Statistics & Probability Letters 18 (1993) 241–244
North-Holland

15 October 1993

A note on robust estimation of location

Probal Chaudhuri and Debapriya SenGupta
Division of Theoretical Statistics and Mathematics, Indian Statistical Institute, Calcutta, India

Received October 1992
Revised December 1992

Abstract: A modified version of the usual M-estimation problem is proposed, and sample median is shown to be a solution of this problem for a wide range of choices of the score function. It exposes certain universality in the robustness of sample median in the univariate case, and this property continues to hold even in multivariate set-ups if we consider the multivariate L_1-median. Some interesting facts related to this 'modified M-estimation' are discussed, and the consequences of a similar modification of the traditional maximum likelihood approach are explored.

Keywords: Modified M-estimation; score function; median unbiasedness; multivariate L_1-median; modified maximum likelihood.

1. Introduction

A well-known robustness property of sample median as an estimate of univariate location is its 50% breakdown point. Kemperman (1987) showed that this property is retained by the L_1-median of any finite measure on a Banach space. In particular, it implies that the spatial median (or the mediancenter), which has been considered by several people (e.g. Haldane, 1948; Gower, 1974; Brown, 1983, Ducharme and Milasevic, 1987; etc.) analyzing multivariate data, also has 50% breakdown point. While exploring the distributional robustness of statistical procedures, Huber (1981, Chapter 4, section 4.2) noted that for models arising from contaminations of symmettric unimodal densities, univariate median is the universal solution for the 'minimax bias problem'. Bassett (1991) observed that univariate median is the only location estimate that is affine equivariant, monotonic, and has 50% breakdown point, and other estimates of univariate location can have at most any two of these three properties. In this note, we will establish a property of sample median that holds in univariate as well as multivariate set-ups and exposes a new aspect of robustness in this popular location estimate. This property of median has some intriguing connections with M-estimation techniques that are extensively discussed in robust estimation literature. There are many different ways of defining the median of a multivariate distribution. However, throughout this note, we will restrict ourselves to only the L_1-median.

Correspondence to: Probal Chaudhuri, Division of Theoretical Statistics and Mathematics, Indian Statistical Institute, 203 B.T. Road, Calcutta 700035, India.

2. Modified M-estimation problem and sample median as its universal solution

Suppose that we have a set of i.i.d. d-dimensional observations $X_1, \ldots, X_n$, and consider the standard M-estimation set-up in which a location estimate $\hat{\theta}_n$ is defines as

$$\sum_{i=1}^{n} \rho\left(X_i - \hat{\theta}_n\right) = \min_{\theta \in \mathbb{R}^d} \sum_{i=1}^{n} \rho(X_i - \theta). \qquad (1)$$

Here ρ is an appropriate nonnegative loss function. Alternatively, the M-estimate $\hat{\theta}_n$ can be defined implicitly as a solution of the estimating equation

$$\sum_{i=1}^{n} \psi(X_i - \theta) = 0, \tag{2}$$

where ψ (the score function) can be viewed as the derivative of ρ. Note that for $d > 1$, both sides of equation (2) are d-dimensional vectors, and ψ represents the usual gradient vector of ρ. Well-known examples of such M-estimates are maximum likelihood estimates based on location models and the usual sample mean, the latter being a solution of (1) when ρ is the squared error loss. In sharp contrast with sample median, sample mean has 0% breakdown point, and this lack of robustness (i.e. a high degree of susceptibility to even a single outlier) can be attributed to the unboundedness of the ψ function associated with the squared error loss (see Hampel, 1974; Huber, 1981, Chapter 1, sections 1.4 and 1.5). In an attempt to robustify M-estimates of location, various bounded versions of the ψ function have been tried in the literature. Some illuminating discussions on several estimates obtained as solutions of (2) based on bounded ψ can be be found in Huber (1981) and Hampel, Rousseeuw, Ronchetti and Stahel (1986). Stigler (1980), who gave a fascinating historical review of robust M-estimates, mentioned about Daniell Bernoulli's recommendation for certain bounded ψ functions to construct location estimates even in the eighteenth century! However, some common technical difficulties associated with all popular M-estimates constructed via equation (2) using bounded ψ functions are: (a) computational complexities caused by the nonlinear nature of equation (2), (b) the need for having an appropriate initial estimate of the scale parameter as bounded versions of ψ typically depend on the scale, and (c) the lack of objective guidelines regarding the choice of ψ. These problems turn out to be particularly critical and hard to overcome in a multivariate set up.

Suppose now that we have a univariate location problem so that θ is real valued, and let us rewrite equation (2) as

$$\frac{1}{n} \sum_{i=1}^{n} \psi(X_i - \theta) = \operatorname*{mean}_{1 \leqslant i \leqslant n} \psi(X_i - \theta) = 0. \tag{3}$$

If we replace 'mean' by its long time competitor 'median' in (3), we get the following 'modified M-estimation equation'.

$$\operatorname*{median}_{1 \leqslant i \leqslant n} \psi(X_i - \theta) = 0. \tag{4}$$

Note at this point that all of the popular ψ functions that are used in the construction of robust M-estimates of univariate location satisfy $\psi(0) = 0$, $\psi(t) > 0$ only if $t > 0$ and $\psi(t) < 0$ only if $t < 0$. As a matter of fact, frequently used ψ's are anti-symmetric functions. Hence, irrespective of the specific form of ψ, a solution of (4) is

$$\hat{\theta}_n = \operatorname*{median}_{1 \leqslant i \leqslant n} X_i.$$

Consider next multivariate observations and the minimization problem

$$\min_{\theta \in \mathbb{R}^d} \frac{1}{n} \sum_{i=1}^{n} \rho(|X_i - \theta|) = \min_{\theta \in \mathbb{R}^d} \operatorname*{mean}_{1 \leqslant i \leqslant n} \rho(|X_i - \theta|). \tag{5}$$

Here $|\cdot|$ denotes the usual Euclidean norm, and ρ is a smooth increasing and nonnegative loss function defined on the interval $[0, \infty)$. Instead of (5), as noted at the beginning of this section, one can also work with the estimating equation

$$\frac{1}{n} \sum_{i=1}^{n} \rho'(|X_i - \theta|) U(X_i - \theta) = \operatorname*{mean}_{1 \leqslant i \leqslant n} \rho'(|X_i - \theta|) U(X_i - \theta) = 0. \tag{6}$$

Here ρ' is the derivative of ρ, and for $x \in \mathbb{R}^d$, we define $U(x) = |x|^{-1} x$ if $x \neq 0$ and $U(0) = 0$. Now, if (6) is modified as

$$\operatorname*{median}_{1 \leqslant i \leqslant n} \rho'(|X_i - \theta|) U(X_i - \theta) = 0, \tag{7}$$

where 'median' means the L_1-median mentioned in the Introduction, we get a multivariate version of the 'modified M-estimation equation'. Kemperman observed that $Y \in \mathbb{R}^d$ is an L_1-median (which is unique in dimensions $d \geqslant 2$ unless all data points fall on a single straight line) of a set of observed data points $Z_1, \ldots, Z_n$ in $\mathbb{R}^d$, if and only if one of the following two conditions holds.

(a) $Y = Z_i$ for some $1 \leqslant i \leqslant n$ and

$$|U(Z_1 - Y) + \cdots + U(Z_n - Y)| \leqslant 1.$$

(b) $Y \neq Z_i$ for all $1 \leqslant i \leqslant n$ and

$|U(Z_1 - Y) + \cdots + U(Z_n - Y)| = 0.$

This fundamental observation immediately implies that the location estimate

$$\hat{\theta}_n = \underset{1 \leqslant i \leqslant n}{\text{median}}\ X_i$$

is again a solution of (7) irrespective of the specific form of ρ. Note that

$$U\{\rho'(|X_i - \theta|)U(X_i - \theta)\} = U(X_i - \theta)$$

because $\rho'(|X_i - \theta|)$ is a real valued object acting as a scalar multiple here. In particular, if ρ arises as the negative log-likelihood associated with a location problem (univariate or multivariate), the median will be a solution of the 'modified maximum likelihood problem' for a wide range of popular location models. In a sense, what we are observing here points at a finite sample 'median unbiasedness property' (i.e. the fact that $\psi(X_i - \theta)$'s, where $1 \leqslant i \leqslant n$, have their median at the origin whenever $(X_i - \theta)$'s have theirs at the origin) that holds for all commonly used score functions.

3. Some concluding remarks

Instead of replacing (3) by (4) or (6) by (7) in the previous section, one can also try to construct location estimates by solving the minimization problem

$$\min_{\theta \in \mathbb{R}^d}\ \underset{1 \leqslant i \leqslant n}{\text{median}}\ \rho(|X_i - \theta|).$$

Clearly, the above is a modification of (5). In view of the monotonicity of ρ on the interval $[0, \infty)$ assumed in the previous section, estimates that solve this minimization problem are 'LMS' ('least median of squares') type estimates (see e.g. Rousseeuw and Leroy, 1987), and they have some close connections with the 'SHORTH' estimate of location (see e.g. Andrews et al., 1972). Such estimates have received a fair amount of attention in the literature, and it is well-known that they are highly unstable in nature with a much slower rate of convergence (see Rousseeuw, 1984; Shorack and Wellner, 1986; and Kim and Pollard, 1990) than the sample median, which is typically $n^{1/2}$-consistent.

In the case of i.i.d. observations $X_1, \ldots, X_n$ with a common smooth density/mass function $f(x \mid \theta)$, where θ is a real valued unknown (not necessarily location) parameter of interest, our 'modified maximum likelihood equation' takes the form

$$\underset{1 \leqslant i \leqslant n}{\text{median}} \frac{f'(X_i \mid \theta)}{f(X_i \mid \theta)} = 0. \tag{8}$$

Here f' denotes the derivative of f w.r.t. θ. With a fixed x, frequently occurring examples of the function $f(x \mid \theta)$ (the model likelihood) has a maximum at $\theta = t(x)$ (say) such that $f(x \mid \theta)$ is increasing in θ for $\theta < t(x)$ and decreasing in θ for $\theta > t(x)$. It is then obvious from our previous discussion that for such an f, a solution of (8) is

$$\hat{\theta}_n = \underset{1 \leqslant i \leqslant n}{\text{median}}\ t(X_i).$$

Note that $t(X_i)$ is nothing but the usual maximum likelihood estimate of θ based on the single observation X_i, and in many practical situations, it will be quite easy to compute it. For instance, it happens to be so if θ is the location parameter in a location model, or if it is the scale parameter in the gamma or the Weibull model with a known shape parameter. It is now immediate that for a large class of models, the solution of the 'modified maximum likelihood problem' is the straight forward median of the solutions of the standard maximum likelihood problems associated with different data points. Further, in view of the monotonicity assumed in the model likelihood, whenever the score function $\{f'(X_i \mid \theta)\}\{f(X_i \mid \theta)\}^{-1}$ has a continuous distribution with a unique median at zero, the distribution of $t(X_i)$ will have a unique median at θ. Hence, if $t(X_i)$ has an absolutely continuous distribution with density g, the 'modified maximum likelihood estimate' $\hat{\theta}_n$ will be asymptotically normally distributed with mean $= \theta$ and variance $= (4n)^{-1}\{g(\theta)\}^{-2}$.

Acknowledgement

The authors are thankful to the referee for a careful and quick reading of the first draft. Comments from the referee were helpful in preparing the revision.

References

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

Bassett, G.W. (1991), Equivariant, monotonic, 50% breakdown estimators, *Amer. Statist.* **45**, 135–137.

Brown, B.M. (1983), Statistical use of the spatial median, *J. Roy. Statist. Soc. Ser. B* **45**, 25–30.

Ducharme, G.R. and P. Milasevick (1987), Spatial median and directional data, *Biometrika* **74**, 212–215.

Gower, J.C. (1974), The mediancenter, *Appl. Statist.* **23**, 466–470.

Haldane, J.B.S. (1948), Note on the median of a multivariate distribution, *Biometrika* **35**, 414–415.

Hampel, F.R. (1974), The influence curve and its role in robust estimation, *J. Amer. Statist. Assoc.* **69**, 383–393.

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

Huber, P.J. (1981), *Robust Statistics* (Wiley, New York).

Kemperman, J.H.B. (1987), The median of a finite measure on a Banach space, in: Y. Dodge, ed. *Statistical Data Analysis Based on the L_1-Norm and Related Methods* (North-Holland, Amsterdam) pp. 217–230.

Kim, J. and D. Pollard (1990), Cube root asymptotics, *Ann. Statist.* **18**, 191–219.

Rousseeuw, P.J. (1984), Least median of squares regression, *J. Amer. Statistic. Assoc.* **79**, 871–880.

Rousseeuw, P.J. and A.M. Leroy (1987), *Robust Regression & Outlier Detection* (Wiley, New York).

Shorack, G.R. and J.A. Wellner (1986), *Emperical Processes with Applications to Statistics* (Wiley, New York).

Stigler, S.M. (1980), Studies in the history of probability and statistics XXXVIII: R.H. Smith, a Victorian interested in robustness, *Biometrika* **67**, 217–221.