Journal of the Korean Statistical Society (2001), **30**: *1, pp 41–61*

Negative Exponential Disparity Based Deviance and Goodness-of-fit Tests for Continuous Models: Distributions, Efficiency and Robustness

Dong-Bin Jeong[1] and Sahadeb Sarkar[2]

Abstract

The minimum negative exponential disparity estimator (MNEDE), introduced by Lindsay (1994), is an excellent competitor to the minimum Hellinger distance estimator (Beran 1977) as a robust and yet efficient alternative to the maximum likelihood estimator in parametric models. In this paper we define the negative exponential deviance test (NEDT) as an analog of the likelihood ratio test (LRT), and show that the NEDT is asymptotically equivalent to the LRT at the model and under a sequence of contiguous alternatives. We establish that the asymptotic strong breakdown point for a class of minimum disparity estimators, containing the MNEDE, is at least 1/2 in continuous models. This result leads us to anticipate robustness of the NEDT under data contamination, and we demonstrate it empirically. In fact, in the simulation settings considered here the empirical level of the NEDT show more stability than the Hellinger deviance test (Simpson 1989). The NEDT is illustrated through an example data set. We also define a goodness-of-fit statistic to assess adequacy of a specified parametric model, and establish its asymptotic normality under the null hypothesis.

Keywords: Disparity based tests; Hellinger Distance; Likelihood Ratio Test; Minimum Disparity Estimation; Outliers; Residual Adjustment Function.

1. INTRODUCTION

The likelihood ratio tests used routinely for testing in parametric problems have certain asymptotic optimality properties but are not, in general, robust when data contain outliers. For a useful analysis of such data, a careful screening for anomalous observations is a must before applying the likelihood ratio tests.

[1]Department of Statistics, Kangnung National University, Kangnung, 210-702, South Korea.
[2]O.M. Group, Indian Institute of Management Calcutta, Post Box 16757,Calcutta 700 027, India.

Simpson (1989) proposed a more direct procedure in the form of the Hellinger deviance tests. These tests are robust under data contamination and asymptotically equivalent to the likelihood ratio tests under local parametric alternatives. Before Simpson's (1989) work robust versions of Wald (1943) tests and the Rao (1948) tests were studied by many authors in various settings (Beran 1981; Hampel, Ronchetti, Rousseeuw and Stahel 1986).

The Hellinger deviance test is based on the minimum Hellinger distance estimator (MHDE). The MHDE is first-order efficient and yet has certain robustness properties (Beran 1977; Tamura and Boos 1986; Simpson 1987; Donoho and Liu 1988). Robustness of the M-estimation based procedures usually comes at the cost of first order efficiency (Hampel et al. 1986). For the discrete models, Lindsay (1994) introduced a class of density based distances called disparities and defined minimum disparity estimators that are robust and efficient at the model. Basu and Lindsay (1994) studied them under continuous models. The class of disparities includes the Hellinger distance (HD), and the negative exponential disparity (NED). That the minimum negative exponential estimator (MNEDE) is an excellent competitor to the MHDE under the continuous models has been shown by Basu, Sarkar and Vidyashankar (1997). Also see Basu and Sarkar (1994a). Both MHDE and MNEDE are robust against outliers, but only the MNEDE is resistant against inliers, defined as data points with less observed frequencies than expected under the model. Moreover, the MNEDE is second order efficient (Rao 1961) at the model, whereas the MHDE is only first order efficient.

In this paper we study the NED based deviance test, and also a goodness-of-fit test to judge adequacy of a chosen parametric model. However, our main focus is on the deviance test. We present several results on the asymptotic distributions of these tests. A strong breakdown point result for a class of minimum disparity estimators, including the MNEDE, is also given. Theoretical and empirical results show that, like the MNEDE to the MHDE, the negative exponential deviance test (NEDT) is an excellent competitor to the Hellinger deviance test (Simpson 1989) as a direct method for robust inferences.

The remainder of this paper is organized as follows: Section 2 briefly describes the MNEDE and its properties. In Section 3 we define the NEDT and also a goodness-of-fit statistic. Section 4 establishes the asymptotic efficiency of the NEDT at the model and under local alternatives, and also derives the asymptotic normality of the NED based goodness-of-fit test under the null hypothesis. In Section 5 we extend Lindsay's (1994) result to continuous models that the asymptotic strong breakdown point for a class of minimum disparity estimators (including the MNEDE) is 1/2 or larger. Section 6 presents some simulation results indicating that the NEDT is efficient at the model and robust under data contamination. Moreover, in the settings considered here, the level of the NEDT appears to be more stable than the Hellinger deviance test. An application to an example data set is given in Section 7. Section 8 has a concluding discussion. Finally, the Appendix contains proofs of results in Sections 4 and 5.

2. MINIMUM NEGATIVE EXPONENTIAL DISPARITY ESTIMATION

Suppose that we have a random sample $(X_1, X_2, \cdots, X_n)$ from a parametric class of distributions $\Im_\Theta = \{F_\theta, \theta \in \Theta\}$, where Θ is a subset of $\Re^p$. Assume that the distributions F_θ are continuous and have probability density functions f_θ. The density based minimum disparity estimates (Lindsay 1994; Basu and Lindsay 1994) are computed by minimizing a nonnegative measure of discrepancy ρ_G, between a nonparametric density estimate and the model density $\hat{f}_n$, defined by

$$\rho_G(\hat{f}_n, f_\theta) \equiv \int G(\delta(\hat{f}_n, \theta, x))dF_\theta(x) \tag{2.1}$$

where G is a real-valued three times differentiable, strictly convex function on $[-1, \infty)$ with $G(0) = 0$, and $\delta(x) = \delta(\hat{f}_n, \Theta, x) = (\hat{f}_n(x) - f_\theta(x))/f_\theta(x)$ denotes the "Pearson" residual at the value x. One can use a nonparametric kernel density estimator defined by

$$\hat{f}_n(x) \equiv \int w(x; y, h_n)dF_n(y),$$

whereF_n is the empirical distribution function, w is a smooth family of kernel functions like the normal densities with mean y and standard deviation h_n. Note that (2.1) is the same as Csiszar's (1963) f-divergence between $\hat{f}_n$ and f_θ, which is a form of generalization of Kullback-Leibler divergence and was independently introduced by Ali and Silvey (1966).

The function $G(\delta) = (\delta + 1)\log(\delta + 1) - \delta$ generates the likelihood disparity whose minimizer in count data models gives the MLE. The Pearson's chi-square is produced by $G(\delta) = \delta^2$, the power divergence family (Cressie and Read 1984) by $G(\delta) = [(\delta + 1)^{\lambda+1} - 1]/\lambda(\lambda + 1)$, the two times squared HD by $G(\delta) = 2[(\delta + 1)^{1/2} - 1]^2$, and the NED by $G(\delta) = [e^{-\delta} - 2]$. Lindsay (1994) discussed many other important disparities.

A value of θ minimizing (2.1) is taken as the minimum disparity estimator.

In particular, MNEDE is obtained by minimizing

$$\rho_{NED}(\hat{f}_n, f_\theta) = \int \{\exp[-(\frac{\hat{f}_n(x)}{f_\theta(x)} - 1)] - 2\} dF_\theta(x),$$

which is the NED between $\hat{f}_n(x)$ and f_θ. Let ∇ represent the gradient with respect to θ. Under differentiability of the model, the minimum disparity estimating equation takes the form

$$-\nabla\rho_G = \int A(\delta(x))\nabla F_\theta(x) = 0, \quad where \;\; A(\delta) \equiv (\delta+1)[(G'(\delta)] - G(\delta)$$

and $G'(\delta)$ denotes the first derivative of $G(\delta)$. The function $A(\delta)$ is an increasing function on $[-1, \infty)$ and can be standardized, without changing the estimates produced by the disparity, so that for the standardized $A(\delta)$ we have $A(0) = 0$ and $A'(0) = 1$, where $A'(\delta)$ denotes the first derivative of $A(\delta)$. This standardized function is called the residual adjustment function of the disparity and determines most of the theoretical properties of the estimators. For the likelihood disparity the residual adjustment function is $A(\delta) = \delta$, for the squared HD it is $A(\delta) = 2[(\delta+1)^{1/2} - 1]$, and for the NED it is $A(\delta) = 2 - (2+\delta)e^{-\delta}$. For more discussion on how HD and NED react to outliers and inliers in data, see Lindsay (1994), Basu and Lindsay (1994) and Basu et al. (1997). While estimators generated by the HD and NED are both first order efficient, only the MNEDE is also second order efficient. Basu et al. (1997) established efficiency and robustness properties of the MNEDE in continuous models. Using the MNEDE we define the NED based deviance test and a goodness-of-fit test in the next section.

3. THE DEVIANCE AND GOODNESS-OF-FIT TESTS

First we discuss the deviance test. Let Θ_0 be a proper subset of Θ and consider the problem of testing the null hypothesis $H_0 : \theta \in \Theta_0$ against the alternative hypothesis $H_a : \theta \in \Theta \setminus \Theta_0$. The log likelihood ratio statistic is given by $\Lambda = 2n[L_n(\tilde{\theta}) - L_n(\tilde{\theta}_0)]$, where $L_n(\theta) = n^{-1}\sum_{i=1}^{n} log(f_\theta(X_i))$ is the average log likelihood function, and $\tilde{\theta}$ and $\tilde{\theta}_0$ are points obtained by maximizing $L_n(\theta)$ over Θ and Θ_0, respectively. In general, for a disparity ρ_G one can define the disparity test statistic

$$d_G = -2n[\rho_G(\hat{f}_n, f_{\hat{\theta}}) - \rho_G(\hat{f}_n, f_{\hat{\theta}_0})]$$

where $\hat{\theta}$ and $\hat{\theta}_0$ minimize $\rho_G(\hat{f}_n, f_{\hat{\theta}})$ over Θ and Θ_0, respectively. Let $\hat{\theta}_{HD}$ and $\hat{\theta}_{0,HD}$ denote the corresponding minimum disparity estimators for the HD, and

$\hat{\theta}_{NED}$ and $\hat{\theta}_{0,NED}$ for the NED. The likelihood ratio test (LRT) Λ has the property that if Θ_0 is a q-dimensional subset of Θ and Θ_1 is an r-dimensional subset of Θ_0, then the test of Θ_1 against $\Theta \setminus \Theta_1$ can be partitioned into a test of Θ_1 versus $\Theta_0 \setminus \Theta_1$ and a test of Θ_0 versus $\Theta \setminus \Theta_0$. This property of Λ is shared by the deviance tests d_G. If $G(\delta) = 2[(\delta+1)^{1/2} - 1]^2$, the d_G becomes Simpson's (1989) Hellinger deviance test (HDT), defined by $d_{HD} = -2n[\rho_{HD}(\hat{f}_n, f_{\hat{\theta}_{HD}}) - \rho_{HD}(\hat{f}_n, f_{\hat{\theta}_{0,HD}})]$. On the other hand, if $G(\delta) = (e^{-\delta} - 2)$ we have the negative exponential deviance test (NEDT) statistic

$$d_{NED} = -2n[\rho_{NED}(\hat{f}_n, f_{\hat{\theta}_{NED}}) - \rho_{NED}(\hat{f}_n, f_{\hat{\theta}_{0,NED}})].$$

For discrete models, Sarkar, Song and Jeong (1998) considered the NEDT and gave empirical results under the Poisson and geometric cases.

We now construct NED based goodness-of-fit tests for families of distributions in testing the null hypothesis H_0^* that the random sample $(X_1, X_2, \ldots, X_n)$ is generated by a member of $\Im_\Theta$ against the alternative hypothesis H_a^* that H_0^* is not true. For continuous models, Beran (1977, Section 5) discussed the above testing problem using the Hellinger distance. In order to achieve comparable asymptotic properties of the disparities based estimators and test statistics the disparities need to be suitably standardized. Thus, to produce comparable test statistics based on the HD and NED, we work with two times squared HD produced by $G(\delta) = 2[(\delta+1)^{1/2} - 1]^2$, and the NED- form generated by $G(\delta) = [e^{-\delta} - 2]$. Using this standardization and NED in place of HD in the definition of the test statistic of Beran (1977, Theorem 8) we define the test statistic

$$d^*_{NED}(\hat{\theta}_{NED}) = [\{nh_n\rho_{NED}(\hat{f}_n, f_{\hat{\theta}_{NED}})\} - (2^{-1}R_n\|w\|^2)]/[h_n^{1/2}(2^{-1}R_n\|w*w\|^2)^{1/2} \quad (3.1)$$

where R_n is the range of $X_i^{'}$s, w is the kernel function used in computing $\hat{f}_n$, $\|\cdot\|$ denotes the L_2-norm and $w * w$ denotes the convolution of w and w.

For the count data models, Lindsay (1994) discussed the test d_G when the null hypothesis is simple. For discrete multivariate data, Basu and Sarkar (1994b) and Jeong and Sarkar (2000), among others, have studied goodness-of-fit tests based on disparities for both simple and composite hypotheses.

4. ASYMPTOTIC DISTRIBUTIONS

Simpson (1989) showed that the HDT is asymptotically equivalent to the LRT under local parametric alternatives to $H_0 : \theta \in \Theta_0$ and hence enjoys first order efficiency within the parametric model $\Im_\Theta$. This property is shared by the NEDT. Assume that the following regularity conditions hold: $f_\theta(x)$ is twice continuously differentiable with respect to θ, and $\rho_{NED}(g, \theta)$ can be twice differentiable with respect to θ under the integral sign. Using the fact that $A_{NED}(\delta)$ and $A'_{NED}(\delta)(1+\delta)$ are bounded, Basu et al. (1997, p. 358) gave a set of sufficient conditions for the above. From the proof of Theorem 1 of Basu et al., it then follows that, as $||t|| \to 0$,

$$\rho_{NED}(f, f_{\theta+t}) = \rho_{NED}(f, f_\theta) + t^T \rho'_{NED}(f, f_\theta) + 2^{-1} t^T \rho''_{NED}(f, f_\theta) t + o(||t||^2),$$

uniformly in f, where ρ'_{NED} and ρ''_{NED} denote the first and second derivatives with respect to θ. An estimator θ^* is said to be first order efficient (Rao 1973, p. 32) if it satisfies

$$\theta^* = \theta + n^{-1/2} I(\theta)^{-1} Z_n(\theta) + o_p(n^{-1/2}) \tag{4.1}$$

under f_θ, where $I(\theta) = \int u(\theta, x) u(\theta, x)^T dF_\theta(x)$ with $u(\theta, x) = \partial \log f_\theta(x)/\partial\theta$ and $Z_n(\theta) = n^{-1/2} \sum_{i=1}^n u(\theta, X_i)$. From Basu et al. (1997) it follows that the MNEDE will satisfy (4.1) if the following approximation holds:

$$n^{1/2} \rho'_{NED}(\hat{f}_n, f_\theta) = Z_n(\theta) + o_p(1). \tag{4.2}$$

By the results of Basu et al. (1997) and Tamura and Boos (1986), (4.2) holds, for example, for the normal location-scale model. If the model has countable support and $\hat{f}_n$ is the empirical density, then (4.2) holds if $f_\theta^{1/2} u(\theta)$ is also in L_1 (see Lindsay 1994; Simpson 1987). Some common models satisfying the latter conditions are Poisson, geometric and log-series.

For a composite null hypothesis $H_0 : \theta \in \Theta_0$, assume that under H_0 the parameter θ can be written as $\theta = g(v)$ for $v \in N \subset \Re^q$ with $q < p$, and g has a continuous derivative $g'(v)$ of order $p \times q$ with rank q. It is also assumed that the constrained estimator under H_0 has the form $\theta_0^* = g(v^*)$ with v^* satisfying

$$v^* = v + n^{-1/2} \{g'(v) I(g(v)) g'(v)^T\}^{-1} g(v) Z_n(g(v))^T + o_p(n^{-1/2})$$

under $f_{g(v)}$. Then under the above assumptions a simple modification of Theorem 1 of Simpson (1989) gives the following result:

Theorem 1. *For a fixed $\theta_0 \in \Theta_0$ and $c \in \Re^p$ define $\theta_n = \theta_0 + cn^{1/2}$. Then under f_{θ_n} and as $n \to \infty$ we have $d_{NED} = \Lambda + o_p(1)$.*

Theorem 1 implies that, like the HDT, the NEDT has the same asymptotic power as the LRT under local parametric alternatives to H_0, and if dimension of Θ is p and dimension of Θ_0 is $q < p$ then the approximate null distribution of d_{NED} is $\chi^2(p-q)$.

We defined the statistic d^*_{NED} in (3.1) for testing H_0^* versus H_a^*. Following Beran (1977, Sec 5) we present Theorem 2 on its asymptotic null distribution, which may be used to carry out the test. The proof of and assumption (iv) used in Theorem 2 are discussed in the Appendix.

Theorem 2. *Assume the following: (i) w is symmetric about 0 and has a compact support Ψ and w is twice continuously differentiable; (ii) f_θ is supported and positive on a compact interval I; (iii) $nh_n^{1/2} \to 0$, $nh_n^3 \to \infty$ as $n \to \infty$; (iv) conditions (a)-(d), (g)-(i) of Theorem 1 of Basu et al. (1997) hold. Then, as $n \to \infty$, the limiting distribution of $d^*_{NED}(\hat{\theta})$, under f_θ, is $N(0,1)$.*

Choice of the bandwidth h_n in finite samples is very important. This problem has received considerable attention in the literature. It has been studied by Parzen (1962), Härdle et al (1988), Marron (1989), and Hall and Marron (1991) among others. Following Simpson (1989) we have adopted here the approach of Parzen. Parzen obtained the value of h_n that minimizes the integrated mean square error between a kernel density estimate and the true density f. More on this is discussed in Section 6 below.

5. ROBUSTNESS

Here we establish that the *asymptotic strong breakdown point*(Lindsay 1994) of a class of minimum disparity estimators, including the MNEDE, is 1/2 or larger. For count data models, Lindsay gave such a result. We extend his result to continuous models for a class of disparities, for which the function G has the properties that $G(-1) < \infty$ and $G(\delta)/\delta \to 0$ as $\delta \to \infty$. For example, this is true for the NED with $G(\delta) = e^{-\delta} - 2$. Such a function G will be a decreasing function.

Simpson (1989, Theorem 2) gave a breakdown point result for the HDT. We anticipate that a similar result holds true for the NEDT, on the basis of our breakdown point result for the MNEDE. But we do not have a proof at this point. However, our empirical findings support our conjecture on the robustness performance of the NEDT.

We consider a fixed model density f_θ and contamination level ϵ. For a given dataset $(X_1, X_2, \ldots, X_n)$, let $f(x)$ denote a density estimate which is bounded, i.e., $0 \le f(x) \le b \; \forall x$, for some b. If $f(x)$ is a kernel density estimate it will be bounded, for example, for normal family of kernels. Let $\{\xi_j : j = 1, 2, \ldots\}$ be a sequence of points on the support of f_θ. Define $f_j = (1-\epsilon)f + \epsilon k^{\xi_j}$ where k^{ξ_j} is a bounded density function defined on the interval $I^{\xi_j} = (\xi_j - \Delta/2, \xi_j + \Delta/2)$, for a small positive Δ. For example, k^{ξ_j} may be taken to be a uniform density. Also define $f_\epsilon^* = (1-\epsilon)f$, which is no longer a density function if $\epsilon > 0$, but we can still calculate $\rho_G(f_\epsilon^*, f_\theta)$. Note that f_ϵ^* represents the situation when the density estimate is computed by simply discarding the outliers in the sample. Let

$$\delta_j(x) = (f_j(x)/f_\theta(x) - 1), \delta_\epsilon^*(x) = (f_\epsilon^*(x)/f_\theta(x) - 1) \tag{5.1}$$

denote the Pearson residuals corresponding to $f_j(x)$ and $f_\epsilon^*(x)$. Now one can define an outlier sequence through the behavior of the Pearson residuals $\delta_j(x)$ for contaminated data and the density estimate $f(x)$ for uncontaminated data, over the interval I^{ξ_j}.

Definition 1. *The sequence $\{\xi_j\}$ is said to be an outlier sequence if* $\inf_{I^{\xi_j}} \delta_j(x) \to \infty$ *and* $\sup_{I^{\xi_j}} f(x) \to 0$ *as* $j \to \infty$.

In order for the minimum disparity estimation functional, denoted by $T_G(\cdot)$, to be stable under data contamination, we would like the corresponding residual adjustment function $A(\cdot)$ to be stable in the sense of Lindsay (1994, Definition 13), for which the following result gives a set of conditions.

Lemma 1. *If for some $k > 1$, $\int |u(\theta, x)|^k dF_\theta(x) < \infty$, $A(-1) < \infty$ and $A(\delta) = O(\delta^{(k-1)/k})$ as $\delta \to \infty$, then $A(\cdot)$ is outlier stable for the model f_θ.*

The proof is given in the Appendix. Lemma 1, for $k = 2$, implies that if the model f_θ has finite Fisher information, then a residual adjustment function satisfying $A(\delta) = O(\delta^{1/2})$ and $A(-1) < \infty$ is outlier stable. These conditions are satisfied, for example, for the NED and HD.

Next in Theorem 3 we give a breakdown point result for the minimum disparity estimation functional T_G in the following sense (Lindsay 1994, Definition 16):

Definition 2. *The strong breakdown point of $T_G(\cdot)$ at the density f is the supremum of the ϵ-values for which $T_G(f_i) \to T_G(f_\epsilon^*)$ as $j \to \infty$ for any outlier* *sequence* $\{\xi_j\}$.

Theorem 3. *If* $G(-1) < \infty$, $G(\delta)/\delta \to 0$ *as* $\delta \to \infty$, *then under some assumptions (given in the Appendix) on the behavior of* ρ_G, f *and* $\{f_\theta, \theta \in \Theta\}$, *the asymptotic strong breakdown point* ϵ^* *of* $T_G(\cdot)$ *at the model density* f_θ *is 1/2 or larger.*

The proof of Theorem 3 is discussed in the Appendix. Using Theorem 3, under conditions of the dominated convergence, one can derive as $j \to \infty$,

$$\rho_G(f_j, f_{T_G(f_j)}) \to \rho_G(f_\epsilon^*, f_{T_G(f_\epsilon^*)}). \tag{5.2}$$

The convergence in (5.2) implies that for extreme outliers and contamination fraction $\epsilon \leq \epsilon^*$, the distance from the contaminated density f_j to $f_{T_G(f_j)}$ is close to that obtained by simply deleting the outliers from the data. Observe that (5.2) holds, for instance, for a bounded disparity defining function $G(\cdot)$ on $[-1, \infty)$, as in case of the NED and disparities with logistic function based residual adjustment functions $A_\lambda(\delta) = 4\lambda^{-1}[e^{\lambda_t}/(1+e^{\lambda_t}) - 1/2]$. For NED, $G(-\delta) = (e^{-\delta} - 2)$, which is bounded by (e-2) on $[-1, \infty)$. Therefore (5.2) holds by the dominated convergence theorem since $\left|G((f_j/f_{T_G(f_j)}) - 1)f_{T_G(f_j)}\right| \leq (e-2)\left|f_{T_G(f_j)}\right|$ with $f_{T_G(f_j)}$ integrable, and, by Theorem 3, as $j \to \infty$ $G((f_j/f_{T_G(f_j)}) - 1)f_{T_G(f_j)} \to G((f_\epsilon^*/f_{T_G(f_\epsilon^*)}) - 1)f_{T_G(f_\epsilon^*)}$.

6. SIMULATIONS

We have performed a Monte Carlo study to empirically examine the theoretical results on the NEDT, HDT and LRT in the normal settings, including those used by Simpson (1989, Section 4). We have included the HDT in our simulations to make a comparative study. For our simulations we have used FORTRAN and various IMSL subroutines.

We have generated data from contaminated distributions $(1-\epsilon)N(\mu, \sigma^2) + \epsilon N(3, \sigma^2)$, where ϵ denotes the contamination level, and μ is the target parameter. We test H_0 that μ is zero in two cases. In the first case, we assume that σ is known, and in the second, we treat σ as unknown. For data generation, in both cases σ is set equal to one. We set $\mu = 0$ for examining the level performance of the tests, and $\mu = 0.5$ for the power. Results corresponding to $\epsilon = 0$ would indicate efficiency performance of the deviance tests. For checking robustness of the tests we have chosen $\epsilon = 0.025, 0.05$ and 0.10.

For the MNEDE and MHDE we computed the kernel density $\hat{f}_n$ using the

Epanechnikov kernel:$w(x) = 0.75(1 - x^2)$ for $|x| \leq 1$, and 0 otherwise. Parzen (1962) derived the value of h_n that minimizes the integrated mean square error between a kernel density estimate and the true density f. When f is the $N(\mu, \sigma^2)$ density, for the above kernel, h_n is of the form $h_n = (1.71877)(1.364\sigma)n^{-1/5}$. If σ is unknown, then in the above formula of h_n, the robust estimate $s_n = (median(|X_i - median(X_i)|)/0.6745$ can be used in place of σ. In our simulations, the numerical integrals were computed using Simpson's one-third rule, and the Newton-Raphson algorithm was used to solve for roots of the estimating equations. As initial estimates of μ and σ in the iterations we used $\hat{\mu}^{(0)} = median(X_i)$ and $\hat{\sigma}^{(0)} = s_n$. The results presented here are all based on 5000 replications for sample sizes 20, 30, 50 and 100, and for nominal levels 10%, 5% and 1%. For all the tests we used the $\chi^2(1)$ critical values. Empirical level and power of the tests have been presented in Tables $1-4$ for the known σ case, and in Tables $5-8$ for the unknown σ case.

We first discuss the results in Tables $1-4$. From the level values (corresponding to $\mu = 0$) in Table 1, we see that the chi-squared approximation seems to work pretty well for the NEDT. The loss in power (corresponding to $\mu = 0.5$, Table 1) for the NEDT and HDT compared to the LRT is small. The loss in power for the NEDT is generally higher than that for the HDT; but this may be expected since the level values of NEDT are generally lower. Next consider Tables $2-4$. The numbers in Table 2 show that when the contamination proportion is 0.025 the level values of the LRT get inflated, whereas those of the NEDT and HDT remain stable under this level of contamination. The empirical level of the NEDT shows more stability compared to that of the HDT. As sample size increases the level values increase for all the tests. The empirical power is higher for the HDT than the NEDT. Next, from Table 3 we see that as contamination level increases to 0.05, inflation in the level values of the LRT becomes even more pronounced. Again, both NEDT and HDT show their robustness in terms of level performance; in terms of level the NEDT appears to be more resistant against outliers than the HDT, and the HDT achieves higher power than the NEDT. Finally, results in Table 4 magnify the lack of robustness of the LRT, and show that even the NEDT breaks down for a contamination proportion of 0.10 under the considered settings. That the breakdown point of the HDT depends on the values of the underlying distribution parameters has been shown by Simpson (1989, p. 111, Example 1). We expect the same to be true for the NEDT as well.

We now discuss the results in Tables $5-8$. The findings are similar to what we have observed in Tables $1-4$. The level values of the tests under no contamination in Table 5 (for $\mu = 0$) are generally higher than their counterparts in Table 1. The level values of the LRT in Tables $6-8$ show slower rate of breakdown in this unknown variance case. On the other hand, the level values of the NEDT and HDT show that their resistance against outliers breaks down under smaller level of contamination, compared to the known variance case.

In terms of stability of empirical level under data contamination the NEDT appears to perform better than the HDT. On the other hand, in terms of loss of power under contamination the latter seems better. However, for sample sizes larger than 30, the loss of power for the NEDT becomes comparable while the difference between empirical levels of the NEDT and HDT gets increasingly prominent as the contamination level increases. Therefore, in the face of the possibility of contamination in data having more than thirty observations, the NEDT may be preferable to the HDT.

7. AN EXAMPLE

We consider a dataset on telephone-line faults analyzed previously by Welch (1987) and Simpson (1989). The data came from an experiment to test a method of reducing faults on telephone lines, and consisted of test and control fault rates in fourteen matched pairs of areas. Assuming the additivity of the treatment effect for the inverse fault rates Welch performed a matched-pairs test using the median. Table 4 of Simpson (1989) shows the ordered differences between the inverse test rates and the inverse control rates. Pair 1 difference (-988) in that table seems to be unusually large in magnitude. By modeling the data as a random sample from $N(\mu, \sigma^2)$ distribution with both μ, σ^2 unknown, Welch tested $H_0 : \mu = 0$ against $H_a : \mu > 0$ with or without pair 1 using his proposed randomized median test, and the LRT which is equivalent to the one sided matched-pairs t test. Welch performed the tests both for the full data (i.e., pair 1 present) and reduced data (i.e., pair 1 absent). He found that for the full and reduced data the p-values of LRT were 0.33 and 0.004 respectively, and for the randomized median test they were .06 and .035 respectively.

For the above one-sided alternative hypothesis a signed disparity test is appropriate and its asymptotic distribution is the same as the signed likelihood ratio test (see Simpson 1989, p.109). We have carried out the signed NEDT using the biweight kernel to compare our results with those of the signed HDT of Simpson (1989). The estimated bandwidth value for the full data comes out to be 202.16. The null and unconstrained MNEDEs of (μ, σ) are (0, 167.1) and (101.5, 120.1), respectively. On the other hand, Simpson's null and unconstrained MHDEs of (μ, σ) are (0, 169.5) and (116.8, 144.6), respectively. In Figure 1 we present the bi-weight kernel density estimate, along with the normal densities corresponding to the unconstrained MLE, and the unconstrained and null MNEDEs and MHDEs. In the figure the six densities are denoted by Kernel, ML, MNED, MNED(Null), MHD and MHD(Null), respectively. It shows that the unusual pair 1 observation has been basically disregarded in the HD and NED estimation.

The signed NEDT value is 3.86 with p-value = 0.0010, obtained by comparison with the $t(13)$ distribution. For the reduced data, the signed NEDT value is 3.82 with p-value = 0.0011. The signed HDT value and associated p-value were 2.74 and 0.0085, respectively, for the full data; and they were 2.72 and 0.0093, respectively, for reduced data. Judging by the p-values for the full and reduced data, the effect of pair 1 on the signed HDT or signed NEDT is little, being more negligible for the latter. In the presence of pair 1, a much smaller p-value for the signed NEDT suggests that the signed NEDT has rejected the null hypothesis more comfortably than the signed HDT.

8. DISCUSSION

In this paper we have shown that, like the MNEDE to the MHDE, the NEDT is a very attractive alternative to the HDT as an efficient as well as robust procedure. In fact, the NED test procedure appears to have more robust level performance than the HDT in the settings considered in our simulations. Results from application of the tests to the example data considered also support a slightly better performance of the NEDT. Our breakdown point result for the MNEDE and the empirical results for the NEDT lead us to strongly believe that a breakdown point result for the NEDT similar to that for the HDT (Simpson 1989) holds. We hope to present a proof in a future article. We have also constructed a goodness-of-fit test statistic for continuous models and established its asymptotic normality under the null hypothesis.

APPENDIX

Assumption (iv) in Theorem 2. Let $\{\phi_n\}$ denote any sequence of estimators converging to the true parameter $\theta \in \Theta$ in probability, and let $\{\alpha_n\}$ denote any sequence of positive real numbers going to infinity and $l_B(x)$ the indicator function for a set B of real numbers. The conditions (a)-(d), (g)-(i) of Theorem 1 of

Basu et al (1997) are listed below (for a discussion of these conditions see Basu et al (1997, p. 362)):
(a) For every $\{\phi_n\}$ defined above, $\int |\ddot{f}_{\phi_n}(x) - \ddot{f}_{\theta_n}(x)|dx$ converges to zero in probability.
(b) For every $\{\phi_n\}$ defined above, $\int |u^2(\phi_n, x)f_{\phi_n}(x) - u^2(\theta, x)f_\theta(x)|dx$ converges to zero in probability.
(c) $I(\theta) < \infty$, and $\int |u^2(\theta, x+a)f_\theta(x) - u^2(\theta, x)f_\theta(x)|dx \to 0$, as $|a| \to 0$.
(d) $\limsup_{n\to\infty} \sup_{y\in\{y:y=h_n z, z\in\Psi\}} \int |f_\theta^{''}(x+y)u(\theta, x)|dx < \infty$, where $f_\theta^{''}(x)$ denotes the second derivative of $f_\theta(x)$ with respect to x.
(g) For every $\{\alpha_n\}$ defined above, $n \sup_{t\in\Psi} P(|X_1 - h_n t| > \alpha_n) \to 0$.
(h) For every $\{\alpha_n\}$ defined above, $n^{-1/2}h_n^{-1}(\int |u(\theta, x)l_{\{x:|x|\le\alpha_n\}}|dx) \to 0$.
(i) For every $\{\alpha_n\}$ defined above, $\sup_{|x|\le\alpha_n} \sup_{t\in\Psi}\{f_\theta(x+h_n t)/f_\theta(x)\} = O(1)$.

Proof of Theorem 2. We write $\hat{\theta}$ for the MNEDE $\hat{\theta}_{NED}$, for brevity. Under the given assumptions, from the results of Basu et al. (1997) we have, under $f_\theta, n^{1/2}(\hat{\theta} - \theta) = O_p(1), n^{-1/2}\rho_{NED}^{'}(\hat{f}_n, f_\theta) = O_p(1), \rho_{NED}^{''}(\hat{f}_n, f_\theta) = O_p(1)$, and hence a Taylor series approximation gives: $nh_n^{1/2}[\rho_{NED}(\hat{f}_n, f_{\hat{\theta}}) - \rho_{NED}(\hat{f}_n, f_\theta)] = o_p(1)$. Since $R_n - \mu(I) = o_p(n^{-1})$, where $\mu(I)$ denotes the length of the interval I, it then suffices to show that, under f_θ,

$$d_{NED}^*(\theta) = [\{nh_n\rho_{NED}(\hat{f}_n, f_\theta)\} - (2^{-1}\mu(I)\|w\|^2)]/[h_n^{1/2}(2^{-1}\mu(I)\|w*w\|^2)^{1/2}] \to N(0,1) \quad (A.1)$$

in distribution. We now establish (A.1). Observe that for the NED with $G(\delta) = (e^{-\delta} - 2)$, $G^{'}(0) = 0$, $G^{''}(0) = 1$ and $G^{'''}(\delta)$ is bounded on $[-1, \infty)$. For a fixed x, letting $y = \hat{f}_n(x)$ and $y_0 = f_\theta(x)$, by using a Taylor series expansion of $G((y/f_\theta(x)) - 1)$ as a function of y at $y = y_0$ we obtain

$$nh_n^{1/2}\rho_{NED}(\hat{f}_n, f_\theta) = nh_n^{1/2}\int_I 2^{-1}(\hat{f}_n - f_\theta)^2\left(\frac{1}{f_\theta}\right) - \int_I 6^{-1}(\hat{f}_n - f_\theta)^3[G^{'''}(\frac{f_n^*}{f_\theta} - 1)\frac{1}{f_\theta^2}] \quad (A.2)$$

where $f_n^*(x)$ is a point between $\hat{f}_n(x)$ and $f_\theta(x)$. Consider the first term on the right hand side of (A.2). Under the assumptions, from Corollary 1 of Rosenblatt (1975) it follows that

$$[\{nh_n^{1/2}\int_I(\hat{f}_n - f_\theta)^2\left(\frac{1}{f_\theta}\right)\} - (h_n^{-1/2}2^{-1}\mu(I)\|w\|^2)]/[(2^{-1}\mu(I)\|w*w\|^2)^{1/2}] \to N(0,1)$$

in distribution, under f_θ. Next consider the second term on the right hand side of (A.2) and it can be shown to be $o_p(1)$ using the facts that $G'''(\cdot)$ is bounded on $[-1,\infty)$ and $1/f_\theta$ bounded on I, and modifying the arguments used in the proof of Theorem 8 of Beran (1977). Therefore, (A.1) follows and this completes the proof.

Proof of Lemma 1. Modification of Lindsay's (1994) arguments in the proof of his Lemma 9 and Proposition 12 establishes that $\{\xi_j\}$ is an outlier sequence if and only if $\inf_{I^{\xi_j}}[1/f_\theta(x)] \to \infty$ and $\sup_{I^{\xi_j}} f(x) \to 0$ as $j \to \infty$; and that for an outlier sequence $\{\xi_j\}$ $\rho_G(f_j, f_\theta) \to \rho_G(f_\epsilon^*, f_\theta)$ as $j \to \infty$. Now, $[|\nabla f_\theta(x)|(f_\theta(x))^{(1-k)/k}]^k \to 0$ as $|x| \to \infty$ since by assumption $\int |u(\theta,x)|^k dF_\theta(x) < \infty$. Since $\inf_{I^{\xi_j}}[\delta_j(x)] \to \infty$ as $j \to \infty$, the above implies that $\sup_{I^{\xi_j}}[|\nabla f_\theta \delta_j(x)|\ (f_\theta(\delta_j(x))^{(1-k)/k}] \to 0$ as $j \to \infty$. Next observe that

$$\int A(\delta_j(x))\nabla F_\theta(x) - \int A(\delta_{\epsilon^*}(x))\nabla F_\theta(x) = \int_{I^{\xi_j}} A(\delta_j(x))\nabla F_\theta(x) - \int_{I^{\xi_j}} A(\delta_{\epsilon^*}(x))\nabla F_\theta(x).$$

The proof then follows using arguments similar to those used in the proof of Lindsay's (1994, Proposition 14).

Assumptions for Theorem 3. (a) $\rho_G(f_j, f_\theta)$ and $\rho_G(f_\epsilon^*, f_\theta)$ are continuous in θ, and the latter has a unique minimum at $T_G(f_\epsilon^*) = b^*$. (b) $\rho_G(f_j, f_\theta)$ converges to $\rho_G(f_\epsilon^*, f_\theta)$ as $j \to \infty$, uniformly in θ, for any compact set B of parameter values containing b^*. (c) For each $0 < \gamma < 1$ there exists a subset S of the support of $\{F_\theta : \theta \in \Theta\}$ such that $f(S) \equiv \int_S dF(x) > 1 - \gamma^2$, where F is the distribution function of f, and $\{\theta : f_\theta(S) \equiv \int_S dF_\theta(x) \geq \gamma\}$ is a compact set.

Proof of Theorem 3. Under the above assumptions (a)-(c) of Theorem 3, the proof follows from a simple modification of arguments used in Lindsay (1994, Lemmas 20, 21 and Proposition 22).

Table 1. Empirical level and power of LRT, NEDT and HDT for the uncontaminaed data($\epsilon = 0$), when normal variance is known.

Sampling distribution:	$N(0,1)$			$N(0.5,1)$		
Nominal level:	0.01	0.05	0.01	0.10	0.05	0.01
$n = 20$						
LRT	10.26	5.12	0.98	72.78	61.14	37.38
HDT	10.26	4.34	0.84	70.72	57.76	33.38
NEDT	9.20	4.20	0.70	67.62	55.82	29.76
$n = 30$						
LRT	10.32	5.02	1.12	86.14	78.02	56.68
HDT	9.82	4.90	1.04	84.40	75.36	53.32
NEDT	9.68	4.74	0.84	82.62	73.22	50.74
$n = 50$						
LRT	10.28	4.86	0.88	97.18	94.14	83.32
HDT	9.78	4.80	0.82	96.62	93.18	80.78
NEDT	9.66	4.78	0.80	95.76	92.80	80.22
$n = 100$						
LRT	10.30	5.16	0.80	99.98	99.92	99.50
HDT	9.78	4.78	0.78	99.98	99.90	99.20
NEDT	9.84	4.92	0.78	99.98	99.92	99.92

Table 2. Empirical level and power of LRT, NEDT and HDT for the uncontaminaed data($\epsilon = 0.025$), when normal variance is known.

Sampling distribution:	$0.975N(0,1) + 0.025N(3,1)$			$0.975N(0.5,1) + 0.025N(3,1)$		
Nominal level:	0.01	0.05	0.01	0.10	0.05	0.01
$n = 20$						
LRT	15.96	9.84	3.16	78.98	69.74	47.84
HDT	11.76	5.84	1.24	74.10	62.56	38.98
NEDT	8.60	3.74	0.56	68.50	54.82	28.60
$n = 30$						
LRT	16.64	9.80	3.10	91.04	84.94	68.22
HDT	11.14	6.04	1.38	87.82	79.66	59.34
NEDT	9.30	4.50	0.72	84.10	74.40	49.84
$n = 50$						
LRT	17.74	11.08	3.80	98.88	97.38	90.64
HDT	11.72	6.24	1.40	98.04	95.40	86.62
NEDT	9.66	4.78	0.84	97.00	93.54	80.96
$n = 100$						
LRT	21.58	14.56	5.28	100.00	99.98	99.86
HDT	13.40	7.70	1.94	100.00	99.98	99.68
NEDT	11.28	5.90	1.34	99.98	99.96	99.44

Table 3. Empirical level and power of LRT, NEDT and HDT for the uncontaminaed data($\epsilon = 0.05$), when normal variance is known.

Sampling distribution:	$0.95N(0,1) + 0.05N(3,1)$			$0.95N(0.5,1) + 0.05N(3,1)$		
Nominal level:	0.01	0.05	0.01	0.10	0.05	0.01
$n = 20$						
LRT	23.30	15.78	6.40	84.46	76.88	57.90
HDT	13.62	6.84	1.98	77.64	66.82	43.92
NEDT	9.64	4.42	0.78	71.32	58.06	31.64
$n = 30$						
LRT	26.14	17.96	7.82	94.50	90.20	77.30
HDT	13.70	8.04	2.02	90.34	83.58	65.22
NEDT	10.38	5.40	1.10	86.38	77.14	53.96
$n = 50$						
LRT	31.50	22.48	10.78	99.58	98.86	95.24
HDT	15.80	8.98	2.74	98.88	97.04	89.66
NEDT	11.60	6.14	1.24	97.48	94.94	84.00
$n = 100$						
LRT	45.56	35.08	19.04	100.00	100.00	99.96
HDT	20.76	13.32	4.40	100.00	100.00	99.84
NEDT	14.88	8.30	2.22	99.98	99.94	99.56

Table 4. Empirical level and power of LRT, NEDT and HDT for the uncontaminaed data($\epsilon = 0.10$), when normal variance is known.

Sampling distribution:	$0.90N(0,1) + 0.10N(3,1)$			$0.90N(0.5,1) + 0.10N(3,1)$		
Nominal level:	0.01	0.05	0.01	0.10	0.05	0.01
$n = 20$						
LRT	41.20	32.44	18.22	91.68	86.94	72.56
HDT	19.00	11.64	3.78	82.66	73.90	53.60
NEDT	12.44	6.70	1.50	75.60	63.80	37.66
$n = 30$						
LRT	49.60	40.20	24.26	97.84	96.08	79.22
HDT	21.42	13.72	4.98	93.84	89.02	74.30
NEDT	14.98	8.06	1.92	89.02	81.36	60.90
$n = 50$						
LRT	62.46	53.44	35.62	99.88	99.76	98.88
HDT	26.54	18.08	7.06	99.46	98.68	94.22
NEDT	16.94	10.12	3.16	98.46	96.44	88.82
$n = 100$						
LRT	84.36	77.66	61.68	100.00	100.00	100.00
HDT	39.70	29.00	13.42	100.00	100.00	100.00
NEDT	24.96	16.18	5.52	100.00	100.00	99.76

Table 5. Empirical level and power of LRT, NEDT and HDT for the uncontaminaed data($\epsilon = 0$), when normal variance is known.

Sampling distribution:	$N(0,1)$			$N(0.5,1)$		
Nominal level:	0.01	0.05	0.01	0.10	0.05	0.01
$n = 20$						
LRT	11.84	5.96	1.42	72.40	59.84	34.08
HDT	10.86	5.90	1.82	69.46	57.00	34.60
NEDT	8.44	4.70	0.66	63.02	55.44	32.66
$n = 30$						
LRT	11.24	5.80	1.34	85.64	76.68	53.16
HDT	10.02	5.12	1.40	87.70	74.24	51.56
NEDT	9.68	4.88	0.74	81.80	74.14	48.06
$n = 50$						
LRT	10.78	5.64	1.16	96.96	93.60	81.18
HDT	9.90	5.10	1.10	96.36	92.52	79.52
NEDT	9.82	4.80	0.86	95.26	90.62	73.04
$n = 100$						
LRT	10.58	4.98	0.98	100.00	99.92	99.34
HDT	9.48	4.46	0.86	99.98	99.90	99.08
NEDT	9.88	4.78	0.90	100.00	99.96	99.68

Table 6. Empirical level and power of LRT, NEDT and HDT for the uncontaminaed data($\epsilon = 0.025$), when normal variance is known.

Sampling distribution:	$0.975N(0,1) + 0.025N(3,1)$			$0.975N(0.5,1) + 0.025N(3,1)$		
Nominal level:	0.01	0.05	0.01	0.10	0.05	0.01
$n = 20$						
LRT	13.02	6.36	1.38	76.60	64.58	37.52
HDT	12.38	6.64	2.02	73.34	61.12	38.16
NEDT	7.04	3.26	0.64	71.76	59.94	34.66
$n = 30$						
LRT	12.98	6.74	1.42	89.42	81.72	58.64
HDT	10.90	6.20	1.54	87.24	78.88	56.60
NEDT	8.50	4.06	0.70	83.56	71.90	43.58
$n = 50$						
LRT	13.44	7.88	1.74	98.50	96.38	86.32
HDT	11.56	6.46	1.54	97.96	94.90	83.98
NEDT	8.98	4.38	0.82	96.92	93.04	77.66
$n = 100$						
LRT	16.78	10.20	2.82	100.00	99.98	99.84
HDT	13.44	7.90	1.92	100.00	99.98	99.60
NEDT	10.84	5.36	1.06	99.98	99.96	99.28

Table 7. Empirical level and power of LRT, NEDT and HDT for the uncontaminaed data($\epsilon = 0.05$), when normal variance is known.

Sampling distribution:	$0.95N(0,1)+0.05N(3,1)$			$0.95N(0.5,1)+0.05N(3,1)$		
Nominal level:	0.01	0.05	0.01	0.10	0.05	0.01
$n = 20$						
LRT	15.68	7.82	1.86	80.44	69.54	41.54
HDT	14.46	7.86	2.56	80.82	68.28	45.90
NEDT	8.90	4.16	0.72	78.36	65.02	40.08
$n = 30$						
LRT	17.42	10.00	2.22	92.36	86.06	64.56
HDT	14.66	8.62	2.30	90.22	83.22	62.28
NEDT	10.22	5.12	1.06	86.46	75.96	48.78
$n = 50$						
LRT	21.94	13.50	3.56	99.28	97.98	90.50
HDT	17.50	10.02	2.76	98.88	96.92	88.72
NEDT	12.44	6.46	1.20	97.72	95.26	82.26
$n = 100$						
LRT	34.58	23.68	8.32	100.00	100.00	99.94
HDT	24.68	15.26	5.48	100.00	100.00	99.82
NEDT	17.74	9.94	2.38	99.98	99.98	99.60

Table 8. Empirical level and power of LRT, NEDT and HDT for the uncontaminaed data($\epsilon = 0.10$), when normal variance is known.

Sampling distribution:	$0.90N(0,1)+0.10N(3,1)$			$0.90N(0.5,1)+0.10N(3,1)$		
Nominal level:	0.01	0.05	0.01	0.10	0.05	0.01
$n = 20$						
LRT	25.16	14.28	3.76	86.78	77.50	49.98
HDT	22.50	13.84	4.70	83.54	74.14	49.90
NEDT	14.76	7.38	1.34	82.40	72.08	46.98
$n = 30$						
LRT	32.50	20.92	5.96	96.32	92.24	75.50
HDT	26.90	18.20	5.62	94.64	89.66	72.88
NEDT	19.20	10.20	2.30	91.78	84.14	59.20
$n = 50$						
LRT	45.94	32.82	12.78	99.80	99.54	96.20
HDT	36.14	25.96	9.96	99.58	98.92	94.92
NEDT	27.26	16.44	4.24	99.46	98.66	94.08
$n = 100$						
LRT	72.68	61.08	33.98	100.00	100.00	100.00
HDT	57.84	45.62	23.48	100.00	100.00	100.00
NEDT	43.16	31.18	12.34	100.00	100.00	99.84

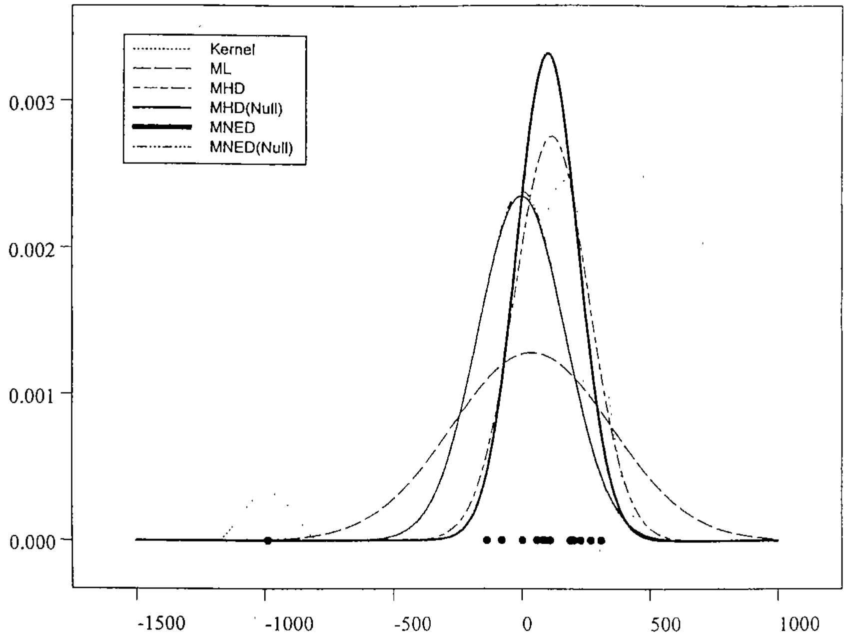

Figure 1. Six Density Estimates for the Telephone-Line Faults Data

ACKNOWLEDGEMENTS

The research of the second author was supported in part by the CMDS Funds at the Indian Institute of Management Calcutta. The authors would like to thank Dr. Ayanendranath Basu for some helpful discussions.

REFERENCES

Ali, S. M. and Silvey, S. D. (1966). "A General Class of Coefficients of Divergence of One Distribution from Another," *Journal of the Royal Statistical Society*, B, **28**, 131-142.

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. and Sarkar, S. (1994a). "The Trade-off Between Robustness and Efficiency and the Effect of Model Smoothing in Minimum disparity Inference," *Journal of Statistical Computation and Simulation*, **50**, 173-185.

Basu, A. and Sarkar, S. (1994b). "On Disparity Based Goodness-of-fit Tests for Multinomial Models," *Statistics & Probability Letters*, **19**, 307-312.

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.

Beran, R. J. (1977). "Minimum Hellinger Distance Estimates for Parametric Models," *Annals of Statistics*, **5**, 445-463.

Beran, R. J. (1981). "Efficient Robust Tests in Parametric Models," *Zeischrift fur Wahrscheeinlichkeitstheorie und Verwandte Gebiete*, **57**, 73-86.

Csiszar, I. (1963). "Eine Informations Theoretische Ungleichung und ihre Anwendung auf den Beweis der Ergodizitat von Markoffschen Ketten," *Publ. Math. Inst. Hungar. Acad. Sci.*, **3**, 85-107.

Cressie, N. and Read, T. R. C. (1984). "Multinomial Goodness-of-Fit Tests," *Journal of the Royal Statistical Society B*, **46**, 440-464.

Donoho, D. L. and Liu, R. C. (1988). "The "Automatic" Robustness of Minimum Distance Functionals," *Annals of Statistics*, **16**, 552-586.

Hall, P. and Marron, J. S. (1991). "Lower Bounds for Bandwidth Selection in Density Estimation", *Probability Theory and Related Fields*, **90**, 149-173.

Hampel, F. R., Ronchetti, E. M. Rousseeuw, P. M. and Stahel, W. A. (1986). *Robust Statistics: The Approach Based on Influence Functions*, New York, Wiley.

Härdle, W., Hall, P. and Marron, J. S. (1988). "How Far are Automatically Chosen Regression Smoothing Parameters from Their Optimum?", *Journal of American Statistical Association*, **83**, 86-95.

Jeong, D. B. and Sarkar, S. (2000). "Negative Exponential Disparity Based Family of Goodness-of-Fit Tests for Multinomial Models," *Journal of Statistical Computation and Simulation*, **65**, 43-61.

Lindsay, B. G. (1994). "Efficiency Versus Robustness: The Case for Minimum Hellinger Distance and Related Methods," *Annals of Statistics*, **22**, 1081-1114.

Marron, J. S. (1989). "Comments on a Data Based Bandwidth Selector", *Computational Statistics and Data Analysis*, **8**, 155-170.

Parzen, E. (1962). "On Estimation of a Probability Density Function and Its Mode," *Annals of Mathematical Statistics*, **33**, 1065-1076.

Rao, C. R. (1948)."Large Sample Tests of Statistical Hypotheses Concerning Several Parameters with Applications of Problems of Estimation," *Proceedings of the Cambridge Philosophical Society*, **44**, 50-57.

Rao, C. R. (1961). "Asymptotic Efficiency and Limiting Information," *Proceedings of the Fourth Berkeley Symposium*, **1**, 531-546.

Rao, C. R. (1973). *Linear Statistical Inference and Its Applications (2nd ed.)*, New York: John Wiley.

Rosenblatt, M. (1975). "A Quadratic Measure of Deviation of Two-Dimensional Density Estimates and a Test of Independence," *Annals of Statistics*, **3**, 1-14.

Sarkar, S., Song, K. and Jeong, D. B. (1998). "Penalizing the Negative Exponential Disparity in Discrete Models," *The Korean Communications in Statistics*, **5**, 517-529.

Simpson, D. G. (1987). "Minimum Hellinger Estimation for the Analysis of Count Data," *Journal of the American Statistical Association*, **82**, 802-807.

Simpson, D. G. (1989). "Hellinger Deviance Tests: 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.

Wald, A. (1943). "Tests of Statistical Hypotheses Concerning Several Parameters When the Number of Observations Is Large," *Transactions of the American Mathematical Society*, **54**, 426-482.

Welch, W. J. (1987). "Randomizing the Median in Matched-Pairs Designs," Biometrika, **74**, 609-614.