The Generalized Jackknife: Finite Samples and Subsample Sizes

TREVOR SHAROT*

This paper is concerned with bias-reduction properties of that generalized jackknife which pertains to a single-estimator, single-sample situation. Jackknifing is a technique for reducing bias by exploiting the dependence of the bias on sample size. In practice, this is carried out by reestimating the unknown parameter(s) with only part of the sample and combining the new and original estimates suitably weighted to produce cancellation in their biases. The size of the subsample affects both the bias and variance of the jackknife. While few results are available, "minimal data omission" seems sensible on grounds of variance. This paper shows the same is always true on grounds of bias.

1. INTRODUCTION

The basic jackknife statistic, due to Quenouille [6], is

$$\hat{\theta} = (nt_0 - (n-r)\bar{t}_1)/r \ , \qquad (1.1)$$

where t_0 is a consistent, though possibly biased, estimator of θ based on a sample of size n, and t_1 is the same estimator based on a subsample of size $n - r$; possibly, the averaging is over some or all of the $\binom{n}{n-r}$ such subsamples. If $n - r = n/j$ for some integer $j \geqslant 2$, then j is the number of disjoint subsets, and the average may be taken over these alone.

Two examples in particular were suggested by Quenouille. The earlier of the two [5] uses $r = n/2$ with averaging over the resulting pair of disjoint subsets a and b, giving

$$\hat{\theta} = 2t_0 - \tfrac{1}{2}(t_{1a} + t_{1b}) \ . \qquad (1.2)$$

The second [6] uses $r = 1$ with averaging over all n subsets, giving

$$\hat{\theta} = nt_0 - (n-1)\bar{t}_1 \ . \qquad (1.3)$$

However, Miller [4, p. 1595] commented (in present notation), "There are no mathematical guidelines yet on the relative choice of j and n but there may be external design reasons for their choice."

Subsequently, Rao and Webster [7], following from Durbin [3], showed that in ratio estimation the bias of $\hat{\theta}$ is decreasing in j, so that $j = n$ $(r = 1)$ is optimum. In addition, under appropriate distributional assumptions, the same was shown to apply to the variance.

In Schucany, Gray and Owen [8], the issue was broadened by an exposition of the generalized jackknife, which includes the jackknife as a special case. In present notation, the first-order jackknife $\hat{\theta}^{(1)} = \hat{\theta}$ in (1.3). Writing t_i for the estimator based on $n - ir$ observations, the second-order jackknife is

$$\hat{\theta}^{(2)} = [n^2t_0 - 2(n-r)^2t_1 + (n-2r)^2t_2]/2r^2 \ , \qquad (1.4)$$

and higher orders are possible, if impractical. It will be shown that $r = 1$ is always optimum (in a certain sense) for $\hat{\theta}^{(k)} (k \geqslant 1)$ on grounds of bias, comparing, in particular, the bias properties of $\hat{\theta}^{(1)}$ and $\hat{\theta}^{(2)}$.

2. THE GENERALIZED JACKKNIFE

In the form proposed in [8], the kth-order generalized jackknife is, for general r,

$$\hat{\theta}^{(k)} = \frac{\begin{vmatrix} t_0 & t_1 & \cdots & t_k \\ 1/n & 1/(n-r) & \cdots & 1/(n-kr) \\ \vdots & \vdots & & \vdots \\ 1/n^k & 1/(n-r)^k & \cdots & 1/(n-kr)^k \end{vmatrix}}{\begin{vmatrix} 1 & 1 & \cdots & 1 \\ 1/n & 1/(n-r) & \cdots & 1/(n-kr) \\ \vdots & \vdots & & \vdots \\ 1/n^k & 1/(n-r)^k & \cdots & 1/(n-kr)^k \end{vmatrix}} \ . \qquad (2.1)$$

A closed form for $\hat{\theta}^{(k)}$ will now be derived. Let $\hat{\theta}^{(k)} = N/D$. Then,

$$\begin{aligned} D &= [n(n-r)\ldots(n-kr)]^{-k} \begin{vmatrix} n^k & (n-r)^k & \cdots & (n-kr)^k \\ n^{k-1} & (n-r)^{k-1} & \cdots & (n-kr)^{k-1} \\ \vdots & \vdots & & \vdots \\ 1 & 1 & \cdots & 1 \end{vmatrix} \\ &= [n(n-r)\ldots(n-kr)]^{-k}V \ , \quad \text{say} \ , \end{aligned} \qquad (2.2)$$

where V is a Vandermonde determinant which may be expanded as the product of all differences of the form

$$\begin{aligned} &(n - \ell r) - (n - mr) \\ &\quad = (m-\ell)r \ ; \quad 0 \leqslant \ell \leqslant k-1 \ , \quad \ell + 1 \leqslant m \leqslant k \end{aligned}$$

(See [1, p. 41–2]).

The products involving a given ℓ equal

$$(k-\ell)r(k-1-\ell)r \ldots (1)r = (k-\ell)!r^{k-\ell} \ ,$$

* Trevor Sharot is statistician, Statistics Department, Audits of Great Britain, Ltd., Audit House, Eastcote, Ruislip, Middlesex HA4 9LT, England. This paper was written at Department of Statistics, University of Aberdeen. The author would like to thank the referee for many helpful comments and corrections.

© Journal of the American Statistical Association
June 1976, Volume 71, Number 354
Theory and Methods Section

so V equals

$$[k!(k-1)!\dots 1!]r^{k(k+1)/2} ,$$

and

$$D = [n(n-r)\dots(n-kr)]^{-k} \cdot [k!(k-1)!\dots 1!]r^{k(k+1)/2} . \quad (2.3)$$

Next, we extract the same factor

$$[n(n-r)\dots(n-kr)]^{-k}$$

from N. The top row of the resulting determinant M is

$$(n^k t_0, (n-r)^k t_1, \dots, (n-kr)^k t_k) ,$$

and its cofactors are Vandermonde determinants. The ith of these, F_i, say, $(i = 0, \dots, k)$ is obtained from V by omitting the first row and $(i+1)$th column, so the products involving this column,

$$(k-i)(k-1-i)\dots(1)\cdot r^k \quad (i=0)$$

$$(k-i)(k-1-i)\dots(1)\cdot(1)(2)\dots(i)\cdot r^k \quad (1 \leqslant i \leqslant k-1 , \quad \text{for } k \geqslant 2)$$

$$(1)(2)\dots(i)\cdot r^k \quad (i=k) ,$$

disappear. Hence,

$$F_i = V/(k-i)!i!r^k .$$

Thus, expanding M by the top row,

$$\begin{aligned} M &= n^k F_0 t_0 - (n-r)^k F_1 t_1 + \dots + (-1)^k (n-kr)^k F_k t_k \\ &= \frac{V}{r^k k!}\left\{ n^k \binom{k}{0} t_0 - (n-r)^k \binom{k}{1} t_1 + \dots + (-1)^k (n-kr)^k \binom{k}{k} t_k \right\} . \quad (2.4) \end{aligned}$$

Finally, therefore,

$$\begin{aligned} \hat{\theta}^{(k)} &= \frac{M}{V} = \frac{1}{r^k k!} \sum_{j=0}^{k} (-1)^j \binom{k}{j} (n-jr)^k t_j \\ &= \frac{1}{p^k k!} \sum_{j=0}^{k} (-1)^j \binom{k}{j} (1-jp)^k t_j , \quad (2.5) \end{aligned}$$

where $p = r/n$.
This is the required closed form. Clearly, also,

$$E(\hat{\theta}^{(k)} - \theta) = \frac{1}{p^k k!} \sum_{j=0}^{k} (-1)^j \binom{k}{j} (1-jp)^k E(t_j - \theta) . \quad (2.6)$$

3. BIAS REDUCTION

We now assume that the bias of t_0 may be arbitrarily closely represented as a sum of powers of $1/n$, i.e.,

$$E(t_0 - \theta) = \sum_{i=0}^{\infty} a_i n^{-x_i} ; \quad (3.1)$$

$x_i > 0$ (not necessarily integer), all i. Then,

$$\begin{aligned} E(t_j - \theta) &= \sum_{i=0}^{\infty} a_i (n-jr)^{-x_i} \\ &= \sum_{i=0}^{\infty} a_i n^{-x_i} (1-jp)^{-x_i} . \quad (3.2) \end{aligned}$$

Let the attenuation of a particular bias term $a_i n^{-x_i}$ in (3.1) resulting from jackknifing be $\alpha(x_i, p, k)$. Then,

$$E(\hat{\theta}^{(k)} - \theta) = \sum_{i=0}^{\infty} a_i \alpha(x_i, p, k) n^{-x_i} .$$

Now, from (2.6),

$$\begin{aligned} \alpha(x, p, k) &= \frac{1}{p^k k!} \sum_{j=0}^{k} (-1)^j \binom{k}{j} (1-jp)^k (1-jp)^{-x} \\ &= \frac{1}{p^k k!} \sum_{j=0}^{k} (-1)^j \binom{k}{j} (1-jp)^{k-x} \quad (3.3) \\ &= \frac{1}{p^k k!} [\Delta^k h^{k-x}]_{h=1} , \quad (3.4) \end{aligned}$$

where Δ is the backward difference operator of step p, i.e.,

$$\Delta f(h) = f(h) - f(h-p) .$$

Asymptotically ($n \to \infty$ or $p \to 0$), it follows that

$$\begin{aligned} \alpha(x, 0, k) &= \frac{1}{k!}\left[\frac{d^k}{dh^k} h^{k-x}\right]_{h=1} \\ &= \frac{1}{k!}(k-x)\dots(1-x)[h^{-x}]_{h=1} ; \end{aligned}$$

therefore,

$$\alpha(x, 0, k) = \frac{(k-x)\dots(1-x)}{k!} . \quad (3.5)$$

Results (3.4) and (3.5) are presented on their own merits, as generalizations of Theorem 1 in Adams, *et al.* [2]; their p is here k.

We now return to (3.3) to prove the following theorem.

Theorem:

$$\alpha > 0 \Leftrightarrow \partial\alpha/\partial p > 0 .$$

Hence, minimum p (i.e., $r = 1$) minimizes $|\alpha|$ for all x and k. It follows that $r = 1$ gives maximum attenuation of every bias term in (3.1).

Proof: Expand $(1-jp)^{k-x}$ as a power series in p:

$$(1-jp)^{k-x} = \sum_{m=0}^{\infty} (-1)^m \binom{k-x}{m} j^m p^m . \quad (3.6)$$

The coefficient of p^{m-k} in (3.3) is, therefore,

$$c_{m-k} = \frac{1}{k!} \sum_{j=0}^{k} (-1)^j \binom{k}{j} (-1)^m \binom{k-x}{m} j^m . \quad (3.7)$$

But,

$$\frac{(-1)^k}{k!} \sum_{j=0}^{k} (-1)^j \binom{k}{j} j^m = S_m{}^k , \quad (3.8)$$

which are the Stirling numbers of the second kind. Substituting (3.8) into (3.7),

$$c_{m-k} = (-1)^{m-k} \binom{k-x}{m} S_m{}^k \, . \tag{3.9}$$

Since $S_m{}^k = 0$ when $m < k$, write $i = m - k$ giving

$$\alpha(x, p, k) = \sum_{i=0}^{\infty} c_i p^i \, , \tag{3.10}$$

where, from (3.9),

$$c_i = (-1)^i \binom{k-x}{k+i} S_{k+i}{}^k \, , \tag{3.11}$$

or

$$c_i = (-1)^i (k-x)(k-1-x)\ldots(1-i-x) \frac{S_{k+i}{}^k}{(k+i)!} .$$

Now let I = integer part of x. The Stirling numbers $S_{k+i}{}^k$ are all positive integers, so the sign of c_i is, for $x < k$,

$$(-1)^i(+1)^{k-I}(-1)^{i+I} = (-1)^I \, ,$$

and for $x > k$,

$$(-1)^i(+1)^0(-1)^{k-i} = (-1)^k \, ,$$

or, in general,

$$(-1)^{\min(I,k)} \, . \tag{3.12}$$

Since this is the same for all i, it is also, from (3.10) the sign of α, i.e.,

$$\alpha > 0 \Leftrightarrow \min\,(I, k) \text{ is even} \, .$$

The corresponding result for $\partial\alpha/\partial p$ is easily derived as follows. Equation (3.6) and, hence, (3.10) always converge in practice, since $jp \leqslant kp = kr/n < 1$, because $n - kr > 0$, so differentiating term-by-term,

$$\frac{\partial \alpha}{\partial p} = \sum_{i=1}^{\infty} i c_i p^{i-1} \, .$$

Since the sign of ic_i is that of c_i, we have

$$\alpha > 0 \Leftrightarrow \min\,(I, k) \text{ is even} \Leftrightarrow \partial\alpha/\partial p > 0 \, , \tag{3.13}$$

which completes the proof.

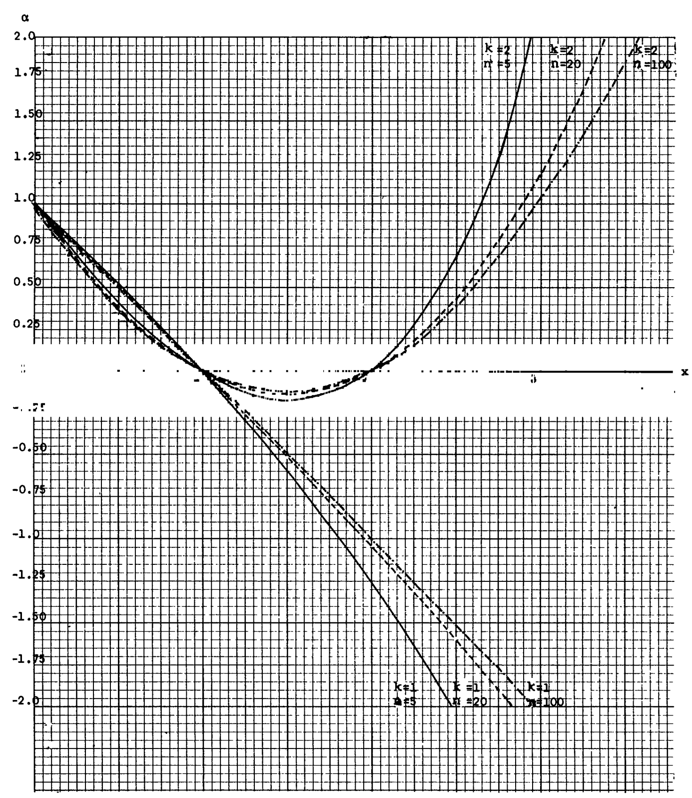
Corollary: The attenuation of every term in (3.1) increases with n, for fixed k and r.

4. DISCUSSION

The behavior of α is, thus, as follows. As a function of x, it is unity when $x = 0$ (asymptotic bias is unchanged), oscillates in sign up to $x = k$ with zeros at $x = 1, 2, \ldots, k$ (these integer-powered biases thus disappearing), and then diverges to $+\infty$ or $-\infty$, depending which way it crosses the axis at $x = k$. As p decreases (r decreasing or n increasing), the amplitude of the oscillations is reduced, so the attenuation increases.

The behavior of $\alpha(x, p, k)$ for $\hat{\theta}^{(1)}$ and $\hat{\theta}^{(2)}$ is plotted against x in the figure for $n = 5$, 20 and 100 with $r = 1$. The asymptotic superiority of $\hat{\theta}^{(2)}$, which was discussed in [2], can be seen to hold in finite samples up to negligibly high orders of bias. These curves also illustrate the sensitivity of α to the choice of r, since they hold for any r and n such that $p = 0.2$, 0.05 and 0.01, respectively. So for $n = 100$, $r = 20$, 5, and 1, respectively. The superiority of $r = 1$ over $r = 5$ is slight, though it will normally be considerable on grounds of variance.

Jackknife Attenuation Curves



5. CONCLUSION

In conclusion, $r = 1$ is always optimal for the purpose of attenuating every term of (3.1). It is perfectly possible to construct special cases of (3.1) for which jackknifing nevertheless increases the overall bias, or for which $\hat{\theta}^{(1)}$ is superior to $\hat{\theta}^{(2)}$ (by allowing both positive and negative a_i to appear). Under these conditions, bias reduction is only guaranteed asymptotically. Where a_i are available from theoretical considerations, such a problem may be identified and dealt with on its own merits. Where the a_i are unknown and the sample is finite, we have demonstrated the following desirable properties. First, $r = 1$ is effective for *all* powers of bias up to a limit between $n^{-\frac{3}{2}}$ and n^{-2}, depending on $n (n \geqslant 5)$. Second, $\hat{\theta}^{(2)}$ is superior to $\hat{\theta}^{(1)}$ for all these powers and effective up to a limit between $n^{-2.3}$ and n^{-3}. In this sense, the jackknives may be considered robust estimators.

[*Received September 1974. Revised June 1975.*]

REFERENCES

[1] Aitken, A.C., *Determinants and Matrices*, London: Oliver & Boyd Ltd., 1939.

[2] Adams, J.E., Gray, H.L. and Watkins, T.A., "An Asymptotic Characterization of Bias Reduction by Jackknifing," *Annals of Mathematical Statistics*, 42 (October 1971), 1606–12.

[3] Durbin, J., "A Note on the Application of Quenouille's Method of Bias Reduction to the Estimation of Ratios," *Biometrika*, 46 (December 1959), 477–80.

[4] Miller, R.G., "A Trustworthy Jackknife," *Annals of Mathematical Statistics*, 35 (December 1964), 1594–605.

[5] Quenouille, M.H., "Approximate Tests of Correlation in Time Series," *Journal of the Royal Statistical Society, Ser. B*, 11 (Part 1, 1949), 68–84.

[6] ———, "Notes on Bias Reduction," *Biometrika*, 43 (December 1956), 353–60.

[7] Rao, J.N.K. and Webster, J.T., "On Two Methods of Bias Reduction in the Estimation of Ratios," *Biometrika*, 53 (December 1966), 571–7.

[8] Schucany, W.R., Gray, H.L. and Owen, D.B., "On Bias Reduction in Estimation," *Journal of the American Statistical Association*, 66 (September 1971), 524–33.