PROBABILITY AND STATISTICS
Essays in Honor of Franklin A. Graybill
J.N. Srivastava (Editor)

ON A UNIFICATION OF BIAS REDUCTION AND NUMERICAL APPROXIMATION

H. L. Gray

Department of Statistical Science
Southern Methodist University
Dallas, Texas
U.S.A.

In this paper a slight, but important, extension of the generalized jackknife is given. By showing that the error in a numerical approximation is simply the degenerate case of bias in an estimator, it is demonstrated that a large body of theory of numerical analysis can be encompassed by the jackknife theory. Numerous examples are given which include parameter estimation, spectral approximation, spectral estimation and the approximation of tail probabilities.

INTRODUCTION

In [2] the relationship between the so called e_n-transformation or ε-algorithm and the generalized jackknife statistic was discussed and many new results concerning the latter were obtained. In this paper we expand that discussion somewhat and establish a much more general relationship between the jackknife and general numerical methods. That is, by considering numerical approximations as degenerate estimators it is pointed out that the error in a numerical approximation is a special case of the concept of bias in an estimator. In this way it is seen that a large body of the theory of numerical analysis can be considered as just a special case of the generalized jackknife, i.e. the degenerate case. From this point of view, it follows that the generalized jackknife can be considered as the natural extension of one of the more fundamental ideas in the theory of numerical approximation.

Although for the most part this paper is expository and the tools employed are not new, they are utilized in such a way as to suggest a more general applicability of the bias reduction technique employed in the generalized jackknife. In order to demonstrate the validity of this last remark, we first define the generalized jackknife, and then a simple example is given which exemplifies it as a bias reduction method. Following this example, it is shown that such well known results as Simpson's rule, Romberg quadrature, Weddle's rule, Newton's rule, Newton-Cotes, Lagrange interpolation, and the e_n-transform are all simple applications of the generalized jackknife and are in fact even more simple applications of the generalized jackknife than the first, admittedly trivial, example. Finally two somewhat more complicated examples are given, approximating and estimating the spectral density and approximating tail probabilities.

The following definition is a rather simple, but significant, extension of the one given in [6] for the generalized jackknife.

Definition 1.
Let $\hat{\theta}_1(n), \hat{\theta}_2(n), \ldots, \hat{\theta}_{k+1}(n)$ be k+1 estimators (possibly degenerate) each of which depend on n and let a_{ij} and C_j, $j=1,2,\ldots,k+1$, be real or complex numbers. Then the generalized jackknife $G(\cdot;a_{ij},C_j)$ is defined by

$$G(\hat{\theta}_1,\hat{\theta}_2,\ldots,\hat{\theta}_{k+1};a_{ij},C_j) = \frac{H_{k+1}(\hat{\theta}_j;a_{ij})}{H_{k+1}(C_j;a_{ij})} \quad , \qquad (1)$$

$$\text{where}\quad H_{k+1}(Z_j;a_{ij}) = \begin{vmatrix} Z_1 & Z_2 & \cdots & Z_{k+1} \\ a_{11} & a_{12} & \cdots & a_{1,k+1} \\ \vdots & & & \\ a_{k,1} & & \cdots & a_{k,k+1} \end{vmatrix}$$

and $H_{k+1}(C_j;a_{ij}) \neq 0$.

Now let us define $B_j(n,\theta)$ by

$$E[\hat{\theta}_j(n)] = C_j\theta + B_j(n,\theta) . \tag{2}$$

We will only have need for two sets of C_j's in (1),namely $\{C_j = 1\}$ and the set $\{C_1 = 1,\ C_j = 0,\ j \geq 2\}$. Hence we will restrict ourselves hereafter to those sets of C_j. In either event, $B_1(n,\theta)$ is the bias in the estimator $\theta_1(n)$. One should note that if θ_1 is degenerate that

$$E[\hat{\theta}_1(n)] = \hat{\theta}_1(n) = C_1\theta + B_1(n,\theta) = \theta + B_1(n,\theta) \tag{3}$$

and hence

$$\hat{\theta}_1 - \theta = B_1(n,\theta) . \tag{4}$$

In this case (the degenerate one) $B_1(n,\theta)$ is usually referred to as the error in $\hat{\theta}_1$. Thus the problem of reducing the error in an approximation can be looked at as a special case of the problem of reducing the bias in an estimator. We will therefore generally refer to the quantity in (4) as bias but it should be understood that when the θ_i are degenerate the word "error" is more conventional.

Although the introduction of C_j in Definition 1 adds some utility, it does not effect the bias reduction property of the generalized jackknife. That is, the following theorem still obtains.

Theorem 1. Let $\hat{\theta}_1,\ \hat{\theta}_2,\ldots,\hat{\theta}_{k+1}$ be k+1 estimators, such that

$$E[\hat{\theta}_j - C_j\theta] = \sum_{i=1}^{\infty} a_{ij}(n)b_i(\theta) . \tag{5}$$

Then for every set of C_j and $a_{ij} = a_{ij}(n)$ such that (1) is defined,

we have

$$E[G(\hat{\theta}_1,\hat{\theta}_2,\ldots,\hat{\theta}_{k+1};a_{ij},C_j)] = \theta + \frac{H_{k+1}(\varepsilon_j;a_{ij}(n))}{H_{k+1}(C_j;a_{ij}(n))} \quad , \tag{6}$$

where

$$\varepsilon_j = \sum_{i=k+1}^{\infty} a_{ij}(n)b_i(\theta) .$$

The proof of the above theorem is the same as its counterpart in (4). It holds for all C_j but, as stated previously, we will restrict ourselves to the sets of C_j already mentioned.

Corollary. If $a_{ij}(n) \equiv 0$ when $i = k+1,\ldots$, then $G(\hat\theta_1,\hat\theta_2,\ldots,\hat\theta_{k+1};a_{ij},C_j)$ is an unbiased estimator for θ, i.e.,

$$E\left[\, G(\hat\theta_1,\hat\theta_2,\ldots,\hat\theta_{k+1};a_{ij},C_j)\right] = \theta \;. \qquad (7)$$

In order to actually make use of Theorem 1 it is necessary to obtain k+1 estimators and their bias expansions. There is fortunately a standard way to obtain these estimators in the usual statistical setting. In the approximation area, their method of obtainment is on the surface more varied, but at the proper level of understanding, no different. In the next section, we review the standard approach for selecting the $\hat\theta_i$ when they are not degenerate. We then demonstrate how this bias reduction technique extends to the degenerate case.

ANALYSIS

In general the problem of selecting k+1 estimators in (1) is more a problem of selecting one estimator, and "perturbing" it properly to obtain the other k estimators, than selecting k+1 distinct estimators. In the case of the jackknife, for example, one usually selects the $\hat\theta_j$ as follows. Let $\hat\theta$ be a given estimator such that (5) holds when $j = 1$ and let $C_j \equiv 1$. Then define

$$\hat\theta_1 = \hat\theta(X_1,X_2,\ldots,X_n)$$

$$\hat\theta_2 = \overline{\hat\theta^{i}} = \overline{\hat\theta}(X_1,X_2,\ldots,X_{i-1},X_{i+1},\ldots,X_n) \qquad (8)$$

$$\hat\theta_3 = \overline{\hat\theta^{i_1,i_2}} = \overline{\hat\theta}\,(X_1,\ldots,X_{i_1-1},X_{i_1+1},\ldots,X_{i_2-1},X_{i_2+1},\ldots,X_n),$$

etc., where the bar denotes the average over the indicated possible subsamples. Clearly if (5) holds for $j = 1$, then it holds for $j = 2,3,\ldots,k+1$ and because of (8), in this case we can shorten the notation to

$$G^{(k)}(\hat\theta;a_i(n-j+1)) = \frac{H_{k+1}(\hat\theta;a_i(n-j+1))}{H_{k+1}(1;a_i(n-j+1))} \qquad (9)$$

$i = 1,2,\ldots,k$, $j = 1,2,\ldots,k+1$ or simply $G^{(k)}(\hat\theta)$, which is the common notation for the generalized jackknife.

The procedure is the same in the degenerate case. That is, select an approximation $\hat\theta(h)$ (here we have denoted n by h since in the degenerate case the quantity perturbed is not the sample size) and obtain the other k approximations by varying h. The following simple examples should clarify the preceeding notions. One should keep in mind that in every example given, the "estimator" obtained is a form of the generalized jackknife.

Example 1.

Let $\bar X^3$ be an estimator for μ^3 based on the random sample $X_1,X_2,\ldots,X_n$ from a distribution with mean μ and variance σ^2. Then

$$E[\bar X^3] = \mu^3 + \frac{3\mu\sigma^2}{n} + \frac{1}{n^2}\,E[(X-\mu)^3] \;.$$

Therefore from the corollary we use k = 2 and find

$$G^{(2)}(\bar{X}^3) = \frac{H_3(\bar{X}^3;a_{ij})}{H_3(1;a_{ij})} = \frac{\begin{vmatrix} \bar{X}^3 & \overline{(\bar{X}^i)^3} & \overline{(\bar{X}^{i,j})^3} \\ \frac{1}{n} & \frac{1}{n-1} & \frac{1}{n-2} \\ \frac{1}{n^2} & \frac{1}{(n-1)^2} & \frac{1}{(n-2)^2} \end{vmatrix}}{\begin{vmatrix} 1 & 1 & 1 \\ \frac{1}{n} & \frac{1}{n-1} & \frac{1}{n-2} \\ \frac{1}{n^2} & \frac{1}{(n-1)^2} & \frac{1}{(n-2)^2} \end{vmatrix}}$$

$$= \bar{X}^3 - \frac{3\bar{X}}{n(n-1)} \sum_{i=1}^{n} (X_i-\bar{X})^2 + \frac{1}{n(n-1)(n-2)} \sum_{i=1}^{n} (X_i-\bar{X})^3 \quad ,$$

and

$$E[G^{(2)}(\bar{X}^3) = \mu^3.$$

Example 2. (Simpson's Rule)

As possibly a more fundamental example we consider the problem of approximating

$$\theta = \int_a^b f(x)dx \tag{10}$$

by the trapezoidal rule, $T(h)$, i.e.,

$$\hat{\theta}(h) = T(h) = \frac{h}{2}\{f(a)+2f(a+h) + \dots + 2f(a+(m-1)h) + f(a+mh)\}, \tag{11}$$

where $a + mh = b$. Now if f is analytic over $[a,b]$ the bias in $T(h)$ can be shown to be given by

$$T(h) - \theta = b_1h^2 + b_2h^4 + \dots + b_mh^{2m} + R_m \ , \tag{12}$$

where the b_i do not depend on h and $R_m = O(h^{2m+2})$. Equation (12) is referred to as the Euler-Maclaurin summation formula. In light of Equation (12) and Theorem 1 there are a number of ways we could select a second approximation to reduce the bias in $T(h)$. One of the more natural choices is $T(h/2)$. From (12)

$$T(\frac{h}{2}) = \theta + \frac{b_1h^2}{4} + \frac{b_2h^4}{16} + \dots + \frac{b_mh^{2m}}{4^m} + R_m \ , \tag{13}$$

and the first order generalized jackknife is then given by

$$G(T(h), T(\tfrac{h}{2}); 2^{-2i(j-1)}, 1) = \frac{\begin{vmatrix} T(h) & T(\frac{h}{2}) \\ 1 & \frac{1}{4} \end{vmatrix}}{\begin{vmatrix} 1 & 1 \\ 1 & \frac{1}{4} \end{vmatrix}}$$

$$= \frac{4}{3}[T(\tfrac{h}{2}) - \tfrac{1}{4}T(h)]$$

$$= \frac{h}{6}[f(a) + 4f(a+\tfrac{h}{2}) + 2f(a+h) + 4f(a + \tfrac{3h}{2}) + \ldots + f(a+mh)]. \quad (14)$$

Equation 14 is of course better known as Simpson's rule, a result familiar to all students of elementary calculus. Numerous additional quadrature formulas could be obtained in precisely the same way. For example Weddle's rule, Newton-Cotes method, Newton's rule, etc. could all be obtained by simply jackknifing T(h) using different partitions of the interval to produce the required additional approximations. The next example demonstrates this notion more fully.

Example 3 (Romberg Quadrature)

Consider further the problem of reducing the bias in the trapezoidal approximation of the integral in (10). Example 2 established that Simpson's rule, as well as several common quadrature methods, are in fact first order jackknives of the trapezoid rule. The observation can be extended to the higher order jackknife. The k+1 approximations required could be obtained by again using (12) and noting that

$$T(\frac{h}{2^j}) = \theta + \frac{b_1 h^2}{2^{2j}} + \frac{b_2 h^4}{2^{4j}} + \ldots + \frac{b_m h^{2m}}{2^{2mj}} + R_m , \quad (15)$$

$$j = 0,1,2,\ldots,k.$$

Then the k-th order jackknife is given by

$$G(\hat{\theta}_1,\ldots,\hat{\theta}_{k+1}; a_{ij}, 1) = \frac{H_{k+1}(T(\frac{h}{2^{j-1}}); 2^{-2i(j-1)})}{H_{k+1}(1; 2^{-2i(j-1)})}$$

$$= \frac{\begin{vmatrix} T(h) & T(\frac{h}{2}) & \cdots & T(\frac{h}{2^k}) \\ 1 & 2^{-2} & \cdots & 2^{-2k} \\ \vdots & & & \\ 1 & 2^{-2k} & \cdots & 2^{-2k(k+1)} \end{vmatrix}}{\begin{vmatrix} 1 & 1 & \cdots & 1 \\ 1 & 2^{-2} & \cdots & 2^{-2k} \\ \vdots & & & \\ & {}^{-2k} & & {}^{-2k(k+1)} \end{vmatrix}} . \quad (16)$$

The approximation in (16) can be written in a recursive form and evaluated very efficiently. Thus the step size, h, can be progressively reduced until $G(\hat{\theta}_1,\ldots,\hat{\theta}_{k+1};a_{ij},1)$ gives the desired accuracy for θ. For these reasons it is a very popular method of numerical quadrature and is more commonly referred to as Romberg quadrature.

The first few examples in numerical analysis we have given are numerical approximation of integrals. This was because of the Euler-Maclaurin formula and it was not meant to suggest that the application of the jackknife principle is limited to that arena. The key of course is the bias expansion. The remaining examples are all somewhat distinct.

Example 4 (Lagrange Interpolation)

Consider the problem of approximating f(x) given several values of f in the neighborhood of x, i.e., $f(x_1),f(x_2),\ldots,f(x_{k+1})$. Now suppose f has a Taylor expansion about x, valid in an interval containing the x_i. Then

$$f(x_1) = f(x) + (x_1-x)f'(x) + \ldots + \frac{(x_1-x)^k}{k!} f^{(k)}(x) + R_k(x,x_1)$$

$$f(x_2) = f(x) + (x_2-x)f'(x) + \ldots + \frac{(x_2-x)^k}{k!} f^{(k)}(x) + R_k(x,x_2)$$

$$\vdots$$

$$f(x_{k+1}) = f(x) + (x_{k+1}-x)f'(x) + \ldots + \frac{(x_{k+1}-x)^k}{k!} f^{(k)}(x) + R(x,x_{k+1}).$$

Thus we have for k-th order jackknife,

$$G[f(x_1),\ldots,f(x_{k+1});(x_j-x),1] = \frac{\begin{vmatrix} f(x_1) & f(x_2) & \cdots & f(x_{k+1}) \\ (x_1-x) & (x_2-x) & \cdots & (x_{k+1}-x) \\ (x_1-x)^k & (x_2-x)^k & \cdots & (x_{k+1}-x)^k \end{vmatrix}}{\begin{vmatrix} 1 & 1 & \cdots & 1 \\ (x_1-x) & (x_2-x) & \cdots & (x_{k+1}-x) \\ \vdots & & & \\ (x_1-x)^k & (x_2-x)^k & \cdots & (x_{k+1}-x)^k \end{vmatrix}}$$

which, when expanded, is the well known Lagrange formula for interpolation.

Example 5 (e_p-Transformation or ε-algorithm)

This example was discussed in some detail in [2] regarding the relationship between the jackknife and the e_p-transform. We repeat the example here with some modification since the picture is now more clear. In the following we see that the relationship between the e_p-transform and the jackknife is a fundamental one, but no more so than the relation between jackknifing and Lagrange interpolation or Romberg integration or even Simpson's rule.

Consider the problem of approximating

$$\theta = \sum_{k=C}^{\infty} a_k$$

by the natural approximation

$$\theta(m) = \sum_{k=C}^{m} a_k \; . \tag{17}$$

Now suppose that the error in (17) eventually satisfies a p-th order homogeneous linear difference equation with constant, but unknown, coefficients, i.e., for $m>M$

$$\Delta^p(\theta(m)-\theta) + \alpha_1 \Delta^{p-1}(\theta(m)-\theta) + \dots + \alpha_p(\theta(m)-\theta) = 0 \; . \tag{18}$$

Rewriting (18) we have (using a backward difference),

$$\theta(m) = \theta + b_1 a_m + b_2 a_{m-1} + \dots + b_p a_{m-p+1} \; . \tag{19}$$

Of course, since the b_i are unknown, θ cannot be calculated directly from (19).

However, the generalized jackknife of $\theta(m)$ can be computed by noting that

$$\theta(m-j) = \theta + b_1 a_{m-j} + b_2 a_{m-j-1} + \dots + b_p a_{m-j-p+1} \; . \tag{20}$$

$$j = 0,1,\dots,p \; .$$

Then from (1) and (20), adopting the notation of (9), we have

$$G^{(p)}(\theta(m);a_{m-i-j+2},1) = \frac{\begin{vmatrix} \theta(m) & \theta(m-1) & \dots & \theta(m-p) \\ a_m & a_{m-1} & \dots & a_{m-p} \\ \vdots & & & \\ a_{m-p+1} & & \dots & a_{m-2p+1} \end{vmatrix}}{\begin{vmatrix} 1 & 1 & \dots & 1 \\ a_m & a_{m-1} & \dots & a_{m-p} \\ \vdots & & & \\ a_{m-p+1} & a_{m-p} & \dots & a_{m-2p+1} \end{vmatrix}} \; , \tag{21}$$

and for $m-p+1 > M$ it follows from (7) and Theorem 1 that

$$G^{(p)}(\theta(m);a_{m-i-j+2},1) = \theta = \sum_{k=C}^{\infty} a_k \; . \tag{22}$$

In the form (21), the generalized jackknife is known as the e_p-transformation. The transformation has been used extensively in numerical analysis for increasing the rate of convergence of a sequence. It can be evaluated very efficiently by the so called ε-algorithm. In the numerical example which follows we make use of the assumption of Equation (19). However, it should be stressed that the major application of the e_p-transformation is to sequences for which (19) is only "approximately" true, see [5].

Let

$$\theta(m) = \sum_{k=1}^{m} \frac{k \cos\pi(k-1)}{2^{k-1}}$$

If one notes that for all m, $\frac{m \cos\pi(m-1)}{2^{m-1}}$ is a solution of a second order difference equation with constant coefficients then it follows that $\theta(m)-\theta$ satisfies (19) for $p = 2$ and $m > 3$. Thus we obtain

$$\sum_{k=1}^{\infty} \frac{k\cos\pi(k-1)}{2^{k-1}} = G^{(2)}(\theta(4);a_{kj},1) = \frac{\begin{vmatrix} \frac{1}{4} & \frac{3}{4} & 0 \\ -\frac{1}{2} & \frac{3}{4} & -1 \\ \frac{3}{4} & -1 & 1 \end{vmatrix}}{\begin{vmatrix} 1 & 1 & 1 \\ -\frac{1}{2} & \frac{3}{4} & -1 \\ \frac{3}{4} & -1 & 1 \end{vmatrix}} = \frac{4}{9} .$$

The sum of the infinite series has therefore been obtained by jackknifing $\theta(4)$, where

$$\theta(4) = \sum_{k=1}^{4} \frac{k\cos\pi(k-1)}{2^{k-1}} .$$

In this example the a_m were real. Although not widely known, if the a_m are complex, the e_p-transform, as defined by (19), is still valid and enjoys many of the same properties,see [4]. This observation will be used in the next example.

Example 6 (Spectral Density of an ARMA process)

Let X_t be ARMA(p,q) with autocorrelation ρ, i.e.,

$$X_t - \phi_1 X_{t-1} - \dots - \phi_p X_{t-p} = a_t - \psi_1 a_{t-1} - \dots - \psi_q a_{t-q} ,$$

where a_t is zero mean white noise. Then for $m > q$

$$\rho(m) - \phi_1 \rho(m-1) - \dots - \phi_p \rho(m-p) = 0 . \tag{23}$$

Moreover for $|f| \le .5$ the spectral density of X_t, $S(f)$, is by definition

$$S(f) = \sum_{-\infty}^{\infty} e^{-2\pi ifk} \rho(k)$$

$$= 1 + 2 \,\mathrm{Re} \sum_{k=1}^{\infty} e^{-2\pi i f k} \rho(k) \,, \tag{24}$$

since $\rho(k) = \rho(-k)$. Now let

$$S_m(f;j) = \sum_{k=-j}^{m} e^{-2\pi i f k} \rho(k) \tag{25}$$

and

$$S(f;j) = \sum_{k=-j}^{\infty} e^{-2\pi i f k} \rho(k) \,. \tag{26}$$

From (24) for $j > 0$

$$S(f) = 1 + 2 \,\mathrm{Re}(S(f;j) - S_0(f;j)) \,. \tag{27}$$

A natural approximation for $S(f)$ is then given by the partial sum

$$S_m(f) = 1 + 2 \,\mathrm{Re}(S_m(f;j) - S_0(f;j)) \,. \tag{28}$$

But from (23) it is easy to show that for $m > q$

$$S_m(f;j) = S(f;j) + b_1\alpha_m(f) + b_2\alpha_{m-1}(f) + \ldots + b_p\alpha_{m-p+1}(f), \tag{29}$$

where

$$\alpha_m(f) = e^{-2\pi i f m} \rho(m) \,.$$

But (29) is exactly the same form as (19), and as in (21), for $m > q+p-1$, we obtain

$$G^{(p)}(\theta_m;\alpha_{m-i-j+1},1) = S(f;j) \,, \tag{30}$$

where

$$\theta_m = S_m(f;j)$$

and j is chosen positive and sufficiently large that $G^{(p)}$ is defined, i.e., for $m = q + p$, $j = p - q$. Then

$$S(f) = 1 + 2 \,\mathrm{Re}(G^{(p)}(\theta_m;\alpha_{m-i-j+2},1) - S_0(f;j)) \,. \tag{31}$$

Thus the spectral density of an ARMA(p,q) process has been obtained by jack-knifing the finite sum in (28). It should be noted that we could have taken $j=1$ in (27) and avoided what might seem as needless confusion introduced by the notation of (27). The astute reader may have noticed however that introducing j in (27) and using the fact that $\rho(m) = \rho(-m)$ allows us to compute the jackknife from fewer values of $\rho(m)$. This is important when $\rho(m)$ must be estimated. In that event, it can be shown that if ρ is estimated by the sample autocorrelation, then (31) yields the method of moments ARMA spectral estimator. On the other hand if $q = 0$ and the estimates of ρ that arise from (23) using the Burg (Marple) estimates for the ϕ_i are used, then the resulting spectral estimator is the Burg (Marple) spectral estimator. This shows that ARMA spectral estimation method is a bias reduction method and in that sense suggests that it is more closely related to tapering than windowing.

Example 7 (Tail Probabilities)

In this example we will show how the jackknife can be used to obtain an approximation to a tail probability.

Let

$$F(\infty;t) = \int_t^\infty f(u)du \quad . \tag{32}$$

Then a natural approximation to $F(\infty;t)$ is

$$F(x;t) = \int_t^x f(u)du \tag{33}$$

and the error, $E(x)$, is given by

$$E(x) = F(x;t) - F(\infty;t) = -\int_x^\infty f(u)du \ . \tag{34}$$

Suppose that for some n and some set of constants, $\{a_i\}$, that

$$E^{(n)}(x) + a_n E^{(n-1)}(x) + \ldots + a_1 E(x) = 0 \ , \tag{35}$$

i.e. that the error satisfies a linear differential equation with constant coefficients of order n. Then, substituting $F(x;t) - F(\infty;t)$ in (35) and rearranging we have (analogous to (20))

$$F(x;t) = F(\infty;t) + b_1 f(x) + b_2 f^{(1)}(x) + \ldots + b_n f^{(n-1)}(x) \quad , \tag{36}$$

where

$$b_m = \frac{a_{m+1}}{a_1} \quad , \quad m = 1,2,\ldots,n-1, \quad b_n = -\frac{1}{a_1} \quad .$$

But now we have a bias expansion for our approximation and it can be used in a variety of ways to determine a generalized jackknife. For example, from (36), for $m = 1,2,\ldots,$ we can write

$$f^{(m-1)}(x) = b_1 f^{(m)}(x) + \ldots + b_n f^{(m+n-1)}(x) \ . \tag{37}$$

From (36) and (37) the n-th order jackknife is then (shortening the notation)

$$G^{(n)}[F(x;t)] = \frac{\begin{vmatrix} F(x;t) & f(x) & \ldots & f^{(n-1)}(x) \\ f(x) & f^{(1)}(x) & \ldots & f^{(n)}(x) \\ \vdots & & & \\ f^{(n-1)}(x) & f^{(n)}(x) & \ldots & f^{(2n-1)}(x) \end{vmatrix}}{\begin{vmatrix} 1 & 0 & \ldots & 0 \\ f(x) & f^{(1)}(x) & \ldots & f^{(n)}(x) \\ \vdots & & & \\ f^{(n-1)}(x) & f^{(n)}(x) & \ldots & f^{(2n-1)}(x) \end{vmatrix}} \ . \tag{38}$$

Note that this is the first example of the jackknife which makes use of our extended definition, i.e. in this case $C_1 = 1$ and $C_j = 0$, $j \geq 2$.

From Theorem 1 it then follows that

$$G^{(n)}[F(x;t)] = \int_t^\infty f(u)du$$

for every t. In fact we can take $x = t$ to obtain

$$\int_t^\infty f(u)du = G^{(n)}[F(x;x)] = \frac{\begin{vmatrix} 0 & f(x) & \cdots & f^{(n-1)}(x) \\ f(x) & f^{(1)}(x) & \cdots & f^{(n)}(x) \\ \vdots & & & \vdots \\ f^{(n-1)}(x) & \cdots & & f^{(2n-1)}(x) \end{vmatrix}}{\begin{vmatrix} f^{(1)}(x) & \cdots & f^{(n)}(x) \\ \vdots & & \\ f^{(n)}(x) & \cdots & f^{(2n-1)}(x) \end{vmatrix}} . \quad (39)$$

The result of (39) can be shown to be of value for a much larger class of functions than those satisfying (35) since the jackknife need not eliminate all of the bias to be of value. It can be shown under rather general conditions that the generalized jackknife of (39) converges to the tail probability with n, i.e.,

$$\lim_{n\to\infty} G^{(n)}[F(x,x)] = \int_x^\infty f(u)du, \quad (40)$$

see (3). Thus the generalized jackknife of the rather natural approximation of the tail probability produces an approximating function to that probability that converges to the tail probability as the order of the jackknife increases.

In order to exemplify (39) and (40), let

$$f(x;\alpha,\beta) = \begin{cases} \dfrac{1}{\Gamma(\alpha+1)\beta^{\alpha+1}} x^\alpha e^{-x/\beta} & , \quad x \geq 0 \\ 0 & , \quad x < 0 \end{cases}$$

and consider

$$F(\infty;t) = \int_t^\infty f(x;\alpha,\beta)dx .$$

Table 1 tabulates $G^{(n)}[F(x;x)]$ as a function of n when $\beta = 2$ and $\alpha = (m/2) -1$, m an integer, i.e. in the chi-square case. However the table is representative of the behavior of the jackknife approximation for a variety of distributions. The convergence with n can clearly be seen in the table. If one notes that for m even, f satisfies a homogeneous linear differential equation with constant coefficients, then that behavior is also clear in the table. For example $G^{(n)}(F(x;x))$ is exact for $m = 4$ and $n \geq 2$. If $m = 6$, the approximation is no longer exact at $n = 2$, but is exact for $n \geq 3$. In any event the approximation improves as n increases and as t increases. The latter behavior is due to the fact that we took $x = t$. That is, taking $x = t$ will limit the usefulness of the approximation to the tails of the distribution.

CONCLUDING REMARKS

We have demonstrated that if we consider the error in a numerical approximation as a degenerate random variable then numerical error is simply a special case of statistical bias. With this connection made, general methods for bias reduction should translate into general methods for error reduction and vice-versa. We have shown this to be the case and demonstrated the value of the observation in both arenas. Although the extension of the generalized jackknife given here is a simple one, it is an important one, as the final example shows.

REFERENCES

1. Gray, H. L., Atchison, T. A. and McWilliams, G. V. (1971), "Higher Order G-Transformations,", SIAM J. Numer. Anal., Vol. 8, No. 2, 365-81.

2. Gray, H. L., Watkins, T. A. and Adams, J. E. (1972), "On the jackknife statistic, its extensions and its relation to e_n-transformation," Annals of Math. Stat. Vol. 43, No. 1, 1-30.

3. Gray, H. L., Lewis, T. O. (1971), "Approximation of Tail Probabilities by Means of the B_n-Transformation," JASA Vol. 66, No. 336, pp. 897-899.

4. Morton, M. J. and Gray, H. L. (1984), "The G-Spectral Estimator," JASA, Vol. 79, No. 387, pp. 692-701.

5. Shanks, D. (1955), "Non-linear transformation of divergent and slowly convergent sequences", J. Math. Phys.. V. 34, pp. 1-42, MR 28 #1736.

6. Schucany, W. R., Gray, H. L. and Owen, D. B. (1971), "On Bias Reduction in Estimation," J. Amer. Stat. Assoc., Vol. 66, No. 335, 524-33.

Original paper received: 07.03.85
Final paper received: 01.09.86

Paper recommended by D.C. Boes