Parametric Location Estimation using L2 Distance

William C. WOJCIECHOWSKI and David W. SCOTT

Many robust location estimators exist and have been studied using real and simulated data. In this paper, simulations are used to examine the properties of a robust estimator based on L_2 distance. The proposed estimator's performance is compared to other popular methods.

KEY WORDS: L_2 distance, Robust estimator, M-estimator, location estimator, two-component normal mixture.

1. INTRODUCTION

This paper presents simulation results for a parametric location estimator based on L_2 distance. The L_2 distance estimators are studied in detail by Scott (1999), who proposes the moniker L_2E (L2 Estimation). We build upon Scott's work by examining the properties of an L_2E location estimator when the true density differs from the assumed (model) density. The L_2E procedure is compared to the estimators considered by Stigler (1977). While this paper utilizes simulated data, Stigler compares the estimators using 24 real datasets. Wojciechowski and Scott (1999) further Stigler's real data results by including the L_2E method.

The L_2E method is a member of the Basu et al. (1999) family of M-estimators. This family is indexed by a parameter ($\alpha \geq 0$), which controls the trade-off between *robustness* and efficiency. Corresponding to $\alpha = 0$, is maximum likelihood estimation (MLE). Therefore, $\alpha = 0$ is the most efficient, but the least robust. The L_2E estimator corresponds to $\alpha = 1$ and is less efficient, but more robust than MLE.

2. ESTIMATORS

2.1 Previous Location Estimators

The mean and median are omnipresent estimators. Their inclusion provides a familiar basis for comparison. Another popular location estimator is the trimmed mean and details are found in Andrews et al. (1972, p. 7). The versions used are 10%, 15%, and 25%. They are respectively termed trim.10 , trim.15 , and trim.25 . Three robust M-estimators are included and their details are described in Andrews et al. (1972, pp. 13,15). These estimators are referred to by the developer's name, huber , andrews , or tukey . Please see Huber (1981) for details of robust M-estimators. The estimator designated edgeworth is a linear combination of the lower quartile, median, and upper quartile. Edgeworth (1893) is credited for its development. The outmean averages the discarded observations from a 25% trimmed mean calculation. Finally, Hogg's T1 (Hogg, 1974) estimator is referred to as hogg .

2.2 Parametric L_2 Estimators

This section presents the general L_2E method for a single parameter. The development is prosaic, but reveals the simplicity of the estimator.

Let $f(x;\theta)$ be a probability density function that depends on the parameter θ. The density, $f(x;\theta)$, is the model. Let $g(x)$ be the true probability density function of a continuous random variable X. The L_2E estimator is defined as

$$\begin{aligned}
\hat{\theta} &= \arg\min_{\theta} \int [f(x;\theta) - g(x)]^2 dx \\
&= \arg\min_{\theta} \int f(x;\theta)^2 dx - 2\int f(x;\theta) g(x) dx \\
&\quad + \int g(x)^2 dx,
\end{aligned}$$

Since $\int g(x)^2$ does not depend on θ and $\int f(x;\theta)g(x)dx = E_g[f(X;\theta)]$ the L_2E estimate may be rewritten as

$$\hat{\theta} = \arg\min_{\theta} \int f(x;\theta)^2 dx - 2E_g[f(X;\theta)].$$

The expected value depends on an unknown quantity and must be approximated. We use a substitution based on Rudemo's (1982) solution to a non-parametric density estimation problem

$$E_g[f(X;\theta)] \approx \frac{1}{n}\sum_{i=1}^{n} f(x_i;\theta),$$

where the x_i are the observations from a sample of size n. Altogether, the L_2E estimator is

$$\hat{\theta} = \arg\min_{\theta} \int f(x;\theta)^2 dx - \frac{2}{n}\sum_{i=1}^{n} f(x_i;\theta). \qquad (1)$$

A numeric optimization routine is used to find the value of $\hat{\theta}$. Making the appropriate changes to Equation (1), the L_2E estimator easily adapts to multiparameter and multivariate problems. For discrete random variables, replace integrals with sums.

For this study, $f(x;\theta)$ is the normal density with $\theta = (\mu,\sigma)$. In this case, the integral in Equation (1) exists in closed form. Both μ and σ are estimated and the value of $\hat{\mu}$ is the location estimate. The L_2E estimator is referred to as l2e.norm , indicating the normal density model.

William C. Wojciechowski is graduate student and David W. Scott is Professor, Department of Statistics, Rice University, Houston, TX 77005-1892. This research was supported in part by the National Science Foundation grant DMS 96-26187 and the Department of Defense contract MDA 904-95-C-2203.

Considering the case of $g(x) = f(x;\theta)$, the L_2E estimator is unbiased. When both model and truth are the normal density, the asymptotic variance of $\sqrt{n}\hat{\mu}$ is $1.54\sigma^2$. This yields an unimpressive efficiency of 65%. This paper examines the bias and variance of an L_2E location estimator when truth and model *do not* coincide.

3. SIMULATION

We consider estimating the location when the true density deviates from the model density. The model is the normal density, $f(x;\theta) = \phi(x;\mu,\sigma)$, and the true density is a two-component normal mixture

$$g(x) = (1-\rho)\phi(x;\mu,\sigma) + \rho\phi(x;\mu+\delta,\epsilon\sigma),$$

where $-\infty < \delta < \infty$, $\epsilon \geq 0$, and $0 \leq \rho \leq 1$. The second component is considered to be the deviation from the model. Shifts in the mean ($\delta \neq 0\,, \epsilon = 1$) and in the standard-deviaton ($\delta = 0\,, \epsilon \neq 1$) are simulated. This facilitates both symmetric and asymmetric contaminations. Simultaneous mean and standard-deviation shifts are not considered.

The simulation parameter sets are $\delta = (2,4)$, $\epsilon = (3,10)$, $\rho = (0.05, 0.10, 0.15, 0.25)$, $\mu = 0$, and $\sigma = 1$. The sample sizes are $n = (5, 10, 20, 40, 100, 1000)$ and 1000 simulations were performed for each combination of parameters.

4. NUMERICAL SUMMARIES

The performance of each estimator is measured relative to the others. Comparisons of mean squared error (MSE), variance, and bias are presented in the proceeding sections. The summary values are the relative index and the relative rank. The details of both are found in Stigler (1977). While the relative index accounts for magnitude, the relative rank only considers the rank, respectively denoted by $\bar{\varepsilon}$ and $\bar{\rho}$. The relative index is greater than equal to 0 and the relative rank is between 1 and 12. For both measures, lower values indicate better relative performance.

To obtain $\bar{\varepsilon}$, several intermediate calculations are necessary. Let μ_j be the true value for the jth dataset and let $\hat{\mu}_{ij}$ be the value of the ith estimator for the jth dataset. The average absolute error (ω_j) is calculated for each dataset and is defined by

$$\omega_j = \frac{1}{12}\sum_{i=1}^{12} |\hat{\mu}_{ij} - \mu_j|.$$

Next, the relative error (ε_{ij}) of estimator i for dataset j is calculated and is defined by

$$\varepsilon_{ij} = \frac{|\hat{\mu}_{ij} - \mu_j|}{\omega_j}.$$

Finally, the index of relative error ($\bar{\varepsilon}_i$) is calculated for the ith estimator by averaging across datasets. It is defined by

$$\bar{\varepsilon}_i = \frac{1}{n}\sum_{j=1}^{n} \varepsilon_{ij}.$$

When calculating $\bar{\varepsilon}_i$, n is the total number of datasets. There are 96,000 simulated datasets. When considering the shift types seperately, n is 48,000.

Unlike the index of relative error, the index of relative rank does not consider the magnitude of the errors and only depends upon the ranks. To begin with, the absolute errors $|\hat{\mu}_{ij} - \mu_j|$ are calculated and ranked within each dataset. Denote the rank for estimator i and dataset j by ρ_{ij}. If dataset j is considered, the estimator with the smallest absolute error has $\rho_{ij} = 1$ and the estimator with the largest absolute error has $\rho_{ij} = 13$. The index of relative rank for estimator i is then computed by averaging across datasets and is defined by

$$\bar{\rho}_i = \frac{1}{n}\sum_{j=1}^{n}\rho_{ij}.$$

The value of n is determined identically to $\bar{\varepsilon}_i$.

5. RESULTS

5.1 MSE Results

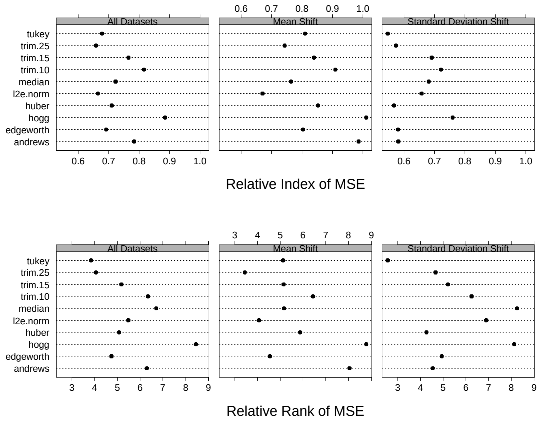

Figure 1. Relative index and relative rank of MSE.

The MSE results are graphically summarized in Figure 1. The top three graphs plot the relative index of MSE for each of the estimators. Two estimators are absent from the plot, the mean and outmean . Their values dominate the scale. Instead, their numerical summaries are presented in Table 1.

Table 1. Relative Index/Relative Rank of MSE.

Estimator	*All*	*Standard Deviation Shift*	*Mean Shift*
Mean	1.40/9.90	1.52/10.29	1.28/9.51
Outmean	3.22/11.94	4.12/12.00	2.33/11.89

When observing Figure 1, several conclusions can be made. The tukey , huber , andrews , and edgeworth estimators are average for a mean shift, but exhibit much better performance for a standard-deviation shift. The trim.10 and trim.25 are average performers for both shift types.

The trim.25 is shown to behave well for both shift types. The median is uniformly average. With all due respect, the hogg estimator has the unfortunate distinction of always being last, excepting the relative rank for a standard deviation shift. However, the hogg estimator is better than the mean and outmean . The l2e.norm estimator exhibits a behaviour that is different than the other estimators. It is relatively better for the mean shift than for the standard-deviation shift. Considering all of the datasets, the tukey , trim.25 , l2e.norm , huber , and edgeworth are among the better performers. The next two sections examine variance and bias.

5.2 Standard-Deviation Results

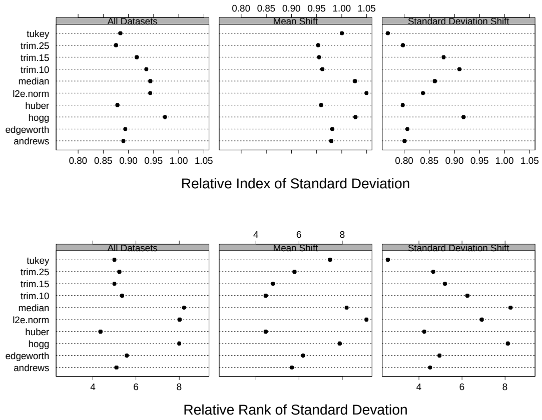

Figure 2. Relative index and relative rank of standard-deviation.

The standard-deviations of the estimators are used to compare the variation. Figure 2 displays the results. The mean and outmean results are found in Table 2.

Table 2. Relative Index/Relative Rank of Standard Deviation.

Estimator	*All*	*Standard Deviation Shift*	*Mean Shift*
Mean	1.17/7.39	1.38/10.29	0.97/1.14
Outmean	1.69/10.75	2.24/12.00	4.49/9.50

The following observations are seen in Figure 2. The tukey estimator is better for a standard-deviation shift than a mean shift. The trim.25 and huber performs well for both shifts. The trim.15 and trim.15 are less variable for a mean shift than a standard-deviation shift. The median and l2e.norm are more variable for a mean shift than a standard-deviation shift. The l2e.norm is last for a mean shift. The hogg estimator exhibits the same relative performance as in the preceeding section. The edgeworth and andrews estimators are average for a mean shift and are slightly better for a standard-deviation shift. For all of the datasets, the l2e.norm estimator has relatively poor variance.

5.3 Bias Results

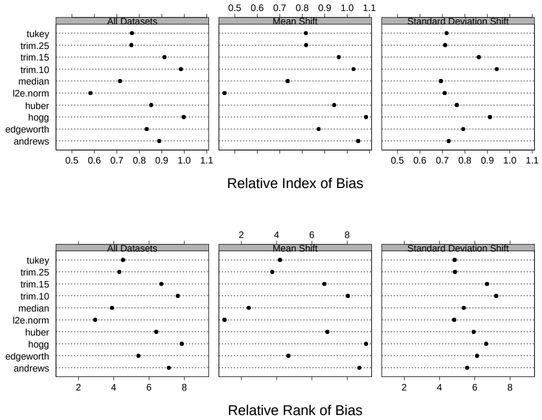

Figure 3. Relative index and relative rank of bias.

The results from the preceeding sections indicate the L_2E estimator has bias properties that offset its poor variational qualities. The graphs in Figure 3 display the bias properties of the L_2E estimator. As with the MSE analysis, the L_2E estimator has the advantage when there is a shift in the mean. When considering the standard-deviation shift, the L_2E estimator is seen to be one of the better performers.

Table 3. Relative Index/Relative Rank of Bias.

Estimator	*All*	*Standard Deviation Shift*	*Mean Shift*
Mean	1.44/9.79	1.54/9.04	1.35/10.55
Outmean	2.26/11.42	2.63/10.83	1.88/12.00

6. CONFIDENCE INTERVALS

Thinking about adding section here.

7. CONCLUSIONS

Overall, the L_2E estimator performs well when compared to the other estimators of location. While the L_2E variance is relatively large, the bias properties offset this disadvantage. We feel this study and L_2E's flexibility show promise for this robust estimator. (add more here)

[Received June 1, 1993. Revised November 2, 1993.]

REFERENCES

Andrews, D. F., Bickel, P. J., Hampel, F. R., Huber, P. J., Rogers, W. H., and Tukey, J. W., (1972), *Robust Estimates of Location: Survey and Advances,* Princeton University Press.

Basu, A., Harris, I. R., Hjort, N. L., Jones, M. C., (1998), , "Robust and efficient estimation by minimising a density power divergence," *Biometrika,* 85(3), 549–559

Edgeworth, F. Y., (1893), "Exercises in the calculation of errors," *Philosophical Magazine,* 36, 98–111.

Huber, P. J., (1981), *Robust Statistics,* John Wiley & Sons.

Hogg, R. V., (1974), "Adaptive robust procedures: A partial review and some suggestions for future applications and theory," *Journal* *of the American Statistical Association,* 69, 909–922.

Rudemo, M., (1982), "Empirical choice of histograms and kernel density estimators," *Scandinavian Journal of Statistics,* 9, 65–78.

Scott, D. W., (1999), "Parametric modeling by minimum L_2 error," submitted.

Stigler, S. M. (1977), "Do robust estimators work with real data?," *The Annals of Statistics,* 5(6), 1055–1098.

Wojciechowski, W. C., Scott, D. W., (1999) "Robust location estimation with L_2 distance," in *Proceedings of the 31st symposium on the interface* 29(1), 292–295.