ON THE POWER BREAKDOWN OF THE NEGATIVE EXPONENTIAL DISPARITY TESTS

PARTHA SARATHI DEY[1] AND AYANENDRANATH BASU[2]*

Summary

The generalized negative exponential disparity, discussed in Bhandari *et al.* (Robust inference in parametric models using the family of generalized negative exponential disparities, 2006, ANZJS, **48**, 95–114), represents an important class of disparity measures that generates efficient estimators and tests with strong robustness properties. In their paper, however, Bhandari *et al.* failed to provide a sharp lower bound for the power breakdown point of the corresponding tests. This was acknowledged by the authors, who indicated the possible existence of a sharper bound, but noted that they did not "have a proof at this point". In this paper we provide an improved bound for this power breakdown point, and show with an example how this can enhance the existing results.

Key words: generalized negative exponential disparity; power breakdown; robustness.

1. Background

Bhandari, Basu & Sarkar (2006) proposed the class of generalized negative exponential disparity tests and demonstrated their robust properties; however, they failed to provide a sharp lower bound to their power breakdown point. In this note we derive this sharp bound.

Let $\mathcal{F}_\Theta = \{F_\theta : \theta \in \Theta \subset \mathbb{R}^p\}$ be the model and G be the true distribution; let f_θ and g be the densities of F_θ and G. We assume that $\mathcal{G}$, the class of all distributions having densities with respect to a dominating measure, is convex and contains both $\mathcal{F}_\Theta$ and G. Let $g_n(\cdot)$ be a kernel density estimate (with distribution function G_n) constructed from a random sample $X_1, \ldots, X_n$ obtained from G (for a discrete model, $g_n(x)$ is the relative frequency at x). Let $\delta(x) = (g_n(x) - f_\theta(x))/f_\theta(x)$ be the Pearson residual at x. For a real, convex, thrice-differentiable function C defined on $[-1, \infty)$ with $C(0) = 0$, define the disparity between g_n and f_θ as

$$D_C(g_n, f_\theta) = \int C(\delta) f_\theta,$$

where the integral is with respect to the dominating measure. The family of generalized negative exponential disparities $\mathrm{GNED}_\lambda(g_n, f_\theta)$, $\lambda \geq 0$, corresponds to the set of functions

$$C_\lambda(\delta) = \begin{cases} (e^{-\lambda\delta} - 1 + \lambda\delta)/\lambda^2, & \text{if } \lambda > 0 \\ \delta^2/2 & \text{if } \lambda = 0. \end{cases} \tag{1}$$

The ordinary negative exponential disparity corresponds to $\lambda = 1$.

Let the hypotheses of interest be $H_0 : \theta \in \Theta_0$ versus $H_1 : \theta \in \Theta - \Theta_0$, $\Theta_0 \subset \Theta$. Then the generalized negative exponential disparity test statistics are given by

$$\mathrm{GNEDT}_\lambda = 2n\big(\mathrm{GNED}_\lambda\big(g_n, f_{\theta^*_{G_n}}\big) - \mathrm{GNED}_\lambda\big(g_n, f_{\theta_{G_n}}\big)\big) = 2n\frac{e^\lambda}{\lambda^2} N_\lambda(G_n). \tag{2}$$

Here, $\theta_{G_n}, \theta^*_{G_n}$ are defined by the relations $\mathrm{GNED}_\lambda(g_n, f_{\theta_{G_n}}) = \min_{\theta\in\Theta} \mathrm{GNED}_\lambda(g_n, f_\theta)$ and $\mathrm{GNED}_\lambda(g_n, f_{\theta^*_{G_n}}) = \min_{\theta\in\Theta_0} \mathrm{GNED}_\lambda(g_n, f_\theta)$. From (1) and (2) it can be seen that

$$N_\lambda(G) = \rho_\lambda\big(g, f_{\theta^*_G}\big) - \rho_\lambda\big(g, f_{\theta_G}\big) \text{ where } \rho_\lambda(g, f) = \int e^{-\lambda g/f} f.$$

Let $H = (1-\alpha)G + \alpha K$, $\alpha \in [0, 1]$, be a contaminated version of G. Define the quantities

$$N_{\lambda,\min} = \inf_{F\in\mathcal{G}} N_\lambda(F) \quad \text{and} \quad \alpha_1(G, N_\lambda) = \inf\big\{\alpha : \inf_{K\in\mathcal{G}} N_\lambda((1-\alpha)G + \alpha K) = N_{\lambda,\min}\big\}.$$

The quantity $\alpha_1(G, N_\lambda)$ is the power breakdown point of the GNEDT_λ test; it is the smallest contamination fraction that may drive the P-value of the test to its minimum possible value.

2. The New Breakdown Bound

Notice that $N_\lambda(H) = \rho_\lambda(h, f_{\theta^*_H}) - \rho_\lambda(h, f_{\theta_H})$. Let L be a lower bound for $\rho_\lambda(h, f_{\theta^*_H})$, and U be an upper bound for $\rho_\lambda(h, f_{\theta_H})$. Because $N_\lambda(H) \geq L - U$, power breakdown cannot occur at a contamination level α for which $L - U > 0 = N_{\lambda,\min}$. Thus, finding a sharp lower bound for the power breakdown point amounts to determining sharp upper and lower bounds U and L. Notice that

$$\rho_\lambda\big(h, f_{\theta_H}\big) \leq \rho_\lambda\big(h, f_{\theta_G}\big) \leq \bigg(\int f_{\theta_G} \exp\bigg(-\frac{\lambda g}{f_{\theta_G}}\bigg)\bigg)^{(1-\alpha)} = \rho_\lambda\big(g, f_{\theta_G}\big)^{(1-\alpha)} = U,$$

and

$$\begin{aligned}
\rho_\lambda\big(h, f_{\theta^*_H}\big) &= \int f_{\theta^*_H} \exp\bigg(-\lambda\frac{(1-\alpha)g}{f_{\theta^*_H}}\bigg) \exp\bigg(-\lambda\frac{\alpha k}{f_{\theta^*_H}}\bigg) \\
&\geq \bigg(\int f_{\theta^*_H} \exp\bigg(-\lambda\frac{(1-\alpha)g}{f_{\theta^*_H}}\bigg)\bigg) \exp\left(-\lambda\alpha \frac{\displaystyle\int k \exp\Big(-\frac{(1-\alpha)g}{f_{\theta^*_H}}\Big)}{\displaystyle\int f_{\theta^*_H} \exp\Big(-\lambda\frac{(1-\alpha)g}{f_{\theta^*_H}}\Big)}\right) \\
&\geq \rho_\lambda\big(g, f_{\theta^*_G}\big) \exp\bigg(-\frac{\lambda\alpha}{\rho_\lambda\big(g, f_{\theta^*_G}\big)}\bigg) \\
&= L.
\end{aligned}$$

The statement $L - U > 0$ is equivalent to

$$\alpha < \left(\log \rho_\lambda\left(g, f_{\theta_G^*}\right) - \log \rho_\lambda\left(g, f_{\theta_G}\right)\right) \left(\frac{\lambda}{\rho_\lambda\left(g, f_{\theta_G^*}\right)} - \log \rho_\lambda\left(g, f_{\theta_G}\right)\right)^{-1}.$$

Thus power breakdown does not occur whenever $\alpha < B_\lambda(N_\lambda(G))$, where

$$B_\lambda(x) = \inf_{\substack{e^{-\lambda} \le z \le y \le 1 \\ y - z = x}} \frac{\log y - \log z}{\frac{\lambda}{y} - \log z} = \inf_{e^{-\lambda} \le z \le 1 - x} \frac{\log(z + x) - \log z}{\frac{\lambda}{z + x} - \log z}, \; 0 \le x \le 1 - e^{-\lambda}.$$

As opposed to the bound $B_\lambda(N_\lambda(G))$, the original Bhandari *et al.* bound is given by

$$C_\lambda(N_\lambda(G)) = \frac{N_\lambda(G)}{(\lambda + 1) - e^{\lambda}}.$$

It is easily seen that the bound B_λ is uniformly better than C_λ for all $\lambda \ge 0$. The graphs of B_λ and C_λ for $\lambda = 1$ are presented in Figure 1. It illustrates the superiority of B_λ, which is well approximated by a straight line in this case.

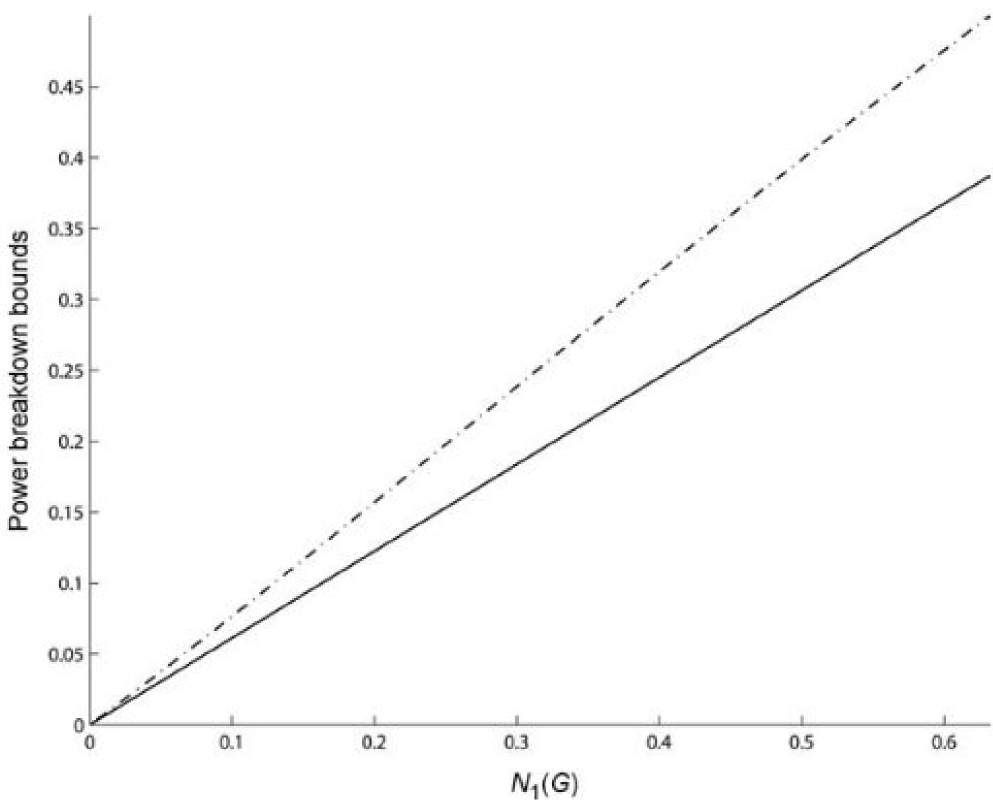

Figure 1. The power breakdown bounds B_1 (dot–dashed line) and C_1 (solid line) plotted against $N_1(G)$.

TABLE 1

Lower bounds for the power breakdown point of GNEDT$_\lambda$ *using* B_λ *for testing* $H_0: \eta = 0$, *with unknown* σ^2 *under the normal model. For each block, the figures for the Bhandari et al. (2006) formula* C_λ *are given in the second line*

		λ									
(η, σ^2)	HDT	0.25	0.50	0.75	0.90	1.00	1.10	1.25	1.50	2.00	4.00
(3,1)	0.127	0.124	0.191	0.232	0.248	0.257	0.264	0.269	0.272	0.259	0.177
		0.116	0.169	0.191	0.196	0.198	0.198	0.197	0.191	0.176	0.116
(3,2)	0.035	0.056	0.095	0.122	0.135	0.141	0.146	0.151	0.154	0.148	0.089
		0.053	0.085	0.102	0.108	0.111	0.112	0.113	0.113	0.105	0.065

HDT: Hellinger deviance test.

3. A Numerical Example

Let F_θ be the $N(\eta, \sigma^2)$ distribution with $\theta = (\eta, \sigma^2)$; we want to test $H_0: \eta = 0$ versus $H_1: \eta \neq 0$, σ^2 unknown. For the true distribution F_θ, we calculate the B_λ bounds for the generalized negative exponential disparity tests for several values of $\theta = (\eta, \sigma^2)$ and λ. The lower bound of the power breakdown of the Hellinger deviance test (HDT; Simpson 1989) and the C_λ bounds are also calculated for comparison. Some of these numbers are presented in Table 1. The numbers clearly show the superiority of the new bound. Many other combinations of θ and λ also led to improvements of similar magnitude, but the results have not been presented for the sake of brevity.

References

BHANDARI, S.K., BASU, A. & SARKAR, S. (2006). Robust inference in parametric models using the family of generalized negative exponential disparities. *Aust. N. Z. J. Stat*. **48**, 95–114.

SIMPSON, D.G. (1989). Hellinger deviance tests: Efficiency, breakdown points and examples. *J. Amer. Statist. Assoc*. **84**, 107–113.