A PROPERTY OF THE JACKKNIFE ESTIMATION OF THE VARIANCE
WHEN MORE THAN ONE OBSERVATION IS OMITTED

BY

R. P. BHARGAVA

TECHNICAL REPORT NO. 152
JANUARY 1980

PREPARED UNDER THE AUSPICES OF
NATIONAL SCIENCE FOUNDATION GRANT
MCS 78-07736

Ingram Olkin, Project Director

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA
A PROPERTY OF THE JACKKNIFE ESTIMATION OF THE VARIANCE

WHEN MORE THAN ONE OBSERVATION IS OMITTED

by

R. P. Bhargava

TECHNICAL REPORT NO. 152

January 1980

PREPARED UNDER THE AUSPICES OF

NATIONAL SCIENCE FOUNDATION GRANT

MCS 78-07736

Ingram Olkin, Project Director

Also prepared under National Science Foundation Grant MPS 77-16974, Technical Report No. 140. This work supported in part by the National Research Council of Canada under grant A9263.

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA
A PROPERTY OF THE JACKKNIFE ESTIMATION OF THE VARIANCE
WHEN MORE THAN ONE OBSERVATION IS OMITTED

By

R. P. Bhargava

Abstract

Let $X_1, \ldots, X_n$ be independent and identically distributed random variables. Let $\psi: \mathbb{R}^{n-r} \rightarrow \mathbb{R}$ be a symmetric function of the $n-r$ arguments. Let set $A \subset \{1, \ldots, n\}$ and $|A|$ denote the number of elements in $A$. Let $|A| = n-r$, $1 \leq r \leq n-1$. Let

$$S_A = \psi(X_A), \quad \bar{S} \equiv \frac{1}{m} \sum_{A \subset \{1, \ldots, n\}, |A| = n-r} S_A,$$

and

$$V \equiv \sum_{A \subset \{1, \ldots, n\}, |A| = n-r} (S_A - \bar{S})^2.$$

We assume $E S_A^2 < \infty$.

In this paper we obtain the following results:

(a) $\operatorname{Var} \bar{S} \leq \frac{n-r}{n} \operatorname{Var} S_A$,

(b) For $n \geq 2$ and $r \geq 1$, we have

$$EV \geq \binom{n}{r} \frac{r}{n} \operatorname{Var} S_A \quad \text{and} \quad E \frac{n-r}{m} \frac{V}{r} \geq \operatorname{Var} \bar{S}.$$

Efron and Stein (1978) obtained results (a) and (b) for $r = 1$.

Key Words: Bias reduction; conservative estimate of variance; Jackknife; Pseudo-values; Robustness.
A PROPERTY OF THE JACKNIFE ESTIMATION OF THE VARIANCE
WHEN MORE THAN ONE OBSERVATION IS OMITTED

by

R. P. Bhargava
Stanford University
and
Ontario Institute for Studies in Education

1. Introduction.

Let $X_1, \ldots, X_n$ be independent and identically distributed observable random variables. Let $\psi: \mathbb{R}^{n-r} \to \mathbb{R}$ be a symmetric function in the sense that its value is invariant under permutation of the arguments. Let set $A \subset \{1, \ldots, n\}$ and let the number of elements in $A$, denoted by $|A|$, be $n-r$ (i.e., $|A| = n-r$), $1 \leq r \leq n-1$. Define $X_A = \{X_i : i \in A\}$. Let

$$S_A = \psi(X_A), \quad (1.1)$$

$$\overline{S} = \frac{1}{m} \sum_{A \subset \{1, \ldots, n\}} S_A, \quad \text{where} \quad m = \binom{n}{r}, \quad (1.2)$$

and

$$V = \sum_{A \subset \{1, \ldots, n\}, \quad |A| = n-r} \left( S_A - \overline{S} \right)^2 \quad (1.3)$$

We assume $E S_A^2 < \infty$.

Partially supported by the National Research Council of Canada Grant No. 9263.
In this paper we prove the following results:

(a) \[ \text{Var} \; \overline{S} \leq \frac{n-r}{n} \text{Var} \; S_A, \] \hspace{1cm} (1.4)

(b) For \( n \geq 2 \), \( r \geq 1 \),

\[ \mathbb{E} V \geq \binom{n}{r} \frac{r}{n} \text{Var} \; S_A \quad \text{and} \quad \mathbb{E} \frac{n-r}{r} \frac{V}{m} \geq \text{Var} \; \overline{S}. \] \hspace{1cm} (1.5)

Efron and Stein (1978) proved results a) and b) for \( r = 1 \) using an ANOVA-like decomposition of \( S_A \) which was an extension of the "Hajek projection", Hajek (1968). Hoeffding (1948) and Colin Mallow (1975) (unpublished report) developed closely related ideas. Stein's presentation of results of Efron and Stein (1978) in class at Stanford differed slightly from Efron and Stein (1978). In the following we use Stein's approach to prove results (1.4) and (1.5). These results are of use in Jackknife statistics.

In section 2 we give ANOVA-like decomposition of \( S_A \). In section 3.1 we prove result a). In section 3.2 we prove result b) for \( 1 \leq r < n-r \). (See Theorem 1). In section 3.3 we prove result b) for \( r \geq n-r \geq 1 \). (See Theorem 2). Theorems 1 and 2 imply (1.5).

2. **ANOVA Decomposition of \( S_A \).**

Let \( X_B = \{X_i : i \in B\} = \{X_i\} i \in B \) and \( \mathcal{X}_B \) denote the space of random variables depending only on \( X_B \) which have finite second moment. *Let \( T_B(X_B) \) be the orthogonal projection of \( S_A, A \subset \{1, \ldots, n\}, |A| = n-r, *

*Footnote: In this paper \( P \subset Q \) means \( P \) is a subset of \( Q \) but \( P \neq Q \), and \( P \subseteq Q \) means \( P \) is a subset of \( Q \) and \( P \) can be equal to \( Q \).*
B \subseteq A, \text{ on the orthogonal complement } \sum_{C \subseteq B} \mathbb{E}_C \text{ in } \mathbb{E}_B. \text{ (Note that } \sum_{C \subseteq B} \mathbb{E}_C = \sum_{C \subseteq B} a_C Z_C, \text{ where } a_C \in \mathbb{R}, Z_C \in \mathbb{E}_C). \text{ Stein showed in class that } S_A = \sum_{B \subseteq A} T_B(X_B), \text{ where } T_B(X_B) \text{ depends on } X_B \text{ and is the same for all } S_A, A \supseteq B, |A| = n-r. \text{ Note }

T_B(X_B) = \sum_{C \subseteq B} (-1)^{|B-C|} \mathbb{E}_{C \subseteq B} \cdot S_A,

(2.1)

where \( v(B-C) \) denotes the number of elements in the set \( B-C \). Geometrically this is obvious. (See Efron and Stein (1978)). Note \( \mathbb{E} T_B(X_B) = 0 \) for \( B \neq \emptyset \) and

\[
\text{Cov}(T_{B_1}(X_{B_1}), T_{B_2}(X_{B_2})) = \mathbb{E} T_{B_1}(X_{B_1}) T_{B_2}(X_{B_2}) = 0,
\]

(2.2)

for \( B_1 \neq B_2, B_1 \subseteq \{1, \ldots, n\}, B_2 \subseteq \{1, \ldots, n\} \). Equation (2.2) follows using (2.1). Further note that if for sets \( B \) and \( B' \), \( |B| = |B'| \) then \( T_B(X_B) \) and \( T_{B'}(X_{B'}) \) are identically distributed. We denote \( \text{Var} T_B(X_B) = \text{Var} T_{B'}(X_{B'}) = \sigma^2_{|B|} \).

3. Proof of Main Results (a), (b_1) and (b_2).

3.1. In this section we want to prove result (a), i.e.,

\[
\text{Var} \bar{S} \leq \frac{n-r}{n} \text{Var} S_A.
\]

(3.1)

Proof. Recall
\[ S_A = \sum_{B \subseteq A} T_B(X_B), \quad E T_B^2(X_B) = \nu_B^2, \quad m = \binom{n}{r}, \quad (3.2) \]

and

\[ \text{Var} \ S_A = \sum_{B \subseteq A} \sigma_B^2 = \sum_{k=1}^{n-r} \binom{n-r}{k} \sigma_k^2 \quad (3.3) \]

Now

\[ \bar{S} = \frac{1}{m} \sum_{A \subseteq \{1, \ldots, n\}} \sum_{|B| \leq n-r} \frac{1}{m} \sum_{A \subseteq \{1, \ldots, n\}} T_B(X_B) \]

\[ = \frac{1}{m} \sum_{A \subseteq \{1, \ldots, n\}} \frac{1}{r} \sum_{|A| \leq n-r} \sum_{r=1}^{n-|B|} T_B(X_B) \quad (3.4) \]

\[ \text{Var} \ \bar{S} = \frac{1}{m^2} \sum_{k=1}^{n-r} \binom{n}{k} (n-k)^2 \sigma_k^2 \]

\[ = \sum_{k=1}^{n-r} \binom{n-k}{n} \binom{n-k-1}{n-1}^2 \cdots \binom{n-k-r+1}{n-r+1} \frac{1}{k! (n-k)!} \sigma_k^2 \]

\[ = \sum_{k=1}^{n-r} \frac{n-k}{n} \frac{n-k-1}{n-1} \cdots \frac{n-k-r+1}{n-r+1} \frac{1}{k! (n-k)!} \sigma_k^2 \]

\[ \leq \sum_{k=1}^{n-r} \frac{n-2}{n} \cdots \frac{n-r}{n-r+1} \sum_{k=1}^{n-r} \binom{n-r}{k} \sigma_k^2 = \frac{n-r}{n} \text{Var} \ S_A. \quad (3.5) \]

3.2. In this section we first prove three lemmas and then result \( (b_1) \).

**Lemma 1.** Let \( X_1, \ldots, X_m \) be identically distributed and
\[ E \sum_{j=1}^{m} X_j \sum_{j=1}^{m} X_j = E X_1 \sum_{j=1}^{m} X_j, \ i, i' = 1, \ldots, m. \] Then

\[ E \sum_{i=1}^{m} (X_i - \bar{X})^2 = m E X_1^2 - E X_1 \sum_{i=1}^{m} X_i, \tag{3.6} \]

where \( m \bar{X} = \sum_{j=1}^{m} X_j. \)

Proof.

\[ \sum_{i=1}^{m} (X_i - \bar{X})^2 = \frac{1}{2m} \sum_{i=1}^{m} \sum_{i'=1}^{m} (X_i - X_{i'})^2 \]

\[ = \frac{1}{2m} \sum_{i=1}^{m} \left[ m X_1^2 - 2 X_1 \sum_{i'=1}^{m} X_{i'} + \sum_{i'=1}^{m} X_{i'}^2 \right] \tag{3.7} \]

and

\[ E \sum_{i=1}^{m} (X_i - \bar{X})^2 = \frac{1}{2m} \sum_{i=1}^{m} \left[ m E X_1^2 - 2 E X_1 \sum_{i'=1}^{m} X_{i'} + m E X_1^2 \right] \]

\[ = m E X_1^2 - E X_1 \sum_{i'=1}^{m} X_{i'}, \tag{3.8} \]

Lemma 2. For \( 1 \leq k \leq n-r-t, \) we have

\[ \binom{n-r}{k} \geq \frac{n-r}{n-r-t} \binom{n-r-t}{k}. \tag{3.9} \]

Proof.

\[ \binom{n-r}{k} = \frac{n-r}{n-r-k} \binom{n-r-1}{k} \geq \frac{n-r}{n-r-1} \binom{n-r-1}{k}. \tag{3.10} \]
The result follows by applying inequality (3.9) t-1 times, t ≤ n-r.

Remark. To prove results (b₁) and (b₂) we assume without loss of generality that $E \mathbf{S}_A = 0$, i.e., $T_\varphi(X_\varphi) = 0$.

**Lemma 3.** Let $A \subset \{1, \ldots, n\}$, $A' \subset \{1, \ldots, n\}$, $|A| = |A'| = n-r$ and $|A \cap A'| = n-r-s$. Then

$$E \mathbf{S}_A \mathbf{S}_{A'} = \sum_{k=1}^{n-r-s} \binom{n-r-s}{k} \sigma_k^2.$$  \hspace{1cm} (3.11)

**Proof.**

$$E \mathbf{S}_A \mathbf{S}_{A'} = E \sum_{B \subset A} \sum_{B' \subset A'} T_B(X_B) T_{B'}(X_{B'})$$

$$= \sum_{B \subset A \cap A'} \sigma_B^2 = \sum_{k=1}^{n-r-s} \binom{n-r-s}{k} \sigma_k^2.$$ \hspace{1cm} (3.12)

Now we prove (b₁).

**Theorem 1.** Let $1 \leq r < n-r$. Then

$$E V = E \sum_{A \subset \{1, \ldots, n\}} \frac{(n-r)_r}{r^n} \text{Var} \mathbf{S}_A$$

$$A \subset \{1, \ldots, n\}$$

$$|A| = n-r$$

and

$$E \frac{n-r}{\mu} V \geq \text{Var} \mathbf{S},$$ \hspace{1cm} (3.14)

where $\mu = \binom{n}{r}$. 

6
\[ EV = E \sum (S_A - \bar{S})^2 = \frac{1}{2m} \sum (S_A - S_{A'})^2 \]

\[ A \subset \{1, \ldots, n\} \quad A, A' \subset \{1, \ldots, n\} \]

\[ |A| = n-r \quad |A'| = n-r \]

\[ = m E S_A^2 - E S_A \sum S_{A'} \quad \text{(by Lemma 1),} \]

\[ A' \subset \{1, \ldots, n\} \]

\[ |A'| = n-r \]

\[ = (m-1)E S_A^2 - \left[ \binom{n-r}{1} \sum_{k=1}^{n-r-1} k \sigma_k^2 \right] (\text{by Lemma 3),} \]

\[ \binom{n-r}{2} \sum_{k=1}^{n-r-2} k \sigma_k^2 \]

\[ \vdots \]

\[ \binom{n-r}{r} \sum_{k=1}^{n-2r} k \sigma_k^2 \]

\[ \geq (m-1) \sum_{k=1}^{n-r} k \sigma_k^2 \]

\[ - \left[ \binom{n-r}{1} \sum_{k=1}^{n-r-1} k \frac{n-r-1}{n-1} \right] \binom{n-r}{2} \sum_{k=1}^{n-r-2} k \sigma_k^2 \]

\[ + \binom{n-r}{2} \sum_{k=1}^{n-r-2} k \sigma_k^2 \]

\[ \vdots \]

\[ \binom{n-r}{r} \sum_{k=1}^{n-2r} k \sigma_k^2 \]

\[ \text{(by Lemma 2),} \]

\[ = (m-1) - \binom{n-r-1}{1} - \binom{n-r-1}{2} - \cdots - \binom{n-r-1}{r} \text{Var } S_A \]

\[ = (m-1) \frac{n-1}{n-r-1} \text{Var } S_A \]

\[ = \binom{n}{r} \frac{r}{n} \text{Var } S_A . \]
This proves (3.13). (3.14) follows from (3.13) and (3.5).

3.3. We now prove result (b)</code>.

Theorem 2. Let \( r \geq n-r \geq 1 \). Then

\[
E V = E \sum_{A \subseteq \{1, \ldots, n\}} \left( \bar{S}_A - \bar{S} \right)^2 \geq \binom{n}{r} \frac{r}{n} \text{Var } S_A
\]

\( A \subset \{1, \ldots, n\} \)

\( |A| = n-r \)

and

\[
E \frac{n-r}{r} \frac{V}{m} \geq \text{Var } \bar{S},
\]

where \( m = \binom{n}{r} \).

Proof. (3.22) and (3.23) are true for \( n-r = 1 \). In the following we therefore assume \( r \geq n-r \geq 2 \). From (3.16),

\[
E V = m \ E S_A^2 - E S_A \ \sum_{A' \subset \{1, \ldots, n\},} \ \left| A' \right| = n-r
\]

\[
\geq (m-1) E S_A^2 - \left[ \binom{n-r}{r} \binom{r}{1} \binom{n-r-1}{k} \sum_{k=1}^{r} \left( \frac{n-r-1}{k} \sigma_k^2 \right) \right.
\]

\[
+ \left. \binom{n-r}{r} \binom{r}{2} \sum_{k=1}^{n-r-2} \left( \frac{n-r-2}{k} \sigma_k^2 \right) \right] ...
\]

\[
+ \left. \binom{n-r}{r} \binom{r}{n-r-1} \sum_{k=1}^{1} \left( \frac{n-r-(n-r-1)}{k} \sigma_k^2 \right) \right]
\]

8
\[ \geq (m-1) E S_A^2 - \left( \binom{n-r}{n-r-1}^2 \binom{n-r-1}{r} \frac{1}{n-r} \sum_{k=1}^{n-r} \binom{n-r}{k} \sigma_k^2 \right) \]

\[ + \binom{n-r}{n-r-2} \binom{n-r-2}{1} \frac{1}{n-r} \sum_{k=1}^{n-r} \binom{n-r}{k} \sigma_k^2 \]

\[ \cdots \]

\[ + \binom{n-r}{n-r-1} \binom{r}{n-r-1} \frac{1}{n-r} \sum_{k=1}^{n-r} \binom{n-r}{k} \sigma_k^2 \]

\[ = m E S_A^2 - \left[ 1 + \binom{n-r-1}{r} \binom{n-r-2}{r} + \binom{n-r-1}{r} \binom{n-r-2}{r} + \cdots + \binom{n-r-1}{0} \binom{r}{n-r-1} \right] E S_A^2 \]

\[ = \left( \binom{n}{r} - \binom{n-1}{r-1} \right) E S_A^2 \]

\[ = \binom{n}{r} \frac{r}{n} \text{var} S_A. \]

This proves (5.22). (5.23) follows from (5.22) and (5.5).

Note Theorems 1 and 2 imply (1.5).

4. **Acknowledgment.**

I am grateful to Professor C. Stein for introducing me to this subject.
References


