JACKKNIFING R-ESTIMATORS

William R. Schucany
Southern Methodist University

Simon J. Sheather
University of New South Wales

SMU/DS/TR-220

April 1988

Jackknifing R-Estimators

William R. Schucany
Department of Statistical Science
Southern Methodist University
Dallas TX 75275
USA

Simon J. Sheather
Australian Graduate School of Management
University of New South Wales
Kensington NSW 2033
AUSTRALIA

Abstract:

Sufficient conditions are given for the consistency of the jackknife variance estimator for R-estimators of location in the one- and two-sample problems. An efficient algorithm for computing the variance estimator is presented. Some Monte Carlo evidence of the small sample efficiency is reported. Extensions of the results to R-estimators in the linear model context are discussed.

Keywords:

Differentiability, Hodges-Lehmann, rank estimator, standard error.

1. Introduction

The properties of the jackknife and its associated estimator of standard error have been reported for several large classes of statistics. A variety of consistency results have been obtained for U-statistics [Arvesen (1969)], maximum likelihood and M-estimators [Brillinger (1977) and Reeds (1978)], L-statistics [Parr and Schucany (1982)] and V-statistics [Sen (1977)]. The fundamental issue of smoothness of the functional, T, of the empirical distribution, F_n, that is being jackknifed is addressed by Parr (1985). He unifies many of the existing results by stating sufficient conditions involving Fréchet differentiability of T. It is noteworthy that various authors compile related results from M-, L- and R-estimators [see for example Serfling (1980, Chapter 7, 8 and 9) or Fernholz (1983, Chapter 5)] and yet the R-estimators have not been covered as one of the classes of estimators that behave properly under jackknifing.

One possible reason for this is that the sample median, which is sometimes viewed as the "smoking gun" against the jackknife, is an R-estimator for the one-sample location problem with sign-test scores. Consider the random sample $X_1, \ldots, X_n$ from the distribution F. Under the assumption that F is absolutely continuous and symmetric, the best known R-estimator is the one based on Wilcoxon scores. This estimator is usually referred to as the Hodges-Lehmann estimator after Hodges and Lehmann (1963). The widely used computational form of this statistic is the median of the pairwise averages, $(X_i + X_j)/2$, $1 \le i \le j \le n$. This involvement of the "median" may be enough to lead many to suspect that this Hodges-Lehmann estimator is subject to the same inconsistency of the jackknife variance estimator that plagues the sample median.

One purpose of the present note is to examine the conditions under which R-estimators will admit consistent jackknife estimators of variance. It will be seen that the familiar Hodges-Lehmann estimator may be jackknifed to yield a consistent estimate of its standard error. Some Monte Carlo results in small samples are quite promising. A natural question to address before continuing is whether an estimate of standard error is actually needed.

In the one-sample problem one often has no pressing need to estimate the standard error of an R-estimator. Given symmetry, the distribution-free nature of the associated signed-rank test yields an exact confidence interval from the ordered pairwise averages. However, there are a number of circumstances in which it is necessary to estimate standard errors. For example, suppose we have independent sets of observations leading to estimates of θ_1, θ_2, θ_3 and that we are interested in $\Delta = \theta_1 - 2\theta_2 + \theta_3$. Then the exact confidence intervals for θ_1, θ_2 and θ_3 cannot be translated into an exact confidence interval for Δ, except by being very conservative. Approximate methods have to be used and one of the most direct ones is based on the standard error of the estimate of Δ, calculated in the obvious way using the standard errors of the individual estimates. In addition, when the problem involves R-estimation in regression models, convenient approaches to interval estimation like those in the one-sample problem do not apply. The large sample normality and formulas for the asymptotic standard errors have formed the basis for current approaches [see for example Hettmansperger (1984, Chapter 5)]. The asymptotic standard error of the Hodges-Lehmann is proportional to $\int f^2$, where $f = F'$ is the density of the additive error term. Nonparametric estimation schemes for this integral of the squared density all involve selection of a smoothing parameter with all of the accompanying difficulties [see for example, Koul, Sievers and McKean (1987), Schweder (1975) and Sheather (1987)]. Consequently in this setting a reliable and less involved estimator of standard error would be a welcome addition. Finally, even though with adequate computing support the bootstrap is an obvious candidate, the current view [see Hall (1988)] is that studentized quantities lead to better approximate confidence intervals. In other words, to use the percentile-*t* method one must produce bootstrap replicates of coefficients that have been standardized by estimates of standard error. Therefore, this refined and more accurate version of the bootstrap does not obviate the need to estimate standard errors.

2. Conditions for Consistency

Consider the one-sample location problem in which $X_1, X_2, \ldots, X_n$ are independent and identically distributed from F, an absolutely continuous distribution with unknown centre of symmetry θ_0. Let $R_i(\theta) =$ rank of $|X_i - \theta|$ among the n absolute residuals. A signed-rank test statistic has the form

$$V_n(\theta) = \frac{1}{n}\sum_{i=1}^{n}\phi\{\frac{R_i(\theta)}{n+1}\ \mathrm{sgn}\ (X_i - \theta)\}, \qquad (2.1)$$

where $\phi(u)$ is a non negative and non decreasing score function on $0 < u < 1$ such that $\int_0^1 \phi(u)\,\mathrm{du} < \infty$ and $\int_0^1 \phi^2(u)\,\mathrm{du} < \infty$. Further, the definition of ϕ is extended to (-1, 1) by $\phi(-u) = -\phi(u)$ so that ϕ is odd. The R-estimator, $\hat{\theta}$, corresponding to (2.1) is a root of

$$V_n(\theta) = 0. \tag{2.2}$$

The fact that V_n is discontinuous implies that an exact root does not exist and $\hat{\theta}$ is taken to be the value at which V_n changes sign. If there are multiple roots, they form an interval and the choice is the midpoint of this interval.

Hettmansperger (1984, pp.99–100) proves that the defining equation (2.2) leads to an asymptotically equivalent equation,

$$\int_{-\infty}^{\infty} \phi\{F(x) - F(-x+2\theta)\}\, dF(x) = 0. \tag{2.3}$$

The location functional such that $\theta_0 = T(F)$ is defined implicitly as the solution of (2.3). An R-estimator is given by $T(F_n)$, where F_n is the empirical distribution function.

To discuss the jackknife we require some additional notation for an empirical distribution function based on the $(n-1)$ observations with X_i held out. For a review of the jackknife see Miller (1974). For a recent bibliography see Frangos (1987). Let

$$F_{ni}(x) = (n-1)^{-1} \sum_{\substack{j=1 \\ j\neq i}}^{n} I(X_j \le x), \quad i = 1, 2, \ldots, n$$

and then the usual jackknife estimate of the variance of $T(F_n)$ may be written

$$S_J^2 = (n-1)^{-1} \sum_{i=1}^{n} [T(F_{ni}) - n^{-1} \sum_{j=1}^{n} T(F_{nj})]^2. \tag{2.4}$$

Under smoothness conditions on T and F, specified below in (i) and (ii), we have $\sqrt{n}\,[T(F_n) - T(F)] \to N(0, \sigma^2)$ in distribution and $S_J^2 \to \sigma^2$ with probability 1 as $n \to \infty$.

Working with the sup norm, $||F - G|| = \sup_{-\infty < x < \infty} |F(x) - G(x)|$, Fernholz (1983) establishes Hadamard differentiability of certain R-estimator functionals, and then asymptotic normality as a consequence. With the additional condition that its asymptotic variance functional is continuous at F we may employ Theorem 2.1 of Sen (1988) to conclude that S_J^2 converges almost surely to σ^2. This theorem of Sen relaxes the requirements for strong Fréchet differentiability by Parr (1985). In particular this result holds if we assume that

(i) ϕ is continuous, odd, increasing and piecewise differentiable with bounded, piecewise continuous derivative, ϕ', and there exists an $m > 0$ such that $\phi'(u) \geq m$ for all u in some neighbourhood of zero; and

(ii) F is symmetric about θ_0 with bounded and continuous density function, f.

Only condition (ii) is slightly more restrictive than is required by Fernholz (1983, p.100), but these are sufficient to ensure that the Hadamard differential exists and is continuous. Sen (1988) requires the continuity at F of

$$T_1(G) = \int \psi^2 (x; G)\, d\, G(x), \tag{2.5}$$

where the influence curve here is

$$\psi(x; F) = \frac{\phi\{2F(x) - f\, 1\}.}{2\int \phi'\{2F(y) - 1\} f^2(y)\, \mathrm{dy}} \tag{2.6}$$

Clearly, if ψ is a continuous functional, then T_1 is also. We do not need to restrict the class of R-estimators sufficiently to yield continuity in (2.6). However, it may be easier to verify than (2.5). For the case of immediate interest $\phi(u) = u$ and (2.6) reduces to $\{F(x) - ½\}/\int f^2(y)\, \mathrm{dy}$. This is clearly bounded and continuous. This latter property gives the Hodges-Lehmann estimator a degree of robustness not shared by the median. Interestingly, another robustness feature, the "change-of-variance" discussed by Hampel et al. (1986), also involves smoothness of the asymptotic variance functional, (2.5). Other popular R-estimators which satisfy the conditions in (i) are those associated with Winsorized Wilcoxon scores and Winsorized normal scores. Finally, it should be noted that the sign-test score function associated with the median as an R-estimator does not satisfy the smoothness conditions in (i).

In the two-sample problem we have independent samples $X_1, X_2, \ldots, X_n$ and $Y_1, Y_2, \ldots, Y_n$ from $F(x)$ and $G(x) = F(x - \Delta)$, respectively. The conditions for strong consistency of the jackknife estimate of the variance of an R-estimate in this case are the same as (i) and (ii) expect that symmetry of F is no longer required.

General results on strong consistency of the jackknife standard error estimate for non linear statistical functionals have been restricted to the i.i.d. case. The extension of these results to the regression case requires a further approximation, since F_n is based on residuals which are dependent. There is no reason to suspect that the strong consistency result will not hold in this case.

3. Small Sample Implementation

To investigate the finite sample efficiency of the jackknife estimator of variance, S_J^2, a computationally efficient algorithm for the one- and two-sample settings can be described. In other words, when the pairwise averages or differences are used to compute the R-estimator, then these same quantities permit the quick identification of the values of the "leave-out" estimators, $T(F_{ni})$. The implementation will be discussed in the context of the Hodges-Lehmann estimator the one-sample problem and some illustrative simulation results described.

The first step is to calculate and store the $M = n(n+1)/2$ pairwise averages. In two vectors of the same length save the indices i and j that correspond to the pair of ordered X values averaged at each of the M entries. In an unpublished University of New South Wales Working Paper, Robinson and Sheather show that it is possible to reduce the number of averages by about a factor greater than two because only certain middle values are candidates for the median. However, except for large-scale simulation studies, the bookkeeping of the pertinent indices may outweigh the greater computational efficiency that follows from only needing to calculate and sort about half as many pairwise averages.

The second step is to sort the M averages. A routine such as (SVRGP) in IMSL also returns an array of indices that permits the permutation of the two vectors of indices to follow the rearrangement of the pairwise averages. In other words, the original X values that enter each ordered average can still be identified easily by reference to the two permuted vectors of indices.

The third step is to select the n values of the Hodges-Lehmann estimator corresponding to the reduced samples associated with successively deleting each X_i. For each subsample there are $M_1 = M - n$ pairwise averages to consider. The savings that may be realized involve using the existing array of ordered pairwise averages and ignoring those n averages involving the particular X_i. The Hodges-Lehmann estimate, $\hat{\theta}$, for the full sample is the median of the M pairwise averages. The "delete-one" values, $\hat{\theta}_i$, are each medians of subsets of M_1 pairwise averages. For a fixed index, i_0, to be deleted, the vectors of indices are examined from the beginning (which corresponds to the minimum pairwise average). If neither index equals i_0, then the corresponding average is counted as belonging to the relevant subset. In this way the middle value(s) can be identified efficiently; no recalculation, no additional sorting and only integer comparisons and counting are needed.

The final step is the calculation of the jackknife standard error estimate, $SE_J = S_J/\sqrt{n}$, where S_J^2 is from (2.4) with $T(F_{ni}) = \hat{\theta}_i$.

Some simple Monte Carlo runs for comparison with bootstrap estimates of standard error, SE_B, are reported in Table 1. For comparison with asymptotic values and comparability across distributions, the numbers reported in Table 1 are standard deviations (and their estimates) divided by $1/\sqrt{12n}\int f^2$. At $n = 10$ and $n = 18$ the jackknife calculations are particularly simple because M and $M_1 = M - n$ are both odd and so each of the pertinent medians is a single pairwise average rather than the midpoint of two. For these special cases calculations for the jackknife are about 50 times as fast as the bootstrap with $B = 100$. To obtain stable estimates of average values 10,000 repetitions were used. Normal deviates, z, and Uniform deviates, u, were generated directly from IMSL routines; the Laplace variates by the inverse edf transformation and Slash by z/u. Matched pairs t-tests for all differences between average values of SE_J and SE_B were highly significant.

The excellent agreement between the averages of SE_J and standard deviation of $\hat{\theta}$ calculated from the 10,000 samples is obvious. Clearly the jackknife is not as sensitive to heavy tails as the bootstrap, e.g., slash inflates SE_B and its standard deviation. However, the jackknife does give up something in statistical efficiency for distributions with light to moderate tail weight.

Acknowledgements

The research of the first author was partially supported by a grant from the US Office of Naval Research. This research was completed while the first author was visiting the Institute of Advanced Studies, Australian National University.

Table 1

Jackknife and bootstrap Monte Carlo estimates of moments of estimators for the standard error of the Hodges-Lehmann estimator

		Uniform	Normal	Laplace	Slash
$n = 10$					
$\hat{\theta}$	S.E.	1.150	1.015	1.062	1.272
$SE_J(\hat{\theta})$	Mean	1.138	1.019	1.077	1.261
	(S.D.)	(.437)	(.394)	(.503)	(1.233)
$SE_B(\hat{\theta})$	Mean	1.049	0.999	1.199	8.534
	(S.D.)	(.209)	(.255)	(.426)	(74.53)
$n = 18$					
$\hat{\theta}$	S.E.	1.109	1.019	1.048	1.114
$SE_J(\hat{\theta})$	Mean	1.125	1.010	1.046	1.094
	(S.D.)	(.337)	(.308)	(.378)	(.511)
$SE_B(\hat{\theta})$	Mean	1.105	1.003	1.096	1.661
	(S.D.)	(.178)	(.209)	(.297)	(3.913)

Estimated standard error of means = S.D./100.

References

Arvesen, J.N. (1969). Jackknifing *U*-statistics. *Ann. Math. Statist., 40,* 2076–2100.

Brillinger, D.R. (1977). Approximate estimation of the standard errors of complex statistics based on sample surveys. *New Zealand Statistician, 11,* 35–41.

Fernholz, L.T. (1983). *von Mises Calculus for Statistical Functionals.* New York: Springer-Verlag.

Frangos, C.C. (1987). An updated bibliography on the jackknife method. *Commun. Statist., A, 16,* 1543–1584.

Hall, P. (1988). Theoretical comparison of bootstrap confidence intervals. *Ann. Statist.*, to appear.

Hampel, F.R., Ronchetti, E.M., Rousseeuw, P.J. and Stahel, W.A. (1986). *Robust Statistics.* New York: Wiley.

Hettmansperger, T.P. (1984). *Statistical Inference Based on Ranks.* New York: Wiley.

Hodges, J.L. and Lehmann, E.L. (1963). Estimates of location based on rank tests. *Ann. Math. Statist., 34,* 598–611.

Koul, H.L., Sievers, G.L. and McKean, J.W. (1987). An estimator of the scale parameter for the rank analysis of linear models under general score functions. *Scand. J. Statist.*, 14, 131–141.

Miller, R.G., Jr. (1974). The jackknife: a review. *Biometrika, 61*, 1–15.

Parr, W.C. (1985). Jackknifing differentiable statistical functionals. *J. R. Statist. Soc., B, 47*, 56–66.

Parr, W.C. and Schucany, W.R. (1982). Jackknifing *L*-statistics with smooth weight functions. *J. Amer. Statist. Assoc., 77*, 629–638.

Reeds, J.A. (1978). Jackknifing maximum likelihood estimates. *Ann. Statist., 6*, 727–739.

Schweder, T. (1975). Window estimation of the asymptotic variance of rank estimators of location. *Scand. J. Statist., 2*, 113–126.

Sen, P.K. (1977). Some invariance principles relating to jackknifing and their rôle in sequential analysis. *Ann. Statist., 5*, 316–329.

Sen, P.K. (1988), Functional jackknifing: rationality and general asymptotics. *Ann. Statist.*, 16, 450–469.

Serfling, R.J. (1980). *Approximation Theorems of Mathematical Statistics*. New York: Wiley.

Sheather, S.J. (1987). A new method of estimating the asymptotic standard error of the Hodges-Lehmann estimator based on generalized least squares. *Austral. J. Statist., 29*, 66–83.