The Residual Adjustment Function and Weighted Likelihood: A Graphical Interpretation of Robustness of Minimum Disparity Estimators

Chanseok Park[a], Ayanendranath Basu[b,*] and Bruce G. Lindsay[a,†]

[a]*Department of Statistics, Pennsylvania State University, University Park, PA 16802, USA*

[b]*Applied Statistics Unit, Indian Statistical Institute, Calcutta 700 035, India*

Running Title: A graphical interpretation of robustness

Abstract

The minimum disparity estimators of Lindsay (1994) combine full asymptotic efficiency and attractive robustness properties and hence are useful practical tools. The residual adjustment function (RAF) introduced and used by Lindsay in this context helps to graphically interpret the robustness of the estimators, but this representation is not completely satisfactory since the domain of the RAF is infinite. In this paper we develop another graphical representation to summarize the behavior of the minimum disparity estimators in relation to maximum likelihood which gives useful insights.

Keywords: residual adjustment function; Pearson residual; Neyman residual.

---

1. Introduction

Suppose that $X_1, \ldots, X_n$ represent a random sample from a discrete distribution with density $f_\theta(x)$. Without loss of generality, let the sample space be $\mathcal{X} = \{0, 1, 2, \ldots\}$. Let $d(x)$ be the empirical density at x (relative frequency at x) based on the given sample. Define the Pearson residual at x to be

$$\delta(x) = \frac{d(x) - f_\theta(x)}{f_\theta(x)}.$$

---

*He is currently visiting the Department of Statistics, Pennsylvania State University.

†His research was supported by NSF Grant 9870193.

Let $G(\cdot)$ be a thrice differentiable convex function on $[-1, \infty)$ with $G(0) = 0$. Lindsay (1994) defined a disparity — a measure of discrepancy between the data and the model density — based on the function $G(\cdot)$, as

$$\rho_G(d, f_\theta) = \sum_{x \in \mathcal{X}} G(\delta(x)) f_\theta(x). \tag{1}$$

Notice that the range of Pearson residual δ is $[-1, \infty)$. Under certain regularity conditions, all minimum disparity estimators are first order efficient; in addition many of them have attractive robustness properties. A similar representation with integrals replacing sums and involving kernel density estimates can be used to define disparities in the continuous case.

Let $g'(\cdot)$ and $g''(\cdot)$ represent the first two derivatives of the function $g(\cdot)$ with respect to its argument. Under differentiability of the model, the minimization of the disparity measure (1) corresponds to solving an estimating equation of the form

$$-\nabla\rho = \sum_{x \in \mathcal{X}} A(\delta(x)) \nabla f_\theta(x) = 0, \tag{2}$$

where $A(\delta) = (1 + \delta)G'(\delta) - G(\delta)$ and ∇ represents the gradient with respect to θ. The function $A(\delta)$ can be centered and rescaled, without changing the estimating properties of the disparity, so that we may $A(0) = 0$ and $A'(0) = 1$. Minimum disparity estimators have received wide attention in statistical inference because of their ability to reconcile the properties of robustness and asymptotic efficiency. See Lindsay (1994) for more details of the method, and Basu *et al.* (1997) for a comprehensive discussion including some of the later work.

2. Residual adjustment function and weighted likelihood

We will call the centered and scaled function $A(\delta)$ the residual adjustment function (RAF) of the disparity. Now, consider the likelihood disparity (LD)

$$\text{LD}(d, f_\theta) = \sum_{x \in \mathcal{X}} d(x) \log\left(d(x)/f_\theta(x)\right),$$

which is minimized by the maximum likelihood estimator (MLE) of θ. In order to achieve the appropriate standardization so that $A(\delta)$ has the aforementioned properties, we will modify the above form to

$$\mathrm{LD}(d, f_\theta) = \sum_{x\in\mathcal{X}} \left[d(x)\log\left(d(x)/f_\theta(x)\right) + f_\theta(x) - d(x)\right].$$

Thus the $G(\cdot)$ function for likelihood corresponds to $G_{\mathrm{LD}}(\delta) = (\delta+1)\log(\delta+1)-\delta$. The corresponding estimating equation has the form

$$-\nabla\mathrm{LD}(d, f_\theta) = \sum \delta(x)\nabla f_\theta(x) = 0,$$

so that $A_{\mathrm{LD}}(\delta) = \delta$. The RAF's for the Hellinger distance and the negative exponential disparity, two other prominent disparities in the minimum disparity literature, are given by $A_{\mathrm{HD}}(\delta) = 2[(\delta + 1)^{1/2} - 1]$, and $A_{\mathrm{NED}}(\delta) = 2 - (2+\delta)e^{-\delta}$, with the corresponding $G(\cdot)$ functions being $G_{\mathrm{HD}}(\delta) = 2[(\delta+1)^{1/2} - 1]^2$ and $G_{\mathrm{NED}}(\delta) = e^{-\delta} + \delta - 1$, respectively.

It is clear that since the forms of the estimating equation (2) are otherwise equivalent, the robustness and efficiency properties of the minimum disparity estimators are governed to a large extent by the form of the RAF $A(\delta)$. For example, large outliers, which manifest themselves as large positive values of δ, are much better controlled by those disparities for which $A(\delta) \to \delta$ as $\delta \to \infty$. The Hellinger distance and the negative exponential disparity are two such examples. See, for example, Lindsay (1994, Figures 4 and 5) for the graphs of the RAF's of various disparities.

As a graphical tool, the residual adjustment function is not completely satisfactory in explaining the robustness features of the minimum disparity estimators. There is a weighted likelihood formulation of the minimum disparity estimating equations derived from equation (2) (see Basu and Lindsay 1994; Markatou 1996; Basu *et al.* 1997; Markatou *et al.* 1998). This representation is perhaps a little more intuitive, because this brings us closer to the familiar fold of likelihood. However its usefulness is also tempered by the fact that, as in the case of the residual adjustment function, the domain of the function is unbounded.

In the following we present a graphical approach which gives a simple and natural summarization of the estimation method in relation to maximum likelihood. Consider the estimating equation (2). Since $A(0) = 0$, it is equivalent to look at the equation

$$\sum_{\delta\neq 0} A(\delta(x))\nabla f_\theta(x) = 0. \tag{3}$$

Our aims are to (a) appropriately scale the $A(\delta)$ function so that the resulting equation is consistent with (3) and the scaled version is useful in terms of providing meaningful description of how the estimating equation compares with that of MLE; and (b) provide a simple graphical representation of such a scaled measure visually explaining the above description. Notice that we can write equation (3) as

$$\sum_{\delta \neq 0} w(\delta(x))\delta(x)\nabla f_\theta(x) = 0,$$

where

$$w(\delta(x)) = \frac{A(\delta(x))}{\delta(x)}. \tag{4}$$

We choose $w(\delta)$ as the scaling that we will use to describe that comparison between the minimum disparity procedure and maximum likelihood.

Before we proceed to our graphical presentation, some discussion of the usefulness of the above weight function is useful. There may be several possible ways to scale the function $A(\delta)$ that are consistent with equation (3). In particular, one could use

$$w(\delta) = \frac{A(\delta) - A(c)}{\delta - c},$$

with $w(c) = A'(c)$. Which choice of weight is right in the sense that a weight greater than or less than certain constants is meaningful in comparisons with the MLE, and why should we choose the weight in (4) in particular?

Suppose $\tau(\cdot)$ is the true density and $\theta = T(\tau)$ where $T(\cdot)$ is the minimum disparity functional. The influence function $T'(\cdot)$ of the functional $T(\cdot)$ at y is given by

$$T'(y) = \frac{A'(\delta(y))u_\theta(y)}{\sum A'(\delta(x))\tau(x)u_\theta^2(x) - \sum A(\delta(x))\nabla^2 f_\theta(x)},$$

where $u_\theta(x) = \nabla \log f_\theta(x)$. This looks like the influence function of the MLE except that (i) it has weight $A'(\delta)$ instead of 1 in the score portion and (ii) it has weight $A(\delta)$ instead of δ in the second term of the denominator. Term (ii) for MLE, represents the change in information that arises in moving $f_\theta(\cdot)$ to $\tau(\cdot)$. Thus if we consider "information constant" neighborhoods of the model, $A'(\delta)$ represents the change in weight. However, although $A'(\delta)$ itself is in many ways the best tool in the comparison of the minimum disparity method with maximum likelihood, it is not possible to rewrite the estimating equation (3) using $A'(\delta)$. But when we use the weight $A(\delta)/\delta$, the weights are centered around $\delta = 0$, which seems to be a natural place to center the weights. Secondly for small δ, $A(\delta)/\delta = A'(\delta/2) + o(\delta^2)$. Finally, for most examples it seems that $A(\delta)/\delta$ and $A'(\delta)$ are similar in sign and information. Given these arguments, $w(\delta) = A(\delta)/\delta$ seems to be the best choice for a weight function.

One can interpret $w(\delta)$ as the degree of downweighting (overweighting if $w(\delta) > 1$) provided by the method in relation to maximum likelihood; the latter corresponds to $w(\delta) \equiv 1$, identically. To represent the weight function (4) graphically, our obvious difficulty is that the domain of the function $w(\delta)$ is unbounded. To overcome this problem, we take the following route. In analogy with the Pearson residual, define the Neyman residual as

$$\delta_N(x) = \frac{d(x) - f_\theta(x)}{d(x)}.$$

Notice that the Neyman residual takes values on $(-\infty, 1)$. To distinguish between the residuals, we will refer to the usual Pearson residual as $\delta_P(x)$. We now define a *combined* residual $\delta_c(x)$ as

$$\delta_c(x) = \begin{cases} \delta_P(x) & : \quad d \leq f_\theta \\ \delta_N(x) & : \quad d > f_\theta \end{cases} .$$

The Pearson and Neyman residuals are related as

$$\delta_P(x) = \frac{\delta_N(x)}{1 - \delta_N(x)},$$

so that the weight $A(\delta_P)/\delta_P$ becomes, in terms of the Neyman residuals,

$$\frac{1 - \delta_N}{\delta_N} A\big(\frac{\delta_N}{1 - \delta_N}\big).$$

Note also that the combined residual δ_c takes values on $[-1, 1)$. Using these definitions we then redefine the weight function as

$$w_c(\delta_c) = \begin{cases} \dfrac{A(\delta_c)}{\delta_c} & : \quad -1 \leq \delta_c < 0 \\ A'(0) & : \quad \delta_c = 0 \\ \dfrac{1 - \delta_c}{\delta_c} A\big(\dfrac{\delta_c}{1 - \delta_c}\big) & : \quad 0 < \delta_c < 1 \\ A'(\infty) & : \quad \delta_c = 1 \end{cases} . \tag{5}$$

Thus, on the positive side of the δ-axis we can interpret (5) to be representing the weights (4) as functions of Pearson residuals but in the Neyman scale. Although the domain of δ_c values is a right open interval, we define $w_c(\cdot)$ at $\delta_c = 1$ through continuity for the sake of completeness and ease of interpretation.

Notice that the above results in a smooth weight function $w_c(\cdot)$ which is differentiable at zero (the point where the Neyman and Pearson residuals are merged along the δ-axis). Notice further that wherever the disparity downweights the corresponding values relative to maximum likelihood, the graph lies below the horizontal line $w_c(\delta_c) = 1$. This is true for the case of inliers as well (observations with less data than expected under the model). For the likelihood disparity $w_c(\delta_c) \equiv 1$, identically.

Graphical representations (via the function defined in (5)) of different families and detailed analysis of their implications will be carried out in the following section. Here we summarize the spirit of the method by presenting the graphs of the RAF's of the likelihood disparity, the Hellinger distance (HD), the negative exponential disparity (NED), the Pearson's chi-square (PCS), and the Neyman's chi-square (NCS) as well as the corresponding weight functions — given by (5) — in Figure 1. The RAF's for the PCS and the NCS are given by $A_{\rm PCS}(\delta) = \delta + \delta^2/2$ and $A_{\rm NCS}(\delta) = \delta/(1+\delta)$ with the corresponding $G(\cdot)$ functions being $G_{\rm PCS}(\delta) = \delta^2/2$ and $G_{\rm NCS}(\delta) = \delta^2/[2(\delta+1)]$, respectively.

*** Figure 1 is around here. ***

There are three important features of the weight function that become directly visible in Figure 1(b): the behavior of the weight function near zero, to the left of zero and to the right of zero. We first consider the interpretation of the behavior near $\delta = 0$. The requirement that $A'(0) = 1$ guarantees that the weight function passes through the horizontal line $w_c(\delta_c) = 1$ at $\delta_c = 0$. The derivative $w_c'(\delta_c)$ equals $A''(0)/2$ at $\delta_c = 0$; thus for functions with negative values of the curvature parameter $A''(0)$, the weight curve dips below the horizontal line as δ moves away from zero in the positive direction and rises above if $A''(0)$ is positive. If $A''(0) = 0$, as in the case of the NED, the weight curve is tangent to the horizontal line at $\delta_c = 0$. It follows from Lindsay (1994) that $A''(0) = 0$ implies second order efficiency in the sense of Rao (1961, 1962). More generally, curvature at zero is a measure of second order deficiency. When $A''(0) = 0$, the second derivative of $w_c(\delta_c)$ exists at $\delta_c = 0$ and equals $A'''(0)/3$.

The interpretation of the right tail, $\delta_c > 0$, is based on the fact that $\delta_c(x)$ close to 1 means that there is much more data in the cell than the model f_θ predict. For the HD, the NCS and the NED the weights converge to zero at $\delta_c = 1$, while for the PCS this goes to ∞. Compared to the HD, the NED provides lesser downweighting for moderate outliers (a consequence of its second order efficiency property) but higher downweighting for more extreme outliers. In Section 4, we will give an example that illustrates the effect of downweighting outliers.

Turning to the left tail, values of $\delta_c(x)$ near -1 imply that there is significantly less data at x than predicted by the model. Such inliers can be created by small sample sizes, or in the case of infinite sample spaces will occur at every empty cell. The PCS and NED downweight inliers (the graph remains below $w_c(\delta_c) = 1$ on the left of $\delta_c = 0$), but for the HD inliers get higher weight, becoming twice that of LD when $\delta_c = -1$ (corresponding to empty cells). For the NCS this weight is actually infinite, signifying that the NCS is not even defined when there is a single empty cell. We note that the weight comparisons are made with the maximum likelihood estimator. The latter, although not outlier robust, is inlier robust in the sense that the inference is not critically affected by inliers. However many of the well known robust disparities such as the Hellinger distance end up greatly magnifying the effect of inlying cells, often leading to poor properties of the corresponding methods in small samples, where the design inevitably leads to several inliers and empty cells. An example will be given in Section 4. We therefore believe that the best weight curves are nearly flat to the left of zero. Another approach to tackling this problem is to modify the distance to include an empty cell penalty. Harris and Basu (1994) and Basu *et al.* (1996) take this approach and show that this can dramatically improve the methods — demonstrating the importance of inlier robustness. For yet another approach at controlling inliers, see Park *et al.* (1995).

3. Graphical representation

In this section we graphically represent the weight function in (5) for several different families of disparities. The six families looked at in this investigation include:

(a) the power divergence family (PD; Cressie and Read 1984);

(b) the blended weight Hellinger distance (BWHD; Lindsay 1994; Basu and Lindsay 1994; Basu and Sarkar 1994);

(c) the blended weight chi-square distance (BWCS; Lindsay 1994; Basu and Sarkar 1994);

(d) the generalized negative exponential disparity (GNED; Bhandari *et al.* 2000);

(e) the generalized Kullback-Leibler divergence (GKL; Park and Basu 2000);

(f) the robustified likelihood divergence (RLD; Basu, Chakraborty and Sarkar 2000).

The graphs of the weight function in (5) for the above families are presented in Figure 2. Given the data $d(x)$ and the model density $f_\theta(x)$, these six families, each indexed by a single parameter α, are given by

$$\mathrm{PD}_\alpha(d,f_\theta) = \sum_{x\in\mathcal{X}}\Big[\frac{1}{\alpha(\alpha+1)}d(x)\Big\{\big(\frac{d(x)}{f_\theta(x)}\big)^\alpha - 1\Big\} + \frac{f_\theta(x)-d(x)}{\alpha+1}\Big], \qquad \alpha\in\mathbb{R}$$

$$\mathrm{BWHD}_\alpha(d,f_\theta) = \frac{1}{2}\sum_{x\in\mathcal{X}}\Big[\frac{d(x)-f_\theta(x)}{\alpha\sqrt{d(x)}+\bar\alpha\sqrt{f_\theta(x)}}\Big]^2, \qquad \bar\alpha = 1-\alpha \qquad \alpha\in[0,1]$$

$$\mathrm{BWCS}_\alpha(d,f_\theta) = \frac{1}{2}\sum_{x\in\mathcal{X}}\frac{\big(d(x)-f_\theta(x)\big)^2}{\alpha\, d(x)+\bar\alpha\, f_\theta(x)}, \qquad \bar\alpha = 1-\alpha \qquad \alpha\in[0,1]$$

$$\mathrm{GNED}_\alpha(d,f_\theta) = \sum_{x\in\mathcal{X}}\Big[\frac{f_\theta(x)}{\alpha^2}\big\{\exp\big(\alpha-\alpha\, d(x)/f_\theta(x)\big)-1\big\} + \frac{d(x)-f_\theta(x)}{\alpha}\Big], \qquad \alpha\ge 0$$

$$\begin{aligned}\mathrm{RLD}_\alpha(d,f_\theta) &= \sum_{\bar\alpha\le d/f_\theta<1/\bar\alpha}\Big[d(x)\log\big(d(x)/f_\theta(x)\big)+f_\theta(x)-d(x)\Big] \qquad \alpha\in(0,1]\\ &+\sum_{d/f_\theta<\bar\alpha}\Big[d(x)\log\bar\alpha+\alpha f_\theta(x)\Big]-\sum_{d/f_\theta\ge 1/\bar\alpha}\Big[d(x)\log\bar\alpha+\frac{\alpha}{\bar\alpha}f_\theta(x)\Big], \qquad \bar\alpha=1-\alpha\end{aligned}$$

$$\mathrm{GKL}_\alpha(d,f_\theta) = \sum_{x\in\mathcal{X}}\Big[\frac{d(x)}{1-\alpha}\log\big(\frac{d(x)}{f_\theta(x)}\big)-\big(\frac{d(x)}{1-\alpha}+\frac{f_\theta(x)}{\alpha}\big)\log\big(\alpha\frac{d(x)}{f_\theta(x)}+1-\alpha\big)\Big]. \quad \alpha\in(0,1]$$

We have, in our graphs, concentrated on α in the range $[-2,1]$ for the PD, $[0,2]$ for the GNED, $(0,1]$ for RLD, and on $[0,1]$ for the other three families. Although the actual ranges of possible values of the parameter is larger than those considered in some cases, there are no practical reasons for considering the members outside this range. Standard distributions generated by specific values of α in each of these families are indicated in Table 1, together with the $G(\delta)$ and $A(\delta)$ functions of these families.

For comparison, we present the second order efficient member of each of the families in the corresponding graphs. For the PD and GKL families, the likelihood disparity itself is the second order efficient element. For the RLD family, all members are second order efficient. The second order efficient member of BWHD is also third order efficient ($A'''(0)=0$), so the corresponding weight curve has a second order contact with the $w_c = 1$ line. The graphs show some interesting characteristics which cannot be graphically described via the RAF curves. Particularly noticeable is the phenomenon that for the BWCS, BWHD, and GNED families, the weights converge to zero in the limit as $\delta_c \to 1$ even for such disparities which would appear to be highly nonrobust when viewed through the RAF. For the RLD family, all weights converge to zero as $\delta_c \to 1$ so long as $\alpha < 1$. Same is the story with the GKL family as long as $\alpha > 0$. In Figure 1 (b), notice that the weight functions for the PCS and the NCS are straight lines on the negative side and positive side of the δ_c axis respectively. This is intuitively expected since δ_c in the negative side of the axis equals the Pearson residual, while being the Neyman residual on the positive side of the axis.

Different families treat inliers differently. The PD family, the BWHD family, and the GKL family have problems with inliers in the sense that all the robust members of these families (say those with curvature smaller than zero) inflate the weight of the inliers. As such empty cell penalties or other inlier corrections (Harris and Basu 1994, Basu *et al.* 1996, Park *et al.* 1995, Basu and Basu 1998) may be expected to improve the small sample performance of the corresponding estimators and tests under the model.

The BWCS and the GNED families, however, have large degree of natural protection against inliers. In the GNED family, for example, there is no serious inlier problem even for highly robust members of this family. For the BWCS family, robust members such as the SCS (symmetric chi-square, corresponding to $\alpha = 1/2$) appear fairly stable against inliers. For the RLD family, there is never any overweighting either for inliers and outliers, the weight function always downweights. Notice that $\delta_N = \alpha$ when $\delta_P = \bar{\alpha}/\alpha = (1-\alpha)/\alpha$, so that in our graphical study of the RLD disparities, the weights represent downweighting for $|\delta_c| > \alpha$. While we have chosen the one parameter formulation here, one could choose the RLD as a function of two parameters α, β such that:

$$A(\delta) = \begin{cases} -\alpha & : \quad \delta < -\alpha \\ \delta & : \quad -\alpha \le \delta < \beta \\ \beta & : \quad \delta \ge \beta \end{cases} . \qquad (6)$$

In terms of the graph this would mean an element of the positive side component of the graph being combined with an element of the negative side other than its own counterpart. The RLD family also deal with an extended definition of disparities. For the RLD, the RAF is not differentiable at the points $-\alpha$ and $\alpha/\bar{\alpha}$, but is differentiable in a neighborhood of zero.

4. Examples

Finally we looked at two examples to demonstrate some of the effects observed in our graphs. In the first example we study the effect of inliers. The binomial$(3, \theta)$ model is fitted to four sets of data on the cells 0, 1, 2, and 3. The HD, the $\text{BWHD}_{\alpha=0.9}$, the NED and the $\text{GKL}_{\alpha=0.1}$ divergences are chosen to be compared to the MLE. The first data set fits the binomial model exactly and all methods lead to an estimate of $\theta = 0.5$. However for the other three data sets the cell 0 represents an inlier in relation to the binomial model. The HD shifts further than the MLE from 0.5 in either case. The $\text{BWHD}_{0.9}$ exhibits a shift which is even more pronounced. In either case one can see from the graph that these divergences are not inlier robust like the MLE and have greater sensitivity to them. On the other hand the NED which has lower sensitivity to inliers in comparison with the LD, generates estimators which shift less than the MLE in either case. For the GKL divergence, whose treatment of inliers is similar to that of the MLE, the shift is almost identical. This example demonstrates the possible increased variability that inlier sensitive methods may experience.

In the second example we investigated the impact of the weights dipping to zero at the right hand tail of the curve for the families BWCS, BWHD and GNED. We restrict ourselves to these three families in this example to show that some of the apparently nonrobust members of these families are not severely affected by a huge outlier. We consider a part of the drosophila data originally reported by Woodruff *et al.* (1984), and analyzed by Simpson (1987). The frequencies of frequencies of daughter flies having a particular lethal mutation in the X-chromosome are considered where the male parents have been exposed to a certain degree of a chemical. This particular experiment resulted in 23 flies having no such daughters, 7 having one such daughter, 3 having two such daughters while one male had 91 such daughters! A Poisson model is fitted to the data, and notice that even the members of these families with very small values of α successfully downweight the outlier and generate reasonable values of the mean parameter θ. Although the outlier is unusually large, it fails to destabilize these estimators. The corresponding values of the maximum likelihood estimator (MLE) and the minimum PCS estimator, with and without the outlier, are also reported for comparison. The effect of the large outliers on the last two disparities is clear; the results are totally different with and without it.

Table 1: The $G(\cdot)$ function, the $A(\cdot)$ function, and the curvature parameter of different disparities. Here $\bar{\alpha} = 1 - \alpha$.

Disparity	$G(\delta)$	$A(\delta)$	$A''(0)$
PD	$\frac{(\delta+1)^{\alpha+1}-1}{\alpha(\alpha+1)} - \frac{\delta}{\alpha}$	$\frac{(\delta+1)^{\alpha+1}-1}{\alpha+1}$	α
	Note: NCS($\alpha = -2$), KL($\alpha = -1$), HD($\alpha = -\frac{1}{2}$), LD($\alpha = 0$), PCS($\alpha = 1$)		
BWHD	$\frac{\delta^2}{2[\alpha\sqrt{\delta+1}+\bar{\alpha}]^2}$	$\frac{\delta}{[\alpha\sqrt{\delta+1}+\bar{\alpha}]^2} + \frac{\bar{\alpha}}{2}\frac{\delta^2}{[\alpha\sqrt{\delta+1}+\bar{\alpha}]^3}$	$1-3\alpha$
	Note: PCS($\alpha = 0$), HD($\alpha = \frac{1}{2}$), NCS($\alpha = 1$)		
BWCS	$\frac{\delta^2}{2(\alpha\delta+1)}$	$\frac{\delta}{\alpha\delta+1} + \frac{\bar{\alpha}}{2}\left[\frac{\delta}{\alpha\delta+1}\right]^2$	$1-3\alpha$
	Note: PCS($\alpha = 0$), SCS($\alpha = \frac{1}{2}$), NCS($\alpha = 1$)		
GNED	$\frac{e^{-\alpha\delta}-1+\alpha\delta}{\alpha^2}$	$\frac{(\alpha+1)-[(\alpha+1)+\alpha\delta]e^{-\alpha\delta}}{\alpha^2}$	$1-\alpha$
	Note: PCS($\alpha = 0$), NED($\alpha = 1$)		
RLD	$\begin{cases} (\delta+1)\log\bar{\alpha}+\alpha \\ (\delta+1)\log(\delta+1)-\delta \\ (\delta+1)\log(1/\bar{\alpha})-\alpha/\bar{\alpha} \end{cases}$	$\begin{cases} -\alpha & : \quad \delta < -\alpha \\ \delta & : \quad -\alpha \le \delta < \alpha/\bar{\alpha} \\ \alpha/\bar{\alpha} & : \quad \delta \ge \alpha/\bar{\alpha} \end{cases}$	0
	Note: LD($\alpha = 1$)		
GKL	$\frac{\alpha(\delta+1)\log(\delta+1)-(\alpha\delta+1)\log(\alpha\delta+1)}{\alpha(1-\alpha)}$	$\frac{1}{\alpha}\log(\alpha\delta+1)$	$-\alpha$
	Note: LD($\alpha = 0$), KL($\alpha = 1$)		

Table 2: Data sets for binomial Bin$(3, \theta)$ model.

Frequency	Sample (x_i)				Estimates				
	0	1	2	3	ML	HD	$\text{BWHD}_{\alpha=0.9}$	NED	$\text{GKL}_{\alpha=0.1}$
Data I	10	40	40	10	0.5	0.5	0.5	0.5	0.5
Data II	0	40	40	20	0.6	0.635	0.833	0.589	0.603
Data III	0	40	40	10	0.556	0.593	0.799	0.542	0.558
Data IV	0	43	43	13	0.56	0.577	0.624	0.551	0.562

Table 3: The estimated parameters under the Poisson model for the drosophila data. The estimated parameters under the LD are $\hat{\theta} = 3.059$ and $\hat{\theta} = 0.394$ with and without the outlier. The same are $\hat{\theta} = 32.565$ and $\hat{\theta} = 0.424$ for the PCS.

BWHD	α	0.01	0.1	0.2	1/3	1/2	0.7	0.9
	$\hat{\theta}$	0.423	0.416	0.407	0.391	0.364	0.305	0.166
BWCS	α	0.01	0.1	0.2	1/3	1/2	0.7	0.9
	$\hat{\theta}$	0.423	0.417	0.409	0.398	0.381	0.353	0.297
GNED	α	0.01	0.1	0.2	1/3	1/2	1	2
	$\hat{\theta}$	0.424	0.422	0.419	0.416	0.412	0.397	0.361

References

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

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

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

Basu, A., Harris, I. R., and Basu, S. (1997). Minimum distance estimation: the approach using density based distances. In G. S. Maddala and C. R. Rao, editors, *Handbook of Statistics Vol. 15, Robust Inference*, pages 21–48. Elsevier Science, New York, NY.

Basu, A., Chakraborty, B., and Sarkar, S. (2000). Robustification of the MLE without loss in efficiency. Unpublished Manuscript.

Bhandari, S., Basu, A., and Sarkar, S. (2000). Robust estimation in parametric models using the family of generalized negative exponential disparities. Technical Report #7/2000, Stat Math Unit, Indian Statistical Institute, Calcutta 700 035, India.

Cressie, N. and Read, T. R. C. (1984). Multinomial goodness-of-fit tests. *Journal of the Royal Statistical Society B*, **46**, 440–464.

Harris, I. R. and Basu, A. (1994). Hellinger distance as a penalized log likelihood. *Communications in Statistics: Simulation and Computation*, **23**, 1097–1113.

Jeong, D. and Sarkar, S. (2000). Negative exponential disparity based family of goodness-of-fit tests for multinomial models. *Journal of Statistical Computation and Simulation*. To appear.

Lindsay, B. G. (1994). Efficiency versus robustness: The case for minimum Hellinger distance and related methods. *Annals of Statistics*, **22**, 1081–1114.

Markatou, M. (1996). Robust statistical inference: Weighted likelihood or usual M-estimation? *Communications in Statistics: Theory and Methods*, **25**, 2597–2613.

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). The generalized Kullback-Leibler divergence and robust inference. Unpublished Manuscript.

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

Rao, C. R. (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.

Simpson, D. G. (1987). Minimum Hellinger distance estimation for the analysis of count data. *Journal of the American Statistical Association*, **82**, 802–807.

Woodruff, R. C., Mason, J. M., Valencia, R., and Zimmering, A. (1984). Chemical mutagenesis testing in drosophila — I: Comparison of positive and negative control data for sex-linked recessive lethal mutations and reciprocal translocations in three laboratories. *Environmental Mutagenesis*, **6**, 189–202.

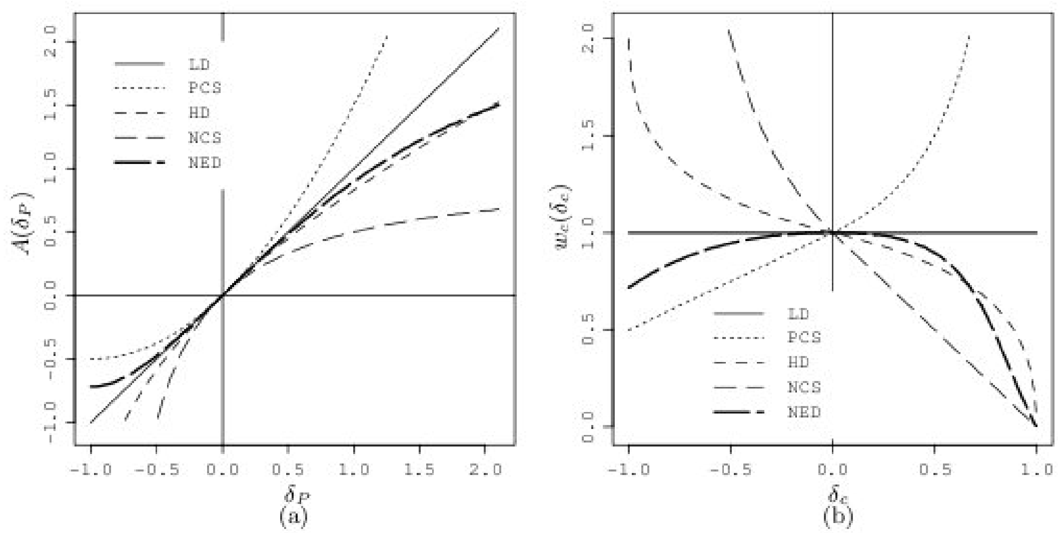

Figure 1: RAF's and weight functions of LD, PCS, HD, NCS and NED.

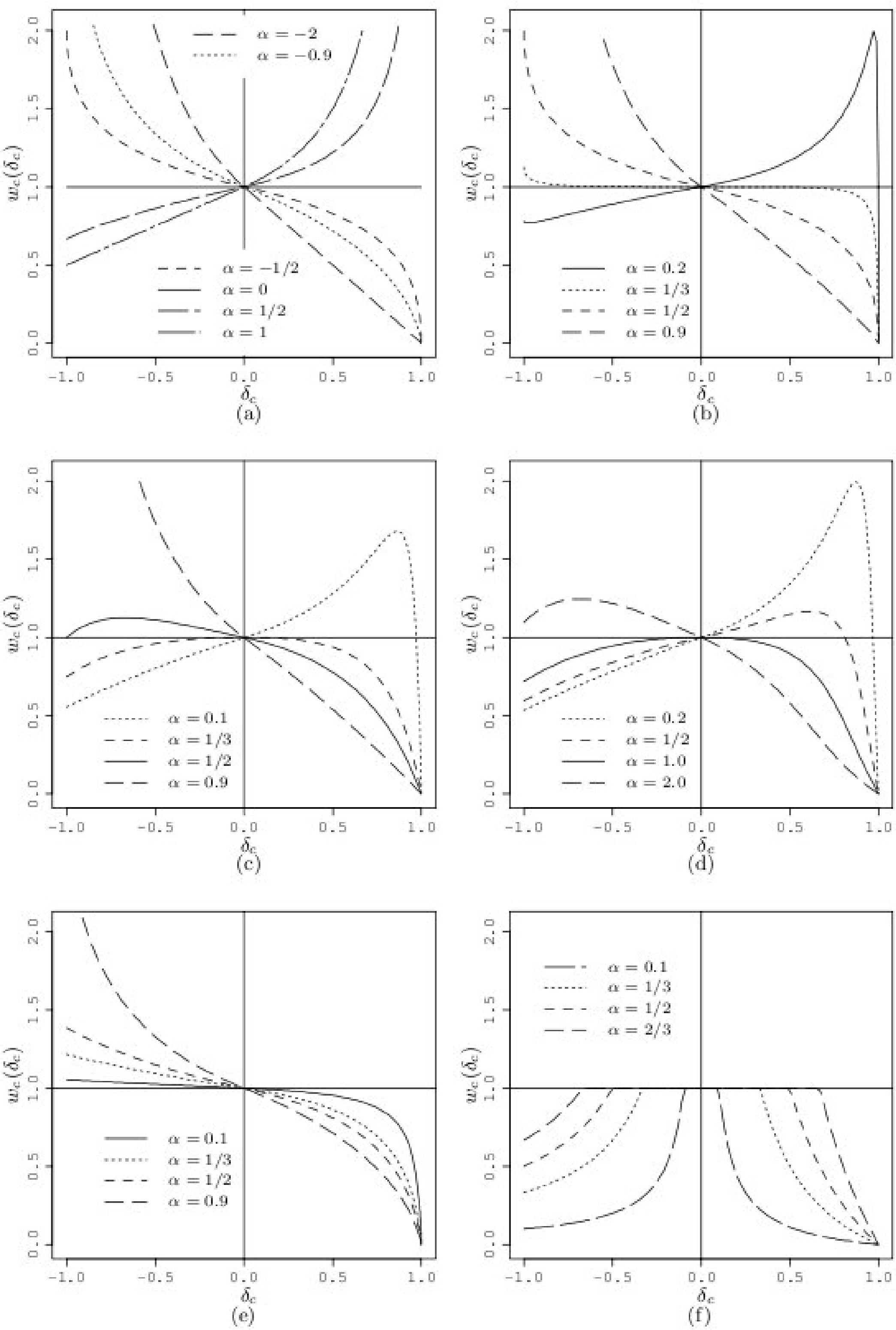

Figure 2: Weight functions of several disparities. (a) PD; (b) BWHD; (c) BWCS; (d) GNED; (e) GKL; (f) RLD.