alpha(x)f_{\theta^*}^{1-\alpha}(x)u_{\theta^*}^2(x)+\sum t^\alpha(x)f_{\theta^*}^{1-\alpha}(x)u_{\theta^*}'(x)],\end{aligned}$$

where $\theta^*=T_\alpha(t)$, so that $Var_t(T_\alpha'(X,t,\theta^*))=Var_t(K_\alpha(X,\theta^*))/J_\alpha^2(\theta^*)$ (see Basu et al. 1997).

PROPOSITION 3. *Let* t *be as defined in Proposition 2, and* ξ *and* α*, belonging to their respective spaces, be such that the conditions of Proposition 2 hold. Then* $Var_t(T'(X,t,\theta_0))$ *is approximately equal to* $[(1-\epsilon)I(\theta_0)]^{-1}$.

PROOF. With t as given above,

$$\begin{aligned} J_\alpha(\theta_0) &= -(1-\epsilon)^\alpha[(1-\alpha)\sum f_{\theta_0}(x)u^2_{\theta_0}(x)+\sum f_{\theta_0}(x)u'_{\theta_0}(x)] \\ &\quad -(1-\alpha)\left([(1-\epsilon)f_{\theta_0}(\xi)+\epsilon]^\alpha f^{1-\alpha}_{\theta_0}(\xi)u^2_{\theta_0}(\xi)-(1-\epsilon)^\alpha f_{\theta_0}(\xi)u^2_{\theta_0}(\xi)\right) \\ &\quad -\left([(1-\epsilon)f_{\theta_0}(\xi)+\epsilon]^\alpha f^{1-\alpha}_{\theta_0}(\xi)u'_{\theta_0}(\xi)-(1-\epsilon)^\alpha f_{\theta_0}(\xi)u'_{\theta_0}(\xi)\right) \\ &\approx \alpha(1-\epsilon)^\alpha I(\theta_0) \end{aligned}$$

Also, along the lines of the proof of Proposition 2, one can check that $E_t[K(X,\theta_0)] \approx 0$. Then

$$\begin{aligned} Var_t(K_\alpha(X,\theta_0)) &\approx \alpha^2 \sum u^2_{\theta_0}(x)t^{2\alpha-1}(x)f^{2-2\alpha}_{\theta_0}(x) \\ &= \alpha^2(1-\epsilon)^{2\alpha-1}\sum u^2_{\theta_0}(x)f_{\theta_0}(x) \\ &\quad +\alpha^2\left(u^2_{\theta_0}(\xi)[(1-\epsilon)f_{\theta_0}(\xi)+\epsilon]^{2\alpha-1}f^{(2-2\alpha)}_{\theta_0}(\xi)\right. \\ &\quad \left. -(1-\epsilon)^{2\alpha-1}u^2_{\theta_0}(\xi)f_{\theta_0}(\xi)\right) \\ &\approx \alpha^2(1-\epsilon)^{2\alpha-1}I(\theta_0) \end{aligned}$$

Combining these, the required result holds. □

Notice that for the generalized Hellinger divergence, $\theta^* = T_\alpha((1-\epsilon)f_{\theta_0} + \epsilon\chi_{\xi_j})$ converges to θ_0 as $j \to \infty$ for an outlier sequence $\{\xi_j\}$. Thus

$$T_\alpha((1-\epsilon)f_{\theta_0} + \epsilon\chi_\xi) \approx T_\alpha(f_{\theta_0}) = \theta_0$$

for a ξ with $f_{\theta_0}(\xi)$ (and $u^2_{\theta_0}(\xi)f^{1-\alpha}_{\theta_0}(\xi)$ and $u'_{\theta_0}(\xi)f^{1-\alpha}_{\theta_0}(\xi)$) sufficiently small. Since under the conditions of Propositions 2 and 3,

$$Var_t(T'(X,t,\theta_0)) \times \nabla^2\rho(t,f_\theta)|_{\theta=\theta_0} \approx (1-\epsilon)^\alpha I(\theta)/[(1-\epsilon)I(\theta)] = (1-\epsilon)^{\alpha-1},$$

whenever $Var_t(T'(X,t,\theta^*))$ and $\nabla^2\rho(t,f_\theta)|_{\theta=\theta^*}$ are close, respectively, to $Var_t(T'(X,t,\theta_0))$ and $\nabla^2\rho(t,f_\theta)|_{\theta=\theta_0}$, $c_\alpha(t)$ itself will be approximately equal to $(1-\epsilon)^{\alpha-1}$ for such a ξ.

As an example look at the binomial $(20, p)$ model. Let $t(x)=(1-\epsilon)f_{p_0}(x)+\epsilon\xi_{20}(x)$, where $p_0 = 0.1$ and $\epsilon = 0.19$. Consider testing the hypothesis $H_0 : p = p^*$, where $p^* = T_{HD}(t)$. Direct calculation of the inflation factor via (16) gives $c_{0.5}(t) = 1.11111$ which is equal, at least up to five places after the decimal sign, to $(1-\epsilon)^{\alpha-1} = (0.9)^{-1}$ for $\alpha = 0.5$.

References

BASU, A., BASU, S. and CHAUDHURI, G. (1997). Robust minimum divergence procedures for count data models. *Sankhyā Ser. B*, **59**, 11-27.

BASU, A. and LINDSAY, B.G. (1994). Minimum disparity estimation for continuous models. *Ann. Inst. Statist. Math.* **46**, 683-705.

BASU, A. and SARKAR, S. (1994a). The trade-off between efficiency and robustness and the effect of model smoothing in minimum disparity inference. *J. Statist. Comput. Simul.* **50**, 173-185.

BASU, A. and SARKAR, S. (1994b). Minimum disparity estimation in the errors-in-variables model. *Statist. Probab. Lett.* **20**, 69-73.

BERAN (1977). Minimum Hellinger distance estimates for parametric models. *Ann. Statist.* **5**, 445-463.

BILLINGSLEY (1986). *Probability and Measure.* Wiley, New York.

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

ESLINGER, P.W. and WOODWARD, W.A. (1991). Minimum Hellinger distance estimation for normal models. *J. Statist. Comput. Simul.* **39**, 95-114.

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

MARKATOU, M., BASU, A. and LINDSAY, B.G. (1998). Weighted likelihood equations with bootstrap root search. *J. Amer. Statist. Assoc.*, **93**, 740-750.

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 tests; efficiency, breakdown points and examples. *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, IL 61820, U.S.A.*

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

AYANENDRANATH BASU
APPLIED STATISTICS UNIT
INDIAN STATISTICAL INSTITUTE
203, BARRACKPORE TRUNK ROAD, KOLKATA-700 108
E-mail: ayanbasu@isical.ac.in