A NEW METHOD FOR CONSTRUCTING APPROXIMATE CONFIDENCE INTERVALS FOR M-ESTIMATES

by

Dennis D. Boos

Department of Statistics
North Carolina State University

Institute of Statistics Mimeo Series #1198

September, 1978

A New Method for Constructing Approximate
Confidence Intervals for M-estimates

by

Dennis D. Boos

The empirical function used to define M-estimates of location is similar to a distribution function when ψ is nondecreasing. This similarity allows approximate confidence intervals to be constructed from the "percentiles" of the defining function.

KEY WORDS: M-estimates; Confidence intervals; Quantiles; t statistic.

1. INTRODUCTION

Let $X_1,\ldots,X_n$ be a sample from a distribution F and define the location "parameter" θ to be the solution of

$$\int_{-\infty}^{\infty} \psi(x-\theta)dF(x) = 0 \ . \qquad (1.1)$$

An M-estimate for θ is the solution $\hat{\theta}$ of the empirical analogue to (1.1)

$$\frac{1}{n}\sum_{i=1}^{n} \psi(X_i-\hat{\theta}) = 0 \ . \qquad (1.2)$$

Asymptotic properties of $\hat{\theta}$ are well-known and the Princeton study Andrews, *et al.* (1972) suggests that $\sqrt{n}\ (\hat{\theta}-\theta)$ approaches normality fairly quickly. Huber (1970), Gross (1976, 1977), and Shorack (1976) have constructed approximate confidence intervals for θ based on studentization of $\sqrt{n}\ (\hat{\theta}-\theta)$ by estimates of the asymptotic standard deviation.

In this paper a new method of constructing approximate confidence intervals for θ is proposed for the special class of monotone nondecreasing, right continuous ψ functions. The method exploits the fact that $\lambda_{F_n}(c) = -n^{-1}\Sigma\psi(X_i-c)$ is like a distribution function (i.e., nondecreasing and right continuous). In particular, the endpoints of the proposed confidence interval are "percentiles" of λ_{F_n}.

Section 2 gives a motivating example and Section 3 provides the basic ideas and method. In Section 4 Monte Carlo results and comparisons with other results are mentioned. Section 5 shows how to extend to the regression situation and Section 6 is a short summary.

2. MOTIVATION FROM QUANTILE ESTIMATION

Let $F^{-1}(p) = \inf\{x:F(x) \geq p\}$. Consider estimation of the pth quantile $F^{-1}(p), 0<p<1$, from a sample $X_1,\ldots,X_n$ having distribution F. If F_n is the usual empirical df, then

$$P(F_n^{-1}(a) \leq F^{-1}(p) < F_n^{-1}(b)) = P(a \leq F_n(F^{-1}(p)) < b) ,$$

using the fact that all df's G satisfy $G^{-1}(t) \leq x$ iff $t \leq G(x)$. For independent X_i the statistic $nF_n(F^{-1}(p))$ is binomial $(n, F(F^{-1}(p)))$. Thus the normal approximation to the binomial and the assumption $F(F^{-1}(p)) \approx p$ lead us to choose

$$-a = b = p + \sqrt{\frac{p(1-p)}{n}}\, z_{\alpha/2}$$

for an approximate $(1-\alpha)$ confidence interval for $F^{-1}(p)$ (z_α is the $100(1-\alpha)$th percentile of the standard normal.) Although exact non-parametric procedures exist for iid samples from continuous distributions, the above method generalizes to quantile estimation in more complicated situations, e.g., stratified sampling from finite populations. The important point for the present discussion is that M-estimation can use the same idea with F_n replaced by λ_{F_n} .

3. APPROXIMATE CONFIDENCE INTERVALS FOR θ

Let $\psi(t)$ be nondecreasing, right continuous, and strictly positive (negative) for large positive (negative) values of t. Two families of such ψ are "Hubers" $\psi(x) = \max(-k, \min(k,x))$ and "vth power" $\psi(x) = |x|^v \operatorname{sgn}(x)$, $0<v\leq 1$. For df's G define

$$\lambda_G(c) = -\int_{-\infty}^{\infty} \psi(x-c)\,dG(x) \qquad -\infty<c<\infty$$

and

$$\lambda_G^{-1}(t) = \inf\{c\colon \lambda_G(c) \geq t\} \quad t \in (\inf_{-\infty<x<\infty} \lambda_G(x),\ \sup_{-\infty<x<\infty} \lambda_G(x)).$$

The parameter and estimate are defined by $\theta = \lambda_F^{-1}(0)$ and $\hat{\theta} = \lambda_{F_n}^{-1}(0)$. Similar to the case of df's it follows that $\lambda_G^{-1}(t) \leq x$ iff $t \leq \lambda_G(x)$ and thus

$$P(\lambda_{F_n}^{-1}(a) \leq \theta < \lambda_{F_n}^{-1}(b)) = P(a \leq \lambda_{F_n}(\theta) < b)$$

$$= P\left(\frac{\sqrt{n}\,a}{\sigma_n} \leq \frac{\sqrt{n}\,\lambda_{F_n}(\theta)}{\sigma_n} < \frac{\sqrt{n}\,b}{\sigma_n}\right),$$

where $\sigma_n^2 = (n-1)^{-1}\Sigma\psi^2(X_i-\hat{\theta})$ is a reasonable estimate of $E\psi^2(X_1-\theta)$. The statistic of interest,

$$T = \frac{\sqrt{n}\,\lambda_{F_n}(\theta)}{\sigma_n} = \sqrt{n} \cdot \frac{\frac{1}{n}\sum_{i=1}^{n} \psi(X_i-\theta)}{\sqrt{\frac{1}{n-1}\sum_{i=1}^{n} \psi^2(X_i-\hat{\theta})}},$$

has a form very close to a t statistic based on the rv's $\psi(X_i-\theta)$. If the X_i are symmetric about θ and $\psi(x) = -\psi(-x)$, then we expect T to be close to a t distribution with $n-1$ degrees of freedom. Choosing $-\sqrt{n}\,a/\sigma_n = t_{\alpha/2} = \sqrt{n}\,b/\sigma_n$, our proposed approximate $(1-\alpha)$ confidence interval is

$$\left[\lambda_{F_n}^{-1}(-t_{\alpha/2}\sigma_n/\sqrt{n}),\lambda_{F_n}^{-1}(t_{\alpha/2}\sigma_n/\sqrt{n})\right]. \qquad (3.1)$$

Under suitable regularity conditions, the asymptotic width of (3.1) is comparable to methods based on studentizing $\sqrt{n}\,(\hat{\theta}-\theta)$, i.e.,

$$\sqrt{n}\left[(\lambda_{F_n}^{-1}\,(t_{\alpha/2}\sigma_n/\sqrt{n}) - \lambda_{F_n}^{-1}(-t_{\alpha/2}\sigma_n/\sqrt{n}))\right]$$

$$\xrightarrow{P} 2z_{\alpha/2}(E\psi^2(X_1-\theta)/\left[E\psi'(X_1-\theta)\right]^2)^{\frac{1}{2}} .$$

It is often desirable that location estimates satisfy

$$\hat{\theta}(aX_1,\ldots,aX_n) = |a|\,\hat{\theta}(X_1,\ldots,X_n) .$$

For M-estimates the usual procedure is to replace $\psi(x)$ by $\psi_{\hat{\sigma}}(x) = \psi(x/\hat{\sigma})$, where $\hat{\sigma}$ is a suitable scale estimate, or solve simultaneous equations as in Huber's Proposal 2. The above methods carry through exactly and the analogous statistic of interest is

$$T_{\hat{\sigma}} = \sqrt{n}\cdot\frac{\frac{1}{n}\sum_{i=1}^{n}\psi\left(\frac{X_i-\theta}{\hat{\sigma}}\right)}{\sqrt{\frac{1}{n-1}\sum_{i=1}^{n}\psi^2\left(\frac{X_i-\hat{\theta}}{\hat{\sigma}}\right)}} \qquad (3.2)$$

4. COMPARISONS AND MONTE CARLO RESULTS

For small samples the form (3.2) is more appealing than $\sqrt{n}\,(\hat{\theta}-\theta)/\hat{S}$, where $\hat{S}$ is an estimate of the asymptotic standard deviation

$\left[\text{Var IC}(X_1)\right]^{\frac{1}{2}}$, for the following reason. Although $\hat{\theta}-\theta$ is approximated by $n^{-1}\Sigma IC(X_i)$, Boos (1977) shows that this approximation is at best $O_p(n^{-1})$ since $n\left[\hat{\theta}-\theta-n^{-1}\Sigma IC(X_i)\right]$ has a limit distribution. Thus, proximity of $\sqrt{n}(\hat{\theta}-\theta)/\hat{S}$ to a t distribution depends on the t-like statistic $n^{-\frac{1}{2}}\Sigma IC(X_i)/\hat{S}$ *and* the approximation of $\hat{\theta}-\theta$ by $n^{-1}\Sigma IC(X_i)$. In fact Gross (1976) prefers to avoid use of the t distribution. On the other hand Shorack (1976) seems to get very good t approximations for certain Hampels.

In order to spot check the performance of the approximate confidence intervals based on (3.2), a small Monte Carlo study was performed. In Table 1 is found the empirical error probabilities and $\sqrt{n}$ times the expected confidence interval lengths (ECIL) for 10,000 Monte Carlo "samples" generated by the McGill "Super-Duper" random number generator. A different set of 10,000 samples was used for each distribution - normal, logistic, D-EXP = double exponential, T3 = t distribution with 3 degrees of freedom, slash = a standard normal deviate divided by an independent uniform (0,1) deviate, and for each sample size, $n = 10$ and $n = 20$. Only crude Monte Carlo techniques were used, so considerable error may exist in the 3rd decimal of the empirical probabilities and in the 2nd decimal of the ECIL. This is exemplified by the mean whose exact error probability we know to be .05 for the normal. SQRT is the M-estimator based on $\psi(x) = |x|^{\frac{1}{2}}\text{sgn}(x)$ and $Hk = 1.0, 1.5$ are Hubers with $k = 1.0, 1.5$ using a normalized interquartile range as an estimate of scale. $Hk^* = 1.5$ is Huber's Proposal 2 with $k = 1.5$. For both $n = 10$ and $n = 20$ the true levels are generally conservative, but $Hk = 1.5$ and $Hk^* = 1.5$ are fairly close to .05 except for the slash distribution and each has reasonably short ECIL. It is mildly surprising that the mean is so

Table 1. Empirical Error Probabilities and Expected 95-Percent Confidence Interval Lengths (multiplied by $\sqrt{n}$)

	n = 10					n = 20				
Estimator	Normal	Logistic	D-Exp	T3	Slash	Normal	Logistic	D-Exp	T3	Slash
					a. Empirical Error Probabilities					
Mean	.054	.048	.045	.039	.022	.055	.046	.049	.042	.020
SQRT	.039	.033	.030	.029	.017	.048	.043	.040	.040	.026
Hk=1.0	.046	.040	.035	.036	.031	.055	.048	.045	.048	.043
Hk=1.5	.059	.050	.046	.044	.033	.056	.050	.048	.049	.040
Hk*=1.5	.060	.053	.048	.046	.036	.058	.053	.050	.051	.039
					b. Expected 95-Percent Confidence Interval Lengths (multiplied by $\sqrt{n}$)					
Mean	4.39	4.34	4.25	6.88	193.79	4.13	4.09	4.06	6.62	128.90
SQRT	4.93	4.69	4.34	6.73	84.35	4.45	4.17	3.73	5.76	32.15
Hk=1.0	4.97	4.65	4.14	6.28	14.60	4.41	4.08	3.54	5.34	11.36
Hk=1.5	4.51	4.30	3.94	5.95	14.61	4.21	3.99	3.62	5.37	12.22
Hk*=1.5	4.45	4.22	3.88	5.84	14.81	4.16	3.93	3.60	5.32	12.62

Table 2. Monte Carlo Estimates of the $100\cdot(1-\alpha)$th Percentiles of $T_{\hat{\sigma}}$

	n = 10					n = 20				
Estimator	Normal	Logistic	D-Exp	T3	Slash	Normal	Logistic	D-Exp	T3	Slash
					α = .01					
Mean	2.90	2.78	2.65	2.57	2.30	2.57	2.50	2.49	2.41	2.11
SQRT	2.60	2.52	2.42	2.40	2.21	2.48	2.42	2.39	2.34	2.18
Hk=1.0	2.59	2.54	2.47	2.51	2.41	2.54	2.48	2.41	2.49	2.41
Hk=1.5	2.81	2.73	2.64	2.67	2.48	2.60	2.52	2.45	2.51	2.39
Hk*=1.5	2.87	2.80	2.69	2.66	2.53	2.64	2.54	2.47	2.54	2.40
					α = .025					
Mean	2.31	2.23	2.19	2.13	1.91	2.14	2.06	2.09	2.02	1.80
SQRT	2.13	2.08	2.04	2.02	1.86	2.07	2.02	2.00	2.00	1.84
Hk=1.0	2.22	2.16	2.11	2.12	2.05	2.14	2.08	2.04	2.07	2.02
Hk=1.5	2.35	2.26	2.22	2.20	2.09	2.16	2.10	2.07	2.08	2.00
Hk*=1.5	2.36	2.30	2.24	2.21	2.10	2.19	2.13	2.10	2.11	2.01
					α = .05					
Mean	1.86	1.82	1.80	1.77	1.62	1.76	1.71	1.74	1.70	1.55
SQRT	1.75	1.73	1.69	1.66	1.58	1.70	1.67	1.67	1.67	1.59
Hk=1.0	1.84	1.80	1.77	1.75	1.73	1.76	1.74	1.72	1.73	1.71
Hk=1.5	1.91	1.86	1.83	1.81	1.76	1.78	1.73	1.71	1.73	1.70
Hk*=1.5	1.93	1.89	1.84	1.82	1.77	1.80	1.75	1.75	1.75	1.72

close to .05 from normal to T3, though the ECIL are expectedly large for heavy tails. SQRT seems to perform worst over all.

Table 2 represents Monte Carlo estimates of the percentiles of $T_{\hat{\sigma}}$. The percentiles tend to be larger than those of a t distribution for the normal and generally smaller for the heavier-tailed distributions. For $n = 20$ and $\alpha = .05$ all estimates except for SQRT and the mean evaluated at the slash distribution are very close to $t_{.05} = 1.73$ (we should note that the method of calculating the estimated percentiles resulted in considerable error in the second decimal place).

5. REGRESSION

The Huber (1973) regression model is

$$X_i = \sum_{j=1}^{p} c_{ij}\theta_j + U_i$$

where $E(U_i) = 0$ and the c_{ij} are known coefficients. Let $(\hat{\theta}_1,\ldots,\hat{\theta}_p,\hat{\sigma})$ be solutions of

$$\sum_{i=1}^{n} \psi_\sigma(X_i - \sum_{k=1}^{p} c_{ik}\theta_k)c_{ij} = 0 \quad j = 1,p,$$

$$\frac{1}{(n-p)} \sum_{i=1}^{n} \psi_\sigma^2(X_i - \sum_{k=1}^{p} c_{ik}\theta_k) = \beta .$$

Define

$$Q_{n,r}(t) = -\sum_{i=1}^{p} \psi_{\hat{\sigma}}(X_i - \sum_{\substack{k=1 \\ k\neq r}}^{p} c_{ik}\hat{\theta}_k - c_{ir}t)c_{ir} \qquad r = 1,p .$$

Then $\theta_r = Q_{n,r}^{-1}(0)$ and

$$P(Q_{n,r}^{-1}(a) \leq \theta_r < Q_{n,r}^{-1}(b)) = P(a \leq Q_{n,r}(\theta_r) < b).$$

By Taylor expansion in $\hat{\theta}_r - \theta_r$ we find

$$Q_{n,r}(\theta_r) = -\sum_{i=1}^{n} \psi_{\hat{\sigma}}(X_i - \sum_{k=1}^{p} c_{ik}\hat{\theta}_k + c_{ir}(\hat{\theta}_r - \theta_r))c_{ir}$$

$$= -\sum_{i=1}^{n} \psi_{\hat{\sigma}}(X_i - \sum_{i=1}^{p} c_{ik}\hat{\theta}_k)c_{ir} + \frac{1}{\hat{\sigma}} \sum_{i=1}^{n} \psi'_{\hat{\sigma}}(X_i - \sum_{k=1}^{p} c_{ik}\hat{\theta}_k + c_{ir}\theta^*)c_{ir}^2(\hat{\theta}_r - \theta_r) \; .$$

The first term in the above expression is 0 and $\theta^* \xrightarrow{p} 0$ if $\hat{\theta}_r \xrightarrow{p} \theta_r$. Thus, under suitable regularity conditions, $n^{-\frac{1}{2}}Q_{n,r}(\theta_r)$ is asymptotically normal with mean 0 and variance

$$AV_r = \left(\frac{1}{n} \sum_{i=1}^{n} c_{ir}^2\right)^2 E\psi^2(U/\sigma)\left[\left(\frac{C^T C}{n}\right)^{-1}\right]_{rr}$$

An approximate confidence interval for θ_r is

$$\left[Q_{n,r}^{-1}(-t_{\alpha/2}\sigma_n\sqrt{n}) \; , \; Q_{n,r}^{-1}(t_{\alpha/2}\sigma_n\sqrt{n})\right] \; ,$$

where

$$\sigma_n^2 = \left(\frac{1}{n} \sum_{i=1}^{n} c_{ir}^2\right)^2 \left(\frac{1}{n-p} \sum_{i=1}^{n} \psi_{\hat{\sigma}}^2 \left(X_i - \sum_{k=1}^{p} c_{ik}\hat{\theta}_k\right)\right) \left[\left(\frac{C^T C}{n}\right)^{-1}\right]_{rr} \; .$$

The advantage of this method over the usual methods is not clear (The simplicity of the location model is gone!). Note though, that use of ψ' is not required and that for least absolute value regression the above method circumvents an estimate of $f\left[F^{-1}(\frac{1}{2})\right]$.

6. SUMMARY AND CONCLUSIONS

A new procedure for constructing confidence intervals for a location parameter has been proposed which exploits the monotonicity of a class of ψ functions. The distributional problem is reduced to consideration of a t-like statistic and Monte Carlo results verify that "Hubers" perform fairly well over a range of distributions and for samples of size $n = 10$ and $n = 20$.

REFERENCES

Andrews, D. F., *et al.* (1972), *Robust Estimation of Location: Survey and Advances*, Princeton, N. J.: Princeton University Press.

Boos, Dennis D. (1977), "Limiting Second Order Distributions for First Order Functionals, with Application to L- and M-Statistics," Institute of Statistics Mimeo Series #1152, North Carolina State University, Raleigh, N. C.

Gross, Alan M. (1976), "Confidence Interval Robustness with Long-Tailed Symmetric Distributions," *Journal of the American Statistical Association*, 71, 409-416.

____________ (1977), "Confidence Intervals for Bisquare Regression Estimates," *Journal of the American Statistical Association*, 72, 341-354.

Huber, Peter J. (1970), "Studentizing Robust Estimates," in *Nonparametric Techniques in Statistical Inference*, ed. Madan L. Puri, Cambridge: Cambridge University Press, 453-463.

____________ (1973), "Robust Regression: Asymptotics, Conjectures, and Monte Carlo," *Annals of Statistics*, 1, 799-821.

Shorack, Galen R. (1976), "Robust Studentization of Location Estimates," *Statistica Neerlandica*, 30, 119-142.