

Covariances of Symmetric Statistics

RICHARD A. VITALE

University of Connecticut

Communicated by C. R. Rao

We examine the second-order structure arising when a symmetric function is evaluated over intersecting subsets of random variables. The original work of Hoeffding is updated with reference to later results. New representations and inequalities are presented for covariances and applied to U -statistics. © 1992 Academic Press, Inc.

1. INTRODUCTION

The importance of the symmetric use of sample observations was apparently first observed by Halmos [6] and, following the pioneering work of Hoeffding [7], has been reflected in the active study of U -statistics (e.g., Hoeffding [8], Rubin and Vitale [11], Dynkin and Mandelbaum [2], Mandelbaum and Taquq [10] and, for a lucid survey of the classical theory Serfling [12]). An important feature is the ANOVA-type expansion introduced by Hoeffding and its application to finite-sample inequalities, Efron and Stein [3], Karlin and Rinnot [9], Bhargava [1], Takemura [14], Vitale [16], Steele [13]). The purpose here is to survey work on this latter topic and to present new results. After some preliminaries, Section 3 organizes for comparison three approaches to the ANOVA-type expansion. The next section takes up characterizations and representations of covariances. These are then used to tighten an important inequality of Hoeffding and, in the last section, to study the structure of U -statistics.

2. NOTATION AND PRELIMINARIES

The general setup is a supply $X_1, X_2, \dots, X_n$ of iid random variables (or random elements of any appropriate space) and a symmetric function h of

Received February 26, 1991; revised August 8, 1991.

AMS 1980 subject classifications: 62J10, 60E15, 62E30, 62G05, 62H20.

Key words and phrases: ANOVA, correlation, covariance, Hájek projection, symmetric statistic, U -statistic.

m arguments, $m \leq n$. Evaluations of h are made over m -sets of $\{X_1, \dots, X_n\}$ and $h(X_{i_1}, \dots, X_{i_m})$, $1 \leq i_1 < i_2 < \dots < i_m \leq n$, is denoted by $h(W_I)$, where $I = \{i_1, i_2, \dots, i_m\}$. $|I|$ stands for cardinality of I . In applications, h often serves as an estimator of Eh but here we assume that $Eh = 0$ and additionally $0 < Eh^2 < \infty$.

3. DECOMPOSITION

Symmetric statistics (e.g., U -statistics; Section 6) are generally built up from evaluations of a kernel function h , and their properties reflect the properties of the kernel. The decomposition of h itself into orthogonal components, an ANOVA-like analysis, is fundamental to the theory. We follow Hoeffding [7].

Set $h_k(X_1, \dots, X_k) \equiv E[h(X_1, \dots, X_m) \mid X_1, \dots, X_k]$, $k = 1, \dots, m$, and note that it inherits symmetry from h . Define

$$\begin{aligned} g_1(X_1) &= h_1(X_1) \\ g_2(X_1, X_2) &= h_2(X_1, X_2) - g_1(X_1) - g_1(X_2) \\ g_3(X_1, X_2, X_3) &= h_3(X_1, X_2, X_3) - \sum_{i=1}^3 g_1(X_i) - \sum [i_1 < i_2] g_2(X_{i_1}, X_{i_2}) \\ &\vdots \\ g_m(X_1, \dots, X_m) &= h(X_1, \dots, X_m) - \sum_{i=1}^m g_1(X_i) - \sum [i_1 < i_2] g_2(X_{i_1}, X_{i_2}) \\ &\quad - \dots - \sum [i_1 < i_2 < \dots < i_{m-1}] g_{m-1}(X_{i_1}, \dots, X_{i_{m-1}}). \end{aligned}$$

An equivalent way of writing the g_k 's follows from observing the pattern

$$\begin{aligned} g_1(X_1) &= h_1(X_1) \\ g_2(X_1, X_2) &= h_2(X_1, X_2) - h_1(X_1) - h_1(X_2) \\ g_3(X_1, X_2, X_3) &= h_3(X_1, X_2, X_3) - \sum [i_1 < i_2] h_2(X_{i_1}, X_{i_2}) + \sum [i] h_1(X_i) \\ &\vdots \end{aligned}$$

and reading off the inclusion-exclusion form (here and below we use bracket notation to emphasize or restrict ranges of summation):

$$g_k(X_1, \dots, X_k) = g_{|I|}(X_I) = \sum [\phi \neq J \subseteq I] (-1)^{|I| - |J|} h_{|J|}(X_J). \quad (3.1)$$

The h_k 's can be recaptured via $h_{|I|}(X_I) = \sum [\phi \neq J \subseteq I] g_{|J|}(X_J)$, $|I| = k$, and in particular

$$h(X_1, \dots, X_m) = \sum [\phi \neq J \subseteq \{1, \dots, m\}] g_{|J|}(X_J). \quad (3.2)$$

EXAMPLE. Let $m = 2$ and $h(X_1, X_2) = \frac{1}{2}(X_1 - X_2)^2 - \sigma^2$, where $\mu = EX_1$, $\sigma^2 = \text{Var } X_1$. Then $h_1(X_1) = \frac{1}{2}(X_1 - \mu)^2 - \sigma^2 = g_1(X_1)$ and $g_2(X_1, X_2) = -(X_1 - \mu)(X_2 - \mu)$. Note the distributional dependence of the decomposition.

We next record some important properties.

THEOREM 3.1. *For a symmetric h and the g_k 's as defined above,*

$$Eg_k(X_1, \dots, X_k) = 0, \quad k = 1, \dots, m \quad (3.3)$$

and

$$E[g_k(X_1, \dots, X_k) \mid X_1, \dots, X_{k-1}] = 0, \quad k = 2, \dots, m, \quad (3.4)$$

which can be subsumed under

$$E[g_{|I|}(X_I) \mid X_{I'}] = 0 \quad \text{if } I \not\subseteq I'. \quad (3.5)$$

Here $I' = \phi$ means an unconditional expectation. Further,

$$\text{Cov}[g_{|I|}(X_I), g_{|I'|}(X_{I'})] = 0 \quad \text{if } I \neq I'. \quad (3.6)$$

The decomposition (3.2) is unique subject to (3.4).

We defer the proof. The decomposition can be viewed geometrically as an organization of projections in a Hilbert space (Dynkin and Mandelbaum [2], Takemura [14]). Specifically, regard h as an element of $L_2(X_1, \dots, X_m)$, the square integrable functions of $X_1, \dots, X_m$. Let Q_i denote the projection operator associated with the subspace $L^{(i)} = L_2(X_1, \dots, X_{i-1}, X_{i+1}, \dots, X_m)$. That is $(Q_i f)(X_1, \dots, X_{i-1}, X_{i+1}, \dots, X_m) = E[f(X_1, \dots, X_m) \mid X_1, \dots, X_{i-1}, X_{i+1}, \dots, X_m]$. It is straightforward to verify that

(1) the Q_i commute

(2) $Q_{i_1} \cdots Q_{i_k}$ is the projection onto $L^{(i_1)} \cap L^{(i_2)} \cap \cdots \cap L^{(i_k)}$.

In particular, $(Q_{k+1} \cdots Q_m f)(X_1, \dots, X_k) = E[f(X_1, \dots, X_m) \mid X_1, \dots, X_k]$ and $Q_1 \cdots Q_m f = Ef$.

With $f = h$, one can verify that

$$g_k(X_1, \dots, X_k) = (I - Q_1)(I - Q_2) \cdots (I - Q_k) Q_{k+1} \cdots Q_m h, \quad k = 1, \dots, m-1,$$

and

$$g_m(X_1, \dots, X_m) = (I - Q_1)(I - Q_2) \cdots (I - Q_m)h.$$

More generally,

$$g_k(X_{i_1}, \dots, X_{i_k}) = (I - Q_{i_1})(I - Q_{i_2}) \cdots (I - Q_{i_k})Q_{i_{k+1}} \cdots Q_{i_m}h.$$

The decomposition follows from expanding the identity

$$h = I^m h = [(I - Q_1) + Q_1][(I - Q_2) + Q_2] \cdots [(I - Q_m) + Q_m]h.$$

A third approach (Rubin and Vitale [11]) uses tensor product methods. Again let $h \in L_2(X_1, \dots, X_m)$. Let $e_0(X_1) \equiv 1$, $e_1(X_1), e_2(X_1), \dots$, be an orthonormal basis for $L_2(X_1)$. Then an orthonormal basis for $L_2(X_1, \dots, X_m)$ comprises products of the form

$$e_{j_1}(X_1) \cdot e_{j_2}(X_2) \cdots e_{j_m}(X_m) \equiv e_j(X_1, \dots, X_m)$$

(NB: j here is a vector, $j = (j_1, \dots, j_m)$). Consequently,

$$h(X_1, \dots, X_m) = \sum [j] c_j e_j(X_1, \dots, X_m) \quad (3.7)$$

for certain coefficients c_j that are invariant under permutations of j (owing to the symmetry of h). Note that the assumption $Eh = 0$ translates here to

$$c_{0,0,\dots,0} = 0. \quad (3.8)$$

It follows from (2.7) that generally

$$h_k(X_1, \dots, X_k) = \sum [j_{k+1} = \cdots = j_m = 0] c_j e_j(X_1, \dots, X_m) \quad (3.9)$$

with easily deducible variants, e.g., $h_1(X_2) = \sum [j_1 = j_3 = \cdots = j_m = 0] c_j e_j(X_1, \dots, X_m)$. Thus

$$\begin{aligned} g_1(X_1) &= h_1(X_1) = \sum [j_2 = \cdots = j_m = 0] c_j e_j(X_1, \dots, X_m) \\ g_2(X_1, X_2) &= h_2(X_1, X_2) - h_1(X_1) - h_1(X_2) \\ &= \sum [j_1 \neq 0, j_2 \neq 0, j_3 = \cdots = j_m = 0] c_j e_j(X_1, \dots, X_m) \end{aligned}$$

and in general,

$$\begin{aligned} g_k(X_1, \dots, X_k) &= \sum [j_1 \neq 0, \dots, j_k \neq 0, j_{k+1} = \cdots = j_m = 0] c_j e_j(X_1, \dots, X_m) \\ g_k(X_{i_1}, \dots, X_{i_k}) &= \sum [j_{i_1} \neq 0, \dots, j_{i_k} \neq 0, \text{other } j_i\text{'s} = 0] c_j e_j(X_1, \dots, X_m). \end{aligned} \quad (3.10)$$

It is useful to observe that in the basis expansion of h , g_k comprises precisely the k -fold products, i.e., those in which the multi-index $j = (j_1, \dots, j_m)$ has exactly k non-zero entries. For later purposes, we introduce the notation $z(j)$ and $nz(j)$ for the number of zero and non-zero components, respectively, in j .

Proof of Theorem 3.1. We verify (3.5) first. The tensor product approach is convenient. Taking, without loss of generality, $I = \{1, 2, \dots, k\}$, we have a typical term in $g_k(X_1, \dots, X_k)$ to be $e_{j_1}(X_1)e_{j_2}(X_2) \cdots e_{j_k}(X_k)$, where $j_1 \neq 0, j_2 \neq 0, \dots, j_k \neq 0$. If, for instance, $1 \notin I'$, then $E[e_{j_1}(X_1)e_{j_2}(X_2) \cdots e_{j_k}(X_k) | X_{I'}] = E[e_{j_1}(X_1)] \cdot E[e_{j_2}(X_2) \cdots e_{j_k}(X_k) | X_{I'}] = 0$ and (3.5) follows by linearity. For (3.6), observe that if $I = \{i_1, \dots, i_k\}$ and $I' = \{i'_1, \dots, i'_k\}$, then $\text{Cov}[g_{|I|}(X_I), g_{|I'|}(X_{I'})]$ contains terms of the form $E[e_{j_1}(X_{i_1}) \cdots e_{j_k}(X_{i_k})e_{j'_1}(X_{i'_1}) \cdots e_{j'_k}(X_{i'_k})]$, which vanishes unless pairings exhaust the product into squares $e_j^2(X_i)$, in which case the expectation is unity.

To verify the uniqueness of the decomposition, suppose that h is expressed as in (3.2) and (3.7). Applying $E[\cdot | X_{\{1, \dots, k\}}]$ to both representations and invoking (3.4) yields

$$\begin{aligned} & \sum [j_{k+1} = 0, \dots, j_m = 0] c_j e_j(X_1, \dots, X_m) \\ &= \sum [\phi \neq J \subseteq \{1, \dots, k\}] E[g_{|J|}(X_J | X_{\{1, \dots, k\}})] \\ &= \sum [\phi \neq J \subseteq \{1, \dots, k\}] g_{|J|}(X_J). \end{aligned}$$

Setting $k = 1, 2, \dots$ provides the determining expansions (3.10).

Remarks. (1) A particular aspect of the decomposition was emphasized by Hájek [5] and has come to be known as the *Hájek projection*. One may ask, what is the best approximation to $h(X_1, \dots, X_m)$ (in the L_2 sense) of the form $f(X_1) + \cdots + f(X_m)$? The answer is apparent from what we have said and has $f(X_1) = E[h(X_1, \dots, X_m) | X_1] = h_1(X_1)$.

(2) For the case $m = 2$, the basis $e_0, e_1, e_2, \dots$ can be taken to be the eigenfunctions of the Hilbert-Schmidt operator induced by $h(x_1, x_2)$ (e.g., Gregory [4], Vitale [15], Serfling [12]). This specification is not needed in the discussion to follow.

(3) For an alternate representation related to (3.1), see Karlin and Rinott [9, Theorem 2.1].

4. COVARIANCES AND REPRESENTATIONS

From the symmetry of h it is evident that the covariance of $h(X_I)$ and $h(X_J)$ depends on I and J only through $|I \cap J|$. Accordingly, we introduce the main focus of our investigation.

DEFINITION. The covariance sequence $\{r_k\}_{k=0}^m$ associated with h is given by $r_k = \text{Cov}\{h(X_I), h(X_J)\}$, where $|I \cap J| = k$.

Note that $r_0 = 0$ and $r_m = \text{Var } h$. It is central to the theory that the intermediate values are highly structured. We begin with a result that implies their nonnegativity.

THEOREM 4.1. For $k = 1, \dots, m$, $r_k = \text{Var } h_k = (1/\binom{m}{k}) \sum [j] \binom{z(j)}{m-k} c_j^2$.

Remark. Traditionally r_k is defined to be $\text{Var } h_k$ (i.e., a "conditional" variance) and then shown to be a covariance (Hoeffding [7], Karlin and Rinott [9]). Our point of view is in line with certain applications to which we turn later.

Proof. By definition, $r_k = \text{Cov}[h(X_1, \dots, X_k, Y_1, \dots, Y_{m-k}), h(X_1, \dots, X_k, Z_1, \dots, Z_{m-k})]$, where $X_1, X_2, \dots, X_k, Y_1, \dots, Y_{m-k}, Z_1, \dots, Z_{m-k}$ are iid. Applying (3.7) yields $r_k = \sum [j, j'] c_j c_{j'} E[e_j(X_1, \dots, X_k, Y_1, \dots, Y_{m-k}) e_{j'}(X_1, \dots, X_k, Z_1, \dots, Z_{m-k})]$. The expectation is one or zero according to whether $j_i = j'_i = 0$ for all i between $k+1$ and m inclusive. Thus $r_k = \sum [j_{k+1} = \dots = j_m = 0] c_j^2$. By the symmetry of c_j in j , we also have $r_k = \sum [j_1 = j_2 = \dots = j_{m-k} = 0] c_j^2$ and in fact a version for each permutation of j . Averaging these versions yields the symmetrized expression

$$r_k = \frac{1}{\binom{m}{k}} \sum [j] \binom{z(j)}{m-k} c_j^2. \quad (4.1)$$

Finally, from the orthonormal expansion (3.9) we have $\text{Var } h_k = \sum [j_{k+1} = \dots = j_m = 0] c_j^2$.

Parallel to the variance sequence for the h_k 's, we have $\{v_k = \text{Var } g_k\}$, $k = 1, \dots, m$.

THEOREM 4.2. $v_k = (1/\binom{m}{k}) \sum [z(j) = k] c_j^2$.

Proof. From (3.10), we have $\text{Var } g_k(X_1, \dots, X_k) = \sum [j_1 \neq 0, \dots, j_k \neq 0, j_{k+1} = \dots = j_m = 0] c_j^2$, which symmetrizes to the asserted quantity.

The two sequences contain exactly the same information, as shown in the following inversion formulas.

THEOREM 4.3 (Hoeffding [7, Lemma 5.1]).

$$r_k = \sum_{l=1}^k \binom{k}{l} v_l, \quad k = 1, \dots, m, \quad (4.2)$$

and

$$v_k = \sum_{l=1}^k \binom{k}{l} (-1)^{k-l} r_l, \quad k = 1, \dots, m. \quad (4.3)$$

Proof. From the orthogonal decomposition of h_k , we have $r_k = \text{Var } h_k(X_I) = \sum [\phi \neq J \subseteq I] v_{|J|} = \sum_{l=1}^k \binom{k}{l} v_l$. Then (4.3) is a standard inversion, which can also be verified directly.

COROLLARY. $\sum_{l=1}^k \binom{k}{l} (-1)^{k-l} r_l \geq 0$, $k = 1, \dots, m$.

We note in passing that the property $\sum_{l=1}^{m-s} \binom{k}{l} (-1)^{k-l} r_{l+s} \geq 0$, $k = 1, \dots, m$, $s = 1, \dots, m-1$ (Karlin and Rinott [9]) can be derived from the corollary.

Owing to the corollary, it is clear that a covariance sequence cannot be arbitrary. However, it is worth emphasizing that the v sequence *can* be arbitrarily constituted so long as the values are nonnegative. This can be seen clearly in the tensor product approach and Theorem 4.2; the terms of (3.7) merely have to be arranged to yield the desired values for the v_k . This observation, an independence principle, immediately provides a characterization of the entire set of covariance sequences. In a sense, the collection is relatively small.

THEOREM 4.4. *The set of all covariance sequences which arise as h and the distribution of the X_i 's (but not m) are allowed to vary is precisely the positive linear hull of the vectors $r^{(1)}, r^{(2)}, \dots, r^{(m)}$, where $r_k^{(l)} = \binom{k}{l} / \binom{m}{l}$, $k = 1, \dots, m$. The vector $r^{(l)}$ corresponds to the situation $h = g_l$, $v_l = 1$.*

COROLLARY. *The set of all sequences with normalization $r_m = 1$ (i.e., correlation sequences) is the $m-1$ dimensional simplex with vertices $r^{(l)}$, $l = 1, \dots, m$. Such a sequence is expressible as a convex combination $r = \sum_{l=1}^m a_l r^{(l)}$, $a_l \geq 0$, $\sum a_l = 1$.*

Remark. These results show that to verify certain general properties of covariance sequences it suffices to consider the vertex sequences $r^{(l)}$, $l = 1, \dots, m$ (see, for instance, Theorems 5.2 and 6.2 below).

We conclude this section by observing that covariance sequences are nicely behaved in that they are nonnegative definite. This conforms to our interpretation of r_k as a covariance, but a slight twist is needed.

Given $r_1, \dots, r_m$ ($r_0 = 0$) define the bi-infinite sequence $\hat{r}$ via $\hat{r}_k = r_{m-k}$ for $k = 0, \dots, m$, $\hat{r}_k = 0$ for $k > m$, and $\hat{r}_k = \hat{r}_{-k}$ for $k < 0$.

THEOREM 4.5. *$\hat{r}$ is nonnegative definite, and consequently there is a bounded and absolutely continuous measure μ on the circle C such that $\hat{r}_k = \int_C e^{ik\theta} \mu(d\theta)$ for all k and thus $r_k = \int_C e^{i(m-k)\theta} \mu(d\theta)$, $k = 0, \dots, m$.*

Proof. We use an indirect time series argument. Let $\dots, X_{-2}, X_{-1}, X_0, X_1, X_2, \dots$ be an iid sequence and define $Y_j = h(X_j, X_{j+1}, \dots, X_{j+m-1})$,

$-\infty < j < \infty$. Then the Y sequence is wide-sense stationary with $\hat{r}_k = \text{Cov}\{Y_j, Y_{j+k}\}$. Consequently the representation holds (with μ absolutely continuous) by virtue of $\hat{r}_k = 0$ for $|k| > m$. The representation for r_k follows by substitution.

5. INCREASE IN COVARIANCE

Intuition suggests that r_k should be non-decreasing in k , and this can be easily established. Hoeffding [7] showed the stronger result that

$$\frac{r_k}{k} \text{ is non-decreasing in } k, \quad (5.1)$$

which can be read as an increase in covariance *per shared argument*. In view of the fact that (5.1) is key to the bias of jackknife estimate of variance (and other inequalities), it is worthwhile to re-visit it. We give a further strengthening on the basis of two new parameters.

DEFINITIONS. The *index* of h is $\min\{k | v_k > 0\}$. The *dual index* of h is $\max\{k | v_k > 0\}$.

The index and dual index thus bracket those k such that there are non-trivial k -fold products in the expansion (3.7).

THEOREM 5.1. *Let h be as above. Then its index $c = \min\{k | v_k > 0\}$ is equivalently $\min\{k | r_k > 0\}$ and $r_c = v_c$. However, generally the dual index $c' = \max\{k | v_k > 0\} \neq \max\{k | r_k > 0\}$.*

Proof. From (4.2), $r_k = \sum_{l=1}^k \binom{k}{l} v_l$, $k = 1, \dots, m$, and consequently index c implies $r_1 = r_2 = \dots = r_{c-1} = 0$ and by inspection $r_c = v_c > 0$. Whatever the dual index c' , which may range from 1 to m , $\text{Var } h > 0$ means $r_m > 0$ and hence $\max\{k | r_k > 0\} = m$ always.

Now we turn to the refinement of (5.1).

THEOREM 5.2. *Let h have index c and dual index c' . Then*

$$\frac{r_k}{\binom{k}{c}} \text{ (interpreted as 0 for } k < c \text{) is nondecreasing in } k. \quad (5.2)$$

Conversely, the condition $0 < r_k/k = r_{k'}/k'$ occurs either for no $k < k'$ or for all such pairs, the latter being equivalent to $c = c' = 1$.

Proof. By Theorem 4.4, it is enough to establish (5.2) for each $r^{(l)}$ with $c \leq l$. Thus, we need that, for $c \leq l \leq k$, $[(\binom{k}{l})/(\binom{m}{l})]/(\binom{k}{c})$ is nondecreasing in k , and this can be verified directly. For the next part, observe that the condition implies $0 < r_k^{(c)}/k = r_{k'}^{(c)}/k'$, i.e., $0 < [(\binom{k}{c})/(\binom{m}{c})]/k = [(\binom{k'}{c})/(\binom{m}{c})]/k'$, which is equivalent to $0 < (k-1)!/(k-c)! = (k'-1)!/(k'-c)!$. This requires that $c' = 1$. Conversely, if $c' = c = 1$, r is some multiple of $r^{(1)}$; recall that $r_k^{(1)} = k/m$, and hence r_k/k is independent of k .

For an alternate sharpening of (5.1), see Karlin and Rinott [9, Theorem D].

6. U -STATISTICS

The sample variance based on $X_1, \dots, X_n$ can be written as $\sigma^2 = (1/\binom{n}{2}) \sum [i < j] (X_i - X_j)^2/2$. Hoeffding [7] observed that many statistics possess a similar structure and can be brought under a single definition.

DEFINITION. The U -statistic based on the symmetric kernel h is $U_n = (1/\binom{n}{m}) \sum [I] = m] h(X_I)$.

As a symmetric function itself, U_n can be treated by the foregoing analyses of decomposition and variance.

THEOREM 6.1 (Hoeffding [7, p. 213]).

$$\text{Var } U_n = \frac{1}{\binom{n}{m}} \sum_{k=1}^m \binom{m}{k} \binom{n-m}{m-k} r_k. \quad (6.1)$$

Further, if h has decomposition (3.2), then U_n can be written as a sum of orthogonal U -statistics,

$$U_n = \sum_{k=1}^m \binom{m}{k} U_{kn}, \quad (6.2)$$

where $U_{kn} = (1/\binom{n}{k}) \sum [I] = k] g_k(X_I)$. It follows that $\text{Var } U_{kn} = v_k/\binom{n}{k}$ and that (6.1) has the alternate form

$$\text{Var } U_n = \sum_{k=1}^m \binom{m}{k}^2 v_k / \binom{n}{k}. \quad (6.3)$$

Proof. Equation (6.1) can be computed directly by observing that the number of m subsets of $\{1, \dots, n\}$ which share exactly k elements with a fixed m -set is $\binom{m}{k} \binom{n-m}{m-k}$. For (6.2), we insert the decomposition for h to

get $U_n = (1/\binom{n}{m}) \sum [|I| = m] \sum [\phi \neq J \subseteq I] g_{|J|}(X_J)$. The number of times $g_k(X_1, \dots, X_k)$ occurs in the sum is exactly the number of m -sets which contain $\{1, \dots, k\}$, i.e., $\binom{n-k}{m-k}$. Hence, $U_n = (1/\binom{n}{m}) \sum_{k=1}^m \binom{n-k}{m-k} \sum [|J| = k] g_k(X_J)$, which can be re-written as (6.2). The orthogonality of the U_{kn} and (6.3) follow directly from the orthogonality of the g_k 's.

The behavior of $\text{Var } U_n$ as $n \rightarrow \infty$ is of interest and Hoeffding [7, Theorem 5.2] established the following features:

$$\frac{m^2}{n} r_1 \leq \text{Var } U_n \leq \frac{m}{n} r_m; \quad (6.4)$$

$$n \text{ Var } U_n \text{ is decreasing in } n, \text{ starting with value } mr_m \text{ at } n=m \text{ and tending to } m^2 r_1 \text{ as } n \rightarrow \infty; \quad (6.5)$$

if h is of index c , then (6.4) may be replaced by

$$\frac{m}{c} K_n(m, c) r_c \leq \text{Var } U_n \leq K_n(m, c) r_m, \quad (6.6)$$

$$\text{where } K_n(m, c) = \frac{1}{\binom{n}{m}} \sum_{k=c}^m \binom{m-1}{k-1} \binom{n-m}{m-k}.$$

In view of the theoretical and practical importance of $\text{Var } U_n$, we turn to a refinement of (6.4) beyond (6.6) by introducing dual index information. This will be useful where, for example, the index (strictly) exceeds one so that $m^2 r_1 = 0$ in (6.5) is not informative.

THEOREM 6.2. *Suppose that h has index c and dual index c' . Then*

$$\frac{\binom{m}{c}^2}{\binom{n}{c}} r_c + \frac{\binom{m}{c'}}{\binom{n}{c'}} \left[r_m - \binom{m}{c} r_c \right] \leq \text{Var } U_n \leq \frac{\binom{m}{c}^2}{\binom{n}{c}} r_c + \frac{\binom{m}{c''}}{\binom{n}{c''}} \left[r_m - \binom{m}{c} r_c \right], \quad (6.7)$$

where $c'' = \min\{c', c+1\}$. Both bounds are tight in the sense that for each there is a single U -statistic yielding equality for all n . Further, if it is merely known that $c \leq \text{index}(h) \leq \text{dual index}(h) \leq c'$, then the bounds hold as well.

Proof. Recall that $\text{Var } U_n = \sum_{k=1}^m ((\binom{m}{k})^2 / \binom{n}{k}) v_k = \sum_{k=c}^{c'} ((\binom{m}{k})^2 / \binom{n}{k}) v_k$ and $r_k = \sum_{l=c}^{c'} \binom{k}{l} v_l$, $k=1, \dots, m$, which implies that $r_c = v_c$ and $r_m = \sum_{l=c}^{c'} \binom{m}{l} v_l$.

Case 1. $c = c'$. Equation (6.7) is easily seen to hold with equality throughout.

Case 2. $c < c'$. The bounds can be derived as extreme values of $\text{Var } U_n$ when r_c and r_m are specified. Set $b_k = \binom{m}{k} v_k$, $k = c, \dots, c'$, which provides

$$\text{Var } U_n = \sum_{k=c}^{c'} \frac{\binom{m}{k}}{\binom{n}{k}} b_k. \quad (6.8)$$

This is to be minimized under the constraints $b_c = \binom{m}{c} v_c = \binom{m}{c} r_c$ and $b_c + b_{c+1} + \dots + b_{c'} = \sum_{k=c}^{c'} \binom{m}{k} v_k = r_m$. The linear programming solution can be read off directly: place as much "b" mass as possible at $l = c'$. Thus $b_c = \binom{m}{c} v_c$, $b_{c'} = r_m - b_c$, $b_l = 0$, $c < l < c'$. Inserted into (6.8), these values yield the lower bound in (6.7). In the other direction, it is best to place as much b mass in low indices, and this yields $b_{c''} = r_m - b_c$, $b_l = 0$, $l \neq c, c''$, which provides the upper bound (6.7). The same arguments establish the last statement of the theorem.

Remarks. (1) As mentioned, the bounds are tight. Necessary and sufficient conditions can be extracted from the proof. The lower bound is attained iff $v_c, v_{c'} \neq 0$ and $v_l = 0$, $l \neq c, c'$. The upper bound is attained iff $v_c, v_{c''} \neq 0$ and $v_l = 0$, $l \neq c, c''$.

(2) The lower bound in (6.7) is sharper than those of both (6.4) and (6.6). It is an interesting exercise to see the slack. Using only the first term of (6.7) and comparing it to (6.6) yields the assertion

$$\frac{m}{c} \frac{1}{\binom{n}{m}} \sum_{k=c}^m \binom{m-1}{k-1} \binom{n-m}{m-k} \leq \frac{(c/m) \binom{n}{m} \binom{m}{c}^2}{\binom{n}{c}}.$$

Setting $k' = k - 1$, $m' = m - 1$, and $n' = n - 1$ and simplifying expressions provides

$$\sum_{k'=c-1}^{m'} \binom{m'}{k'} \binom{n'-m'}{m'-k'} \leq \binom{m'}{c-1} \binom{n'-(c-1)}{m'-(c-1)}.$$

This can be argued as follows: consider a set of n' objects and a fixed subset S of size m' . The LHS counts the number of subsets of size m' which share at least $c - 1$ elements with S . The RHS overcounts the same collection of subsets by counting the number of ways of forming subsets of size m' by first choosing $c - 1$ elements from S and then choosing $m' - (c - 1)$ other elements arbitrarily (in or out of S).

(3) The upper bound in (6.7) is sharper than that of (6.6). Suppose that $c < m$ ($c = m$ can be done directly). Then

$$\begin{aligned} & \frac{\binom{m}{c}^2}{\binom{n}{c}} r_c + \frac{\binom{m}{c+1}}{\binom{n}{c+1}} \left[r_m - \binom{m}{c} r_c \right] \\ & \leq \frac{\binom{m}{c}^2}{\binom{n}{c}} r_c + \frac{\binom{m}{c}}{\binom{n}{c}} \left[r_m - \binom{m}{c} r_c \right] = \frac{\binom{m}{c}}{\binom{n}{c}} r_m \leq \left(\frac{m}{n} \right) r_m. \end{aligned}$$

(4) Finally we observe that the upper and lower bounds (6.7) are of the same form, their second terms bearing rates $n^{-c'}$ and $n^{-c''}$, respectively.

We conclude by giving a refinement of (6.4) and showing that $\{\text{Var } U_n\}$ has a property of complete monotonicity.

THEOREM 6.3. *Suppose that h has index c and dual index c' . Then $\binom{n}{c} \text{Var } U_n$ is a decreasing sequence with value $\binom{m}{c} r_m$ at $n=m$ and limit $\binom{m}{c}^2 r_c$ as $n \rightarrow \infty$. Moreover, $\binom{n}{c} \text{Var } U_n$ is a completely monotonic sequence. More precisely, letting $n = m + p$,*

$$\binom{m+p}{c} \text{Var } U_{m+p} = \binom{m}{c} \left[\binom{m}{c} r_c + \int_0^\infty e^{-px} w_h(x) dx \right], \quad p \geq 0, \quad (6.9)$$

where $w_h(x) = \sum_{k=c}^{c'} \binom{m-c}{k-c-1} \binom{m}{k} (m-k+1) v_k e^{-(m-k)x} (e^x - 1)^{k-c-1}$ is a non-negative function which decreases exponentially to zero like $e^{-(m+1-c')x}$.

Proof. The first statement follows from (6.8). Observe that (6.9) is direct for $c' = c$ with $w_h \equiv 0$. For $c' > c$, write

$$\begin{aligned} \binom{m+p}{c} \text{Var } U_{m+p} &= \binom{m+p}{c} \sum_{k=c}^{c'} \frac{\binom{m}{k}^2}{\binom{m+p}{k}} v_k \\ &= \binom{m}{c}^2 v_c + \binom{m}{c} \sum_{k=c+1}^{c'} \binom{m}{k} \\ &\quad \times v_k \frac{(m-c) \cdots (m-k+1)}{(m+p-c) \cdots (m+p-k+1)}. \end{aligned}$$

By using the identity $(m-c) \cdots (m-k+1)/(m+p-c) \cdots (m+p-k+1) = \int_0^\infty e^{-px} (u_{m-c} * \cdots * u_{m-k+1})(x) dx$ in which $*$ denotes convolution and $u_a(x) = 0$ for $x < 0$, $= ae^{-ax}$ for $0 \leq x$, the representation (6.9) follows from

superposition and recalling that $r_c = v_c$. The remaining properties of w_h can be checked directly.

ACKNOWLEDGMENTS

This work was supported in part by NSF Grants DMS-8603944 and DMS-9002665 and ONR Grant N00014-90-J-1641. An invitation to present an earlier version (Vitale [17]), and kind hospitality, were extended by the organizers of the Miniconference on Multiple Stochastic Integration, Polynomial Chaos, and Their Applications (Case Western Reserve University and Cleveland State University, 1985).

REFERENCES

- [1] BHARGAVA, R. P. (1983). A property of the jackknife estimation of the variance when more than one observation is omitted. *Sankhyā Ser. A* **45** 112–119.
- [2] DYNKIN, E. B., AND MANDELBAUM, A. (1983). Symmetric statistics, Poisson point processes, and multiple Wiener integrals. *Ann. Statist.* **11** 739–745.
- [3] EFRON, B., AND STEIN, C. (1981). The jackknife estimate of variance. *Ann. Statist.* **9** 586–596.
- [4] GREGORY, G. G. (1977). Large sample theory for U -statistics and tests of fit. *Ann. Statist.* **5** 110–123.
- [5] HÁJEK, J. (1968). Asymptotic normality of simple linear rank statistics under alternatives. *Ann. Math. Statist.* **39** 325–346.
- [6] HALMOS, P. R. (1946). The theory of unbiased estimation. *Ann. Math. Statist.* **17** 34–43.
- [7] Hoeffding, W. (1948). A class of statistics with asymptotically normal distribution. *Ann. Math. Statist.* **19** 293–325.
- [8] Hoeffding, W. (1961). The strong law of large numbers for U -statistics. Mimeograph Series 302, Institute of Statistics, University of North Carolina.
- [9] KARLIN, S., AND RINOTT, Y. (1982). Applications of ANOVA type decompositions for comparisons of conditional variance statistics including jackknife estimates. *Ann. Statist.* **10** 485–501.
- [10] MANDELBAUM, A., AND TAQQU, M. S. (1984). Invariance principle for symmetric statistics. *Ann. Statist.* **12** 483–496.
- [11] RUBIN, H., AND VITALE, R. A. (1980). Asymptotic distribution of symmetric statistics. *Ann. Statist.* **8** 165–170.
- [12] SERFLING, R. J. (1980). *Approximation Theorems of Mathematical Statistics*. Wiley, New York.
- [13] STEELE, J. M. (1986). An Efron–Stein inequality for nonsymmetric statistics. *Ann. Statist.* **14** 753–758.
- [14] TAKEMURA, A. (1983). Tensor analysis of ANOVA decomposition. *J. Amer. Statist. Assoc.* **78** 894–900.
- [15] VITALE, R. A. (1977). Asymptotic area and perimeter of sums of random convex sets. Technical Summary Report No. 1770, Mathematics Research Center, University of Wisconsin, Madison, WI.
- [16] VITALE, R. A. (1984). An expansion for symmetric statistics and the Efron–Stein inequality. In *Inequalities in Statistics and Probability* (Y. L. Tong, Ed.), IMS Lecture Notes and Monograph Series, Vol. 5, pp. 112–114. Hayward, CA.
- [17] VITALE, R. A. (1985). *Notes on Symmetric Statistics*. Technical Report, Claremont Graduate School.