Computational Issues in Robust Statistics

A.J. Stromberg[1]

[1] Dept. of Statistics, University of Kentucky, 817 Patterson Office Tower, Lexington, KY 40506-0027, USA

Keywords: Computing, Robustness, Breakdown

1 Introduction

Hundreds, and perhaps thousands, of papers have been published in the area of robust statistics, yet robust methods are still not used routinely by most applied statisticians. An important reason for this is the many computational issues in robust statistics.

Most applied statisticians agree conceptually that robust methods are a good idea, but they fail to use them for a number of reasons. Often, software is not available. Other times, like in linear regression, there are so many choices, it is not clear which estimator to use. In still other situations, the data sets are too big for robust techniques to handle. This paper discusses these issues and others.

2 Which Robust Linear Regression Estimator is Best?

Since the original introduction of the M-estimator, approximately twenty robust linear regression estimators have been introduced. They will not be reviewed here. Each new estimator compared favorably with the last, but no one really knows which estimator should be used in practice. Extensive simulations are needed to compare and contrast these estimator and make suggestions on which estimators can be recommended for use in practice.

3 Revisiting the Princeton Robustness Study

Thirty years ago, a group of statisticians got together and concluded that what really matters is the breakdown point. Our computers are likely a thousand times faster today. Are the same conclusions valid? Do we really need a high breakdown point?

4 Robust Estimation for Large Data Sets

Applied statisticians have the impression, perhaps correctly, that robust techniques can only be used on small to moderate sized data sets. Clearly, as the sample size increases and the dimension of the data set increases, the need for robust techniques increases. We need to find techniques that work well for large high dimensional data sets.

5 The Practical Importance of Theoretical Advances

Theoretical advances in robust statistics are clearly publishable in major journals, but the practical importance of these advances are less evident. As an example, consider the max-bias curve. Is a better max-bias curve of any practical significance? What about other theoretical advances?

6 Resampling Methods for Robust Statistics

Estimating standard errors using resampling methods has been largely ignored in the robustness literature. Standard errors based on asymptotic normality are convenient, but for small samples, standard errors based on resampling may be more effective. Can we find methods for getting standard errors based on resampling quickly?

7 Conclusion

Many other computational issues are likely to be discussed at ICORS2002. I encourage all participants to become involved with finding solutions to these problems.

Please fill in this form and mail it together with your abstract.

My abstract fits best to topic number 4. (**"PLEASE INDICATE HERE THE NUMBER FROM THE LIST OF TOPICS BELOW WHICH BEST FITS TO YOUR ABSTRACT"**.).

List of Topics:

1. Algorithms
2. Applications
3. Biostatistics
4. Computing and graphics
5. Data analysis
6. Data mining
7. Economics, finance
8. Efficiency and robustness
9. Functionals and bias
10. Fuzzy statistics
11. Geostatistics
12. Inference for robust methods, model testing
13. Location depth and regression depth
14. Multivariate methods
15. Neural networks
16. Rank-based methods
17. Regression quantiles, trimming
18. Robust covariance
19. Robust designs
20. Robust regression
21. Time series analysis
22. Wavelets
23. Other (please specify)