Søren Johansen

Some Comments on Robustness

Preprint
September
1979
5

Institute of Mathematical Statistics
University of Copenhagen

Søren Johansen

SOME COMMENTS ON ROBUSTNESS

Preprint 1979 No. 5

INSTITUTE OF MATHEMATICAL STATISTICS

UNIVERSITY OF COPENHAGEN

August 1979

Summary

These comments were presented at the 12th European Meeting of Statisticians as a contribution to the discussion in the session on robustness organized by Jana Jurečkova.

Keywords

Robustness; Conditional inference; Outliers; Hyperbolic distributions; Gross errors.

1. The theory of robustness concerns itself with the behaviour of estimators and teststatistics when the underlying model is modified slightly. There seems to be at least two possible formulations of this modification which are not always clearly separated.

The first is the model / supermodel formulation where as an example one can think of the normal distribution as embedded in a family of heavy tailed distributions or in the class of all distributions. The idea then is to modify the estimator ($\overline{X}$) such that it behaves reasonably well under the various alternatives while paying the price of a small loss of efficiency if the model is true.

This first formulation leads to an investigation of the estimator in a neighbourhood of the small model. This is the approach taken by Tukey (1960) and Huber (1964).

The other formulation is that we have a model but want to safeguard ourselves against gross errors or irregular observations. It is not obvious that the best formulation is in terms of a stochastic description of the gross errors. The nature of a gross error is such that we want to set it aside and consider it separately, this is not the usual attitude towards observations from a distribution.

This second formulation points towards a modification of the estimator with the purpose of making it insensitive or more stable to a few wild observations. Thus one would find the best estimator under some restriction, like a bound on the influence curve, and evaluate its performance under the model. This is the approach taken by Hampel (1968).

There seems to be a strong interrelation between the results derived from the two formulations in the sense that stable estimators are often robust against heavy tails and vice versa.

A similar result is not found in sampling inspection, where the two types of contamination have been discussed by Hald (1979). One can either safeguard the sampling plan against a few wrong batches with a too large percentage defective or one can find the optimal plan in the presence of a heavy tail in the prior distribution.

A detailed discussion of various types of models for outliers has been given by Barnett and Lewis (1978).

2. The next remark on gross errors also concerns the adequacy of the usual location - scale model.

If you observe $y = \mu + \sigma x$ then x is the error and y the response. The usual formulation of gross errors is that something went wrong with x.

In reporting data the error often occurs in y. Two typical examples are wrong decimal points and interchange of two digits. This type of error is certainly not location and scale invariant and hence can not be attributed to x. The usual formulation thus seems to be inadequate for this type of error.

3. A final remark on gross errors of the random type has to do with some new models investigated by Barndorff - Nielsen (1977), who found that in describing the distribution of the size of sand-particles a distribution of the type

$$\log f_{\alpha,\beta}(x) = \alpha\sqrt{1 + x^2} + \beta x + c(\alpha,\beta)$$

fitted well.

It is seen that this type of distribution has exponential tails and one can show that it interpolates (in some sense) between the normal distribution and the double exponential distribution.

It is interesting to note that the influence curve for the maximum likelihood estimator remains bounded and can be considered a smooth version of Hubers M-estimate given by $\rho(x) = x^2/2$, $|x| \leqq k$, $\rho(x) = k|x| - k^2/2$, $|x| > k$.

4. My final comment will have to do with the application of ancillarity to robustness, and is taken from the recent book by Fraser (1979). Similar points of view have been expressed by Barnard at various occasions, see Barnard (1974).

The normal distribution has the beautiful property that $(\bar{x}, s^2)$ is sufficient for (μ, σ^2) and independent of $d = (x_1 - \bar{x}, \ldots, x_n - \bar{x})/s$, which is ancillary.

The exponential families generalize the normal distribution in that they insist on the sufficiency, whereas the location-scale models generalize in the other direction and preserve ancillarity.

It seems reasonable to exploit the property of ancillarity in location-scale families as follows: Since the distribution of $d = ((x_i - \bar{x})/s)^n$, does not involve the unknown parameters

(μ,σ^2) we want to establish confidence intervals and estimates conditional on the observed value of d.

This is done as follows, choose a location and scale estimate, $\bar{x}$, s, say, and define $t(x) = (\bar{x} - \mu)\sqrt{n}/s$, $u(x) = s/\sigma$ then a confidence interval for μ can be derived from the distribution of t(x) given d(x), and for σ from the distribution of u(x) given d(x).

This procedure has the important and interesting property that it is independent of the choice of which location and scale estimate we take. If instead of $(\bar{x},s^2)$ we choose $(x_1,|x_2 - x_1|^2)$ and construct the confidence interval for μ from $(x_1 - \mu)/|x_2 - x_1|$ and for σ from $|x_2 - x_1|$ then it is easy to see, that we get the same confidence intervals. The reason for this is that for given value of d(x), there is a one - to - one linear transformation between the first and second set of statistics which implies that the confidence intervals are the same. Thus the procedure gives a uniquely defined confidence interval.

This has the interesting implication for robust procedures that if you prefer to work with, say, an α - trimmed mean, then you loose no efficiency under the true model. Obviously you need not use the α - trimmed mean, since the conditional confidence interval based on it is nothing but $\bar{x} \pm t\, s/\sqrt{n}$, where t has to be chosen in the distribution of $(\bar{x} - \mu)\sqrt{n}/s$ given d.

The same book by Fraser also has some suggestions for robust procedures.

Consider the model

$$f_\lambda\left(\frac{x - \mu}{\sigma}\right)\frac{1}{\sigma}$$

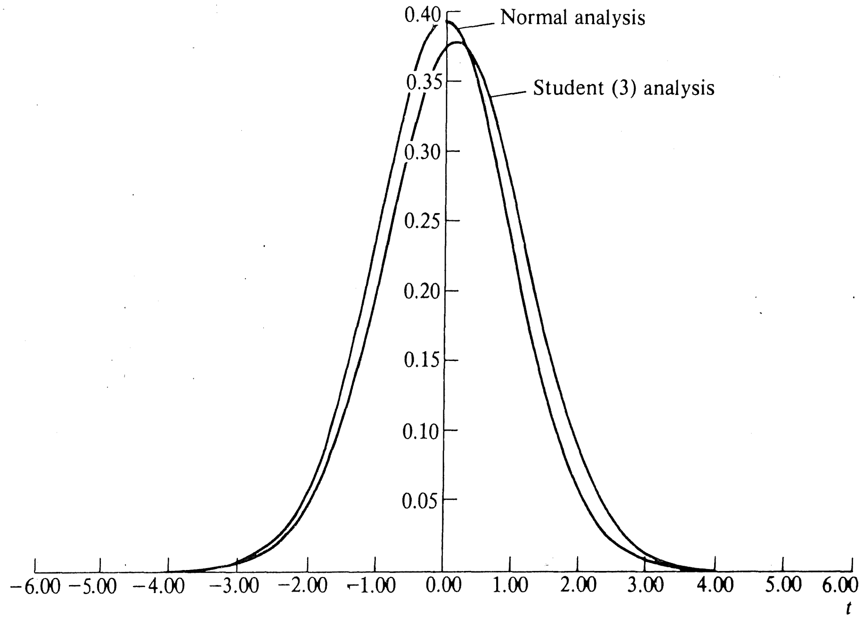
where λ is a shape parameter. As an example one can take a Student (λ) distribution, a Weibull or a gamma distribution. The distribution of $d = ((x_i - \bar{x}) / s)^n$, depends only on λ and it is therefore again relevant to use d to estimate the shape of the error distribution and make inference on (μ,σ) conditional on d. In the following we shall assume λ fixed and known, and show by example how the conditional procedures have some nice robustness and stability properties.

Two sets of 30 random numbers were generated. The first N(10,1) the second Student (3) (10,1.1966). The reason for the scaling is that the probability of the interval [9,11] should be the same in both cases.

We shall apply a normal analysis and a Student (3) analysis to the two samples.

The first figure shows the result for the normal sample

Figure 1



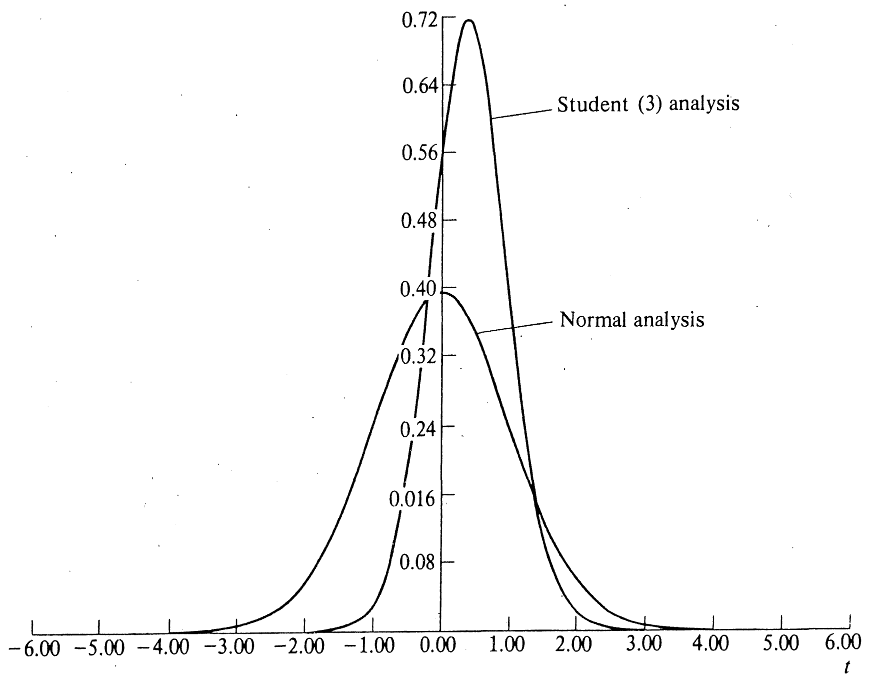
The distribution of t(x) for $\lambda = 3, \infty$; the normal sample.

Fraser: Inference in Linear Models (1979) Figure 2 - 5
Reproduced with the kind permission of the publisher.

Note that the two analyses give almost identical results.

For the Student (3) sample we get the next figure which clearly shows

Figure 2



The distribution of t(x) for $\lambda = 3, \infty$; the Student sample.

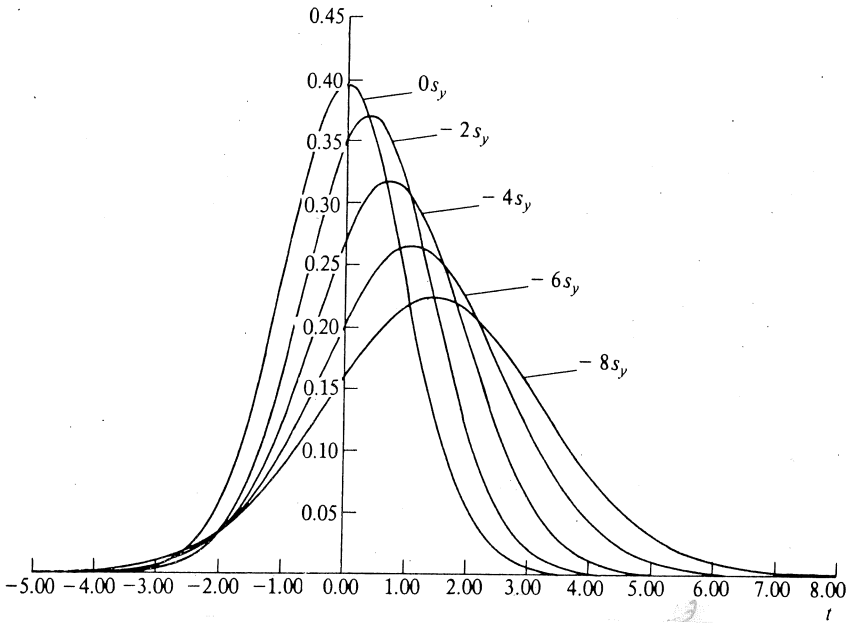
Fraser: Inference in Linear Models (1979) Figure 2 - 8
Reproduced with the kind permission of the publisher.

that the normal analysis is not robust to heavy tails. Comparing the two figures we see that the Student (3) analysis in both cases gave a reasonable analysis of the data whereas the normal analysis does not. Thus the Student (3) analysis is robust.

It is also stable to extreme observations. Consider a normal sample of 30 with one observation moved out various multiples of the standard deviation.

The distribution of $(\bar{x} - \mu)\sqrt{n}/s$ is constructed for each sample and rescaled to be comparable. For the normal analysis we find

Figure 3



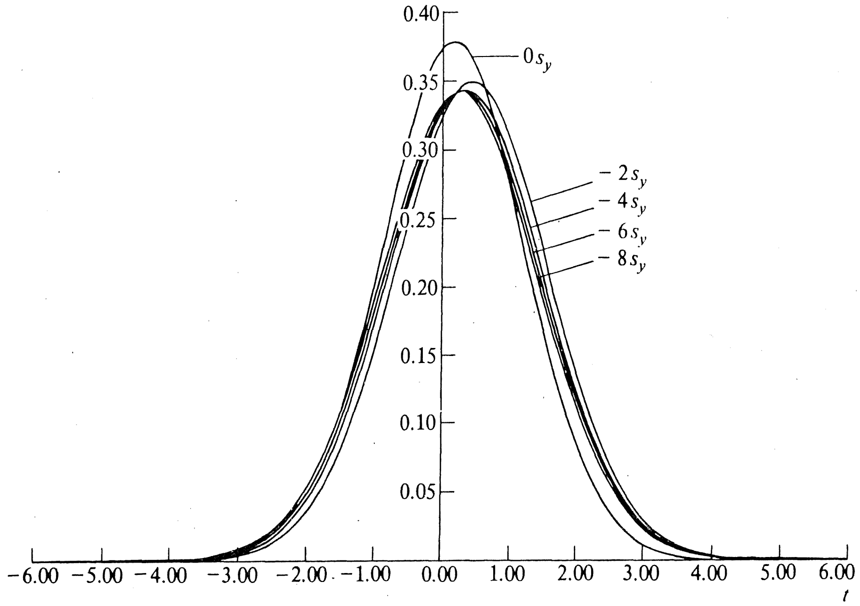
One observation displaced 0, $-2\,s_y$, $-4\,s_y$, $-6\,s_y$, $-8\,s_y$ for a normal sample of 30: the normal analysis t - statistic distribution has been rescaled to be comparable.

Fraser: Inference in Linear Models (1979) Figure 2 - 10
Reproduced with the kind permission of the publisher.

which shows how sensitive the normal analysis is to a single outlier.

The Student (3) analysis, however, is rather insensitive to a single large observation.

Figure 4



One observation displaced 0, $-2\,s_y$, $-4\,s_y$, $-6\,s_y$, $-8\,s_y$ for a normal sample of 30: the Student (3) analysis t-statistic distribution has been rescaled to be comparable.

Fraser: Inference in Linear Models (1979) Figure 2 - 11
Reproduced with the kind permission of the publisher.

As a concluding remark let me point out that there seems to be two conflicting points of view when doing statistical inference. One is to build a model and then analyse it with respect to which estimator and test to use, the other is to start out with the estimator or test statistic without specifying the underlying model precisely.

Sometimes, as in the theory of exponential families, one can show that for a given statistic one can find the model for which this statistic is the relevant to investigate. Thus in this sense there is no conflict but rather a complementarity between the two points of view.

The usual approach to robustness is that of trimming the usual estimators in various ways, the approach outlined above emphasized the model more and derives the robust procedures. It will be interesting to see if these points of view are conflicting or complementary.

5. References

Barnard, G.A. (1974) Conditionality, Pivotals and Robust Estimation (with discussion), in O. Barndorff - Nielsen, P. Blaesild, and G. Schou (eds.) "Proceedings of the Conference on Foundational Questions in Statistical Inference" Department of Theoretical Statistics. Aarhus University, 61 - 80.

Barndorff - Nielsen, O. (1977) Exponentially Decreasing Distributions for the Logarithm of Particle Size. Proc.Roy. Soc. London Ser. A. 353, 401 - 419.

Barnett, V. and Lewis, T. (1978) Outliers in Statistical Data, New York, Wiley.

Fraser, D.A.S. (1979) Inference and Linear Models. McGraw - Hill.

Hald, A. (1978) Statistical Theory of Sampling Inspection by Attributes, Vol. 1. University of Copenhagen.

Hampel, F. (1968) Contributions to the Theory of Robust Estimation, Ph.D. Thesis, University of California, Berkeley.

Huber, P. (1964) Robust Estimation of a Location Parameter, Ann. Math.Statist. 35, 73 - 101.

Tukey, J.W. (1960) A Survey of Sampling from Contaminated Distributions, in Contribtuions to Probability and Statistics, ed. I. Olkin, Stanford University Press.

PREPRINTS 1978

COPIES OF PREPRINTS ARE OBTAINABLE FROM THE AUTHOR OR FROM THE INSTITUTE OF MATHEMATICAL STATISTICS, UNIVERSITETSPARKEN 5, 2100 COPENHAGEN Ø, DENMARK.

No. 1 Tjur, Tue: Statistical Inference under the Likelihood Principle.

No. 2 Hering, Heinrich: The Non-Degenerate Limit for Supercritical Branching Diffusions.

No. 3 Henningsen, Inge: Estimation in M/G/1-Queues.

No. 4 Braun, Henry: Stochastic Stable Population Theory in Continuous Time.

No. 5 Asmussen, Søren: On some two-sex population models.

No. 6 Andersen, Per Kragh: Filtered Renewal Processes with a Two-Sided Impact Function.

No. 7 Johansen, Søren & Ramsey, Fred L.: A Bang-Bang Representation for 3x3 Embeddable Stochastic Matrix.

No. 8 Braun, Henry: A Simple Method for Testing Goodness of Fit in the Presence of Nuisance Parameters.

No. 9 Lauritzen, S.L. & Speed, T.P. & Vijayan, K.: Decomposable Graphs and Hypergraphs.

No. 10 Hald, Anders: On the Statistical Theory of Sampling Inspection by Attributes.

No. 11 Darroch, J.N. & Lauritzen, S.L. & Speed, T. P.: Markov Fields and Log-linear Interaction Models for Contingency Tables.

No. 12 Speed, T.P.: A Note on Nearest-Neighbour Gibbs and Markov Probabilities.

No. 13 Andersson, Steen A.: Canonical Correlations with Respect to a Complex Structure.

PREPRINTS 1979

COPIES OF PREPRINTS ARE OBTAINABLE FROM THE AUTHOR OR FROM THE INSTITUTE OF MATHEMATICAL STATISTICS, UNIVERSITETSPARKEN 5, 2100 COPENHAGEN Ø, DENMARK.

No. 1 Edwards, David: Large Sample Tests for Stationarity and Reversibility in Finite Markov Chains.

No. 2 Andersson, Steen A. : Distribution of Maximal Invariants Using Proper Action and Quotient Measures.

No. 3 Johansen, Søren: A Note on the Welch-James Approximation to the Distribution of the Residual Sum of Squares in a Weighted Linear Regression.

No. 4 Björnsson, Ottó: Four Simple Cahrcterizations of Standard Borel Spaces.

No. 5 Johansen, Søren: Some Comments on Robustness