

Some Perspectives on Location and Scale Depth Functions

Invited discussion of “Location-Scale Depth”, I. Mizera and C. H. Müller, *JASA*, 2004

Robert Serfling*

University of Texas at Dallas

May 2004

Mizera and Müller are to be congratulated heartily for their thoughtful articulation of an intriguing new approach to univariate location and scale estimation. Applying notions of statistical depth functions, an emerging methodology, and linking in a striking way with certain non-Euclidean geometries over the relevant bivariate parameter space, they develop novel “Student depth” location and scale estimators that satisfy quite favorable properties although imposing “considerable training” on the practical user. Mizera and Müller have opened up important new vistas on an “old” problem.

To place the new estimators in perspective, I discuss connections with some familiar points of reference, especially “scores” and “M-estimation”. It is seen, for example, that the new estimators may be represented as the solution of a simple modification of the usual (vector) likelihood equation for the location and scale MLE’s, where “average” is replaced by “bivariate halfspace median”. Various questions emerge, such as: are the new location and scale estimators really very different from two classical estimators, the *median* and *MAD*?

THE SETTING

Depth functions in the sample space originated with the notion of *halfspace depth* (Tukey, 1975), which has become much-studied as a tool in nonparametric multivariate location inference. Subsequently, distinctive other versions of “location” depth and many theoretical and methodological tools have been developed (for details, see Liu, Parelius and Singh,

*Department of Mathematical Sciences, University of Texas at Dallas, Richardson, Texas 75083-0688, USA. Email: serfling@utdallas.edu. Website: www.utdallas.edu/~serfling. Support by NSF Grant DMS-0103698 is gratefully acknowledged.

1999, Zuo and Serfling, 2000, and Mosler, 2002, for example). Innovative extension to univariate multiple regression was accomplished by Rousseeuw and Hubert (1999), whose “regression depth” defined in the (multivariate) *parameter space of regression “fits”* reduces for the location model to an equivalent formulation of halfspace location depth. This was generalized to a wide range of models by Mizera (2002), who introduced equivalent versions of halfspace depth on suitably regular parameter spaces, in particular a notion of “tangent depth”, and employed a “critical functions” approach. The present paper applies this foundation utilizing critical functions of log-likelihood type based on an assumed model shape f . In their equation (3) the authors formulate an f -based “location-scale likelihood depth” $d(\mu, \sigma)$, which although parametric in its definition is used “operationally” in a nonparametric sense. The case of Gaussian f yields the “Student depth”, the central object of the paper under discussion.

REPRESENTATION IN TERMS OF “SCORES”

Characterization of statistical procedures in terms of “scores” is a leading motif in statistical inference. Interpretation or comparison of procedures in terms of the form of the scores and the manner in which they are used carries strong explanatory power and intuitive appeal. Indeed, the authors’ official definition of $d(\mu, \sigma)$ in their equation (4) is a re-expression of their equation (3) in terms of the f -based “score functions” that define the usual likelihood equations for joint estimation of location and scale. While this aspect does not receive any special role in the further development of the paper, it leads to a productive characterization of the Student depth location and scale estimators. Denote the relevant score functions by $\psi(\tau) = -f'(\tau)/f(\tau)$ and $\chi(\tau) = \tau\psi(\tau) - 1$ (this $\chi(\cdot)$ shifts the authors’ by -1). For univariate data y_i , $1 \leq i \leq n$, and model f , the corresponding data-based “scores” are

$$\begin{pmatrix} \psi(\tau_i) \\ \chi(\tau_i) \end{pmatrix}, \quad 1 \leq i \leq n, \quad (1)$$

where $\tau_i = (y_i - \mu)/\sigma$, $1 \leq i \leq n$, denote the observations standardized by μ and σ , and the usual corresponding (vector) likelihood equation for μ and σ is

$$\sum_{i=1}^n \begin{pmatrix} \psi(\tau_i) \\ \chi(\tau_i) \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}. \quad (2)$$

Standardization by “true values” makes, of course, the score vector an unbiased estimate of the origin, and, accordingly, the usual MLE’s are the empirical standardization that makes the *sample mean score vector* equal to the origin.

Now note that $d(\mu, \sigma)$ in the authors’ equation (4) may be expressed in this framework as the *halfspace location depth of the origin* $(0, 0)$ in the set of points (1). Consequently, the “maximal location-scale depth estimators” of (μ, σ) are given by that standardization of the

observations which makes the *sample (bivariate) halfspace median* of the score vectors equal to the origin. That is, the relevant equation for these estimators may be re-expressed as

$$\text{halfspace median } \left\{ \begin{pmatrix} \psi(\tau_i) \\ \chi(\tau_i) \end{pmatrix} \right\} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \quad (3)$$

which simply modifies the likelihood equation (2) by replacing “average” with “halfspace median”. This is one way to see how the new estimators relate to other types, especially the MLE’s.

Of course, while (3) provides an informative perspective, its use for actual computation of the solution presents a complicated inverse problem due to the scores changing with the parameters. This difficulty is alluded to in the discussion of Student depth via the authors’ equation (10), and their solution is to transform the problem to facilitate explicit computation of the Student depth location and scale estimators via bivariate halfspace depth defined on the Klein disk \mathbb{KD} .

VARIATIONS ON THE THEME

The above perspective points the way, via (3), to a formulation of natural competitors to the Student depth estimators by changing the choice of bivariate median, or the choice of scores, or both. Perhaps some of these cases merit detailed investigation.

Other choices of bivariate median. The *spatial median* long-standing in the literature, or the *simplicial median* of Liu (1988), for example, might just as well be used in (3). In particular, with the spatial median (3) becomes like (2) in form,

$$\sum_{i=1}^n \begin{pmatrix} \psi(\tau_i) \\ \chi(\tau_i) \end{pmatrix} / \left\| \begin{pmatrix} \psi(\tau_i) \\ \chi(\tau_i) \end{pmatrix} \right\| = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \quad (4)$$

and thus computationally similar. It may be fruitful to explore for various choices of bivariate median how (3) converts to a corresponding depth problem in the parameter space of (μ, σ) in an alternative geometry, just as the authors have shown (beautifully) for the bivariate halfspace median utilizing the Lobachevski geometry. (They too consider alternatives to the halfspace depth, but in the context of comparison to depths defined over \mathbb{KD} .)

Other choices of score functions. The f -based “likelihood scores” in (1) are adopted in the paper for specific choices of f in an “operational” rather than a parametric modeling sense, as discussed in Section 4. Thus “Student depth”, the focus of the paper, is based on the scores for Gaussian f ,

$$\begin{pmatrix} \tau_i \\ \tau_i^2 - 1 \end{pmatrix}, \quad 1 \leq i \leq n, \quad (5)$$

but is emphasized for its tractability and proposed as a tool having relevance beyond the Gaussian model. For example, as the authors discuss in Section 6, the maximal Student

depth location and scale estimators happen to estimate, in the case of a random sample from Cauchy f , the same population parameters as the Cauchy-based MLE's: the *population median and mean absolute deviation (MAD)*, respectively. Yet, if a broad nonparametric approach is perhaps the goal, then it becomes of interest to consider scores in (1) employing $\psi(\cdot)$ and $\chi(\cdot)$ chosen in certain other ways as well, for example for intuitive appeal or to satisfy minimax-type criteria. With the use of the estimating equation (2), this becomes simply a general formulation of *simultaneous M-estimation* of location and scale (Huber, 1981). On the other hand, if the estimating equation (3) is used instead, then each choice of $\psi(\cdot)$ and $\chi(\cdot)$ generates alternatives to these corresponding M-estimators.

A class of estimators. Thus arises, based on (3), a very broad class of estimators which may be viewed as competitors to the class of M-estimators as a natural point of reference. We inquire how these estimators, especially the Student depth, compare to M-estimators. Broad investigation of the use of (3) versus (2) may be of interest. Can we have an *influence function analysis* as with M-estimates? Can (3) be solved *iteratively*, as an alternative to transforming to \mathbb{KD} ?

STUDENT MEDIAN VERSUS (MEDIAN, MAD)

A classical example of “nonparametric” scores,

$$\psi_0(\tau) = \text{sign}(\tau), \tag{6}$$

$$\chi_0(\tau) = \text{sign}(|\tau| - 1), \tag{7}$$

yields, when used in conjunction with the M-estimation equation (2), well-known and easily understood robust location and scale M-estimators, the *median* and the *MAD*. Can we improve upon these by using equation (3) instead of (2)? Would simplicity be sacrificed without a satisfactory gain in either robustness or efficiency?

Indeed, in this vein, it is of particular interest to compare specifically the Student median and the (median, MAD). For, as the authors point out in Section 3, simplification of the Student depth to location likelihood depth holding scale fixed yields the *median* as maximal location depth estimator, and simplification to scale likelihood depth with location fixed yields the *MAD* as maximal scale depth estimator. This suggests that the Student median *joint* estimators of location and scale might be close variants, perhaps complicated, of the median and MAD. We ask: in what way does the Student median take us beyond just using the median and MAD? Is its performance essentially the same?

SCORE-BASED DEPTH FUNCTIONS

Another way to connect depth-based methods and classical M-estimation is through score-based depth functions. Let ψ and χ be score functions as considered above for general M-estimation of location and scale, respectively, with ψ odd and χ even, and let $m(\cdot)$ and $s(\cdot)$ be

location and scale equivariant functionals, respectively, defined on univariate distributions F . Then corresponding M-estimates of location and scale are the solutions μ and σ , respectively, of the equations

$$\sum_{i=1}^n \psi \left(\frac{y_i - \mu}{s(F_n)} \right) = 0, \quad (8)$$

$$\sum_{i=1}^n \chi \left(\frac{y_i - m(F_n)}{\sigma} \right) = 0, \quad (9)$$

where F_n denotes the usual empirical distribution. In this context, Zhang (2002) defines an associated *outlyingness measure* for μ with respect to data $Y = (y_1, \dots, y_n)$,

$$O_{s\psi 1n}(\mu, Y) = \left| n^{-1} \sum_{i=1}^n \psi \left(\frac{y_i - \mu}{s(F_n)} \right) \right|, \quad (10)$$

and a corresponding *location depth* function, $D_{s\psi 1n}(\mu, Y) = (1 + O_{s\psi 1n}(\mu, Y))^{-1}$. Likewise, he defines a natural outlyingness measure for σ with respect to $\{y_1 - m(F_n), \dots, y_n - m(F_n)\}$ by

$$\tilde{O}_{m\chi 1n}(\sigma, Y) = \left| n^{-1} \sum_{i=1}^n \chi \left(\frac{y_i - m(F_n)}{\sigma} \right) \right|, \quad (11)$$

with $\tilde{D}_{m\chi 1n}(\sigma, Y) = (1 + \tilde{O}_{m\chi 1n}(\sigma, Y))^{-1}$ a corresponding *scale depth* function. Thus makes the ψ -, χ -, m -, and s -based M-estimates interpretable as *maximal depth estimators*.

In particular, the case $\psi_0(z)$ given by (6) gives the univariate halfspace depth with solution $\mu =$ the median, independently of the choice of $s(\cdot)$. The case $\chi_0(z)$ given by (7), along with $m(F_n) =$ the median, yields the solution $\sigma =$ the MAD. Thus (Med, MAD) turn up as depth-based, using separate score-based depth functions for location and scale. In this example the depth function for scale is centered using the maximal depth point of the depth function for location. This coherency is a reasonable rule to follow in general.

Is it perhaps more natural even, as well as more convenient, to estimate location and scale using separate depth functions? For any odd translation equivariant statistic V and even translation invariant statistic W must be uncorrelated. We thus expect estimators for location and scale to extract different pieces of information from the data, and so, in this special case of “bivariate” parameter, it seems quite natural to use componentwise approaches.

EXTENSION TO MULTIVARIATE DATA

The authors mention without details “straightforward but technically somewhat demanding” extension of the Student depth to the multivariate location-scale model. Presumably this

is again to involve a system of f -based likelihood equations for a location vector $\boldsymbol{\mu}$ and a dispersion matrix \mathbf{C} . As an alternative approach, again following Zhang (2002), the above location and scale outlyingness functions for univariate data extend very readily to data $Y^{(d)} = (\mathbf{y}_1, \dots, \mathbf{y}_n)$ in \mathbb{R}^d via *projection pursuit*, taking the maximum outlyingness over all one-dimensional projections. Thus, for location, outlyingness of a vector $\boldsymbol{\mu}$ is given by

$$O_{s\psi dn}(\boldsymbol{\mu}, Y^{(d)}) = \sup_{\|\mathbf{a}\|=1} O_{s\psi 1n}(\mathbf{a}'\boldsymbol{\mu}, \mathbf{a}'Y^{(d)}). \quad (12)$$

Likewise, for dispersion, outlyingness of a covariance matrix \mathbf{C} is given by

$$\tilde{O}_{m\chi dn}(\mathbf{C}, Y^{(d)}) = \sup_{\|\mathbf{a}\|=1} \tilde{O}_{m\chi 1n}(\sqrt{\mathbf{a}'\mathbf{C}\mathbf{a}}, \mathbf{a}'Y^{(d)}). \quad (13)$$

Corresponding depth functions are again defined by $D = (1 + O)^{-1}$, and the associated maximal depth values now may be characterized as *minimax estimators*. In particular, ψ_0 given by (6) now yields the halfspace depth in (the location parameter space) \mathbb{R}^d .

This projection pursuit approach parallels that of Donoho and Gasko (1992), who define location outlyingness in \mathbb{R}^d by

$$O_{dn}(\boldsymbol{\mu}, Y^{(d)}) = \sup_{\|\mathbf{a}\|=1} O_{1n}(\mathbf{a}'\boldsymbol{\mu}, \mathbf{a}'Y^{(d)}), \quad (14)$$

using the classical univariate location outlyingness measure of Mosteller and Tukey (1977) for $Y = (y_1, \dots, y_n)$,

$$O_{1n}(\mu, Y) = \left| \frac{\mu - \text{Med}(Y)}{\text{MAD}(Y)} \right| = \left| \text{median} \left\{ \frac{y_1 - \mu}{\text{MAD}(Y)}, \dots, \frac{y_n - \mu}{\text{MAD}(Y)} \right\} \right|, \quad (15)$$

with $\text{Med}(Y) = \text{median}\{y_i, 1 \leq i \leq n\}$ and $\text{MAD}(Y) = \text{median}\{|y_i - \text{Med}(Y)|, 1 \leq i \leq n\}$.

For the “projection” depth function corresponding to (14), the corresponding *projection median* has asymptotic breakdown point 1/2 (Tyler, 1994), whereas the *halfspace median* has asymptotic breakdown point only 1/3 (Donoho and Gasko, 1992). Both estimators have only moderate relative efficiency in comparison with the sample mean under the normal model, as seen from simulation studies of Zhang (2002) and Zuo (2003), each of whom introduces modified estimators designed to attain higher efficiency while also having breakdown point 1/2. In particular, Zuo (2003) replaces (Med, MAD) in (15) by choices that are still robust but more efficient, while Zhang (2003) constructs smoothed versions of halfspace outlyingness in \mathbb{R}^d and then defines a class of outlyingness functions which are weighted sums of these versions and the Donoho and Gasko (1992) version (14).

These considerations lead to questions in two directions. (i) How does the authors’ intended multivariate extension compare with the above minimax-type approach dealing separately with $\boldsymbol{\mu}$ and \mathbf{C} , especially using a robust and efficient depth function for location?

(ii) Is there a case for replacing the bivariate halfspace median defining Student depth in equation (3) by a Zhang (2002) or Zuo (2003) version that retains the desired equivariance properties while being more robust and more efficient? Would tractability be compromised unduly without sufficient gain? How important is it to preserve “transform to Klein—calculate depth—transform back” as a feature of the estimation method?

SUMMARY

The nature of the authors’ maximal likelihood location-scale depth estimators has been placed in some perspective through comparisons, via (3), with location and scale MLE’s, with the class of M-estimators in general, and with a wide class of alternative versions defined by substituting other bivariate depths for the bivariate halfspace depth. Closeness of the authors’ Student median to the classical (median, MAD) is conjectured. Potential for improvements in both the robustness and the efficiency of the Student median has been indicated. A score-based location and scale depth approach that extends easily via projection pursuit to the case of multivariate data has been described. Various questions and issues have been raised.

The Student depth and concomitant details present a provocative new research direction with important open issues to be resolved. Continuing investigation should yield insights and results of broad interest. Mizera and Müller are to be thanked for a significant contribution that expands the horizons of statistical science in a refreshing and stimulating style.

ADDITIONAL REFERENCES

- Mosler, K. (2002). *Multivariate Dispersion, Central Regions and Depth*. Springer, New York.
- Mosteller, C. F. and Tukey, J. W. (1977). *Data Analysis and Regression*. Addison-Wesley, Reading, Mass.
- Zhang, J. (2002). Some extensions of Tukey’s depth function. *Journal of Multivariate Analysis* **82** 134–165.
- Zuo, Y. (2003). Projection-based depth functions and associated medians. *Annals of Statistics* **31** 1460–1490.