

MINIMUM DISPARITY ESTIMATION: ASYMPTOTIC NORMALITY  
AND BREAKDOWN POINT RESULTS

by

Chanseok PARK and Ayanendranath BASU

---

*Reprinted from the Bulletin of Informatics and Cybernetics  
Research Association of Statistical Sciences, Vol. 36*

—••••—  
FUKUOKA, JAPAN  
2004

# MINIMUM DISPARITY ESTIMATION: ASYMPTOTIC NORMALITY AND BREAKDOWN POINT RESULTS

By

Chanseok PARK\* and Ayanendranath BASU†

## Abstract

The minimum disparity estimators for discrete models introduced by Lindsay (1994, *Ann. Statist.*, **22**, 1081-1114) are useful and practical data analytic tools. But for continuous models, further machinery such as kernel density estimation is needed and general results concerning the efficiency and breakdown results are unavailable. In this paper we establish the asymptotic normality for a subclass of minimum disparity estimators which satisfy certain properties, as well as lay out general results for breakdown point analysis.

*Key Words and Phrases:* Minimum disparity, efficiency, breakdown point, residual adjustment function.

## 1. Introduction

Consider the general setting of inference in a parametric class of distributions  $\mathcal{F}_\Theta = \{F_\theta, \theta \in \Theta\}$ ,  $\Theta \subseteq \mathbb{R}^p$ . Let  $\mathcal{G}$  be the class of all distributions having probability density functions (pdf's) with respect to a dominating measure (e.g., Lebesgue measure). We will assume that the model  $\mathcal{F}_\Theta$  is a subclass of  $\mathcal{G}$ , which also contains the true data generating distribution  $G$ .

In parametric estimation one's goal is to estimate  $\theta$  efficiently when the model is correct and robustly in case the true distribution is in the neighborhood of the model but not necessarily in it. Several authors, Beran (1977), Tamura and Boos (1986), and Simpson (1987, 1989) have demonstrated that the above goal is achievable by using the minimum Hellinger distance estimator. Lindsay (1994) generalized the earlier work based on Hellinger distance (HD) to a general class of disparities generating estimators that are both robust and first order efficient for discrete models. A disparity is a measure of discrepancy between a nonparametric density estimator obtained from data and the model density. In the discrete case one can simply use the empirical density function  $d(\cdot)$  as this nonparametric density estimate where  $d(x)$  is defined to be the relative frequency of the value  $x$  in the sample. For continuous models one requires a nonparametric density estimator such as the kernel density estimator from the data, since the model is now continuous while the data by itself is discrete. Although kernel density estimation is

\* Department of Mathematical Sciences, Clemson University, Clemson, SC 29634, U.S.A. cspark@ces.clemson.edu

† Applied Statistics Unit, Indian Statistical Institute, Calcutta 700 108, India. ayanbasu@isical.ac.in  
This work was done during a visit of the second author to the Department of Statistics, Pennsylvania State University, University Park, PA 16802, U.S.A.

just used as an intermediate tool, it brings in all the associated problems of bandwidth selection and rates of convergence. Unlike the discrete case, there is no general framework providing the asymptotic results for minimum disparity estimation in the continuous case although certain specific cases have been dealt with such as the Hellinger distance (Beran 1977; Tamura and Boos 1986) or the generalized negative exponential disparity (Basu *et al.* 1997; Bhandari *et al.* 2000).

Here we provide a general setup which applies to a large subclass of disparities and under which the appropriate minimum disparity estimators are first order efficient at the model. We also present general conditions under which the breakdown point results of the robust minimum disparity estimators may be established and 50% breakdown at the model is attained.

Basu and Lindsay (1994) proposed a different approach to minimum disparity estimation in continuous models, where they smoothed the model density with the same kernel as well. This approach leads to a general framework of asymptotic inference for minimum disparity estimators but requires the selection of kernels relative to the model for asymptotic efficiency which may not always be possible.

In this paper, we focus on the mathematical properties of minimum disparity estimators such as asymptotic normality and breakdown point results. Numerical studies involving large scale simulations and real data examples will be undertaken in a future paper. However the performance of several of the disparities proposed here have already been studied in some detail. See Bhandari *et al.* (2000), Markatou *et al.* (1998), and Park and Basu (2000).

The rest of the paper is organized as follows. We introduce and describe minimum disparity estimation in Section 2. The consistency and the asymptotic normality of the minimum disparity estimators are shown in Section 3. The breakdown point analysis is given in Section 4. Section 5 presents several disparities in this class.

## 2. Minimum disparity estimation

For the parametric setup of Section 1, let  $X_1, X_2, \dots, X_n$  be a random sample from the distribution  $G$ , and let

$$g_n(x) = \frac{1}{nh_n} \sum_{i=1}^n w\left(\frac{x - X_i}{h_n}\right) \quad (1)$$

define a nonparametric density estimator of  $g$  (the density function of the distribution  $G$ ), where  $w$  is a smooth family of *kernel* functions with bandwidth  $h_n$ . For discrete models, we will, without loss of generality, let the sample space be  $\{0, 1, 2, \dots\}$  and take  $g_n$  to be the empirical density function, where  $g_n(x)$  is the relative frequency of the value  $x$  in the sample. Define the *Pearson residual* at a point  $x$  as

$$\delta(x) = \frac{g_n(x) - f_\theta(x)}{f_\theta(x)},$$

where  $f_\theta$  is the density function corresponding to  $F_\theta$ . Lindsay (1994) constructed the *disparity*  $\rho_C(\cdot)$  defined as

$$\rho_C(g_n, f_\theta) = \int C(\delta) f_\theta, \quad (2)$$

where  $C(\cdot)$  be a real-valued, thrice differentiable, strictly convex function on  $[-1, \infty)$  and  $C(\delta) \geq 0$  with the equality holding only when  $\delta = 0$ . Here the integration is with respect to the dominating measure. Notice that the above implies  $C'(\delta) < 0$  for  $\delta < 0$ , and  $C'(\delta) > 0$  for  $\delta > 0$ . Under the assumptions the disparity  $\rho_C(\cdot)$  is nonnegative and equals zero if and only if  $g_n \equiv f_\theta$ .

Examples of disparities include the likelihood disparity (LD) and the twice squared Hellinger distance, defined by  $\text{LD}(g, f_\theta) = \int \{g \log(g/f_\theta) - (f_\theta - g)\}$  and  $\text{HD}(g, f_\theta) = \int (g^{1/2} - f_\theta^{1/2})^2$  respectively. The LD is a version of the Kullback-Leibler divergence, and in the discrete case it is minimized by the maximum likelihood estimator (MLE) of  $\theta$ .

Let  $\nabla$  represent the gradient with respect to  $\theta$ . For any real valued function  $a(x)$  we will let  $a'(x)$  and  $a''(x)$  denote its first and second derivatives with respect to  $x$ . In general terms, the minimization of the disparity  $\rho_C(g_n, f_\theta)$  over  $\theta \in \Theta$  gives the minimum disparity estimator (MDE) corresponding to the function  $C$ . In particular, when the disparity  $\rho_C(\cdot, \cdot)$  is the Hellinger distance, this minimization produces the minimum Hellinger distance estimator (MHDE). Under differentiability of the model, the minimum disparity estimating equation becomes

$$-\nabla \rho_C = \int A(\delta) \nabla f_\theta = 0, \quad (3)$$

where  $A(\delta) \equiv (\delta + 1)C'(\delta) - C(\delta)$ . The function  $A(\delta)$  is increasing on  $[-1, \infty)$ , and without affecting the estimating properties of the disparity  $\rho_C$ , it can be redefined to satisfy  $A(0) = 0$  and  $A'(0) = 1$ . This standardized function  $A(\delta)$  is called the residual adjustment function (RAF) of the disparity. Notice that the estimating equations are otherwise equivalent so that the form of the function  $A(\delta)$  determines the specific properties of the estimator such as how strongly the effects of large probabilistic outliers (which manifest themselves as large positive values of  $\delta$ ) are downweighted. The RAF for the LD and HD are given respectively by  $A(\delta) = \delta$  and  $A(\delta) = 2[(\delta + 1)^{1/2} - 1]$ . The curvature parameter  $A_2 = A''(0)$  is a measure of how fast the function curves away from the line  $A(\delta) = \delta$  at  $\delta = 0$  (Lindsay 1994). Large negative values of  $A_2$  provide greater downweighting effect for large outliers relative to maximum likelihood estimation, while  $A_2 = 0$  indicates a form of second order efficiency of the estimator in the sense of Rao (1961,1962).

The minimum disparity estimator (MDE) is defined as

$$\hat{\theta}_n = \arg \inf_{\theta \in \Theta} \rho_C(g_n, f_\theta).$$

Define the MD estimation functional  $T : \mathcal{G} \rightarrow \Theta$  as

$$T(G) \equiv \arg \inf_{\theta \in \Theta} \rho_C(g, f_\theta),$$

provided such a minimum exists. In case  $T(G)$  is multiple-valued the notation  $T(G)$  will represent one of the possible values chosen arbitrarily. Since  $T(F_\theta) = \theta$ ,  $T(\cdot)$  is Fisher consistent. By the definition above, the MDE is  $T(G_n)$ , where  $G_n$  is the cumulative distribution function (CDF) of the kernel density estimator  $g_n$  based on the data.

For the rest of the paper we assume:

- (a)  $C(\cdot)$  is strictly convex and thrice differentiable,

(b)  $A(\delta)$  satisfies  $A(0) = 0$  and  $A'(0) = 1$ .

Notice that  $A(\cdot)$  is an increasing function and  $A(\cdot)$  is strictly increasing when  $C(\cdot)$  is strictly convex. We now establish two mathematical results useful for our subsequent theoretical development.

LEMMA 2.1. *Let  $D(g, f) = C(g/f - 1)f$ , so that  $\rho_C(g, f) = \int C(g/f - 1)f = \int D(g, f)$ . Then under the assumptions (a) and (b), we have the results:*

$$(i) D(g, f) \leq D(0, f)\mathcal{I}(g \leq f) + D(g, 0)\mathcal{I}(f < g) \leq D(0, f) + D(g, 0),$$

$$(ii) D(g, f) \leq C(-1)f + C'(\infty)g,$$

where  $\mathcal{I}(\cdot)$  is the indicator function.

PROOF. First, for  $g \in [0, f]$  with fixed  $f$ , look at  $D(g, f)$  as a function of  $g$ .

$$\frac{\partial}{\partial g} D(g, f) = C'\left(\frac{g}{f} - 1\right) < 0, \quad \forall g \in (0, f)$$

since  $C(\cdot) < 0$  for  $\delta < 0$ . Hence  $D(g, f) \leq D(0, f)$  for  $g \in [0, f]$ . Note that  $D(g, f) \leq C(-1)f$  for  $\forall g \in [0, f]$ .

Next, for  $f \in (0, g)$  with fixed  $g$ , look at  $D(g, f)$  as a function of  $f$ .

$$\frac{\partial}{\partial f} D(g, f) = -C'\left(\frac{g}{f} - 1\right)\frac{g}{f} + C\left(\frac{g}{f} - 1\right) = -A\left(\frac{g}{f} - 1\right) \leq 0, \quad \forall f \in (0, g)$$

since  $A(\delta)$  is an increasing function with  $A(0) = 0$ . Hence  $D(g, f) \leq D(g, 0)$  for  $f \in [0, g]$ . This proves part (i). Note that

$$D(g, 0) = \lim_{t \rightarrow 0} C(g/t - 1)t = C'(\infty)g,$$

so that part (ii) holds. □

LEMMA 2.2. *Suppose that  $C(-1)$  and  $C'(\infty)$  are finite. Then the disparity  $\rho_C(g, f)$  is bounded above by  $C(-1) + C'(\infty)$ .*

PROOF. This follows easily from Lemma 2.1 (ii). □

### 3. Consistency and asymptotic normality

This section is devoted to discussion of the efficiency of the minimum disparity estimators under certain general conditions which is satisfied by a fairly large class of disparities. The following set of results provides this framework.

THEOREM 3.1. *Assume that (a) the parameter space  $\Theta$  is compact, (b) for  $\theta_1 \neq \theta_2$ ,  $f_{\theta_1}(x) \neq f_{\theta_2}(x)$  on a set of positive dominating measure, (c)  $f_{\theta}(x)$  is continuous in  $\theta$  for almost every  $x$ , and (d)  $C(-1)$  and  $C'(\infty)$  are finite. Then (i) for any  $G \in \mathcal{G}$ , there exists  $\theta \in \Theta$  such that  $T(G) = \theta$ , and (ii) for any  $F_{\theta^*} \in \mathcal{F}_{\Theta}$ ,  $T(F_{\theta^*}) = \theta^*$  is unique.*

PROOF. (i) Existence. Denote  $D(g, f_\theta) = C(g/f_\theta - 1)f_\theta$ . Let  $\{\theta_n : \theta_n \in \Theta\}$  be a sequence such that  $\theta_n \rightarrow \theta$  as  $n \rightarrow \infty$ . Since  $D(g, f_{\theta_n}) \rightarrow D(g, f_\theta)$  by assumption (c) and  $\int D(g, f_{\theta_n})$  is finite by Lemma 2.2 with assumption (d), we have

$$\rho_C(g, f_{\theta_n}) = \int D(g, f_{\theta_n}) \rightarrow \int D(g, f_\theta) = \rho_C(g, f_\theta),$$

by a generalized version of the dominated convergence theorem (Royden 1988, p. 92). Hence  $\rho_C(g, f_t)$  is continuous in  $t$  and achieves an infimum for  $t \in \Theta$  since  $\Theta$  is compact.

(ii) Uniqueness. This is immediate from the identifiability assumption on the parameterization and the unique minimum of  $C(\cdot)$ .  $\square$

REMARK. As in Beran (1977), the above result assumes a compact parameter space  $\Theta$ , but also applies when  $\Theta$  is not compact but can be embedded within a compact set  $\bar{\Theta}$ , and the disparity  $\rho_C(g, \cdot)$  can be extended to a continuous function of  $\theta$  on  $\bar{\Theta}$ . For example, in the location-scale family

$$\left\{ f_\theta(x) = \frac{1}{\sigma} f\left(\frac{x - \mu}{\sigma}\right) : \theta = (\mu, \sigma) \in (-\infty, \infty) \times (0, \infty), f \text{ continuous} \right\},$$

where the parameter space is not compact,  $(\mu, \sigma)$  can be reparameterized as  $\beta = (\beta_1, \beta_2)$ ,  $\mu = \tan(\beta_1)$ ,  $\sigma = \tan(\beta_2)$ , and  $(\beta_1, \beta_2) \in \bar{\Theta} = (-\pi/2, \pi/2) \times (0, \pi/2)$ . It can be easily checked that the disparity can be extended as a continuous function on  $\bar{\Theta} = [-\pi/2, \pi/2] \times [0, \pi/2]$  which is compact, while the minimum must occur in  $\Theta$ . Hence the conclusions of Theorem 3.1 remain valid in this case.

THEOREM 3.2. Let  $G \in \mathcal{G}$  be the true distribution with pdf  $g$ . Given a random sample  $X_1, X_2, \dots, X_n$ , let the kernel density estimate  $g_n$  (with CDF  $G_n$ ) be as defined in (1). Suppose that  $|C'(\cdot)|$  is bounded on  $[-1, \infty]$ . If  $T(G)$  is unique, then under the assumptions of Theorem 3.1 the functional  $T(G_n)$  is continuous at  $T(G)$  in the sense that  $T(G_n)$  converges to  $T(G)$  as  $g_n \rightarrow g$  in  $L_1$ .

PROOF. Suppose that  $g_n \rightarrow g$  in  $L_1$ . For convenience, denote  $\varrho(t) = \rho_C(g, f_t)$  and  $\varrho_n(t) = \rho_C(g_n, f_t)$ . It is easily seen that

$$|\varrho_n(t) - \varrho(t)| \leq \int |C(\delta_n) - C(\delta)| f_t,$$

where  $\delta_n = g_n/f_t - 1$  and  $\delta = g/f_t - 1$ . By the mean value theorem, there exists  $\delta^*$  satisfying

$$C(\delta_n) - C(\delta) = C'(\delta^*)(\delta_n - \delta),$$

where  $\delta^*$  lies between  $\delta_n$  and  $\delta$ . Denote  $K = \max_\delta |C'(\delta)|$ . Then we have

$$|\varrho_n(t) - \varrho(t)| \leq K \int |g_n - g| \quad \text{for all } t \in \Theta.$$

Hence we have

$$\sup_t |\varrho_n(t) - \varrho(t)| \rightarrow 0, \tag{4}$$

as  $g_n \rightarrow g$  in  $L_1$ . Denote  $\theta_n = \arg \inf_t \varrho_n(t)$  and  $\theta = \arg \inf_t \varrho(t)$ . If  $\varrho(\theta) \geq \varrho_n(\theta_n)$ , then  $\varrho(\theta) - \varrho_n(\theta_n) \leq \varrho(\theta_n) - \varrho_n(\theta_n)$ , and if  $\varrho_n(\theta_n) \geq \varrho(\theta)$ , then  $\varrho_n(\theta_n) - \varrho(\theta) \leq \varrho_n(\theta) - \varrho(\theta)$ . Therefore we have

$$|\varrho_n(\theta_n) - \varrho(\theta)| \leq |\varrho_n(\theta_n) - \varrho(\theta_n)| + |\varrho_n(\theta) - \varrho(\theta)| \leq 2 \sup_t |\varrho_n(t) - \varrho(t)|,$$

which implies  $\varrho_n(\theta_n) \rightarrow \varrho(\theta)$  as  $g_n \rightarrow g$  in  $L_1$ . Using this and (4), we obtain

$$\lim_{n \rightarrow \infty} \varrho(\theta_n) = \varrho(\theta). \quad (5)$$

Then we have only to show that  $\theta_n \rightarrow \theta$ . If  $\theta_n \not\rightarrow \theta$ , compactness of  $\Theta$  ensures existence of a subsequence  $\{\theta_m\} \subset \{\theta_n\}$  such that  $\theta_m \rightarrow \theta^* \neq \theta$ , implying  $\varrho(\theta_m) \rightarrow \varrho(\theta^*)$  by continuity of  $\varrho(\cdot)$  (see Theorem 3.1(i)). By (5),  $\varrho(\theta^*) = \varrho(\theta)$ , which contradicts the uniqueness assumption of  $\theta = T(G)$ .  $\square$

**COROLLARY 3.3.** *If the conditions of Theorem 3.2 hold and  $g = f_\theta$ , then  $\theta_n = T(G_n) \rightarrow \theta$  and  $\rho_C(f_{\theta_n}, f_\theta) \rightarrow 0$ .*

**PROOF.** It is easily seen from Theorem 3.2 that  $\theta_n = T(G_n) \rightarrow \theta$  since  $T(F_\theta) = \theta$ . Since  $\theta_n \rightarrow \theta$ , we have  $f_{\theta_n} \rightarrow f_\theta$  and hence  $\int |f_{\theta_n} - f_\theta| \rightarrow 0$  by Glick's theorem (Devroye and Györfi 1985, p. 10). Since  $\rho_C(f_{\theta_n}, f_\theta) = \int C(\delta_n) f_\theta$  with  $\delta_n = f_{\theta_n}/f_\theta - 1$  and  $C(0) = 0$ , there exists  $\delta^*$  such that  $C(\delta_n) = C'(\delta^*)\delta_n$  by the mean value theorem and it follows that

$$\rho_C(f_{\theta_n}, f_\theta) \leq K \int |f_{\theta_n} - f_\theta|,$$

where  $K = \max_\delta |C'(\delta)|$ . This completes the proof.  $\square$

**REMARK.** Suppose the parameter space  $\Theta$  is not compact, and it is difficult to find a compact embedding  $\bar{\Theta}$  such that the disparity extends continuously to all the points of  $\bar{\Theta}$ . In such a situation, a result of Simpson (1987, Theorem 1) may be applicable to show that the existence and continuity results above still remain valid for the class  $\mathcal{G}^*$  of distributions, where any  $G \in \mathcal{G}^*$  satisfies

$$\inf_{t \in \Theta \setminus \Theta^*} \rho_C(g, f_t) > \rho_C(g, f_{\theta^*})$$

for some compact  $\Theta^* \subset \Theta$  and  $\theta^* \in \Theta^*$ . Formally, suppose that  $f_\theta(x)$  is continuous in  $\theta$  for each  $x$ . Then, for each  $G \in \mathcal{G}^*$ , (i)  $T(G)$  exists, and (ii) if  $T(G)$  is unique, then the condition  $g_n \rightarrow g$  in  $L_1$  implies that  $T(G_n) \rightarrow T(G)$  as  $n \rightarrow \infty$ .

In particular, suppose that conditions (b) and (c) of Theorem 3.1 hold and the parametric family satisfies

$$\inf_{t \in \Theta \setminus \Theta^*} \rho_C(f_\theta, f_t) > 0$$

for some compact  $\Theta^* \subset \Theta$ . Then  $g_n \rightarrow f_\theta$  in  $L_1$  implies that  $T(G_n) \rightarrow \theta$  as  $n \rightarrow \infty$ .

Next we introduce some notation. Let  $u_\theta(x) = \nabla \log f_\theta(x)$  be the  $p$ -dimensional vector of likelihood score function, and let  $\nabla^2 f_\theta(x)$  denote the  $p \times p$  matrix of second partial derivatives of  $f_\theta$ . We will denote by  $u_\theta^i(x)$  and  $\nabla^{ij} f_\theta(x)$  the  $i$ -th element of  $u_\theta(x)$  and  $(i, j)$ -th element of  $\nabla^2 f_\theta(x)$  respectively. Then

$$I(\theta) \equiv \int u_\theta u_\theta^T f_\theta$$

is the Fisher information matrix, where the superscript  $T$  denotes transpose. For deriving asymptotic normality of the MDE we will assume that  $f_\theta(x)$  is twice continuously differentiable with respect to  $\theta$ , and  $\rho_C(\cdot, f_\theta)$  can be twice differentiated with respect to  $\theta$  under the integral sign. For the latter one, sufficient conditions are: for any  $\theta \in \Theta$ , there exists  $\epsilon > 0$ , and functions  $K_\theta^i(x)$ ,  $L_\theta^{ij}(x)$  and  $M_\theta^{ij}(x)$ ,  $i, j = 1, \dots, p$ , such that for  $\theta^\dagger$  satisfying  $\|\theta - \theta^\dagger\| < \epsilon$  ( $\|\cdot\|$  being the Euclidean norm),

$$(I) |u_{\theta^\dagger}^i f_{\theta^\dagger}(x)| < K_\theta^i(x), \quad \int K_\theta^i < \infty, \quad i = 1, 2, \dots, p;$$

$$(II) |\nabla^{ij} f_{\theta^\dagger}(x)| < L_\theta^{ij}(x), \quad \int L_\theta^{ij} < \infty, \quad i, j = 1, \dots, p;$$

$$(III) |u_{\theta^\dagger}^i(x) u_{\theta^\dagger}^j(x) f_{\theta^\dagger}(x)| < M_\theta^{ij}(x), \quad \int M_\theta^{ij} < \infty, \quad i, j = 1, \dots, p.$$

**THEOREM 3.4.** *Suppose  $G = F_{\theta_0}$ ,  $g = f_{\theta_0}$ ,  $G_n$  and  $g_n$  are as defined in Theorem 3.2. Let  $\{\phi_n\}$  denote any sequence of estimators such that  $\phi_n = \theta_0 + o_p(1)$ ,  $\{\alpha_n\}$  any sequence of positive real numbers going to  $\infty$ , and  $\mathcal{I}_B$  the indicator function for the set  $B$  of real numbers. In addition to the conditions (I) – (III) above, assume the following:*

$$(a) \int |\nabla^{ij} f_{\phi_n} - \nabla^{ij} f_{\theta_0}| = o_p(1) \text{ and } \int |u_{\phi_n}^i u_{\phi_n}^j f_{\phi_n} - u_{\theta_0}^i u_{\theta_0}^j f_{\theta_0}| = o_p(1), \quad i, j = 1, \dots, p, \text{ for all } \{\phi_n\} \text{ defined above.}$$

$$(b) \text{ The matrix } I(\theta_0) \text{ is finite (element wise), and } \int u_{\theta_0}^i(x+a) u_{\theta_0}^j(x+a) f_{\theta_0}(x) dx - \int u_{\theta_0}^i(x) u_{\theta_0}^j(x) f_{\theta_0}(x) dx \rightarrow 0 \text{ as } |a| \rightarrow 0, \quad i, j = 1, \dots, p.$$

$$(c) \text{ The kernel } w(\cdot) \text{ is a density which is symmetric about } 0, \text{ square integrable, and twice continuously differentiable with compact support } S. \text{ The bandwidth } h_n \text{ satisfies } h_n \rightarrow 0, \quad n^{1/2} h_n \rightarrow \infty, \text{ and } n^{1/2} h_n^4 \rightarrow 0.$$

$$(d) \limsup_{n \rightarrow \infty} \sup_{y \in \mathcal{A}_n} \int |\nabla^{ij} f_{\theta_0}(x+y) u_{\theta_0}(x)| dx < \infty \text{ for } i, j = 1, \dots, p, \text{ where } \mathcal{A}_n = \{y : y = h_n z, \quad z \in S\}.$$

$$(e) n \sup_{t \in S} P(|X_1 - h_n t| > \alpha_n) \rightarrow 0 \text{ as } n \rightarrow \infty, \text{ for all } \{\alpha_n\} \text{ defined above.}$$

$$(f) (n^{1/2} h_n)^{-1} \int |u_{\theta_0} \mathcal{I}_{|x| \leq \alpha_n}| \rightarrow 0, \text{ for all } \{\alpha_n\} \text{ defined above.}$$

$$(g) \sup_{|x| \leq \alpha_n} \sup_{t \in S} \{f_{\theta_0}(x + h_n t) / f_{\theta_0}(x)\} = O(1), \text{ for all } \{\alpha_n\} \text{ defined above.}$$

$$(h) A(\delta), A'(\delta), A'(\delta)(\delta + 1) \text{ and } A''(\delta)(\delta + 1) \text{ are bounded on } [-1, \infty].$$

Then,  $\sqrt{n}(T(G_n) - T(G))$  converges in distribution to  $N(0, I^{-1}(\theta_0))$ , where  $\theta_0 = T(G)$ .

**PROOF.** By condition (c), we have  $\hat{g}_n(x) \xrightarrow{a.s.} f_{\theta_0}(x)$  for every  $x$  and

$$\int |g_n(x) - f_{\theta_0}(x)| dx \rightarrow 0,$$

and hence  $T(G_n) \xrightarrow{P} \theta$ . For convenience, denote  $\varrho_n(\theta) = \rho_C(g_n, f_\theta)$  and  $\hat{\theta} = T(G_n)$ . Since  $\hat{\theta}$  minimizes  $\varrho_n(\cdot)$  over  $\Theta$ , the Taylor series expansion of  $\nabla \varrho_n(\hat{\theta})$  at  $\theta_0$  yields

$$0 = \sqrt{n} \nabla \varrho_n(\hat{\theta}) = \sqrt{n} \nabla \varrho_n(\theta_0) + \sqrt{n} \nabla^2 \varrho_n(\theta^\dagger) (\hat{\theta} - \theta_0),$$



where  $\theta^\dagger$  is a point on the line segment  $L(\theta_0, \hat{\theta})$ . The line segment is defined as the curve joining  $a$  and  $b$ , that is, the continuous function  $f(t) = tb + (1-t)a$  on  $t \in [0, 1]$ . It follows that

$$\sqrt{n}(\hat{\theta} - \theta_0) = -\nabla^2 \varrho_n(\theta^\dagger)^{-1} \sqrt{n} \nabla \varrho_n(\theta_0).$$

Now it suffices to prove that

$$\nabla^2 \varrho_n(\theta^\dagger) \xrightarrow{\mathcal{P}} I(\theta_0) \quad (6)$$

and

$$-\sqrt{n} \nabla \varrho_n(\theta_0) \xrightarrow{\mathcal{D}} N(0, I(\theta_0)). \quad (7)$$

First we prove (6). For convenience, denote  $\delta_n(\theta) = g_n/f_\theta - 1$ . Differentiating with respect to  $\theta$ , we have

$$\nabla^2 \varrho_n(\theta) = - \int A(\delta_n) \nabla^2 f_\theta + \int A'(\delta_n) (\delta_n + 1) u_\theta u_\theta^T f_\theta$$

Let us denote  $B_1 = \sup_\delta |A(\delta)|$  and  $B_2 = \sup_\delta |A'(\delta)(\delta + 1)|$ .  $B_1$  and  $B_2$  are finite from condition (h). It follows that by condition (a)

$$\left| \int A(\delta_n(\theta^\dagger)) (\nabla^2 f_{\theta^\dagger} - \nabla^2 f_{\theta_0}) \right| \leq B_1 \int |\nabla^2 f_{\theta^\dagger} - \nabla^2 f_{\theta_0}| \xrightarrow{\mathcal{P}} 0, \quad (8)$$

and

$$\left| \int A'(\delta_n(\theta^\dagger)) \{ \delta_n(\theta^\dagger) + 1 \} (u_{\theta^\dagger} u_{\theta^\dagger}^T f_{\theta^\dagger} - u_{\theta_0} u_{\theta_0}^T f_{\theta_0}) \right| \leq B_2 \int |u_{\theta^\dagger} u_{\theta^\dagger}^T f_{\theta^\dagger} - u_{\theta_0} u_{\theta_0}^T f_{\theta_0}| \xrightarrow{\mathcal{P}} 0. \quad (9)$$

Since  $A(0) = 0$  and  $\delta_n(\theta^\dagger) \rightarrow 0$  as  $n \rightarrow \infty$ , using the dominated convergence theorem we have

$$\int A(\delta_n(\theta^\dagger)) \nabla^2 f_{\theta_0} \xrightarrow{\mathcal{P}} 0,$$

and hence by (8)

$$\int A(\delta_n(\theta^\dagger)) \nabla^2 f_{\theta^\dagger} \xrightarrow{\mathcal{P}} 0.$$

Using condition (b),  $A'(0) = 1$  and  $\delta_n(\theta^\dagger) \rightarrow 0$  as  $n \rightarrow \infty$ , we have

$$\int A'(\delta_n(\theta^\dagger)) \{ \delta_n(\theta^\dagger) + 1 \} u_{\theta_0} u_{\theta_0}^T f_{\theta_0} \xrightarrow{\mathcal{P}} \int u_{\theta_0} u_{\theta_0}^T f_{\theta_0},$$

by the dominated convergence theorem and hence by (9)

$$\int A'(\delta_n(\theta^\dagger)) \{ \delta_n(\theta^\dagger) + 1 \} u_{\theta^\dagger} u_{\theta^\dagger}^T f_{\theta^\dagger} \xrightarrow{\mathcal{P}} \int u_{\theta_0} u_{\theta_0}^T f_{\theta_0}.$$

Therefore we have (6).

Next we prove (7). Note that

$$-\sqrt{n} \nabla \varrho_n(\theta_0) = \sqrt{n} \int A(\delta_n(\theta_0)) \nabla f_{\theta_0},$$

and from the equations (3.12) and (3.13) of Beran (1977),

$$\sqrt{n} \int \delta_n(\theta_0) \nabla f_{\theta_0} = \sqrt{n} \int (g_n - f_{\theta_0}) u_{\theta_0} \xrightarrow{\mathcal{D}} N(0, I(\theta_0)).$$

Therefore, it is enough to prove that

$$\left| \sqrt{n} \int \{A(\delta_n(\theta_0)) - \delta_n(\theta_0)\} \nabla f_{\theta_0} \right| \xrightarrow{\mathcal{P}} 0. \quad (10)$$

From condition (h)  $A'(\delta)$  and  $A''(\delta)(\delta + 1)$  are bounded. Thus using Lemma 25 of Lindsay (1994), we have a finite  $B$  such that  $|A(r^2 - 1) - (r^2 - 1)| \leq B \times (r - 1)^2$  for all  $r \geq 0$ . Evidently we have

$$\begin{aligned} |A(\delta_n(\theta_0)) - \delta_n(\theta_0)| &= |A(\{(g_n/f_{\theta_0})^{1/2}\}^2 - 1) - [\{(g_n/f_{\theta_0})^{1/2}\}^2 - 1]| \\ &\leq B \{(g_n/f_{\theta_0})^{1/2} - 1\}^2. \end{aligned}$$

Using this and (10), we have

$$\left| \sqrt{n} \int \{A(\delta_n(\theta_0)) - \delta_n(\theta_0)\} \nabla f_{\theta_0} \right| \leq B \sqrt{n} \int \{g_n^{1/2} - f_{\theta_0}^{1/2}\}^2 |u_{\theta_0}|.$$

Now we consider

$$\sqrt{n} \int \{g_n^{1/2} - f_{\theta_0}^{1/2}\}^2 |u_{\theta_0}|.$$

From the inequality  $(a - b)^2 \leq (a - c)^2 + (b - c)^2$  for real numbers  $a, b, c$ , we get Thus using this, we have

$$\sqrt{n} \int \{g_n^{1/2} - f_{\theta_0}^{1/2}\}^2 |u_{\theta_0}| \leq T_1 + T_2,$$

where  $T_1 = \sqrt{n} \int \{g_n^{1/2} - (Eg_n(X))^{1/2}\}^2 |u_{\theta_0}|$  and  $T_2 = \sqrt{n} \int \{f_{\theta_0}^{1/2} - (Eg_n(X))^{1/2}\}^2 |u_{\theta_0}|$ . The term  $T_1$  represents the Hellinger deviation of the estimator  $G_n$  from its mean and its convergence to zero in probability has been established by Tamura and Boos (1986) using conditions (e), (f) and (g). The term  $T_2$  represents the bias in the Hellinger metric and its convergence to zero in probability follows from conditions (c) and (d). This completes the proof.  $\square$

REMARK. (Comments on the conditions in Theorem 3.4): Condition (a) is a simple continuity condition at the model. The second derivative of  $f_{\theta}$  with respect to  $\theta$  is  $L_1$ -continuous and  $E_{\theta}[u_{\theta}^i u_{\theta}^j]$  is continuous in  $\theta$ . This condition is satisfied by the distributions of the exponential family. Condition (b) is satisfied if  $u_{\theta_0}^i(x)$  is uniformly continuous in  $x$  on compact sets. Condition (c) represent the key restriction on bandwidth which may be chosen at the optimal rate  $h_n \sim n^{-1/5}$ . Conditions (e) – (g) have been used and discussed by Tamura and Boos (1986, Theorem 4.1). The condition (h) provides us convenient boundedness conditions leading to easy manipulation of the terms in the normality proof. Notice that in (c) we require  $n^{1/2}h_n^4 \rightarrow 0$  (not  $n^{1/2}h_n^2 \rightarrow 0$ ). This is compatible with the optimal rate  $h_n \sim n^{-1/5}$ . Beran's (1977) proof of Theorem 4 actually uses (in the notation of Beran)  $n^{1/2}c_n^4 \rightarrow 0$ , although his condition (v) mentions  $n^{1/2}c_n^2 \rightarrow 0$ . Similarly the correct condition (c) of Theorem 1 in Basu, *et al.* (1997) should be  $n^{1/2}h_n^4 \rightarrow 0$  as  $n \rightarrow \infty$ .

#### 4. Breakdown point analysis

The breakdown point of a statistical functional is roughly the smallest fraction of contamination in the data that may cause an arbitrarily extreme value in the estimate. Here we establish the breakdown point of the minimum disparity functional  $T(G)$  under general conditions. Consider the contamination model,

$$H_{\epsilon,n} = (1 - \epsilon)G + \epsilon K_n,$$

where  $\{K_n\}$  is a sequence of contaminating distributions. Let  $h_{\epsilon,n}$ ,  $g$  and  $k_n$  be the corresponding densities. Following Simpson (1987), we say there is *breakdown* in  $T$  for  $\epsilon$  level contamination if there exists a sequence  $K_n$  such that  $|T(H_{\epsilon,n}) - T(G)| \rightarrow \infty$  as  $n \rightarrow \infty$ . We write below  $\theta_n = T(H_{\epsilon,n})$ . We assume the following assumptions for the breakdown point analysis. The assumptions reflect the intuitively worst possible choice of the contamination, and the expected behavior or the functional when breakdown *does* and *does not* occur respectively. Assumption **A4** denotes the specific assumption on the disparity.

ASSUMPTION. The contaminating sequence  $\{k_n\}$ , the truth  $g(x)$  and the model  $f_\theta(x)$  satisfy the following:

- A1**  $\int \min\{g(x), k_n(x)\} \rightarrow 0$  as  $n \rightarrow \infty$ . That is, the contamination distribution becomes asymptotically singular to the true distribution.
- A2**  $\int \min\{f_\theta(x), k_n(x)\} \rightarrow 0$  as  $n \rightarrow \infty$  uniformly for  $|\theta| \leq c$ , for any fixed  $c$ . That is the contamination distribution is asymptotically singular to the specified models.
- A3**  $\int \min\{g(x), f_{\theta_n}(x)\} \rightarrow 0$  as  $n \rightarrow \infty$  if  $|\theta_n| \rightarrow \infty$  as  $n \rightarrow \infty$ . That is, large values of the parameter  $\theta$  give distributions which become singular to the true distribution.
- A4**  $C(-1)$  and  $C'(\infty)$  are finite.

THEOREM 4.1. *Under the assumptions **A1** – **A4** above, the asymptotic breakdown point  $\epsilon^*$  of the MDE is at least  $\frac{1}{2}$  at the model.*

PROOF. Let  $\theta_n$  be the minimizer of  $\rho_C(h_{\epsilon,n}, f_\theta)$ . Given a level of contamination  $\epsilon$ , let, if possible, breakdown occurs, that is there exists a sequence  $\{K_n\}$  such that  $|\theta_n| \rightarrow \infty$  where  $\theta_n = T(H_{\epsilon,n})$ . Then

$$\rho_C(h_{\epsilon,n}, f_{\theta_n}) = \int_{A_n} D(h_{\epsilon,n}(x), f_{\theta_n}(x)) + \int_{A_n^c} D(h_{\epsilon,n}(x), f_{\theta_n}(x)),$$

where  $A_n = \{x : g(x) > \max(k_n(x), f_{\theta_n}(x))\}$  and  $D(g, f) = C(g/f - 1) f$ . From **A1**,  $\int_{A_n} k_n(x) \rightarrow 0$ , and from **A3**,  $\int_{A_n} f_{\theta_n}(x) \rightarrow 0$  as  $n \rightarrow \infty$ , so under  $k_n(\cdot)$  and  $f_{\theta_n}(\cdot)$ , the set  $A_n$  converges to a set of zero probability. Thus on  $A_n$ ,  $D(h_{\epsilon,n}(x), f_{\theta_n}(x)) \rightarrow D((1 - \epsilon)g(x), 0)$  as  $n \rightarrow \infty$ , so

$$\left| \int_{A_n} D(h_{\epsilon,n}(x), f_{\theta_n}(x)) - \int_{A_n} D((1 - \epsilon)g(x), 0) \right| \rightarrow 0 \quad (11)$$

by dominated convergence theorem, and by **A1** and **A3** we have

$$\left| \int_{A_n} D((1 - \epsilon)g(x), 0) - \int_{g>0} D((1 - \epsilon)g(x), 0) \right| \rightarrow 0. \quad (12)$$

Using (11) and (12), we have

$$\left| \int_{A_n} D(h_{\epsilon,n}(x), f_{\theta_n}(x)) - \int_{g>0} D((1-\epsilon)g(x), 0) \right| \rightarrow 0$$

Notice that for  $g > 0$

$$D((1-\epsilon)g(x), 0) = \lim_{f \rightarrow 0} D((1-\epsilon)g(x), f) = (1-\epsilon)C'(\infty)g(x)$$

and

$$\int_{g>0} D((1-\epsilon)g(x), 0) = \int D((1-\epsilon)g(x), 0) = (1-\epsilon)C'(\infty).$$

Thus we have

$$\left| \int_{A_n} D(h_{\epsilon,n}(x), f_{\theta_n}(x)) - \int D((1-\epsilon)g(x), 0) \right| \rightarrow 0,$$

and hence

$$\int_{A_n} D(h_{\epsilon,n}(x), f_{\theta_n}(x)) \rightarrow (1-\epsilon)C'(\infty).$$

From **A1** and **A3**,  $\int_{A_n^c} g(x) \rightarrow 0$  as  $n \rightarrow \infty$ , so under  $g(\cdot)$ , the set  $A_n^c$  converges to a set of zero probability. By similar arguments, we get

$$\left| \int_{A_n^c} D(h_{\epsilon,n}(x), f_{\theta_n}(x)) - \int D(\epsilon k_n(x), f_{\theta_n}(x)) \right| \rightarrow 0.$$

Notice that  $\int D(\epsilon k_n(x), f_{\theta_n}(x)) \geq C(\epsilon - 1)$  by Jensen's inequality. It follows that

$$\liminf_{n \rightarrow \infty} \rho_C(h_{\epsilon,n}, f_{\theta_n}) \geq C(\epsilon - 1) + (1-\epsilon)C'(\infty) = a_1(\epsilon),$$

say. We will have a contradiction to our assumption that  $\{k_n\}$  is a sequence for which breakdown occurs if we can show that there exists a constant value  $\theta^*$  in the parameter space such that for the same sequence  $\{k_n\}$ ,

$$\limsup_{n \rightarrow \infty} \rho_C(h_{\epsilon,n}, f_{\theta^*}) < a_1(\epsilon) \quad (13)$$

as then the  $\{\theta_n\}$  sequence above could not minimize  $\rho_C(h_{\epsilon,n}, f_{\theta^*})$  for every  $n$ .

We will now show that equation (13) is true for all  $\epsilon < 1/2$  under the model when we choose  $\theta^*$  to be the minimizer of  $\int D((1-\epsilon)g(x), f_{\theta}(x))$ . For any fixed  $\theta$ , let  $B_n = \{x : k_n(x) > \max(g(x), f_{\theta}(x))\}$ . From **A1**,  $\int_{B_n} g(x) \rightarrow 0$  and from **A2**,  $\int_{B_n} f_{\theta}(x) \rightarrow 0$  as  $n \rightarrow \infty$ . Similarly from **A1** and **A2**,  $\int_{B_n^c} k_n(x) \rightarrow 0$  as  $n \rightarrow \infty$ . Thus under  $k_n(\cdot)$ , the set  $B_n^c$  converges to a set of zero probability, while under  $g(\cdot)$  and  $f_{\theta}(\cdot)$ , the set  $B_n$  converges to a set of zero probability. Thus on  $B_n$ ,  $D(h_{\epsilon,n}(x), f_{\theta}(x)) \rightarrow D(\epsilon k_n(x), 0)$  as  $n \rightarrow \infty$  and

$$\left| \int_{B_n} D(h_{\epsilon,n}(x), f_{\theta}(x)) - \int_{k_n>0} D(\epsilon k_n(x), 0) \right| \rightarrow 0$$

by dominated convergence theorem. Notice that for  $k_n > 0$

$$D(\epsilon k_n(x), 0) = \lim_{f \rightarrow 0} D(\epsilon k_n(x), f) = \epsilon C'(\infty) k_n(x)$$

and

$$\int_{k_n > 0} D(\epsilon k_n(x), 0) = \int D(\epsilon k_n(x), 0) = \epsilon C'(\infty).$$

Similarly we have

$$\left| \int_{B_n^c} D(h_{\epsilon, n}(x), f_\theta(x)) - \int D((1 - \epsilon)g(x), f_\theta(x)) \right| \rightarrow 0.$$

Hence we have

$$\begin{aligned} \lim_{n \rightarrow \infty} \rho_C(h_{\epsilon, n}, f_\theta) &= \epsilon C'(\infty) + \int D((1 - \epsilon)g(x), f_\theta(x)) \\ &\geq \epsilon C'(\infty) + \inf_{\theta} \int D((1 - \epsilon)g(x), f_\theta(x)). \end{aligned} \quad (14)$$

with equality for  $\theta = \theta^*$ . Let  $a_2(\epsilon) = \epsilon C'(\infty) + \int D((1 - \epsilon)g(x), f_{\theta^*}(x))$ . Notice from (14) that among all fixed  $\theta$  the disparity  $\rho_C(h_{\epsilon, n}, f_\theta)$  is minimized in the limit by  $\theta^*$ .

If  $g(\cdot) = f_{\theta_t}(\cdot)$ , that is the true distribution belongs to the model,  $\int D((1 - \epsilon)f_{\theta_t}(x), f_{\theta_t}(x)) = C(-\epsilon)$  which is also the lower bound (over  $\theta \in \Theta$ ) for  $\int D((1 - \epsilon)f_{\theta_t}(x), f_\theta(x))$ . Thus in this case  $\theta^* = \theta_t$ , and from (14),

$$\lim_{n \rightarrow \infty} \rho_C(h_{\epsilon, n}, f_\theta) = C(-\epsilon) + \epsilon C'(\infty). \quad (15)$$

As a result asymptotically there is no breakdown for  $\epsilon$  level contamination when  $a_3(\epsilon) < a_1(\epsilon)$ , where  $a_3(\epsilon)$  is the right hand sides of equation (15). Note that  $a_1(\epsilon)$  and  $a_3(\epsilon)$  are strictly decreasing and increasing respectively in  $\epsilon$ , and  $a_1(1/2) = a_3(1/2)$ , so that asymptotically there is no breakdown and  $\limsup_{n \rightarrow \infty} |T(H_{\epsilon, n})| < \infty$  for  $\epsilon < 1/2$ .  $\square$

## 5. Disparities in this class

The different conditions on the disparities that we have used in the different steps of the efficiency proof may now be consolidated together as follows:

- (a)  $C(\cdot)$  is strictly convex and thrice differentiable,
- (b)  $C(\delta) \geq 0$ , with equality only at  $\delta = 0$  and  $C''(0) = 1$ ,  
(notice that the above implies  $A(0) = C'(0) = 0$  and  $A'(0) = C''(0) = 1$ ).
- (c)  $C'(\delta)$ ,  $A(\delta)$ ,  $A'(\delta)$ ,  $A'(\delta)(\delta + 1)$  and  $A''(\delta)(\delta + 1)$  are bounded on  $\delta \in [-1, \infty]$ .

While these conditions may appear quite restrictive, there is a rich class of disparities satisfying the above. We list some of them below. Each family below is indexed by a single parameter  $\alpha$ .

$$\rho_1(g, f) = \frac{1}{2} \int \frac{(g - f)^2}{\alpha g + \bar{\alpha} f}, \quad \alpha \in (0, 1)$$

$$\rho_2(g, f) = \int \frac{(g - f)^2}{\sqrt{\alpha g^2 + \bar{\alpha} f^2 + \alpha g + \bar{\alpha} f}}, \quad \alpha \in (0, 1)$$

$$\begin{aligned}
&= \frac{1}{\alpha\bar{\alpha}} \int \left[ \sqrt{\alpha g^2 + \bar{\alpha} f^2} - \alpha g - \bar{\alpha} f \right], \\
\rho_3(g, f) &= \frac{1}{2} \int \frac{(g-f)^2}{\sqrt{\alpha g^2 + \bar{\alpha} f^2}}, \quad \alpha \in (0, 1) \\
\rho_4(g, f) &= \int \left[ \frac{f}{\alpha^2} \{ \exp(\alpha - \alpha g/f) - 1 \} + \frac{g-f}{\alpha} \right], \quad \alpha > 0 \\
\rho_5(g, f) &= \int \left[ \frac{4}{\alpha\pi} g \tan \left( \frac{\pi\alpha g-f}{2 g+f} \right) - (g-f) \right], \quad \alpha \in [0, 1) \\
\rho_6(g, f) &= \int \left[ \frac{2}{\alpha\pi} (g-f) \sin \left( \frac{\pi\alpha g-f}{2 g+f} \right) \right], \quad \alpha \in [0, 1] \\
\rho_7(g, f) &= \int \left[ \frac{2}{\alpha\pi} (g-f) \tan \left( \frac{\pi\alpha g-f}{2 g+f} \right) \right], \quad \alpha \in [0, 1)
\end{aligned}$$

where  $\bar{\alpha} = 1 - \alpha$ .

Table 1: The  $C(\cdot)$  and the  $A(\cdot)$  functions corresponding to  $\rho_i$  disparity. Here  $\bar{\alpha} = 1 - \alpha$ .

Disparity	$C(\delta)$	$A(\delta)$
$\rho_1$	$\frac{\delta^2}{2(\alpha\delta+1)}$	$\frac{\delta}{\alpha\delta+1} + \frac{\alpha}{2} \left[ \frac{\delta}{\alpha\delta+1} \right]^2$
$\rho_2$	$\frac{\delta^2}{1+\alpha\delta+\sqrt{\alpha(\delta+1)^2+\bar{\alpha}}}$	$\frac{1}{\alpha} - \frac{1}{\alpha\sqrt{\alpha(\delta+1)^2+\bar{\alpha}}}$
$\rho_3$	$\frac{\delta^2}{2\sqrt{\alpha(\delta+1)^2+\bar{\alpha}}}$	$\frac{\delta}{\sqrt{\alpha(\delta+1)^2+\bar{\alpha}}} + \frac{\alpha}{2} \left[ \frac{\delta}{\sqrt{\alpha(\delta+1)^2+\bar{\alpha}}} \right]^3$
$\rho_4$	$\frac{e^{-\alpha\delta}-1+\alpha\delta}{\alpha^2}$	$\frac{(\alpha+1)-[(\alpha+1)+\alpha\delta]e^{-\alpha\delta}}{\alpha^2}$
$\rho_5$	$\frac{4}{\alpha\pi}(\delta+1) \tan \left( \frac{\alpha\pi}{2} \frac{\delta}{\delta+2} \right) - \delta$	$4 \left( \frac{\delta+1}{\delta+2} \right)^2 \sec^2 \left( \frac{\alpha\pi}{2} \frac{\delta}{\delta+2} \right) - 1$
$\rho_6$	$\frac{2}{\alpha\pi} \delta \sin \left( \frac{\alpha\pi}{2} \frac{\delta}{\delta+2} \right)$	$\frac{2\delta(\delta+1)}{(\delta+2)^2} \cos \left( \frac{\alpha\pi}{2} \frac{\delta}{\delta+2} \right) + \frac{2}{\alpha\pi} \sin \left( \frac{\alpha\pi}{2} \frac{\delta}{\delta+2} \right)$
$\rho_7$	$\frac{2}{\alpha\pi} \delta \tan \left( \frac{\alpha\pi}{2} \frac{\delta}{\delta+2} \right)$	$\frac{2\delta(\delta+1)}{(\delta+2)^2} \sec^2 \left( \frac{\alpha\pi}{2} \frac{\delta}{\delta+2} \right) + \frac{2}{\alpha\pi} \tan \left( \frac{\alpha\pi}{2} \frac{\delta}{\delta+2} \right)$

The family  $\rho_1$  is the blended weight chi-square (BWCS) family (Lindsay 1994). The member of this family for  $\alpha = 1/2$  is the symmetric chi-square (SCS) utilized by Markatou *et al.* (1998) for robust inference. The families  $\rho_2$  and  $\rho_3$  are two other variants of  $\rho_1$  with similar properties. The family  $\rho_4$  is the generalized negative exponential disparity (Bhandari *et al.* 2000) which includes the negative exponential disparity of Lindsay (1994) for  $\alpha = 1$ . The families  $\rho_5$ ,  $\rho_6$  and  $\rho_7$  are based on trigonometric functions and also satisfy the properties. Note that each of these three families contain the SCS family as limiting cases of  $\alpha \rightarrow 0$ . The families  $\rho_5$  and  $\rho_6$  have been proposed

by Park and Basu (2000). In case of both  $\rho_6$  and  $\rho_7$ , although these disparities satisfy conditions (a), (b) and (c), more theoretical and empirical investigations are necessary to get a better idea about the performance of the corresponding estimators. Notice that condition (c) above also implies  $C'(\infty)$  and  $C(-1)$  are finite, the conditions on the disparity that are necessary for establishing the breakdown results.

It would of course be of great interest if one were able to pick out a method as the optimal one from the multitude of methods discussed in this paper. That, however, seems to be an extremely challenging task. Any idea of optimality in choosing a method from any of the classes presented here will have to balance the two conflicting ideas of robustness and small sample efficiency. It may be that the methods on one end of the scale (say, larger values of  $\alpha$ ) will perform well in terms of efficiency, while those in the other end of the scale (say, smaller values of  $\alpha$ ) will perform better in terms of robustness (or vice versa), so that selecting a single optimum will be practically impossible. But it may be possible to provide a band of values of  $\alpha$  for which the resulting methods will represent reasonable combinations of efficiency and robustness. This, however, will require extensive numerical studies, and may be quite specific to the chosen parametric model. At any rate, a selection scheme – data adaptive or otherwise – of choosing a single method as the most suitable one from the the totality of methods described here will certainly require further research.

Another point worth mentioning is that at the current state of knowledge it appears to be very difficult to avoid the nonparametric density estimation inherent in the minimum disparity approach presented here. The only way this may be avoided is if the unknown true density  $g$  shows up as a linear term in the integrand of the disparity, in which case  $dG$  may be approximated with the empirical distribution function. Any other component of the disparity which involves  $g$  through something other than the linear term must be independent of the unknown parameter  $\theta$ . Unfortunately this is the case only for the likelihood disparity which may be written as (ignoring the part of the integrand which integrates identically to zero)

$$\text{LD}(g, f_\theta) = \int g \log(g/f_\theta) = \int (\log g - \log f_\theta) dG.$$

Here the right hand side is the difference of the term  $\int \log g dG$  (which does not involve  $\theta$  and hence does not figure in the minimization) and the term  $\int g \log f_\theta = \int \log f_\theta dG$  (which involves  $g$  only linearly and hence may be approximated by  $\int \log f_\theta dG_n = \frac{1}{n} \sum_{i=1}^n \log f_\theta(X_i)$ ), and thus the minimization procedure can be carried out without any nonparametric density estimation (here  $G_n$  is the empirical distribution function). This, however, is not the case for the Hellinger distance  $\text{HD}(g, f_\theta) = \int (g^{1/2} - f_\theta^{1/2})^2$  or any other robust disparity.

Some related density-based minimum divergence estimators have been discussed in Basu *et al.* (1998) where the proposed estimators have bounded influence functions and the construction of the estimating equations does not require any nonparametric density estimation. However, these divergences are not in the general class of disparities, and they *do* sacrifice full asymptotic efficiency to achieve their robustness. Thus, as of now, nonparametric density estimation seems to be a necessary part of robust and fully efficient parametric estimation.

### References

- Basu, A. and Lindsay, B. G. (1994). Minimum disparity estimation for continuous models: Efficiency, distributions and robustness. *Annals of the Institute of Statistical Mathematics*, **46**, 683–705.
- Basu, A., Sarkar, S., and Vidyashankar, A. N. (1997). Minimum negative exponential disparity estimation in parametric models. *Journal of Statistical Planning and Inference*, **58**, 349–370.
- Basu, A., Harris, I. R., Hjort, N. L., and Jones, M. C. (1998). Robust and efficient estimation by minimising a density power divergence. *Biometrika*, **85**, 549–559.
- Beran, R. J. (1977). Minimum Hellinger distance estimates for parametric models. *Annals of Statistics*, **5**, 445–463.
- Bhandari, S., Basu, A., and Sarkar, S. (2000). Robust estimation in parametric models using the family of generalized negative exponential disparities. Technical Report #2000/7, Stat Math Unit, Indian Statistical Institute, Calcutta 700 035, India.
- Devroye, L. and Györfi, L. (1985). *Nonparametric Density Estimation: The  $L_1$  View*. Wiley.
- Lindsay, B. G. (1994). Efficiency versus robustness: The case for minimum Hellinger distance and related methods. *Annals of Statistics*, **22**, 1081–1114.
- Markatou, M., Basu, A., and Lindsay, B. G. (1998). Weighted likelihood equations with bootstrap root search. *Journal of the American Statistical Association*, **93**, 740–750.
- Park, C. and Basu, A. (2000). Minimum disparity inference based on trigonometric disparities. Technical Report No. ASD/2000/23, Applied Stat. Div., Indian Statistical Institute.
- Rao, C. R. (1961). Asymptotic efficiency and limiting information. In *Proc. Fourth Berkeley Symp.*, volume 1, pages 531–546, Berkeley. University of California Press.
- Rao, C. R. (1962). Efficient estimates and optimum inference procedures in large samples (with discussion). *Journal of the Royal Statistical Society B*, **24**, 46–72.
- Royden, H. L. (1988). *Real Analysis*. Macmillan, New York.
- Simpson, D. G. (1987). Minimum Hellinger distance estimation for the analysis of count data. *Journal of the American Statistical Association*, **82**, 802–807.
- Simpson, D. G. (1989). Hellinger deviance test: efficiency, breakdown points, and examples. *Journal of the American Statistical Association*, **84**, 107–113.
- Tamura, R. N. and Boos, D. D. (1986). Minimum Hellinger distance estimation for multivariate location and covariance. *Journal of the American Statistical Association*, **81**, 223–229.

Received September 29, 2003

Revised June 11, 2004