The Annals of Mathematical Statistics
1971, Vol. 42, No. 5, 1606–1612

AN ASYMPTOTIC CHARACTERIZATION OF BIAS REDUCTION BY JACKKNIFING

BY J. E. ADAMS, H. L. GRAY, AND T. A. WATKINS
Texas Tech University

This paper investigates the effectiveness of the first and second order jackknife estimators, as tools for bias reduction. The biases of the estimators are characterized in terms of the bias of the original estimator. The biases of the two estimators are compared asymptotically with each other and with the bias of the original estimator.

The jackknife estimator introduced by Quenouille (1949) has been the subject of much study in recent years. In many of these studies a question of primary concern is: under what circumstances is the jackknife procedure an effective tool for bias reduction? The purpose of this paper will be to give characterizations of the expected values of the first order jackknife and second order jackknife, as defined in Schucany, Gray and Owen (1971), in terms of the bias of the original estimator, and to determine conditions under which one can compare each of these statistics with the original estimator in order to evaluate their effectiveness in bias reduction. Also relationships between the biases of the first order jackknife and second order jackknife are investigated.

Let θ be an unknown parameter, and let $X_1, X_2, \cdots, X_N$ be N independent, identically distributed observations from the cdf F_θ. Suppose $\hat{\theta} = \hat{\theta}(X_1, X_2, \cdots, X_N)$ is an estimator for θ where $E(\hat{\theta}) \neq \theta$. Now partition the sample into n groups of size k where $N = nk$ and form subsamples of size $N-k$ by deleting one of the groups of observations, and subsamples of size $N-2k$ by deleting two groups of observations. Then we define $\hat{\theta}^i$ as the estimator obtained by evaluating $\hat{\theta}$ on the subsample of size $N-k$ where the ith group of observations has been deleted; $\hat{\theta}^{ij}$, $i \neq j$, as the estimator obtained by evaluating $\hat{\theta}$ on the sample of size $N-2k$ where the ith and jth groups have been deleted.

The first order jackknife, $J^{(1)}[\hat{\theta}]$, and the second order jackknife, $J^{(2)}[\hat{\theta}]$, are then defined by

$$J^{(1)}[\hat{\theta}] = n\hat{\theta} - \frac{n-1}{n}\sum_{i=1}^{n} \hat{\theta}^i \tag{1}$$

and

$$J^{(2)}[\hat{\theta}] = \tfrac{1}{2}\left[n^2\hat{\theta} - 2(n-1)^2 \cdot \frac{1}{n}\sum_{i=1}^{n} \hat{\theta}^i + (n-2)^2 \cdot \frac{1}{n(n-1)}\sum_{i\neq j} \hat{\theta}^{ij}\right]. \tag{2}$$

The definition given by (2) is not the standard one but apparently is the most appropriate one as was shown in Schucany, Gray and Owen (1971).

We now present characterizations for $E[J^{(1)}[\hat{\theta}]]$ and $E[J^{(2)}[\hat{\theta}]]$.

Received September 28, 1970.

THEOREM 1. *Let* $\hat{\theta}^i$ *and* $\hat{\theta}^{ij}$ *be defined as above. Further define*

$$E[\hat{\theta}-\theta] = B(N,\theta) = B(nk,\theta). \tag{3}$$

Then, for $p = 1, 2$

$$E[J^{(p)}(\hat{\theta})] = \theta + B(nk,\theta) + (p-1)(pn-p-1)\Delta B(nk,\theta) + \frac{(n-p)^p}{p}\Delta^p B(nk,\theta), \tag{4}$$

where

$$\Delta^1 B(nk,\theta) = \Delta B(nk,\theta) = B(nk,\theta) - B((n-1)k,\theta)$$

and

$$\Delta^2 B(nk,\theta) = B(nk,\theta) - 2B((n-1)k,\theta) + B((n-2)k,\theta).$$

PROOF. Since $X_1, X_2, \cdots, X_n$ are independent identically distributed random variables then

$$E[\hat{\theta}^i - \theta] = B((n-1)k,\theta) \qquad \text{for all } 1 \leqq i \leqq n$$

and

$$E[\hat{\theta}^{ij} - \theta] = B((n-2)k,\theta) \qquad \text{for all } 1 \leqq i \leqq n, i \leqq j \leqq n, i \neq j.$$

By algebraic manipulation (1) can be written as

$$J^{(1)}[\hat{\theta}] = \hat{\theta} + \frac{n-1}{n}\sum_{i=1}^{n}[\hat{\theta}-\hat{\theta}^i].$$

Thus

$$\begin{aligned} E[J^{(1)}[\hat{\theta}]] &= \theta + B(nk,\theta) + \frac{(n-1)}{n}\sum_{i=1}^{n}[B(nk,\theta) - B((n-1)k,\theta)] \\ &= \theta + B(nk,\theta) + (n-1)\Delta B(nk,\theta). \end{aligned}$$

Similarly (2) can be written as

$$J^{(2)}[\hat{\theta}] = \frac{n}{2(n-1)}\sum_{i\neq j}[\hat{\theta} - 2\hat{\theta}^i + \hat{\theta}^{ij}] + \frac{2}{n}\sum_{i\neq j}[\hat{\theta}^i - \hat{\theta}^{ij}] + \frac{1}{n}\sum_{i=1}^{n}\hat{\theta}^i.$$

We now obtain

$$\begin{aligned} E[J^{(2)}[\hat{\theta}]] &= \frac{n^2}{2}[B(nk,\theta) - 2B((n-1)k,\theta) + B((n-2)k,\theta)] \\ &\quad + 2(n-1)[B((n-1)k,\theta) - B((n-2)k,\theta)] + \theta + B((n-1)k,\theta) \\ &= \theta + B(nk,\theta) - [B(nk,\theta) - B((n-1)k,\theta)] \\ &\quad + 2(n-1)[B(nk,\theta) - B((n-1)k,\theta) - B(nk,\theta) \\ &\quad + 2B((n-1)k,\theta) - B((n-2)k,\theta)] + \frac{n^2}{2}\Delta^2 B(nk,\theta) \\ &= \theta + B(nk,\theta) + (2n-3)\Delta B(nk,\theta) + \frac{(n-2)^2}{2}\Delta^2 B(nk,\theta). \end{aligned}$$

The proof of the theorem is complete.

Throughout the remainder of this paper we will assume $k = 1$ for convenience. There is no loss of generality in this assumption and the results which follow hold equally well for all $1 \leqq k \leqq N$ for which $J^{(p)}[\hat{\theta}]$ is defined.

Theorem 1 illustrates that the biases of the first order jackknife and second order jackknife can be expressed in terms of the bias of $\hat{\theta}$ and the first and second order differences of the bias of $\hat{\theta}$.

Before proceeding, two definitions and a preliminary theorem are in order.

DEFINITION 1. Let $\hat{\theta}_1(n)$ and $\hat{\theta}_2(n)$ be estimators of θ where $B_1(n, \theta) = E[\hat{\theta}_1(n)-\theta] \neq 0$ and $B_2(n, \theta) = E[\hat{\theta}_2(n)-\theta] \neq 0$ and suppose

$$\left|\lim_{n\to\infty}\frac{B_1(n,\theta)}{B_2(n,\theta)}\right| = L. \tag{5}$$

If $L = 1$ then we shall say $\hat{\theta}_2(n)$ and $\hat{\theta}_1(n)$ are same order bias estimators of θ, denoted $\hat{\theta}_2(n)$ S.O.B.E. $\hat{\theta}_1(n)$. If $0 < L < 1$ then we shall say $\hat{\theta}_1(n)$ is a better same order bias estimator than $\hat{\theta}_2(n)$, denoted $\hat{\theta}_1(n)$ B.S.O.B.E. $\hat{\theta}_2(n)$.

DEFINITION 2. Let $B_1(n, \theta)$ and $B_2(n, \theta)$ be as defined in Definition 1 excluding the condition that $B_1(n, \theta) \neq 0$, then if

$$\lim_{n\to\infty}\frac{B_1(n,\theta)}{B_2(n,\theta)} = 0$$

we shall say $\hat{\theta}_1(n)$ is a lower order bias estimator of θ than $\hat{\theta}_2(n)$, denoted $\hat{\theta}_1(n)$ L.O.B.E. $\hat{\theta}_2(n)$.

The following well-known theorem, Bromwich (1926), will be useful in the results which follow.

THEOREM 2. *If $a_n \to 0$, b_n is monotone and converges to zero, and if*

$$\lim_{n\to\infty}\frac{a_n-a_{n-1}}{b_n-b_{n-1}}$$

exists, finite or infinite, then

$$\lim_{n\to\infty}\frac{a_n}{b_n}$$

exists and

$$\lim_{n\to\infty}\frac{a_n}{b_n} = \lim_{n\to\infty}\frac{\Delta a_n}{\Delta b_n}$$

where $\Delta a_n = a_n - a_{n-1}$ and $\Delta b_n = b_n - b_{n-1}$.

We are now able, at least asymptotically, to determine whether or not $J^{(1)}[\hat{\theta}]$ and $J^{(2)}[\hat{\theta}]$ are effective bias removal tools.

THEOREM 3. *Let $B(n, \theta) = E[\hat{\theta}-\theta]$. If there exists $k > 0$ such that*

$$\lim_{n\to\infty} n^k B(n,\theta) = C(\theta) \neq 0 \quad or \quad \pm\infty,$$

then for $p = 1, 2$

(i) *if $k = 1$ or $k = p$,* $J^{(p)}[\hat{\theta}]$ L.O.B.E. $\hat{\theta}$

(ii) *if $k < p+1$ and $k \neq 1$ or p,* $J^{(p)}[\hat{\theta}]$ B.S.O.B.E. $\hat{\theta}$

(iii) *if $k = p+1$,* $J^{(p)}[\hat{\theta}]$ S.O.B.E. $\hat{\theta}$

(iv) *if $k > p+1$,* $\hat{\theta}$ B.S.O.B.E. $J^{(p)}[\hat{\theta}]$.

provided $\lim_{n\to\infty} n^{k+p}\Delta^p B(n, \theta)$ *exists.*

PROOF. Since $C(\theta)$ is finite it follows that $\lim_{n\to\infty} B(n, \theta) = 0$. Also n^{-k} converges to zero monotonically; thus it follows by Theorem 2 that

$$C(\theta) = \lim_{n\to\infty}\frac{B(n,\theta)}{n^{-k}} = \lim_{n\to\infty}\frac{\Delta B(n,\theta)}{\Delta(n^{-k})}.$$

By the same reasoning, $\lim_{n\to\infty} \Delta B(n, \theta) = 0$ and $\Delta(n^{-k})$ converges monotonically to zero implies

$$C(\theta) = \lim_{n\to\infty}\frac{\Delta B(n,\theta)}{\Delta(n^{-k})} = \lim_{n\to\infty}\frac{\Delta^2 B(n,\theta)}{\Delta^2(n^{-k})}.$$

Hence,

$$\lim_{n\to\infty}\frac{\Delta^p B(n,\theta)}{\Delta^p(n^{-k})} = C(\theta), \qquad p = 1, 2. \tag{6}$$

Now from Theorem 1 we have, for $p = 1, 2$,

$$E[J_n^{(p)}[\hat{\theta}]-\theta] = B(n,\theta)+(p-1)(pn-p-1)\Delta B(n,\theta)+\frac{(n-p)^p}{p}\Delta^p B(n,\theta).$$

Therefore,

$$\left|\lim_{n\to\infty}\frac{B(n,\theta)+(p-1)(pn-p-1)\Delta B(n,\theta)+\frac{(n-p)^p}{p}\Delta^p B(n,\theta)}{B(n,\theta)}\right|$$

$$= \left|1+\lim_{n\to\infty}\frac{(p-1)(pn-p-1)\Delta B(n,\theta)}{B(n,\theta)}+\lim_{n\to\infty}\frac{(n-p)^p\Delta^p B(n,\theta)}{pB(n,\theta)}\right|$$

$$= \left|1+\lim_{n\to\infty}\frac{(p-1)(pn-p-1)(n-1)^k\Delta(n^{-k})\left[\frac{\Delta B(n,\theta)}{\Delta(n^{-k})}\right]}{(n-1)^k B(n,\theta)}\right. \tag{7}$$

$$\left.+\lim_{n\to\infty}\frac{(n-p)^{p+k}\Delta^p(n^{-k})\left[\frac{\Delta^p B(n,\theta)}{\Delta^p(n^{-k})}\right]}{p(n-p)^k B(n,\theta)}\right|$$

$$= |k-1| \qquad \text{if } p = 1;$$

$$= \frac{|(k-1)(k-2)|}{2} \qquad \text{if } p = 2.$$

This follows from (6) and the relations

$$\lim_{n\to\infty}(n-1)^{k+1}\Delta(n^{-k}) = -k$$

and

$$\lim_{n\to\infty}(n-2)^{k+2}\Delta^2(n^{-k}) = k(k+1).$$

The results (i), (ii), (iii), (iv) then follow from (7).

Before proceeding to Theorem 4, which is of the same nature as Theorem 3. let us give a brief discussion of the application of such theorems. As is well known, if $B(n, \theta) = C(\theta)/n$ the jackknife will eliminate the bias exactly and in many other cases it will reduce bias. Theorem 3 essentially establishes what the general nature of the bias should be in order for the jackknife to be beneficial in this respect. This is of course clear since if $J^{(p)}[\hat{\theta}]$ L.O.B.E. $\hat{\theta}$, then for any $\varepsilon > 0$ and n sufficiently large

$$|E[J^{(p)}(\hat{\theta})-\theta]| < \varepsilon|E[\hat{\theta}-\theta]| \tag{8}$$

so that one may expect a rather significant decrease in bias. Moreover if $J^{(p)}(\hat{\theta})$ B.S.O.B.E. $\hat{\theta}$ then, for large n, (8) will always hold for $\varepsilon = 1$. Consequently one can still expect $J^{(p)}(\hat{\theta})$ to perform satisfactorily. On the other hand when $J^{(p)}(\hat{\theta})$ S.O.B.E. $\hat{\theta}$ it is quite possible, even for large n, that $J^{(p)}(\hat{\theta})$ will increase the bias although probably not significantly. However when $\hat{\theta}$ L.O.B.E. $J^{(p)}(\hat{\theta})$ one can expect $J^{(p)}(\hat{\theta})$ to significantly increase the bias, at least from a proportionate viewpoint. Particular instances covered by Theorem 3 were noted by Miller (1964), where he observed that $J^{(1)}[\hat{\theta}]$ will reduce bias of order $n^{-\frac{1}{2}}$ but not bias of order n^{-2}. This is of course the cases $k = \frac{1}{2}$ and $k = 2$ with $p = 1$ in Theorem 3 and hence Miller's observations follow from that theorem for large n. Of course in particular instances n may not have to be very large. In the cases in point for instance, the implications of the theorem are correct for $n \geqq 2$.

In order to compare $J^{(1)}[\hat{\theta}]$ with $J^{(2)}[\hat{\theta}]$, we have the following theorem.

THEOREM 4. *Let the conditions of Theorem* 3 *be satisfied with the additional assumptions that* $k \neq 1$. *Then*

(i) *if* $k = 2$, $J^{(2)}[\hat{\theta}]$ L.O.B.E. $J^{(1)}[\hat{\theta}]$

(ii) *if* $k < 4$ *and* $k \neq 2$, $J^{(2)}[\hat{\theta}]$ B.S.O.B.E. $J^{(1)}[\hat{\theta}]$

(iii) *if* $k = 4$, $J^{(2)}[\hat{\theta}]$ S.O.B.E. $J^{(1)}[\hat{\theta}]$

(iv) *if* $k > 4$, $J^{(1)}[\hat{\theta}]$ B.S.O.B.E. $J^{(2)}[\hat{\theta}]$.

PROOF.

$$\left|\lim_{n\to\infty}\frac{B(n,\theta)+(2n-3)\Delta B(n,\theta)+\frac{(n-2)^2}{2}\Delta^2 B(n,\theta)}{B(n,\theta)+(n-1)\Delta B(n,\theta)}\right|$$

$$=\left|\lim_{n\to\infty}\frac{\dfrac{B(n,\theta)+(2n-3)\Delta B(n,\theta)+\frac{(n-2)^2}{2}\Delta^2 B(n,\theta)}{B(n,\theta)}}{\dfrac{B(n,\theta)+(n-1)\Delta B(n,\theta)}{B(n,\theta)}}\right|$$

$$=\left|\frac{\frac{1}{2}(k-1)(k-2)}{1-k}\right| \quad \text{, from the proof of Theorem 3,} \tag{9}$$

$$=\tfrac{1}{2}|k-2|.$$

The results follow from (9).

It is clear from the preceding results that the jackknife procedure can be used with confidence as a bias removal statistic when the order of the bias of the original estimator is approximately known.

In many instances the bias of $\hat{\theta}$ has the form

$$B(\theta, n) = \frac{a_1}{n} + \frac{a_2}{n^2} + \frac{a_3}{n^3} + \cdots. \tag{10}$$

When this is the case $k = 1$ and Theorem 4 gives us no comparison between $J^{(1)}(\hat{\theta})$ and $J^{(2)}(\hat{\theta})$. This problem is eliminated by the next two theorems, the first of which is the case described by (10) when $a_i = 0$, $i = 2, 3, \cdots$, and the second, the case where $a_i \neq 0$ for some $i \geqq 2$. We shall prove the first and only state the second since its proof is along the same line as our previous results.

THEOREM 5. *If $J^{(1)}[\hat{\theta}]$ is an unbiased estimator of θ then $J^{(2)}[\hat{\theta}]$ is an unbiased estimator of θ.*

PROOF. Since $J^{(1)}[\hat{\theta}]$ is unbiased, then by Theorem 1

$$B(n,\theta) + (n-1)\Delta B(n,\theta) = 0. \tag{11}$$

Therefore

$$\Delta[B(n,\theta) + (n-1)\Delta B(n,\theta)] = \Delta(0) = 0.$$

This yields

$$\Delta B(n,\theta) + (n-1)\Delta B(n,\theta) - (n-2)\Delta B(n-1,\theta) = 0$$

$$\Rightarrow 2\Delta B(n,\theta) + (n-2)[\Delta B(n,\theta) - \Delta B(n-1,\theta)] = 0$$

$$\Rightarrow \Delta B(n,\theta) + \frac{(n-2)}{2}\Delta^2 B(n,\theta) = 0 \tag{12}$$

$$\Rightarrow (n-2)\Delta B(n,\theta) + \frac{(n-2)^2}{2}\Delta^2 B(n,\theta) = 0.$$

From Theorem 1 we have

$$\begin{aligned} E[J^{(2)}[\hat{\theta}] - \theta] &= B(n,\theta) + (n-3)\Delta B(n,\theta) + \frac{(n-2)^2}{2}\Delta^2 B(n,\theta) \\ &= [B(n,\theta) + (n-1)\Delta B(n,\theta)] + [(n-2)\Delta B(n,\theta) \\ &\qquad + \frac{(n-2)^2}{2}\Delta^2 B(n,\theta)] \\ &= 0 \end{aligned}$$

by (11) and (12). Hence $J^{(2)}[\hat{\theta}]$ is unbiased.

THEOREM 6. *Let $B(n, \theta)$ and k be defined as in Theorem 3 and assume $k = 1$. If there exists a $k_1 > 0$ such that*

$$\lim_{n\to\infty} n^{k_1}[nB(n,\theta) - C(\theta)] = C_1(\theta) \neq 0 \quad \textit{or} \quad \pm\infty,$$

and $\qquad \lim_{n\to\infty} n^{k_1+2}\Delta^2[nB(n,\theta)]$ *exists,* *then*

(i) *if* $k_1 = 1, J^{(2)}[\hat{\theta}]$ L.O.B.E. $J[\hat{\theta}]$

(ii) *if* $0 < k_1 < 3$ *and* $k_1 \neq 1, J^{(2)}[\hat{\theta}]$ B.S.O.B.E. $J^{(1)}[\hat{\theta}]$

(iii) *if* $k_1 = 3, J^{(2)}[\hat{\theta}]$ S.O.B.E. $J^{(1)}[\hat{\theta}]$, *and*

(iv) *if* $k_1 > 3, J^{(1)}[\hat{\theta}]$ B.S.O.B.E. $J^{(2)}[\hat{\theta}]$.

Concluding remarks. A complete asymptotic characterization of the bias reduction properties of the first and second order jackknife has been established. These results should be useful in aiding the statistician in determining when the jackknife might be useful. This of course is not altogether true since one of the primary uses of the jackknife is in the area of approximate confidence intervals rather than bias reduction. However in this connection the bias reduction is also usually desirable for better positioning of the confidence interval. Moreover there are a number of problems where the bias reduction aspect of the jackknife is the property of interest and where this property also leads to a reduction in mean square error. See Durbin (1959) or Rao and Webster (1966) for instance.

Finally it should be mentioned that the results of this paper can be interpreted as robustness results regarding the bias reduction property of the jackknife. That is, originally the jackknife was a tool developed for reducing bias of the form of (10). The theorems of this paper establish how far the bias can be from the original assumed form and the jackknife remain effective for this purpose. From this standpoint $J^{(2)}(\hat{\theta})$ appears to be more robust than $J^{(1)}(\hat{\theta})$. Thus, when reduction of bias is the purpose of the user, $J^{(2)}(\hat{\theta})$ is superior. On the other hand it may increase the variance to a point of making it unsatisfactory. However, in what little information exists regarding such an increase in variance no appreciable increase has been noted. Unfortunately no general results have been established in regard to these problems and a detailed study of the mean square errors in $J^{(1)}(\hat{\theta})$ and $J^{(2)}(\hat{\theta})$ would be interesting.

REFERENCES

BROMWICH, T. J. (1926). *An Introduction to the Theory of Infinite Series* 2nd ed. Macmillan, New York.

DURBIN, J. (1959). A note on the application of Quenouille's method of bias reduction to the estimation of ratios. *Biometrika* **46** 477–480.

MILLER, R. G., Jr. (1964). A trustworthy jackknife. *Ann. Math. Statist.* **35** 1594–1605.

QUENOUILLE, M. H. (1949). Notes on bias in estimation. *Biometrika* **43** 353–360.

RAO, J. N. K. and WEBSTER, J. T. (1966). On two methods of bias reduction in the estimation of ratios. *Biometrika* **53** 571–577.

SCHUCANY, W. R., GRAY, H. L. and OWEN, D. B. (1971). On bias reduction in estimation. To appear in *J. Amer. Statist. Assoc.*