* This research was supported by the Air Force Office of Scientific Research under Grant AFOSR-75-2796.

Studentizing Robust Estimates

by

Raymond J. Carroll*

Department of Statistics
University of North Carolina at Chapel Hill

Institute of Statistics Mimeo Series No. 1085
August, 1975

STUDENTIZING ROBUST ESTIMATES

by

Raymond J. Carroll*

University of North Carolina

Summary

The consistency of commonly used variance estimates for studentizing robust estimates of location is studied when the underlying distributions are asymmetric. Surprisingly, the popular methods for studentizing M-estimates, jackknifed M-estimates and adaptive trimmed means underestimate the true variance by a factor approaching 50% under the negative exponential distribution. Possible reasons for the failures and suggested modifications are presented.

Key Words and Phrases: Robustness, t-tests, asymmetry, Monte-Carlo, trimmed means, M-estimators, adaptive estimators

AMS (1970) Subject Classifications: Primary 62G35; Secondary 62E25

* This research was supported by the Air Force Office of Scientific Research under Grant AFOSR-75-2796.

Introduction

This project started with the problem of sequential fixed-width interval estimation of the limiting functional of robust estimators. We first compared a number of estimators in many symmetric situations with predictable results (concerning robustness of efficiency) which will not be reported here (see Andrews, et al (1972), Carroll and Wegman (1975)). When the sampling distribution was the negative exponential, the situation was surprisingly degenerate; the modern robust estimators all gave much lower coverage probabilities than advertised. After some study, we discovered that the commonly used variance estimates underestimated the true variance by nearly 50%, so that too few observations were being taken. In the context of the two-sample location problem, this means the Type I error is much higher than desired. Before routine application of robust two-sample tests can be recommended, it is important to discover where the studentization fails and to suggest appropriate modifications to existing procedures.

This paper describes the results of a Monte-Carlo experiment, comparing true and estimated variances in asymmetric situations. We show that in asymmetric cases such as the negative exponential distribution, hard adaption (Hogg (1974)), (common) M-estimators and jackknifed M-estimators *all* fail. Our major initial success was obtained with trimmed means. The reasons for the occurrence of the observed phenomena are investigated, and recommendations for future study are made.

The Estimators

Define $U(\alpha)$ $(L(\alpha))$ as the mean of the largest (smallest) $[\alpha(n+1)]$ order statistics. As a measure of tail length, Hogg proposed

$$Q = (U(.20) - L(.20))/(U(.50) - L(.50)).$$

Hampel's scale is defined by

$$s_n = \text{median}\{|X_i - \text{sample median}|\}/.6754,$$

with the property that $s_n \to 1$ in the normal case. In Table 1 we list the estimators of location used in this study. The variance estimate for the sample mean M is the usual sample variance, while for the trimmed means 5% - 38% it is the jackknifed variance estimate (Shorack (1974)). For the one-step Hampels and Hubers (D10-D20, 12A-25A, HUB) it is (Huber (1970)),

$$\frac{n s_n^2 \int \psi^2\left(\frac{x-T_n}{s_n}\right) dF_n(x)}{(n-1)\left\{\int \psi'\left(\frac{x-T_n}{s_n}\right) dF_n(x)\right\}^2}. \tag{1}$$

For the adaptors (HG1, HG2, 1.81D, 1.90D, 2.00A, 1.81D), the variance estimate of the chosen location estimate is used. For the jackknifed Huber (JHUB), the jackknifed variance estimate is used.

TABLE 1

A description of the location estimates used in the study.

Code	Description
M	Sample mean
5% 10% 25% 38%	α% means an α% symmetrically trimmed mean. 38% has $\alpha = .375$.
HG1	$\begin{cases} 5\% & Q\le 1.81 \\ 10\% \text{ if } & 1.81<Q\le 1.87 \\ 25\% & Q>1.87 \end{cases}$
HG2	$\begin{cases} & 5\% & Q\le 1.80 \\ (.05+(Q-1.80)(4/3))100\% & \text{if} & 1.80<Q\le 1.95 \\ & 25\% & Q>1.95 \end{cases}$
D10 D15 D20	One-step Huber estimates, with start being the sample median, scale s_n (Andrews, et al (1972)). The knot in the ψ function for DK is K/10.
1.81D	$\begin{cases} D20 & Q\le 1.81 \\ D15 & 1.81<Q\le 1.87 \\ D10 & Q>1.87 \end{cases}$
1.90	$\begin{cases} D20 & Q\le 1.90 \\ D15 \text{ if } & 1.90<Q\le 2.05 \\ D10 & Q>2.05 \end{cases}$
HUB	Solution (via method of bisection with 20 iterations) to $\Sigma\psi\left(s_n^{-1}(X_i-t)\right) = 0$, where $\psi(x) = \max\left(-1.5, \min(x,1.5)\right)$.
JHUB	A jackknifed version of HUB, with $n = 50$.
12A 21A 25A	One-step Hampel estimates, with knots given in Andrews, et al (1972). Start is the sample median, scale is s_n .
2.00A	$\begin{cases} 21A \text{ if } Q<2.00 \\ 12A \text{ if } Q\ge 2.00 \end{cases}$
1.81A	$\begin{cases} 25A & Q\le 1.81 \\ 21A \text{ if } & 1.81<Q\le 1.87 \\ 12A & Q>1.87 \end{cases}$

Random Numbers

A congruential random number generator (period > 2^{29}) with a shuffling feature was used to obtain the uniform random variables, while standard normal observations were obtained by the Box-Muller algorithm. The shuffler first generated 300 uniform deviates and then chose one of them at random; this effectively overcomes the undesirable dependence features of the usual congruential generator. No Monte-Carlo swindle was used, so reported results will be accurate to about one decimal place, more then enough to make the conclusions given here.

Sampling Distributions

Ten sampling distributions were studied. If Φ is the standard normal distribution function, the first six had distributions as follows:

Type	Code
$\Phi(x)$	N(0,1)
$.95\Phi(x)+.05\Phi(x-1)$	N(0,1)+.05N(1,1)
$.90\Phi(x)+.10\Phi(x-1)$	N(0,1)+.10N(1,1)
$.95\Phi(x)+.05\Phi(x-3)$	N(0,1)+.05N(3,1)
$.90\Phi(x)+.10\Phi(x-3)$	N(0,1)+.10N(3.1)
Negative exponential with mean 1.25	NE .

Letting X be a standard normal random variable, the last four were generated by

12.5exp(.10X)	Exp(.10X)
4.88exp(.25X)	Exp(.25X)
2.46exp(.50X)	Exp(.50X)
.49exp(X)	Exp(X) .

Samples of size $n = 75$ were taken for each iteration. There were 500 iterations on N(0,1), 1000 iterations on NE and Exp(X), and 250 iterations on the others.

Major Results

This paper focuses on the effects of asymmetry on the consistency of variance estimates of robust estimators. If $n = 75$ is the sample size, N is the number of experiments, and $Y_1, \ldots, Y_N$ the realized values of a particular estimator, the (standardized) Monte-Carlo variance of the estimator is

$$\sigma^2 = \frac{n}{N} \sum_1^n (Y_i - \bar{Y}_N)^2 .$$

The average value of a variance estimate for a particular location estimator is denoted by σ_n^2. Table 2 presents values of σ_n^2/σ^2 for all ten sampling distributions. Table 3 presents values of σ^2 and σ_n^2 for N(0,1), NE, and Exp(X), the first acting as a standard, with the latter two representing "worst case" asymmetric distributions.

TABLE 2

Values of the ratio σ_n^2/σ^2. A consistent variance estimate should have this value near 1.00.

Code	N(0,1)	N(0,1)+.05N(1,1)	N(0,1)+.10N(0,1)	N(0,1)+.05N(3,1)	N(0,1)+.05N(3.1)	NE	Exp(.10X)	Exp(.25X)	Exp(.50X)	Exp(X)
M	.97	.99	.94	1.11	1.00	1.01	.96	.97	.93	1.00
5%	.95	.97	.93	1.14	1.10	.96	.90	1.02	.97	1.05
10%	.94	.95	.92	1.15	1.06	.97	.88	1.04	.97	1.04
25%	.93	.96	.93	1.10	.99	1.00	.86	1.05	.94	.99
38%	.92	1.04	1.00	1.13	1.02	1.02	.88	1.02	.94	.98
HG1	.90	.92	.90	1.06	.90	.53	.84	.97	.70	.62
HG2	.97	.94	.97	.98	.97	.60	1.08	.87	.79	.62
D10	.90	1.03	.94	1.09	.95	.54	.89	.88	.82	.52
D15	.95	1.02	.95	1.12	.97	.59	.92	.88	.75	.53
D20	.97	1.02	.95	1.13	.98	.68	.95	.91	.78	.59
1.81D	.91	1.04	.95	1.11	.99	.59	.91	.90	.82	.54
1.90D	.97	1.02	.95	1.12	.96	.59	.94	.89	.74	.42
HUB	1.07	.99	.84	.93	.93	.60	.88	.81	.67	.52
JHUB	1.00	--	--	--	--	.61	--	--	--	.53
12A	.97	.95	.92	1.12	.98	.49	.90	.84	.75	.44
21A	1.01	.96	.94	1.12	1.01	.62	.91	.90	.81	.50
25A	1.01	.97	.93	1.11	1.03	.70	.93	.93	.86	.55
2.00A	1.01	.96	.94	1.11	.99	.61	.91	.90	.80	.35
1.81A	1.00	.94	.92	1.08	.90	.42	.91	.84	.65	.34

TABLE 3

Values of σ_n^2 and σ^2 for three sampling distributions.

Code	$N(0,1)-\sigma^2$	$N(0,1)-\sigma_n^2$	$NE-\sigma^2$	$NE-\sigma_n^2$	$Exp(X)-\sigma^2$	$Exp(X)-\sigma_n^2$
M	1.03	1.00	1.54	1.55	1.10	1.10
5%	1.09	1.04	1.45	1.38	.58	.61
10%	1.13	1.07	1.35	1.31	.45	.47
25%	1.31	1.22	1.31	1.30	.35	.35
38%	1.46	1.35	1.38	1.40	.36	.35
HG1	1.13	1.02	2.30	1.22	.56	.35
HG2	1.04	1.01	2.07	1.24	.55	.36
D10	1.21	1.09	1.37	.74	.37	.19
D15	1.10	1.04	1.49	.88	.45	.24
D20	1.05	1.02	1.53	1.04	.51	.30
1.81D	1.21	1.10	1.30	.78	.43	.23
1.90D	1.05	1.02	1.71	1.00	.60	.25
HUB	.98	1.04	1.52	.91	.49	.25
JHUB	1.01	1.01	1.43	.87	.47	.25
12A	1.10	1.07	1.57	.77	.44	.19
21A	1.00	1.02	1.66	1.05	.55	.28
25A	1.01	1.01	1.65	1.15	.59	.32
2.00A	1.01	1.02	1.71	1.04	.72	.25
1.81A	1.00	1.01	2.38	1.01	.62	.21

I. (Trimmed Means)

The symmetrically trimmed means M (or 0%), 5% - 38% were the only estimators from Table 1 whose variance estimates consistently estimated the true variance over the whole range of distributions. Hence, if the user has serious concern about asymmetry, trimmed means would be recommended for their robustness of validity.

The first Hogg adaptor (HG1) uses a discrete adaptive approach which chooses among a finite set of possible trimmed means. The particular choice used here fails at NE, Exp(.50X), and Exp(X). This is most unusual, especially in view of the success of individual trimmed means. Note that in Table 3, while the value for σ_n^2 is in line with individual trimmed means, the value for σ^2 is considerably larger in magnitude. The following arguments should indicate in part that this explosion of σ^2 can be expected to occur over the whole range of discrete adaptors. Denote $Q = Q(n,F)$ to emphasize the dependence on the sample size n and the underlying (continuous) distribution function F. Typically $n^{\frac{1}{2}}(Q(n,F) - Q(F)) \xrightarrow{L} N(0, \sigma^2(Q))$, where $Q(F)$ is some constant. Denote the α-trimmed mean by $S(n,\alpha)$ with limit value $S(\alpha)$. Suppose F is symmetric about θ and that one of the "knots" or "change points" in the adaptor is $Q(F)$ (for example, in HG1 these knots are 1.81 and 1.87). Then asymptotically, the adaptor is

$$\begin{aligned} T^*(n,F) &= S(n,\alpha_1) \quad \text{if} \quad Q(n,F) \leq Q(F) \\ &= S(n,\alpha_2) \quad \text{if} \quad Q(n,F) > Q(F). \end{aligned} \tag{2}$$

Since $Q(n,F)$ is even and $S(n,\alpha)$ is odd, $n^{\frac{1}{2}}(Q(n,F) - Q(F))$ and $n^{\frac{1}{2}}(S(n,\alpha) - S(\alpha))$ are asymptotically independent, $S(\alpha) = \theta$ and

$$\Pr\{n^{\frac{1}{2}}(T^*(n,F)-\theta)\leq z\} \approx \tfrac{1}{2}\Big[\Pr\{n^{\frac{1}{2}}(S(n,\alpha_1)-\theta)\leq z\} + \Pr\{n^{\frac{1}{2}}(S(n,\alpha_2)-\theta)\leq z\}\Big] . \quad (3)$$

Hence, $n^{\frac{1}{2}}(T^*(n,F) - \theta)$ does not quite have a limiting normal distribution, although the variance will not have blown up. However, if F is asymmetric, $S(\alpha_1) \neq S(\alpha_2)$ and if the approximation (3) holds, we see that for no θ does $n^{\frac{1}{2}}(T^*(n,F) - \theta)$ have a proper limit distribution (in fact, mass will be placed at $+\infty$ or $-\infty$). These very heuristic arguments show that the studentization of hard adaptors is likely to fail whenever one is unfortunate enough to have chosen $Q(F)$ as a knot. Since $Q(NE) = 1.804$ and $Q(\text{Exp}(.50X)) = 1.82$ (the latter obtained by Monte-Carlo sampling with 250 trials), one sees the probable explanation for the failure of HG1 in these cases. In Monte-Carlo sampling we obtained $Q(\text{Exp}(X)) = 1.96$; intuitively, in this heavily asymmetric situation the choices 10% and 25% are so different as to provide a boost to the variance. However, an adequate explanation is still needed.

It would seem then that a source of difficulty is that $T^*(n,F)$ is not continuous in $Q(n,F)$ for all n. HG2 provides an example which shows that the continuity merely of $\lim_n T^*(n,F) = T^*(F)$ is not sufficient. As a first constraint then, any adaptive estimate should have the above continuity property. Although we experimented with linear functions of order statistics with weight functions depending smoothly on α (such as piecewise linear and quadratic), we had very little success. The problem of constructing smoothly adaptive estimates with reasonably consistent variance estimates in small samples is an open and worthwhile problem.

II. (M-estimators)

The Huber estimate HUB (generated by $\psi(x) = \max(k_1, \min(x,k_2))$, $-k_1 = k_2 = 1.5$) has good robustness and breakdown properties but Carroll (1975a) has shown that if F is asymmetric, $s_n \to \xi$, then there exist non-zero constants a_1, a_2 with

$$(T_n - T(F)) = a_1 n^{-1} \sum_1^n \psi\left[\xi^{-1}(X_i - T(F))\right] + a_2(s_n^{-1} - \xi) + o(n^{-\frac{1}{2}}) \quad \text{(a.s.).} \qquad (4)$$

This means that the variance estimate (1) (suggested by Huber (1970) and used by Gross (1976)) is not consistent for asymmetric distributions. This is clearly reflected in Tables 2 and 3. The studentization suggested by (1) would hold if $a_2 = 0$, which is equivalent to

$$E\left\{(X - T(F))\psi'\left[\xi^{-1}(X - T(F))\right]\right\} = 0. \qquad (5)$$

One possibility would be to choose k_1, k_2 to force (5) to hold, at least asymptotically. Since Huber estimates have such nice properties, some future work should be devoted to finding a satisfactory solution.

That the jackknifed Huber estimate failed was both surprising and disappointing. Some insight into the cause of this phenomena is available. Jaeckel in an unpublished paper indicates that showing something like

$$n^{\frac{1}{2}}\left\{T_n - T(F) - a_1 n^{-1} \sum_1^n \psi\left[\xi^{-1}(X_i - T(F))\right]\right\} \xrightarrow{p} 0 \qquad (6)$$

is the missing step which needs to be verified for his arguments to hold;

(4) shows that (6) fails. Further, it is well-known that the jackknife does not provide a consistent estimate of the variance of the sample median; since the scale s_n is very close to a sample median, it would seem to be the cause of some instability. If this latter point is the root difficulty (a point which needs to be settled), perhaps using a scale with somewhat smoother influence function is mandated. One possibility would be to take the average of the smallest half of $\{|X_1 - \text{median}|, \ldots ,|X_n - \text{median}|\}$. We have done some partial work in this direction with little success.

III. (One-step estimates)

The one-step M-estimates and their adaptors were included for illustrative purposes although they would not be used in obviously asymmetric situations. Carroll (1975b) has shown that an analogue to (4) exists in this case as well, so the failures at NE and Exp(X) are not unexpected. One interesting point to note is the very poor performance of 21A and 25A at the negative exponential; we have no good explanations for this. Gross (1976), in surveying the disappointing performance (in symmetric situations) of jackknifed versions of these estimates, recommends replacing the median by a smooth starting value. The arguments of the previous section indicate an additional need for a smooth scale.

Conclusions

In this paper we have shown that wide classes of robust estimates will not perform adequately in the two-sample location problem in the presence of asymmetry. While trimmed means work quite well, discrete adaptive trimmed means have been shown to degenerate whenever one of the knots is near the limiting value of the tail length functional $Q(n,F)$; evidently only smoothly adaptive trimmed means have any hope of success. M-estimates and their one-step versions have been shown analytically to lead to inconsistent variance estimates if the ψ-function is skew-symmetric; this suggests the possibility of allowing the ψ-function to also depend on the data. It would appear that M-estimates and their jackknifed versions share the same source of difficulty, but this point needs to be clarified.

References

ANDREWS, D. F., BICKEL, P. J., HAMPEL, F. R., HUBER, P. J., ROGERS, W. H., and TUKEY, J. W. (1972). *Robust Estimates of Location: Survey and Advances*. Princeton University Press.

CARROLL, R. J. (1975a). An asymptotic representation for M-estimators and linear functions of order statistics. Institute of Statistics Mimeo Series #1006, University of North Carolina at Chapel Hill.

CARROLL, R. J. (1975b). On the asymptotic normality of stopping times based on robust estimators. Institute of Statistics Mimeo Series #1027, University of North Carolina at Chapel Hill.

CARROLL, R. J. and WEGMAN, E. J. (1975). A Monte-Carlo study of robust estimators of location. Institute of Statistics Mimeo Series #1040, University of North Carolina at Chapel Hill.

GROSS, A. M. (1976). Confidence interval robustness with long-tailed symmetric distributions. *J. Amer. Statist. Assoc. 71*, 409-416.

HOGG, R. V. (1974). Adaptive robust procedures: a partial review and suggestions for future applications and theory. *J. Amer. Statist. Assoc. 69*, 909-923.

HUBER, P. J. (1970). Studentizing robust estimates. *Nonparametric Techniques in Statistical Inference*, ed. M. L. Puri. Cambridge University Press, 453-463.

SHORACK, G. R. (1974). Random means. *Ann. Statist. 2*, 661-675.