Robust Location Estimation with L2 Distance

William C. Wojciechowski

Department of Statistics
Rice University
Houston, TX 77005-1892

David W. Scott

Department of Statistics
Rice University
Houston, TX 77005-1892

Abstract

Many robust location parameter estimators exist and have been studied. The performance of these estimators has been examined using theory, simulations, and real observations. This study considers the eleven robust estimators and the 24 real datasets presented by Stigler in 1977. The datasets contain measurements of physical constants. We propose two robust estimators that are based on L2 distance. The details of these methods are discussed. Using these thirteen methods, a location estimate is obtained for each of the 24 datasets. Performance is measured by the relative closeness to the true value. We then give a comparison based on the relative performance of each estimator.

1 Introduction

In a 1977 paper, Stigler [8] compares 11 robust estimators of location using 24 datasets. Each dataset contains measurements of a physical constant and performance is measured by the relative accuracy of estimating the known values. While the merits of simulations are acknowledged, for brevity those studies are omitted from this paper. Building upon Stigler's work, parametric estimators based on $L2$ distance are added to the list of robust estimators. $L2$ distance estimators are studied by Scott [7] who also coined the name L_2E ($L2$ Estimation). Due to the nature of the data, two parametric L_2E models are considered, the Normal distribution and the Gamma distribution. The results presented in this study were obtained by combining the results from Stigler's paper and the results from the two additional models. Only the necessary aspects of the study will be presented. For further details, the reader is invited to see Stigler [8].

2 Previous Estimators

Details of each of the previous 11 estimators will be omitted. Instead, brief comments and/or references are provided. The furnished citations may also be found in Stigler [8]. The details of the L_2E estimators will be presented in the next section.

The mean and median are two familiar measure of location and are included in the study. There are three versions of the trimmed mean — 10%, 15%, and 25%. The trimmed mean can be found on page 7 in Andrews et al. [1] Huber's P15, Andrews' AMT, and Tukey's Biweight may respectively be found on pages 13, 15, and 15 of Andrews et al. [1]. These three estimators are M-estimators and the details of M-estimators may be found in Huber [5]. Also included is an estimator based on a weighted average of the lower quartile, median, and upper quartile which was developed by Edgeworth [3]. The Outmean is calculated by averaging the discarded observations resulting from calculating the 25% trimmed mean. The final estimator considered by Stigler is Hogg's T1 [4].

3 L_2 Estimators

The L_2 estimators are a special case of the minimum power divergence estimator. Basu et al. [2] develop this family of M-estimators and derive some of its properties. The family is indexed by a parameter, α, which controls a trade-off between robustness and efficiency. When $\alpha = 0$, the estimator corresponds to maximum likelihood estimation and is the most efficient, but is also the least robust. L_2E is the case when $\alpha = 1$. The L_2 estimator is more robust than MLE, but is also less efficient. Basu et al. [2] suggest $0 \leq \alpha \leq 1$ and provide guidelines for selecting an appropriate value.

For completeness, the general L_2E procedure will be presented. For a more detailed discussion and further examples, see Scott [7]. We will begin by describing the general case and proceed further by expounding the details of the Normal and Gamma distributions.

L_2E is a paramater estimation method that involves the minimization of an objective function. Let $f(x;\theta)$ be the density function of a continuous random variable X that depends on a parameter θ. Let θ_0 be the true value of θ. The L_2E estimate $(\hat{\theta})$ is then defined by

$$\begin{aligned}\hat{\theta} &= \arg\min_{\theta} \int [f(x;\theta) - f(x;\theta_0)]^2 dx \\ &= \arg\min_{\theta} \int f(x;\theta)^2 dx - 2\int f(x;\theta)f(x;\theta_0)dx \\ &\quad + \int f(x;\theta_0)^2 dx.\end{aligned}$$

Noting $\int f(x;\theta_0)^2 dx$ does not depend on θ and $\int f(x;\theta)f(x;\theta_0)dx = E[f(X;\theta)]$, the minimization problem can be rewritten as

$$\hat{\theta} = \arg\min_{\theta} \int f(x;\theta)^2 dx - 2E_{\theta_0}[f(X;\theta)].$$

Since the expected value is the only quantity to depend on the true unknown parameter, θ_0, an approximation is necessary. A solution is to use a technique similar to Rudemo's [6]. The proposed substitution is

$$E_{\theta_0}[f(X;\theta)] \approx \frac{1}{n}\sum_{i=1}^{n} f(x_i;\theta),$$

where the x_i are the data points from a sample of size n. Thus, the L_2E estimator for a single parameter is

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

The value of $\hat{\theta}$ is found by using a numeric optimization routine. Applying the appropriate changes to equation (1), the L_2E method easily adapts to a multiparameter scenario where the random variable may be univariate or multivariate.

For the Normal and Gamma densities, the integral in equation (1) exists in closed form. A Normal distribution with a mean of μ and a standard deviation of σ has

$$\int f(x;\mu,\sigma)^2 dx = \frac{1}{2\sigma\sqrt{\pi}}$$

and a Gamma distribution with parameters α and β has

$$\int f(x;\alpha,\beta)^2 dx = \frac{\Gamma(\delta)\xi^{\delta}}{(\Gamma(\alpha)\beta^{\alpha})^2},$$

where $\delta = 2\alpha - 1$ and $\xi = \beta/2$

4 The Study

Using each estimator, an estimate of location is found for each of the 24 datasets. The Normal L_2E estimate of location is found by estimating both μ and σ and then using μ as the estimate. For the Gamma L_2E, both α and β are estimated and the estimate of location is taken to be $\alpha\beta$. Details of the other estimators can be found in Stigler [8].

After obtaining the estimator values, two criteria are used for measuring the relative performance of each estimator. The benchmarks are the index of relative error and the index of relative rank. Both were used by Stigler [8] and they are respectively denoted by $\bar{\varepsilon}$ and $\bar{\rho}$. These summaries are calculated for small datasets ($n \approx 20$) and for large datasets ($n > 50$). There are 20 small datasets and 4 large datasets.

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, $i = 1...13$, $j = 1...24$. The average absolute error (ω_j) is calculated for each dataset and is defined by

$$\omega_j = \frac{1}{13}\sum_{i=1}^{13} |\hat{\mu}_{ij} - \mu_j|.$$

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 either 20 or 4, the number of small and large datasets.

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 $\bar{\varepsilon}_i$.

Estimator	$\bar{\varepsilon}$	s_{ε}
Mean	0.93	0.21
Median	1.14	0.25
Edgeworth	1.01	0.09
Outmean	1.02	0.52
10% Trim	0.91	0.20
15% Trim	0.98	0.12
25% Trim	1.03	0.08
Huber P15	0.92	0.20
Andrews AMT	0.96	0.15
Tukey Biweight	1.02	0.13
Hogg T1	1.01	0.08
L_2E Normal	1.21	0.41
L_2E Gamma	0.87	0.17

Table 1: The average and standard deviation of the relative errors for the thirteen estimators. Twenty small datasets ($n \approx 20$) were used to calculate this table.

5 Results

Displayed in Table 1, are the $\bar{\varepsilon}$ values for the small datasets. An estimator's relative performance is determined by $\bar{\varepsilon}$ being less than or greater than 1. When $\bar{\varepsilon} < 1$, the estimator is performing relatively better than the others and the opposite holds for $\bar{\varepsilon} > 1$. In the second column of Table 1, are the standard deviations of the values use to calculate $\bar{\varepsilon}$. The L_2E Gamma has the lowest $\bar{\varepsilon}$ value, but the L_2E Normal is a relatively poor performer. It should also be noted that the L_2E Gamma is the only estimator with $\bar{\varepsilon} < .90$ and a relatively small s_{ε}.

Table 2 contains the $\bar{\varepsilon}$ values for the large datasets and the the trends are similar to those of Table 1. The Outmean has a relatively small value of $\bar{\varepsilon}$, but notice s_{ε} is the largest.

The index of relative ranks for the small datasets are displayed in Table 3. The second column contains the standard deviation of the values used to calculate $\bar{\rho}$. In this table, the L_2E Gamma is seen to have the lowest value of $\bar{\rho}$ and its s_{ρ} is moderate. Both $\bar{\rho}$ and s_{ρ} are relatively large for the L_2E Normal.

Presented in Table 4, are the index of relative ranks for the large datasets. The Edgeworth estimator is the clear winner with low $\bar{\rho}$ and s_{ρ} values. The L_2E Gamma is seen to be a moderate performer. Having large values for both $\bar{\rho}$ and s_{ρ}, the L_2E Normal is seen to be a relatively poor performer.

Estimator	$\bar{\varepsilon}$	s_{ε}
Mean	0.92	0.20
Median	1.14	0.17
Edgeworth	0.94	0.08
Outmean	0.77	0.50
10% Trim	0.94	0.06
15% Trim	0.98	0.04
25% Trim	1.07	0.10
Huber P15	0.98	0.04
Andrews AMT	1.02	0.13
Tukey Biweight	1.09	0.16
Hogg T1	1.08	0.12
L_2E Normal	1.22	0.29
L_2E Gamma	0.86	0.19

Table 2: The average and standard deviation of the relative errors for the thirteen estimators. Four large datasets ($n > 50$) were used to calculate this table.

Estimator	$\bar{\rho}$	s_{ρ}
Mean	5.98	3.61
Median	8.48	4.49
Edgeworth	7.62	3.25
Outmean	6.05	5.73
10% Trim	5.58	2.46
15% Trim	6.95	1.86
25% Trim	7.85	3.15
Huber P15	6.10	3.09
Andrews AMT	7.28	2.94
Tukey Biweight	7.80	3.35
Hogg T1	8.02	2.69
L_2E Normal	8.60	4.94
L_2E Gamma	4.70	3.69

Table 3: The average and standard deviation of the relative ranks for the thirteen estimators. Twenty small datasets ($n \approx 20$) were used to calculate this table.

Estimator	$\bar{\rho}$	s_ρ
Mean	7.50	5.20
Median	9.25	3.59
Edgeworth	4.75	0.96
Outmean	7.00	6.93
10% Trim	5.25	1.89
15% Trim	6.50	1.73
25% Trim	6.88	4.01
Huber P15	6.00	1.63
Andrews AMT	6.50	3.70
Tukey Biweight	8.00	4.32
Hogg T1	8.12	2.95
L_2E Normal	8.75	5.32
L_2E Gamma	6.50	5.20

Table 4: The average and standard deviation of the relative ranks for the thirteen estimators. Four large datasets ($n > 50$) were used to calculate this table.

6 Summary and Closing Remarks

In comparing the L_2E estimators to the previous estimators considered by Stigler [8], the L_2E Gamma is seen to produce relatively accurate values and the L_2E Normal is seen to be a poor performer. Since both densities have two parameters, the ability of the Gamma density to model skewness of the data may explain its excellent performance. However, the results are limited to these datasets.

This study shows promise for the L_2E method. Moreover, this proposed method is easily adapted to more complicated estimation problems which makes it a viable alternative estimation procedure.

Acknowledgements

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.

References

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

[2] Ayanendranath Basu, Ian R. Harris, Nils L. Hjort, and M. C. Jones. Robust and efficient estimation by minimising a density power divergence. *Biometrika*, 85(3):549–559, 1998.

[3] F. Y. Edgeworth. Exercises in the calculation of errors. *Philosophical Magazine*, 36:98–111, 1893.

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

[5] Peter J. Huber. *Robust Statistics.* John Wiley & Sons, New York, 1981.

[6] M. Rudemo. Empirical choice of histograms and kernel density estimators. *Scandinavian Journal of Statistics*, 9:65–78, 1982.

[7] David W. Scott. Parametric modeling by minimum l_2 error. *submitted*, 1999.

[8] Stephen M. Stigler. Do robust estimators work with real data. *The Annals of Statistics*, 5(6):1055–1098, 1977.