The Annals of Statistics
1981, Vol. 9, No. 3, 586–596

THE JACKKNIFE ESTIMATE OF VARIANCE

BY B. EFRON AND C. STEIN

Stanford University

Tukey's jackknife estimate of variance for a statistic $S(X_1, X_2, \cdots, X_n)$ which is a symmetric function of i.i.d. random variables X_i, is investigated using an ANOVA-like decomposition of S. It is shown that the jackknife variance estimate tends always to be biased upwards, a theorem to this effect being proved for the natural jackknife estimate of Var $S(X_1, X_2, \cdots, X_{n-1})$ based on $X_1, X_2, \cdots, X_n$.

1. Introduction. The Quenouille-Tukey jackknife, as described in Miller (1974), gives useful nonparametric estimates of bias and variance. Suppose $S(X_1, X_2, \cdots, X_n)$ is a statistic of interest,where $X_1, X_2, \cdots, X_n$ are independent and identically distributed observable random variables, and $S(X_1, X_2, \cdots, X_n)$ is invariant under permutation of the arguments. The jackknife estimate of variance, $\widehat{\text{VAR}}\, S(X_1, X_2, \cdots, X_n)$, is defined in terms of the quantities

$$S_{(i)} \equiv S(X_1, X_2, \cdots, X_{i-1}, X_{i+1}, \cdots, X_n), \tag{1.1}$$

the value of S when X_i is deleted from the sample,

$$\widehat{\text{VAR}}\, S(X_1, X_2, \cdots, X_n) \equiv \frac{n-1}{n} \sum_{i=1}^n [S_{(i)} - S_{(\cdot)}]^2, \tag{1.2}$$

where

$$S_{(\cdot)} \equiv \sum_{i=1}^n S_{(i)}/n. \tag{1.3}$$

Formula (1.2) is often used as a variance estimate for the jackknife version of S, defined as $nS - (n-1)S_{(\cdot)}$, but here we will be thinking of it either as a variance estimate for S itself, or perhaps more appropriately for $S_{(\cdot)}$.

Notice that $\widehat{\text{VAR}}$ is defined entirely with respect to samples of size $n - 1$, rather than n. It is useful to think of $\widehat{\text{VAR}}\, S(X_1, X_2, \cdots, X_n)$ as estimating the true variance Var $S(X_1, X_2, \cdots, X_n)$ in two distinct steps, (i) a direct estimate of Var $S(X_1, X_2, \cdots, X_{n-1})$ the variance for sample size $n - 1$, and (ii) a modification to go from sample size $n - 1$ to sample size n. The direct estimate is

$$\widetilde{\text{VAR}}\, S(X_1, X_2, \cdots, X_{n-1}) \equiv \sum_{i=1}^n [S_{(i)} - S_{(\cdot)}]^2, \tag{1.4}$$

and the sample size modification is

$$\widehat{\text{VAR}}\, S(X_1, X_2, \cdots, X_n) = \frac{n-1}{n} \widetilde{\text{VAR}}\, S(X_1, X_2, \cdots, X_{n-1}), \tag{1.5}$$

which together give (1.2). Notice that $\widetilde{\text{VAR}}\, S(X_1, X_2, \cdots, X_{n-1})$ is a function of all n variables $S(X_1, X_2, \cdots, X_n)$.

Our main result is that the jackknife estimate of variance (1.4) is conservative in expectation,

Received March, 1979; revised October, 1979.
AMS 1970 *subject classification*. 62G05.
Key words and phrases. Jackknife, variance estimation, ANOVA decomposition, bootstrap, U statistics.

$$E\{\widetilde{\text{VAR}}\, S(X_1, X_2, \cdots, X_{n-1})\} \geq \text{Var}\, S(X_1, \cdots, X_{n-1}) \tag{1.6}$$

for any symmetric function $S(X_1, X_2, \cdots, X_{n-1})$. As a matter of fact, neither symmetry nor identical distributions for the X_i are essential, as shown in Comment 4 of Section 2. We will also discusss (1.5), and show that under certain conditions, in particular if S is a U statistic, this step produces a further conservative bias in the jackknife variance estimate, though the results are not as satisfactory here.

The main tool in verifying (1.6) is an ANOVA-like decomposition of $S(X_1, X_2, \cdots, X_n)$, described in Section 2, which is a simple extension of the "Hajek projection", Hajek (1968). Colin Mallows, in lectures and an unpublished paper (1975), has developed closely related methods. All of these ideas connect with Hoeffding's (1948) well known work on U statistics. (See also Rubin and Vitale (1980) for a similar development.) A simple formula for the bias in (1.6) is derived from the ANOVA decomposition.

Much of jackknife theory concerns statistics S which are smooth functions of the empirical probability distribution. Section 3 relates this concept to the ANOVA decomposition of S, particularly focusing on *quadratic functionals*, which are useful in understanding the approximations involved in jackknife estimates, both for bias and for variance. This approach is quite similar to that of Hinkley (1978), as are the results of Section 5. The rationale behind (1.5) is examined in Section 4. Section 5 suggests a bias correction technique for the jackknife variance estimate. The "bootstrap", which is a generalization of the jackknife described in Efron (1979a), is examined briefly in Section 6.

2. ANOVA decomposition of $S(X_1, X_2, \cdots, X_n)$. A random variable $S(X_1, X_2, \ldots, X_n)$ which is a function of n independent random variables $X_1, X_2, \cdots, X_n$ can be decomposed into "main effects", "interactions", "higher order interactions", etc., in a manner directly analogous to the decomposition of a complete n-way ANOVA table. Here we do *not* have to assume that the X_i are identically distributed, nor that S is symmetrically defined with respect to its n arguments. The only assumption is that $ES^2 < \infty$. Taking advantage of this wide generality, an extended version of the main result (1.6) is given in Comment 4.

The quantities involved in the decomposition, and their corresponding ANOVA names are

$$\mu = ES, \tag{2.1}$$

grand mean;

$$A_i(x_i) = E\{S \mid X_i = x_i\} - \mu, \tag{2.2}$$

ith main effect;

$$B_{ii'}(x_i, x_{i'}) = E\{S \mid X_i = x_i, X_{i'} = x_{i'}\} - E\{S \mid X_i = x_i\} - E\{S \mid X_{i'} = x_{i'}\} + \mu, \tag{2.3}$$

ii'th second order interaction; etc.

Decomposition Lemma. *The random variable $S(X_1, X_2, \cdots, X_n)$ can be expressed as*

$$\begin{aligned} S(X_1, X_2, \cdots, X_n) = \mu &+ \textstyle\sum_i A_i(X_i) + \sum_{i<i'} B_{ii'}(X_i, X_{i'}) \\ &+ \textstyle\sum_{i<i'<i''} C_{i,i',i''}(X_i, X_{i'}, X_{i''}) + \cdots + H(X_1, X_2, \cdots, X_n), \end{aligned} \tag{2.4}$$

where all $2^n - 1$ random variables on the right side of (2.5) *have mean zero and are mutually uncorrelated with each other.*

Proof. Following through definitions (2.1)–(2.4), the coefficient of μ on the right side of (2.4) is

$$1 - \binom{n}{1} + \binom{n}{2} - \binom{n}{3} \cdots = (1-1)^n = 0. \tag{2.5}$$

Likewise the coefficient of $E\{S|X_i\}$ is $(1-1)^{n-1} = 0$, the coefficient of $E\{S|X_i, X_{i'}\}$ is $(1-1)^{n-2} = 0$, etc. The last term in (2.4) $H(X_1, \cdots, X_n)$, itself has first term $S(X_1, X_2, \cdots, X_n)$, which is the only term not cancelling out. This verifies expression (2.4).

Notice that

$$EA_i(X_i) = E\{E(S|X_i) - \mu\} = 0. \tag{2.6}$$

A similar calculation shows that

$$\begin{aligned} E\{B_{ii'}(X_i, X_{i'})|X_i\} &= E\{C_{ii'i''}(X_i, X_{i'}, X_{i''})|X_i, X_{i'}\} \\ &= \cdots E\{H(X_1, X_2, \cdots, X_n)|X_1, X_2, \cdots, X_{n-1}) = 0, \end{aligned} \tag{2.7}$$

and likewise $E\{B_{ii'}(X_i, X_{i'})|X_{i'}\} = 0$, etc. Together, (2.6) and (2.7) imply that all the random variables on the right side of (2.4) have mean zero and correlation zero, which completes the proof of the lemma. Note that $\mu + \sum_{i=1}^n A_i(X_i)$ is the Hajek projection of $S(X_1, X_2, \cdots, X_n)$, Hajek (1960). Expansion (2.4) is unique in the sense that once given properties (2.6)–(2.7), the terms μ, A_i, $B_{ii'}$, $C_{ii'i''}$, $\cdots$ must be given by expressions (2.1)–(2.3).

We now return to the situation where $X_1, X_2, \cdots, X_n$ are i.i.d., and $S(X_1, X_2, \cdots, X_n)$ is symmetrically defined in its arguments. In this case the functions $A_i(\cdot)$, $B_{ii'}(\cdot, \cdot)$, $\cdots$ do not depend upon the subscripts $i, i', i'', \cdots$, and we can indicate them as $A(\cdot)$, $B(\cdot, \cdot)$, $\cdots$. It will be helpful, for reasons stated in Section 3, to rescale and rename quantities as follows,

$$\begin{aligned} &\alpha_i \equiv \alpha(X_i) \equiv nA(X_i), \qquad \beta_{ii'} \equiv \beta(X_i, X_{i'}) \equiv n^2 B(X_i, X_{i'}), \\ &\gamma_{ii'i''} \equiv \gamma(X_i, X_{i'}, X_{i''}) \equiv n^3\, C(X_i, X_{i'}, X_{i''}), \cdots. \end{aligned} \tag{2.8}$$

Expansion (2.4) now becomes

$$\begin{aligned} S(X_1, X_2, \cdots, X_n) &= \mu + \frac{1}{n}\sum_i \alpha_i + \frac{1}{n^2}\sum_{i<i'} \beta_{ii'} \\ &\quad + \frac{1}{n^3}\sum_{i<i'<i''} \gamma_{ii'i''} + \cdots + \frac{1}{n^n}\eta_{1,2,3,\cdots,n}. \end{aligned} \tag{2.9}$$

Example. If $S(X_1, X_2, \cdots, X_n) = \sum_{i=1}^n (X_i - \bar{X})^2/n$, where $\bar{X} = \sum_{i=1}^n X_i/n$, and if the X_i have mean ξ and variance σ^2, then $\mu = \frac{n-1}{n}\sigma^2$, $\alpha(X_i) = \frac{n-1}{n}[(X_i - \xi)^2 - \sigma^2]$, $\beta(X_i, X_{i'}) = -2(X_i - \xi)(X_{i'} - \xi)$, and all the higher order terms equal zero.

Expansion (2.9), which is similar to a form Colin Mallows has suggested in unpublished lectures, leads to an easy proof of (1.6). Define

$$\sigma_\alpha^2 \equiv \operatorname{Var} \alpha(X_i), \qquad \sigma_\beta^2 \equiv \operatorname{Var} \beta(X_i, X_{i'}), \qquad \sigma_\gamma^2 \equiv \operatorname{Var} \gamma(X_i, X_{i'}, X_{i''}), \cdots. \tag{2.10}$$

Then using the fact that the terms in (2.9) are uncorrelated, we get

$$\operatorname{Var} S(X_1, X_2, \cdots, X_n) = \frac{\sigma_\alpha^2}{n} + \binom{n-1}{1}\frac{\sigma_\beta^2}{2n^3} + \binom{n-1}{2}\frac{\sigma_\gamma^2}{3n^5} + \cdots + \frac{\sigma_\eta^2}{n^{2n}}. \tag{2.11}$$

For any two indices i and i', the difference of the deleted sample values $S_{(i)}$ and $S_{(i')}$ is, by (2.9), equal to

$$S_{(i)} - S_{(i')} = \frac{1}{n-1}[\alpha_{i'} - \alpha_i] + \frac{1}{(n-1)^2}\sum_j^{(i,i')}[\beta_{i'j} - \beta_{ij}]$$

$$(2.12) \qquad + \frac{1}{(n-1)^3} \sum_{j<j'}^{(i,i')} [\gamma_{i'jj''} - \gamma_{ijj'}] + \cdots,$$

where the notation $\sum^{(i,i')}$ indicates summation *avoiding* the values i and i'. This implies that

$$(2.13) \qquad E[S_{(i)} - S_{(i')}]^2 = 2\left[\frac{\sigma_\alpha^2}{(n-1)^2} + \binom{n-2}{1}\frac{\sigma_\beta^2}{(n-1)^4} + \binom{n-2}{2}\frac{\sigma_\gamma^2}{(n-1)^6} + \cdots\right].$$

Since, by elementary algebra, $\widetilde{\text{VAR}}\ S(X_1, X_2, \cdots, X_{n-1}) \equiv \sum_{i=1}^n [S_{(i)} - S_{(\cdot)}]^2 = (1/n) \sum_{i<i'} [S_{(i)} - S_{(i')}]^2$, equation (2.13) gives

$$(2.14) \qquad E\{\widetilde{\text{VAR}}\ S(X_1, X_2, \cdots, X_{n-1})\} = \frac{\sigma_\alpha^2}{n-1} + \binom{n-2}{1}\frac{\sigma_\beta^2}{(n-1)^3} + \binom{n-2}{2}\frac{\sigma_\gamma^2}{(n-1)^5} + \cdots.$$

This, when compared with formula (2.11), for sample size $n-1$, verifies inequality (1.6), and gives a simple expression for the difference between the two sides.

THEOREM 1.

$$(2.15) \qquad E\{\widetilde{\text{VAR}}\ S(X_1, X_2, \cdots, X_{n-1})\} - \text{Var}\ S(X_1, X_2, \cdots, X_{n-1}) = \frac{1}{2}\binom{n-2}{1}\frac{\sigma_\beta^2}{(n-1)^3} + \frac{2}{3}\binom{n-2}{2}\frac{\sigma_\gamma^2}{(n-1)^5} + \cdots,$$

there being $n-2$ terms on the right side of (2.15).

Comment 1. The variances $\sigma_\beta^2, \sigma_\gamma^2, \cdots$ appearing in (2.15) refer to the expansion (2.9) for $S(X_1, X_2, \cdots, X_{n-1})$, *not* for $S(X_1, X_2, \cdots, X_n)$. Section 3 discusses this point in more detail.

Comment 2. For *linear functionals*, i.e., S statistics such that the higher order terms $\beta_{ii'}, \gamma_{ii'i''}, \cdots$ are all zero, the right-hand side of (2.15) equals zero, and so the jackknife variance estimate is unbiased. For a *quadratic functional*, one having all third and higher order terms equal to zero, the bias is $\frac{n-2}{2}\sigma_\beta^2/(n-1)^3$, which equals the contribution of the quadratic term, in (2.11), to Var $S(X_1, X_2, \cdots, X_{n-1})$. In other words, the jackknife variance estimate doubles the quadratic term in expectation. A correction is suggested in Section 5. In general, the quadratic term is doubled, the cubic term tripled. etc.

Comment 3. Let $\mu = ES(X_1, X_2, \cdots, X_{n-1})$, and consider the identity $\sum [S_{(i)} - S_{(\cdot)}]^2 = \sum [S_{(i)} - \mu]^2 - n[S_{(\cdot)} - \mu]^2$. Taking expectations gives

$$(2.16) \qquad E\sum [S_{(i)} - S_{(\cdot)}]^2 = n\ \text{Var}\ S_{(i)} - n\ \text{Var}\ S_{(\cdot)}.$$

Equation (5.21) of Hoeffding (1948), applied with $m = n-1$, gives

$$(2.17) \qquad n\ \text{Var}\ S_{(\cdot)} \leq (n-1)\ \text{Var}\ S_{(i)}.$$

Together, (2.16) and (2.17) imply (1.6).[1] Expansion (2.9) is closely related to Hoeffding's theory; the quantities $\delta_1, \delta_2, \delta_3, \cdots$ which play a crucial role in his proofs are multiples of

[1] We are indebted to Akimichi Takemura for pointing out this connection, and also to Mark Chesters for discussions relating to Comment 4.

σ_α^2, σ_β^2, σ_γ^2, $\cdots$, respectively. Expansion (2.9) is somewhat more convenient to work with, and yields one line proofs of Hoeffding's important theorems 5.1 and 5.2.

Comment 4. We can use expansion (2.4) to prove a more general version of (1.6):

THEOREM 2. *For any statistic* $S(X_1, X_2, \cdots, X_n)$ *having finite second moment, where* S *is not necessarily symmetrically defined in its arguments and the* X_i *are independent but not necessarily identically distributed,*

$$E \sum_{i=1}^n [S_{(i)} - S_{(\cdot)}]^2 \geq \frac{1}{n} \sum_{i=1}^n \operatorname{Var} S_{(i)} \geq \frac{n}{n-1} \operatorname{Var} S_{(\cdot)}. \tag{2.18}$$

PROOF. First assume that $\mu_i \equiv ES_{(i)} = 0$, $i = 1, 2, \cdots, n$. Define Diff $\equiv (n-1) \sum \operatorname{Var} S_{(i)} - n^2 \operatorname{Var} S_{(\cdot)}$ and let I, II, III be the three terms, from left to right, in (2.18). Taking expectations in $\sum [S_{(i)} - S_{(\cdot)}]^2 - \frac{1}{n} \sum S_{(i)}^2 = \{(n-1) \sum S_{(i)}^2 - n^2 S_{(\cdot)}^2\}/n$ gives I–II = Diff$/n$, while, directly, II–III = Diff$/\{n(n-1)\}$. We now show that Diff ≥ 0.

Still assuming $\mu_i = 0$, expansion (2.4) for $S_{(i)}$ can be written

$$S_{(i)} = \sum_{\mathscr{C}} S_{i\mathscr{C}} \tag{2.19}$$

where $\mathscr{C}$ indexes the $2^n - 2$ nonempty proper subsets of $\{1, 2, \cdots, n\}$. For example, with $i = 1$ and $\mathscr{C} = \{2, 3\}$, $S_{1\mathscr{C}} = B_{23}(X_2, X_3)$ in the expansion (2.4) of $S_{(1)}$. The random variables $S_{i\mathscr{C}}$ satisfy (i) $ES_{i\mathscr{C}} = 0$, (ii) $S_{i\mathscr{C}} = 0$ if $i \in \mathscr{C}$, and by (2.8), (iii) $ES_{i\mathscr{C}} S_{i'\mathscr{C}'} = 0$ *if* $\mathscr{C} \neq \mathscr{C}'$.

Define $S_{+\mathscr{C}} \equiv \sum_i S_{i\mathscr{C}}$, and notice that $ES_{+\mathscr{C}} S_{+\mathscr{C}'} = 0$ for $\mathscr{C} \neq \mathscr{C}'$. Therefore $E\, n^2 S_{(\cdot)}^2 = E \sum_{\mathscr{C}} S_{+\mathscr{C}}^2$, and likewise $E(n-1) \sum_i S_{(i)}^2 = E \sum_{\mathscr{C}} [(n-1) \sum_i S_{i\mathscr{C}}^2]$, so

$$\text{Diff} = E \sum_{\mathscr{C}} [(n-1) \sum_i S_{i\mathscr{C}}^2 - S_{+\mathscr{C}}^2]. \tag{2.20}$$

Letting $n_{\mathscr{C}}$ be the number of elements in $\mathscr{C}$, and $\bar{S}_{\mathscr{C}} \equiv S_{+\mathscr{C}}/(n - n_{\mathscr{C}})$,

$$\text{Diff} = E \sum_{\mathscr{C}} [(n_{\mathscr{C}} - 1) \sum_i S_{i\mathscr{C}}^2 + (n - n_{\mathscr{C}}) \sum_{i \notin \mathscr{C}} (S_{i\mathscr{C}} - \bar{S}_{\mathscr{C}})^2], \tag{2.21}$$

which verifies Diff ≥ 0.

Finally, notice that if we drop the assumption that the $S_{(i)}$ have means $\mu_i = 0$, the second and third terms in (2.18) are unchanged, while the first term is increased by the amount $\sum (\mu_i - \mu_\cdot)^2$, $\mu_\cdot \equiv \sum \mu_i/n$. This completes the proof of Theorem 2.

Essentially the same proof yields (2.18) in the following more general context: Let $(\Omega, \mathscr{B}, P)$ be a probability space and $\mathscr{B}_1, \cdots, \mathscr{B}_n$ independent sub σ-algebras of $\mathscr{B}$. For $i \in \{1, \cdots, n\}$ let $\mathscr{B}^{(i)}$ be the smallest σ-algebra containing all $\mathscr{B}_j$ for $j \neq i$, and let $S_{(i)}$ be a $\mathscr{B}^{(i)}$-measurable real random variable with $E\, S_{(i)}^2 < \infty$.

3. Functions of the empirical probability distribution.

Most of the theoretical work relating to the jackknife concerns statistics which are smooth functions $s(\hat{F})$ of the empirical probability distribution $\hat{F}$, putting mass $1/n$ at each value X_i. A typical example is the sample variance $S(X_1, X_2, \cdots, X_n) = \sum_{i=1}^n (X_i - \bar{X})^2/n$, while the "unbiased" version $\sum_{i=1}^n (X_i - \bar{X})^2/(n-1)$ is not of form $s(\hat{F})$.

By considering the case where the sample space $\mathscr{X}$ of the X_i is finite, say $\mathscr{X} = \{1, 2, \cdots, L\}$, we can describe the condition $S(X_1, \cdots, X_n) = s(\hat{F})$ in concrete terms. Define $f_l \equiv \text{Prob}\{X_i = l\}$, $\sum_{l=1}^L f_l = 1$, and let $\mathbf{f} = (f_1, f_2, \cdots, f_L)$ be the vector of probabilities. Likewise, let the empirical probabilities be the observed proportions $\hat{\mathbf{f}}_l = \#\{X_i = l\}/n$, and let $\hat{\mathbf{f}} = (\hat{f}_1, \hat{f}_2, \cdots, \hat{f}_L)$ be the empirical probability vector. The possible values of $\mathbf{f}$ compose the L-dimensional simplex

$$\mathscr{S} \equiv \{\mathbf{v} : v_l \geq 0, \sum_{l=1}^L v_l = 1\}, \tag{3.1}$$

while $\hat{\mathbf{f}}$ can occur only at certain lattice points of $\mathscr{S}$.

In this case, a "smooth function of the empirical probability distribution" is a statistic

$$S(X_1, X_2, \cdots, X_n) = s(\hat{\mathbf{f}}), \tag{3.2}$$

where $s(\cdot)$ is defined continuously on $\mathscr{S}$. The statistic $S \sum_{i=1}^n (X_i - \bar{X})^2/(n-1)$ cannot be of this form since doubling the number of observed X_i's at each value of l changes S without changing $\hat{\mathbf{f}}$. Huber (1977) gives an enlightening description of the continuity properties desirable in a good statistic. A statistic of the form $s(\hat{\mathbf{f}})$ is automatically defined for every sample size, not just for the n we happen to have. This is a handy property for jackknife calculations, where it is necessary to change the sample size, at least by one, to get the variance and bias estimates.

The simplest form of (3.2) is a linear functional,

$$s(\hat{\mathbf{f}}) = s(\mathbf{f}) + (\hat{\mathbf{f}} - \mathbf{f})\mathbf{u} \tag{3.3}$$

where $\mathbf{u} = (u_1, u_2, \cdots, u_L)'$ is a fixed, known vector. Since

$$(\hat{\mathbf{f}} - \mathbf{f})\mathbf{u} = \textstyle\sum_{l=1}^L (\hat{f}_l - f_l)u_l = \frac{1}{n}\sum_{i=1}^n (u_{X_i} - Eu) \tag{3.4}$$

where $Eu \equiv \sum_{l=1}^L f_l u_l$, (3.3) can be written as

$$S(X_1, X_2, \cdots, X_n) = s(\hat{\mathbf{f}}) = \mu + \frac{1}{n}\textstyle\sum_{i=1}^n \alpha_i, \tag{3.5}$$

using the definitions

$$\mu = s(\mathbf{f}) \qquad \text{and} \qquad \alpha_i = \alpha(X_i = l) = u_l - Eu. \tag{3.6}$$

The α_i have mean zero under $\mathbf{f}$, and so (3.5) is of the form (2.9).

The jacknife works perfectly for linear functionals (3.3), in the sense that it produces the obvious unbiased estimate of variance, $\widehat{\text{VAR}}\ S(X_1, X_2, \cdots, X_n) = \sum_{l=1}^L \hat{f}_l(u_l - \bar{u})^2/(n-1)$, where $\bar{u} = \sum_{l=1}^L \hat{f}_l u_l$. In order to examine the effects of nonlinearity on the jackknife, it is natural to consider quadratic functionals, say

$$s(\hat{\mathbf{f}}) = s(\mathbf{f}) + (\hat{\mathbf{f}} - \mathbf{f})\mathbf{u} + \frac{1}{2}(\hat{\mathbf{f}} - \mathbf{f})\mathbf{v}(\hat{\mathbf{f}} - \mathbf{f})', \tag{3.7}$$

$\mathbf{v}$ being a given symmetric $L \times L$ matrix. Hinkley (1978) considers a similar class of functionals.

Some straightforward algebraic manipulation gives expansion (2.9) for a quadratic functional (3.7). Let $\mathbf{1} \equiv (1, 1, \cdots, 1)$, and define

$$\begin{aligned} &\mathbf{U} \equiv \mathbf{u} - \mathbf{1}'\mathbf{f}\mathbf{u}, \qquad \mathbf{V} \equiv \mathbf{v} - \mathbf{1}'\mathbf{f}\mathbf{v} - \mathbf{v}\mathbf{f}'\mathbf{1} + \mathbf{1}'(\mathbf{f}\mathbf{v}\mathbf{f}')\mathbf{1}, \\ &\Delta(l) \equiv V_{ll}/2, \qquad E\Delta \equiv \textstyle\sum_{l=1}^L f_l\Delta(l). \end{aligned} \tag{3.8}$$

Then (3.7) can be written as

$$S(X_1, X_2, \cdots, X_n) = s(\hat{\mathbf{f}}) = \mu^{(n)} + \frac{1}{n}\textstyle\sum_i \alpha^{(n)}(X_i) + \frac{1}{n^2}\sum_{i<i'} \beta(X_i, X_{i'}), \tag{3.9}$$

where

$$\mu^{(n)} = s(\mathbf{f}) + \frac{E\Delta}{n}, \qquad \alpha^{(n)}(l) = U_l + \frac{\Delta(l) - E\Delta}{n}, \qquad \beta(l, l') = V_{l,l'}. \tag{3.10}$$

Comment 5. Letting $\hat{\mathbf{f}} = (1-\epsilon)\mathbf{f} + \epsilon\mathbf{e}_l$, $\mathbf{e}_l$ the lth coordinate vector, we get

$$\left.\frac{ds(\hat{\mathbf{f}})}{d\epsilon}\right|_{\epsilon=0} = U_l, \tag{3.11}$$

so that the coordinates of $\mathbf{U}$ are the *influence function* of $s(\cdot)$, see Huber (1977). Likewise the coordinates of $\mathbf{V}$ are the second order influence function. The normalization in (2.8), by powers of n, results in $\alpha(\cdot)$ approaching the first order influence function as $n \to \infty$, $\beta(\cdot, \cdot)$ approaching the second order influence function, etc. In other words, (2.9) approaches the standard von Mises expansion as $n \to \infty$, see Hinkley (1978).

Comment 6. The jackknife estimate of bias is

$$\widehat{\text{BIAS}}\; S(X_1, X_2, \cdots, X_n) = (n-1)(S_{(\cdot)} - S). \tag{3.12}$$

For the quadratic functional (3.7), equation (2.9) implies that

$$\widehat{\text{BIAS}}\; S(X_1, X_2, \cdots, X_n) = \frac{\Delta.}{n} - \frac{\beta..}{2n}, \tag{3.13}$$

where

$$\Delta_i \equiv \Delta(X_i), \qquad \Delta. \equiv \textstyle\sum_{i=1}^n \Delta_i/n, \qquad \beta.. \equiv \sum_{i<i'} \beta_{ii'}\Big/\binom{n}{2}. \tag{3.14}$$

Equation (3.13) says that the expectation of $\widehat{\text{BIAS}}\; S(X_1, X_2, \cdots, X_n)$, for a quadratic functional, equals $E\Delta/n = \mu^{(n)} - s(\mathbf{f}) = ES(X_1, X_2, \cdots, X_n) - s(\mathbf{f})$, so that $\widehat{\text{BIAS}}$ is itself unbiased for the bias of $S(X_1, X_2, \cdots, X_n)$ in estimating $s(\mathbf{f})$. The variance of $\widehat{\text{BIAS}}$ is

$$\text{Var}\; \widehat{\text{BIAS}} = \frac{\sigma_\Delta^2}{n^3} + \frac{\sigma_\beta^2}{2n^3(n-1)} \tag{3.15}$$

where

$$\sigma_\Delta^2 \equiv \text{Var}\; \Delta_i = \textstyle\sum_{l=1}^L f_l[\Delta(l) - E\Delta]^2. \tag{3.16}$$

Expression (3.15) follows from (3.13) because the Δ_i are mutually uncorrelated with each other and also with all the $\beta_{ii'}$.

Comment 7. Following through definitions (3.8)-(3.10), we see that (3.9) can be written as

$$\begin{aligned} &\mu^{(n)} = \mu^{(\infty)} + \frac{E\beta(X, X)}{2n}, \\ &\alpha^{(n)}(X) = \alpha^{(\infty)}(X) + \frac{\Delta(X) - E\Delta}{n} = \alpha^{(\infty)}(X) + \frac{\beta(X, X) - E\beta(X, X)}{2n}. \end{aligned} \tag{3.17}$$

A quadratic functional S can be defined as any statistic having the form (3.10), with $\mu^{(n)}$ and $\alpha^{(n)}(\cdot)$ obeying (3.17). This definition avoids mentioning the discrete sample space $\mathscr{X}$, and so is preferred for general discussion.

4. Variance relationships between different sample sizes. The rationale behind the sample size modification (1.5) is that for many familiar situations S, the true variance satisfies to a useful degree the approximation

$$\text{Var}^{(n)} = \frac{n-1}{n} \text{Var}^{(n-1)}, \tag{4.1}$$

where $\text{Var}^{(j)} \equiv \text{Var}\; S(X_1, X_2, \cdots, X_j)$. For linear functionals $S = \mu + \frac{1}{n} \sum_{i=1}^n \alpha_i$, (4.1) is an exact equality. Here we will discuss (4.1) for three classes of nonlinear functionals, (i) U statistics, (ii) quadratic functionals (3.9), (3.17), and (iii) "Jth order von Mises series,"[2] $S = \mu + \frac{1}{n} \sum_i \alpha_i + \frac{1}{n^2} \sum_{i<i'} \beta_{ii'} + \cdots$, where the highest term of the series has coefficient $1/n^J$, and the functions $\alpha(\cdot)$, $\beta(\cdot, \cdot)$, $\cdots$ do not change form as the sample size changes.

Theorem 1 can be rewritten, using definition (1.5), as

$$E\,\widehat{\mathrm{VAR}} = \mathrm{Var}^{(n)} + \left\{\frac{n-1}{n}\mathrm{Var}^{(n-1)} - \mathrm{Var}^{(n)}\right\} + O\left(\frac{1}{n^2}\right), \tag{4.2}$$

where $O(1/n^2) \geq 0$. For our three classes of nonlinear functionals, it turns out that $O(1/n^2)$ is actually of order $1/n^2$, or smaller, as n goes to infinity, and that the bracketed term in (4.2) is of order $O(1/n^3)$; see Efron and Stein (1979), Section 4. If S is a U statistic of fixed degree J, then for $n \geq J + 1$, this follows from Section 5 of Hoeffding (1948). As a matter of fact, Hoeffding's results show that the bracketed term is always nonnegative, so that for U statistics we have

$$E\,\widehat{\mathrm{VAR}} \geq \mathrm{Var}^{(n)}. \tag{4.3}$$

Inequality (4.3) also holds for von Mises series (a proof is given in Efron and Stein (1979)), though in this case the bracketed term in (4.2) can be negative. Note: a slight modification of our definition of a von Mises series, to $S = \mu + \frac{1}{n} \sum_i \alpha_i + \frac{1}{n(n-1)} \sum_{i<i'} \beta_{ii'} + \cdots$, makes it into a U statistic. The only reason for not beginning this way at definitions (2.9)–(2.10) is that it makes the connection with polynomial functionals, as at (3.9), (3.17), slightly more complicated.

For quadratic functionals (3.9), (3.17),

$$E\,\widehat{\mathrm{VAR}} = \mathrm{Var}^{(n)} + \frac{1}{n(n-1)}\left\{\frac{n^3 - n^2 - 3n + 1}{n^3 - n^2}\frac{\sigma_\beta^2}{2} + \frac{2\sigma_{\alpha\Delta}}{n} + \frac{\sigma_\Delta^2}{n^2(n-1)}\right\}, \tag{4.4}$$

where $\sigma_\Delta^2 \equiv \mathrm{Var}\,\Delta(X)$ as at (3.16), and $\sigma_{\alpha\Delta} \equiv E\alpha(X)\Delta(X)$, see Efron and Stein (1979). If $\sigma_{\alpha\Delta} \geq 0$ then (4.3) holds for all n; if $\sigma_{\alpha\Delta} < 0$ then (4.3) holds for sufficiently large n, a sufficient condition being $n - [3/(n-1)] > -4\sigma_{\alpha\Delta}/\sigma_\beta^2$.

It is not true, then, that the usual jackknife variance formula (1.2) is always nonnegatively biased for $\mathrm{Var}\,S(X_1, X_2, \cdots, X_n)$. However for smooth functionals the bias terms are of high order, $O(1/n^2)$, and are positive for sufficiently large n. (The results for quadratic functionals can be extended to higher polynomial forms.) Specific analytical and numerical results, for the case of ratio estimation, are given in Rao and Rao (1971). A more important question, which this paper does not address, is the variance of $\widehat{\mathrm{VAR}}$ itself; see Efron (1979).

5. Correcting the bias of the variance estimate. Knowing that the jackknife variance estimate is always biased upwards, it is natural to look for some correction to remove this bias. We will consider only quadratic functionals, (3.9), (3.17), and omit algebraic details, which are straightforward. Hinkley (1978) provides a similar development.

Define $S_{(i,i')}$ to be the value of S when both X_i and $X_{i'}$ are removed from the original sample, and let

$$Q_{ii'} \equiv nS - (n-1)(S_{(i)} + S_{(i')}) + (n-2)S_{(ii')}, \qquad i \neq i'. \tag{5.1}$$

Then

$$Q_{ii'} = \frac{1}{n-2}\left[(\beta_{ii'} - \beta_{i\cdot} - \beta_{i'\cdot} + \beta_{\cdot\cdot}) - \frac{(\Delta_i - \Delta_\cdot) + (\Delta_{i'} - \Delta_\cdot)}{n-1}\right], \tag{5.2}$$

---

[2] A name coined by Colin Mallows, in unpublished lectures.

the dot indicating averaging as at (3.14): $\beta_{i\cdot} \equiv \sum_{j \neq i} \beta_{ij}/(n-1)$, etc. The $Q_{ii'}$ can be used to estimate σ_β^2, and thereby eliminate the leading term in the bias of the jackknife variance estimate, (2.15):

LEMMA. *For quadratic functionals,*

$$\frac{1}{2} E[Q_{12} - Q_{34}]^2 = \frac{\sigma_\beta^2}{(n-1)^2}\left\{1 - \frac{3}{(n-2)^2}\right\} + \frac{2\sigma_\Delta^2}{(n-1)^2}\frac{1}{(n-2)^2} \tag{5.3}$$

and

$$\frac{1}{2} E[Q_{12} - Q_{23}]^2 = \frac{\sigma_\beta^2}{(n-1)^2}\left\{1 - \frac{n-3}{(n-2)^2}\right\} + \frac{\sigma_\Delta^2}{(n-1)^2}\frac{1}{(n-2)^2}. \tag{5.4}$$

The lemma says that

$$\frac{(n-1)^2}{2} E[Q_{ii'} - Q_{jj'}]^2 = \sigma_\beta^2 + O(1/n) \tag{5.5}$$

for any two distinct pairs $(i, i') \neq (j, j')$. Suppose we evaluate $Q_{ii'}$ for all $N = n(n-1)/2$ distinct pairs, and let $\bar{Q}$ be the average of the N values of $Q_{ii'}$. Then

$$\hat{\sigma}_\beta^2 \equiv \frac{(n-1)^2}{N-1} \sum [Q_{ii'} - \bar{Q}]^2 \tag{5.6}$$

is an estimator of σ_β^2 having bias $O(1/n^2)$. The bias corrected estimate of VAR $S(X_1, X_2, \cdots, X_n)$ is

$$\widehat{\text{VAR}}^{(\text{Corr})} S(X_1, X_2, \cdots, X_n) = \widehat{\text{VAR}}\, S(X_1, X_2, \cdots, X_n) - \frac{1}{n(n+1)} \sum_{i<i'} [Q_{ii'} - \bar{Q}]^2, \tag{5.7}$$

where $\widehat{\text{VAR}}$ is given by (1.2), and $\bar{Q} = \sum_{i<i'} Q_{ii'}/[n(n-1)/2]$.

Example. Efron (1979b), pages 462–463, considers the sample correlation coefficient of 15 pairs of numbers, each pair referring to two characteristics of the 1973 entering class at an American law school. The data are in Table 1. In this case the statistic of interest, $S(X_1, X_2, \cdots, X_n)$ is the sample correlation coefficient between the two characteristics. $S = .776$, and (3.12), (1.2), and (5.7) give

$$\begin{aligned} \widehat{\text{BIAS}}\, S(X_1, X_2, \cdots, X_{15}) &= 0.0065 \\ \widehat{\text{VAR}}\, S(X_1, X_2, \cdots, X_{15}) &= .0203 \\ \widehat{\text{VAR}}^{(\text{Corr})} S(X_1, X_2, \cdots, X_{15}) &= .0179. \end{aligned} \tag{5.8}$$

The referee has pointed out that the jackknife itself could be used to remove the bias in $\widehat{\text{VAR}}$. Doing so gives an estimate similar, but not identical, to $\widehat{\text{VAR}}^{(\text{Corr})}$. More ambitious unbiasing methods are also available, see Gray and Schucany (1972).

6. The bootstrap. A more general approach to jackknife-like calculations is de-

TABLE 1.
The average LSAT *score and undergraduate* GPA *at 15 American law schools, entering classes of 1973.*

School #	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
LSAT	576	635	558	578	666	580	555	661	651	605	653	575	545	572	594
GPA	3.39	3.30	2.81	3.03	3.44	3.07	3.00	3.43	3.36	3.13	3.12	2.74	2.76	2.88	2.96

scribed in Efron (1979a), under the rubric "bootstrap". The jackknife and the bootstrap are both examples of *resampling schemes*, in which the statistician attempts to learn the sampling properties of a statistic $S(X_1, X_2, \cdots, X_n)$ by recomputing its value for artificial samples, obtained by distorting the actual sample $X_1, X_2, \cdots, X_n$. Hartigan (1969, 1971, 1975) and Mallows (1975) have described several other interesting resampling schemes.

Using convenient notation, a vector of weights $\mathbf{P}^* = (P_1^*, P_2^*, \cdots, P_n^*)$, with the $P_i^* \geq 0$, $\sum_{i=1}^n P_i^* = 1$ leads to a resampled value of S in which the ith observation has weight P_i^*, compared to its weight $1/n$ in the real sample. We denote the resampled value as $S^* = s(\mathbf{P}^*)$. Here we are assuming that $S(X_1, X_2, \cdots, X_n)$ is of functional form, see Efron (1979a); $\mathbf{P}^*$ is abbreviated notation for the empirical probability distribution $\hat{F}^*$ putting mass P_i^* at X_i. Notice that $s(\mathbf{1}/n) = S(X_1, X_2, \cdots, X_n)$, the observed value.

The bootstrap considers vectors $\mathbf{P}^*$ having the distribution

$$\mathbf{P}^* \sim \frac{\text{Mult}(n, \mathbf{1}/n)}{n}. \tag{6.1}$$

(This should be compared with the jackknife, which uses $\mathbf{P}^*$ equal to all permutations of $(0, 1/(n-1), 1/(n-1), \cdots, 1/(n-1))$.) Here Mult$(n, \mathbf{1}/n)$ indicates a multinomial distribution, n draws, probability $1/n$ for each of the n categories. The vector $\mathbf{P}^*$ has mean vector and covariance matrix

$$E_*\mathbf{P}^* = \mathbf{1}/n, \qquad \text{Cov}_*\mathbf{P}^* = \mathbf{I}/n^2 - \mathbf{1}'\mathbf{1}/n^3. \tag{6.2}$$

The asterisk is a reminder that these calculations have nothing to do with the inherent randomness in the data, but rather with probabilities imposed by the statistician.

The bootstrap estimates of bias and variance are

$$\widehat{\text{BIAS}}^{(B)}\, S(X_1, X_2, \cdots, X_n) = E_* s(\mathbf{P}^*) - s(\mathbf{1}/n) \tag{6.3}$$

and

$$\widehat{\text{VAR}}^{(B)}\, S(X_1, X_2, \cdots, X_n) = \text{VAR}_* s(\mathbf{P}^*),$$

$E_* s(\mathbf{P}^*)$ and $\text{Var}_* s(\mathbf{P}^*)$ being taken with respect to distribution (6.1). The rationale behind these estimates is simply this: if the true probability distribution of the X_i happens to equal the empirical distribution (the distribution which puts mass $1/n$ at each observed X_i) then (6.3) gives exactly the correct bias and variance. The jackknife can be thought of as a "delta method" approximation to the bootstrap, see Section 5 of Efron (1979a). The bootstrap idea can be used to give different, more robust, estimates of bias and variance, see Efron (1979b), but here we will restrict our attention to (6.3), and demonstrate results similar to those obtained for the jackknife. Proofs are contained in Efron and Stein (1979) and will not be given here.

Once again we consider quadratic functionals $S(X_1, X_2, \cdots, X_n) = \mu + 1/n \sum_i \alpha_i + \frac{1}{n^2} \sum_{i<i'} \beta_{ii'}$. Since the bootstrap only involves samples of size n, the same size as the genuine sample, there is no need to consider how μ and α_i depend on n.

THEOREM 3. *For a quadratic functional* $S(X_1, X_2, \cdots, X_n)$,

$$\widehat{\text{BIAS}}^{(B)}\, S(X_1, X_2, \cdots, X_n) = \frac{n-1}{n} \widehat{\text{BIAS}}\, S(X_1, X_2, \cdots, X_n), \tag{6.4}$$

where $\widehat{\text{BIAS}}\, S(X_1, X_2, \cdots, X_n)$ *is the jackknife bias estimate* (3.13), *discussed in Comment* 6, *Section* 3.

THEOREM 4. *For a quadratic functional,*

$$E\left\{\frac{n}{n-1}\widehat{\mathrm{VAR}}^{(B)}\,S(X_1, X_2, \cdots, X_n) - \mathrm{Var}\,S(X_1, X_2, \cdots, X_n)\right\} \tag{6.5}$$
$$= \frac{c_1(n)\sigma_\beta^2 + 4c_2(n)\,\sigma_{\alpha\Delta}}{n^2} + \frac{6c_3(n)\sigma_\Delta^2 + c_4(n)(E\Delta)^2}{n^3},$$

where, as $n \to \infty$,

$$c_i(n) \to 1, \qquad i = 1, 2, 3, 4. \tag{6.6}$$

Specific values for $c_1(n)$, $c_2(n)$, $c_3(n)$, $c_4(n)$ are given in Efron and Stein (1979).

Comment 8. For a linear functional S, (6.5) equals 0. The form of (6.5) facilitates comparison with the corresponding jackknife result (4.4). The right-hand side of (6.5) can be either positive or negative, depending mainly on the sign of $\sigma_{\alpha\Delta}$ and the latter's relative magnitude compared to σ_β^2, cf. the remarks following (4.4).

REFERENCES

EFRON, B. (1979a). Bootstrap methods: Another look at the jackknife. *Ann. Statist.* **6** 1-26.

EFRON, B. (1979b). Computers and the theory of statistics: Thinking the unthinkable. *SIAM Rev.* **21** 460-480.

EFRON, B. and STEIN, C. (1979). The jackknife estimate of variance. Tech. Report No. 120, Dept. Statist., Stanford Univ.

GRAY, H. L. and SCHUCANY, W. R. (1972). *The Generalized Jackknife Statistic*. Dekker, New York.

HAJÉK, J. (1968). Asymptotic normality of simple linear rank statistics under alternatives. *Ann. Math. Statist.* **39** 325-346.

HARTIGAN, J. A. (1969). Using subsample values as typical values. *J. Amer. Statist. Assoc.* **64** 1303-1317.

HARTIGAN, J. A. (1971). Error analysis by replaced samples. *J. Roy. Statist. Soc. Ser. B* **33** 98-110.

HARTIGAN, J. A. (1975). Necessary and sufficient conditions for asymptotic joint normality of a statistic and its subsample values. *Ann. Statist.* **3** 573-580.

HINKLEY, D. V. (1978). Improving the jackknife with special reference to correlation estimation. *Biometrika* **65** 13-21.

HOEFFDING, W. (1948). A class of statistics with asymptotically normal distributions. *Ann. Math. Statist.* **19** 293-325.

HUBER, P. J. (1977). *Robust Statistical Procedures, SIAM*, Philadelphia.

MALLOWS, C. L. (1975). On some topics in robustness. Unpublished report, Bell Laboratories.

MILLER, R. G. (1974). The jackknife—a review. *Biometrika* **61** 1-15.

RAO, S. R. S. and RAO, J. N. K. (1971). Small sample results for ratio estimators. *Biometrika* **58** 625-630.

RUBIN, H. and VITALE, R. A. (1980). Asymptotic distribution of symmetric statistics. *Ann. Statist.* **8** 165-170.

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA 94305