COMMENT ON: LOCATION-SCALE DEPTH

XUMING HE[1] AND STEPHEN PORTNOY[1]

Mizera (2002) developed an innovative unifying approach to likelihood depth for parametric models. The present work by Mizera and Müller further illustrates the power of this approach for location and scale models. We applaud their efforts to consider plausible diagnostics and data analytic procedures based on depth and look forward to an ever increasing understanding of the statistical implications of depth. In this discussion, we would like to raise several questions, primarily concerning interpretation.

1. It is useful to distinguish two slightly different concepts of depth. One is the depth of a point (usually a data point) relative to a given sample of many data points. The other is the depth of a fit, usually represented by a set of parameters, given a sample of observations. In a regular location model, the parameter space and the sample space coincide, and therefore these two notions of depth need not be distinguished. In a more general setting, we think that it is helpful to note the differences. For convenience, let's call them as data depth and parameter depth, respectively.

A salient point to make is that the traditional notion of data depth assumes no explicit parametric structures. For example, Tukey's half-space depth provides an outlyingness measure of any point in the sample space without imposing any structure on the data-generating mechanism. Some authors may have a multivariate location-scatter model in mind when they use data depth of this type, but it is not required at all. Donoho and Gasko (1992) considered both half-space depth and projection depth, and pointed out that the depth contours can reveal the structure of the multivariate distribution, elliptically symmetric or not. On other hand, the parameter depth, a focal point of Mizera and Muller, makes sense only when "a fit" is being evaluated. In theory, we can talk about a nonparametric or semiparametric fit to the data.

---

[1] Professors, Department of Statistics, University of Illinois at Urbana-Champaign, corresponding email: `x-he@uiuc.edu`

Date: May 11, 2004.

Research of both authors partially supported by NSF Grant: DMS-0102411 and by NSA Grant H98230-04-1-0034 .

The regression depth of Rousseeuw and Hubert (1999) does this in the sense that the maximal regression depth estimator is consistent for the conditional median as long as it is linear, *irrespective* of the actual distribution. Thus, we wonder, driven by intellectual curiosity, whether the likelihood-based depth will lead to something interesting if the fit is semiparametric or nonparametric.

2. When considering parametric models, we find it helpful to think about structured data problems. Linear models came to our mind naturally in this direction. In this case, data depth appears less meaningful than parameter depth. If we think about a related notion of ranking, then the literature on quantile regression suggests that we can rank regression fits determined by p data points, where p is the dimensionality of a regression fit, but not an individual data point. If we apply the parameter depth as in the location-scale model, should we include a scale parameter? If the data are heteroscedastic, is one scale parameter sufficient to be meaningful?

For general regression modelling (linear or not), model adequacy is often assessed using residuals. Thus, it would seem reasonable to apply location-scale depth to the residuals, especially if new diagnostics can be developed. Several questions arise. Is there a way to use the location-scale depth plot to assess simultaneously bias (presumably through the location part) and variability (presumably through the scale part)? Can heteroscedasticity (or more general forms of heterogeneity) by detected? Can other measures of location-scale depth be used to assess regression models? Positive answers for these questions would be valuable.

3. In order to see what simultaneous consideration of location and scale offers, it might be useful to consider their implications separately. One question is what information does the scale depth provide that wouldn't be available in, say, a quantile plot. Another concerns the location parameter. For a univariate location model, the parameter depth of a location parameter is simple: the sample median is generally (universally?) taken as the deepest point. The fact that the location component of the Student median from the location-scale depth is different from the median (*e.g.*, see figures 7 and 8) seems rather intriguing. Though the difference often seems small enough so that data with modest sample size might not be able to distinguish it, we would still like to understand what this difference indicates.

Specifically, since the normal likelihood is basic to location-scale depth, we tried to find out whether the location component is providing some measure of fit to normality – or at least to a family of bell-shaped distributions. One apparent feature we looked at is the discrepancy between the Student median, *i.e.*, the deepest point in the (μ, σ) space, and the deepest location value (μ at each σ).

In an effort to get some rough ideas of what is being measured, we did some plots based on the R functions kindly provided by Ivan Mizera. We chose a range of what we thought might be suggestive distributions: standard normal, Laplace (double exponential), some gamma's ($\alpha = 1$, which is the negative exponential, and $\alpha = .2$), some beta's (as in the figures, with parameters: (.2, 1) and (.2, .2)), and some "normal mixtures" including 75: $N(-2, 1)$ & 25: $N(2, 1)$ and .75 $N(-3, 1)$ + .25 $N(3, 1)$. The first mixture is split exactly 75 - 25 with means 2 and -2 (sometimes called a "contamponents" model), while the second is a random mixture having mean -3 with probability 3/4 and mean 3 with probability 1/4. Sample sizes were 100 in all cases. Then the depth estimates were simulated 100 times (1000 times for the normal case). The measure we considered was the discrepancy between the location-scale depth at the Student median and at the sample median:

$$DIF = \max_{\mu,\sigma} depth(\mu, \sigma) - \max_{\sigma} depth(\text{``}median\text{''}, \sigma)\,.$$

Figure 1 gives box plots of the distribution of the DIF. Clearly, the only distributions for which DIF is significantly smaller than in a normal model are those with density tending to infinity quite fast (gamma and beta with exponent: .2 - 1 = -.8) Lack of symmetry seems to have little effect (note especially the normal contamponents model). Note also that the median depth is much closer to the maximal location-scale depth in these very unusual cases: in most cases the median doesn't go nearly as deep into the data (in terms of location-scale depth). This suggests perhaps that for most distributions the perturbation introduced by using the likelihood approach does provide a rather different picture from that of the median.

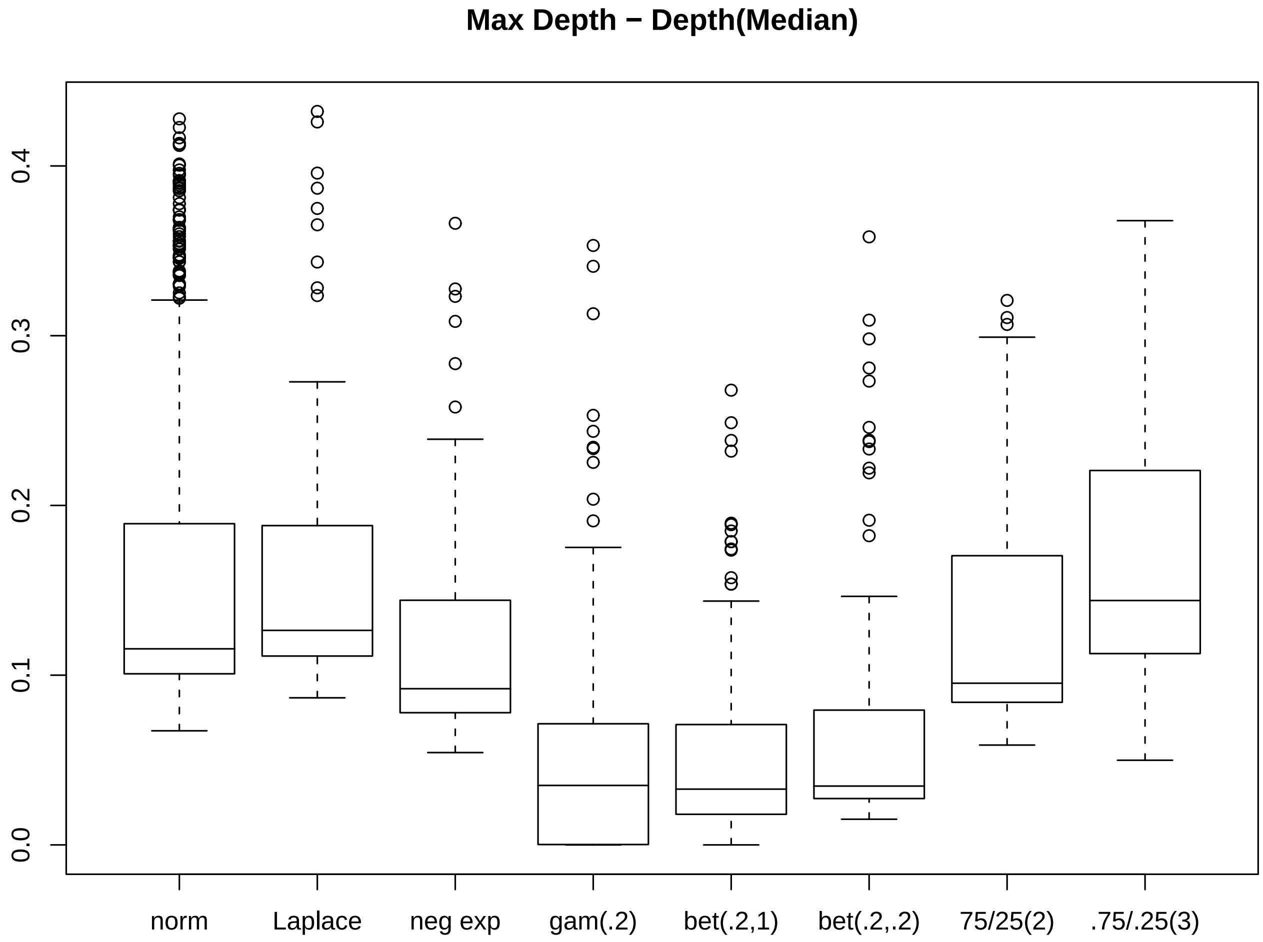
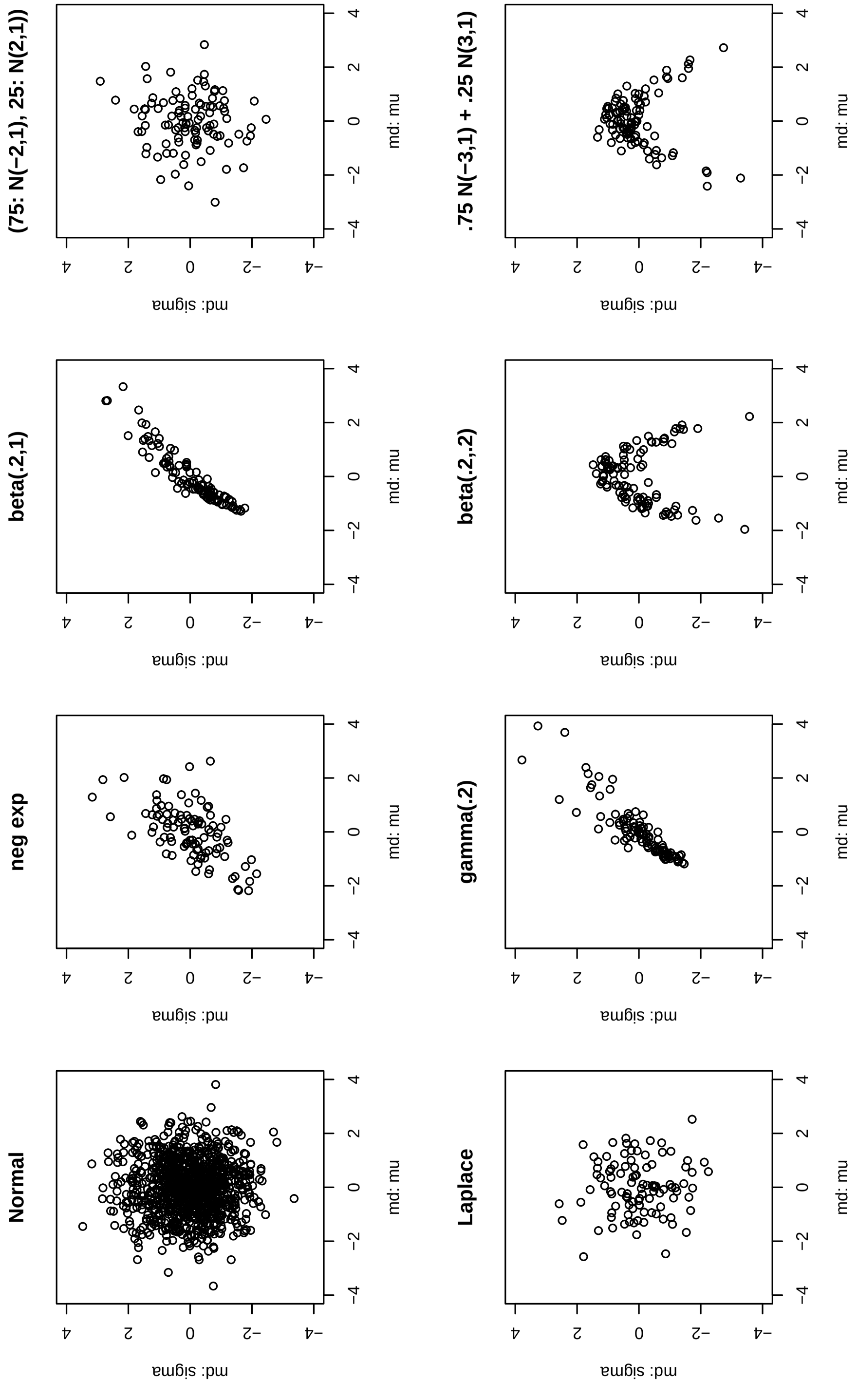


Figure 1: Box plots of the distribution of the difference between the maximal depth and depth at the median maximized over scale.

Figure 2: Scatter plots of maximal location-scale depth estimators, $\hat{\mu}$ and $\hat{\sigma}$.



One explanation might concern the dependence between $\hat{\mu}$ and $\hat{\sigma}$. For distributions nearly symmetric, $\hat{\mu}$ and $\hat{\sigma}$ should be rather independent (at least asymptotically). Figure 2 gives the scatter plots for the simulated samples described above. Here we begin to see some correlation in the highly asymmetric negative exponential, but the bimodal distributions seem to suggest a much more interesting story. The beta(.2, .2) distribution has $\hat{\sigma}$ quadratically related to $\hat{\mu}$, as does the random mixture whose means are separated by 6 standard deviations. In trials with a larger $n = 10,000$, the quadratic relationship is somewhat less clear, but is still statistically present: for this normal mixture, a LS regression using $\hat{\sigma} = a + b\hat{\mu} + c\hat{\mu}^2$ showed c significant at level .00095. This does suggest that the lack of independence might disappear as $n \to \infty$ (though slowly), but this remarkable picture seems to demand explanation.

Interestingly, the concatenation of 75 and 25 normals with means separated by 4 standard deviations fails to show this quadratic behavior. Though the lack of randomness in the choice of mixture produces samples with much less statistical variation, the LS analysis gave a t-value of .313 for the quadratic coefficient, c. We have no insight into why this is so, and would welcome any response.

4. The diagnostic aspect of the Student depth as described in Section 8 of paper is clearly meant to generalize the much simpler traditional QQ-plot. Mizera and Müller showed that the bivariate depth plots for both location and scale look quite interesting. They can easily reveal symmetry, including that present in the core of the data, rather than just in the tails. They are "capable of detecting heavy-tailed behavior too". Whether this really provides an improvement on QQ-plots is not clear. At this point, we would find it hard to convince anyone to learn how to use the Student depth plots just for the univariate data. The value of the work on the Student depth plots probably lies beyond those examples.

In this regard, we feel that the appearance of the location-scale plots can be improved. The lines connecting the data points on the location axis certainly complicate the plot. We don't see how they contribute much to understanding the data. Perhaps they could be eliminated by plotting only the deeper contours. Alternatively, the plot could be based on a continuous (linearly - interpolated) version of the empiric distribution function (instead of the point mass version). Another improvement might be in the use of color. Following Tufte (1990, Chapter 5) we would suggest emulating the cartographers and using pastel shades of blue (deeper) to brown (higher). In fact, there is a geographer's package, RColorBrewer (see: `www.colorbrewer.org`). Clearly, a good bit of more formal experimentation with the plotting will probably be needed before such methods find acceptance by the applied statistics community.

Concluding Remarks: Without a clear understanding of the behavior of location-scale depth (and parameter depth in general), it seems unlikely that the notion will have easy-to-understand statistical implications. We believe that we need to understand better the population version of the maximal depth estimators. In some special cases, like the regression depth of Rousseeuw and Hubert (1999), the maximal depth estimators are indeed consistently estimating a meaningful population quantity: *viz.*, a linear conditional median. More generally, parameter depth will be consistent for some functional of the assumed model. It is not clear that this population quantity will be easily accessible and interpretable in general (especially, outside a specific model), but it might be reasonable in the multivariate location-scale situation. However, this will certainly require extensive comparisons with other depth measures and other robust approaches.

In summary, at present traditional data depth looks more straightforward to us for exploratory data analysis. Though the work here suggests the exciting possibility that likelihood-based parameter depth might also prove useful in a diagnostic mode (in addition to providing robust parameter estimates), we feel that these implications still need to be drawn more completely. In general, the distinction between data depth and parameter depth seems quite important.

Should we say that data depth is traditional fare, and parameter depth is the genetically engineered variety? We do not know yet which one will dominate the future supermarket, but for now it seems clear that both will be on the shelves.

References

[1] Donoho, D. and Gasko, M. (1992), Breakdown properties of location estimates based on halfspace depth and projected outlyingness, *Ann. Statist.*, 20, 1803-1827.

[2] Rousseeuw P. and Hubert, M. (1999). Regression Depth *J. Amer. Statist. Assoc*, 94, 388-402.

[3] Tufte, E. (1990). it Envisioning Information, Graphics Press, Cheshire Connecticut.