The Annals of Statistics
1982, Vol. 10, No. 2, 485–501

APPLICATIONS OF ANOVA TYPE DECOMPOSITIONS FOR COMPARISONS OF CONDITIONAL VARIANCE STATISTICS INCLUDING JACKKNIFE ESTIMATES[1]

BY SAMUEL KARLIN AND YOSEF RINOTT

Stanford University; Stanford University and the Hebrew University of Jerusalem

Variance and bias comparisons are set forth for a sequence of nonlinear statistics based on independent random variables including jackknife and U-statistics of various orders. The analysis relies heavily on an orthogonal decomposition first introduced by Hoeffding in 1948. This ANOVA type decomposition is refined for purposes of discerning higher order convexity properties for an array of conditional variance coefficients. There is also some discussion of two-sample statistics.

1. Introduction, statements of results, and interpretations. Let $X_1, X_2, \cdots, X_n (n = p + k)$ be i.i.d. real random variables. Many natural statistics based on such observations derive from a symmetric function $\varphi(x_1, x_2, \cdots, x_p)$ (i.e., φ is invariant under all permutations of its arguments) including the U-statistics of Hoeffding and the class of Von Mises' statistics. Assume henceforth that $E[\{\varphi(X_1, \cdots, X_p)\}^2] < \infty$. In the course of studies on jackknifing and bootstrapping methodologies, Efron and Stein (1981) (for $n = p + 1$) compared the variance of $\varphi(X_1, X_2, \cdots, X_p)$ with the corresponding expected sample jackknife variance. They discovered the pervasive inequality

$$\operatorname{Var}\{\varphi(X_1, \cdots, X_p)\} \le E\{\textstyle\sum_{i=1}^{p+1} (\varphi_i - \bar{\varphi})^2\}, \tag{1.1}$$

where $\varphi_i = \varphi(X_1, \cdots, X_{i-1}, X_{i+1}, \cdots, X_{p+1})$ and $\bar{\varphi} = 1/(p+1) \sum_{i=1}^{p+1} \varphi_i$ with equality in force when $\varphi(x_1, \cdots, x_p)$ is linear.

The above inequality was the initial impetus for our present study.

The proof of Efron and Stein uses propitiously an ANOVA-type orthogonal decomposition, introduced by Hoeffding (1948, 1961), in studies pertaining to central limit theorems for U-statistics; see Serfling (1980, Chapters 5–7) for an excellent coverage of these developments and further references, and Rubin and Vitale (1980).

This orthogonal representation is of the form

$$\varphi(X_1, \cdots, X_p) = H_0 + \textstyle\sum_{i=1}^{p} H_1(X_i) + \sum_{i<j} H_2(X_i, X_j) + \sum_{i<j<k} H_3(X_i, X_j, X_k) + \cdots, \tag{1.2a}$$

where H_0 is constant and H_0, $H_1(X_i)$, $H_2(X_i, X_j)$, $\cdots$, $H_r(X_{i_1}, X_{i_2}, \cdots, X_{i_r})$, $\cdots$ are all mutually orthogonal with respect to the product measure induced by $(X_1, \cdots, X_p)$. In the nonsymmetric case when X_i are independent but not necessarily identically distributed and φ is not necessarily symmetric, the corresponding representation takes the form

$$\varphi(X_1, \cdots, X_p) = H_0 + \textstyle\sum_{i=1}^{p} H_i(X_i) + \sum_{i<j} H_{ij}(X_i, X_j) + \sum_{i<j<k} H_{ijk}(X_i, X_j, X_k) + \cdots \tag{1.2b}$$

so that there occur generally p one-variable functions $H_i(X_i)$, $p(p-1)/2$ functions $H_{ij}(X_i, X_j)$, etc. In the symmetric i.i.d. case there occur in (1.2) at most $p + 1$ distinct

Received March 1981; revised October 1981.

[1] Supported in part by NIH Grant 5R01 GM10452-18 and NSF Grant MCS79-24310.

AMS 1970 *subject classification.* Primary 62G05.

Key words and phrases. Orthogonal decomposition, jackknife statistics, bias comparisons, absolute monotonicity.

functions, while in the nonsymmetric case there appear up to 2^p distinct functions. Hajek (1968) applied the orthogonal decomposition for a variety of cases based on nonsymmetric statistics $\varphi(\mathbf{X})$, generally of linear rank forms and some classes of Von Mises statistics depending on the empirical distribution functions. The central limit theorem analysis in most cases falls back on the Hajek projection construction highlighting the one-variable term $\sum_{i=1}^{p} H_i(X_i)$. When $\sum_{i=1}^{p} H_i(X_i) = 0$ the limit theorems focus on the second order terms $\sum_{i<j} H_{ij}(X_i, X_j)$ when present, and involve suitable mixtures of Chi squared random variables.

The orthogonal decomposition (1.2) is fundamental for our purposes as well. We will therefore review its structure (Sections 2 and 3) in a generalized format. In dealing with random functions of the form $S(X_1, \cdots, X_p; Y_1, \cdots, Y_q)$ involving two separate groups of i.i.d. random variables (as occurs, for example, in the classical Wilcoxon two-sample statistics) we will require direct product versions of the orthogonal decomposition.

In order to provide perspective to our later presentation, we next give a direct proof of (1.1). Expansion of the right-hand sum in (1.1), under the condition that the X_i are i.i.d. and φ is symmetric, reduces the inequality (1.1) to the convexity relationship $c_p - c_0 \le pc_p - pc_{p-1}$, or $c_{p-1} \le (1/p)c_0 + (p-1)/pc_p$, where by definition

$$(1.3) \qquad c_r = E[\{E(\varphi(X_1, \cdots, X_p) \mid X_1, \cdots, X_r)\}^2], \quad r = 0, 1, 2, \cdots, p.$$

Starting from the identity

$$c_{r+1} - c_r = E([E\{\varphi(X_1, \cdots, X_p) \mid X_1, \cdots, X_{r+1}\} - E\{\varphi(X_1, \cdots, X_p) \mid X_1, \cdots, X_r\}]^2),$$

and now applying Schwarz' inequality with respect to integration on X_1 relying on the symmetry and i.i.d. assumptions, yields that $c_{r+1} - c_r \ge c_r - c_{r-1}$ for each $r = 1, 2, \cdots, p$. Thus, $\{c_r\}$ forms a convex sequence and (1.1) follows.

Actually, the sequence $\{c_r\}$ enjoys the following higher order convexity endowment proved in Section 2:

$$(1.4) \qquad \Delta^k c_r = \sum_{i=0}^{k} \binom{k}{i} (-1)^{i+k} c_{r+i} \ge 0, \quad r = 0, 1, \cdots, p-k.$$

Here, $\Delta^1 c_r = c_{r+1} - c_r$ determines the first order difference as indicated and $\Delta^k c_r = \Delta(\Delta^{k-1} c_r)$ computes the kth successive difference operation. Although c_r is defined only for $r = 0, 1, \cdots, p$, with the aid of the ANOVA decomposition formula (see (2.16)) there exists a natural extension defined for all integers $r = 0, 1, 2, \cdots$ coincident to c_r when $r = 0, 1, 2, \cdots, p$ and satisfying (1.4) for all r. With this extension $\{c_r\}_0^\infty$ forms an absolutely monotone sequence (cf. Widder, 1946, Chapter IV).

In our elaboration of the ANOVA-type orthogonal decomposition for general $X_1, \cdots, X_p$ independent but not identically distributed, we consider the generalized conditional variance functional

$$(1.5) \qquad C_\eta = E\{[E\{\varphi(\mathbf{X}) \mid \mathbf{X}_\eta\}]^2\}$$

defined for each $\boldsymbol{\eta} = (\eta_1, \cdots, \eta_p)$, $\eta_i = 0$ or 1, where the inner expectation is taken with respect to the components corresponding to $\eta_i = 0$ conditioned on the variables X_{η_i} corresponding to $\eta_i = 1$ and the outer expectation is evaluated over the random variables with $\eta_i = 1$.

In the symmetric case, i.e., $\{X_i\}$ i.i.d., C_η reduces to c_r (for $\Sigma\eta_i = r$) as defined in (1.3). In Section 2 we establish the basic property that C_η presents an additive nonnegative set function defined effectively over a Boolean algebra of sets. In fact (Theorem 2.1), the representation $C_\eta = \Sigma_{\varepsilon \le \eta} a_\varepsilon$ holds, where the summation is over vectors $\boldsymbol{\varepsilon} = (\varepsilon_1, \cdots, \varepsilon_p)$, $\varepsilon_i = 0$ or 1 satisfying $\varepsilon_i \le \eta_i$, $i = 1, \cdots, p$ and $a_\varepsilon \ge 0$ for every $\boldsymbol{\varepsilon}$. Thus C_η is represented (up to a constant) as a multivariate distribution function. It will be shown that the inclusion-exclusion identity applied to this distribution encompasses the absolute monotonicity of (1.4). These results rely decisively on the ANOVA decomposition and some direct product extensions.

Extending the Efron-Stein jackknife inequality, Bhargava (1980) ascertained the analog of (1.1) corresponding to the procedure of removing groups of k, rather than one, of the observations. We will refer to this construction as the kth-order jackknife variance estimator for the statistic $\varphi(X_1, \cdots, X_p)$. More specifically, consider a sample $X_1, \cdots, X_n$ $(n = p + k, p$ fixed) of i.i.d. real random variables and form

$$\bar{\varphi}^{(k)} = \frac{1}{\binom{p+k}{p}} \sum_{\mathscr{D}} \varphi(X_{i_1}, \cdots, X_{i_p}), \tag{1.6}$$

where $\mathscr{D}$ is the set of all $\binom{p+k}{p}$ choices of p distinct indices from $\{1, \cdots, n\}$. For $\mathbf{i} = (i_1, \cdots, i_p) \in \mathscr{D}$, we denote

$$\mathbf{X_i} = (X_{i_1}, \cdots, X_{i_p}).$$

In this notation (1.6) becomes

$$\bar{\varphi}^{(k)} = \frac{1}{\binom{p+k}{p}} \sum_{\mathbf{i}\in\mathscr{D}} \varphi(\mathbf{X_i}). \tag{1.7}$$

The kth-order jackknife variance estimator based on φ is defined as

$$V^{(k)}(\varphi) = \frac{1}{\binom{p+k-1}{p}} \sum_{\mathbf{i}\in\mathscr{D}} [\varphi(\mathbf{X_i}) - \bar{\varphi}^{(k)}]^2. \tag{1.8}$$

The normalization $\binom{p+k-1}{p}$ maintains the equation

$$E\{V^{(k)}(L)\} = \operatorname{Var}\{L(X_1, \cdots, X_p)\},$$

for linear functions such as $L(\mathbf{x}) = \varphi(x_1, \cdots, x_p) = \sum_{i=1}^{p} w(x_i)$, so that $V^{(k)}(L)$ is an unbiased estimator in this case.

Bhargava (1980) proved

$$\operatorname{Var}\{\varphi(X_1, \cdots, X_p)\} \le E\{V^{(k)}(\varphi)\}. \tag{1.9}$$

In this context, we will establish the finer relationship described in the following theorem.

THEOREM A. *Let $X_1, \cdots, X_n, n = p + k$, be* i.i.d. *random variables and $\varphi(x_1, \cdots, x_p)$ a real symmetric function of p variables. Let $V^{(k)}(\varphi)$ be the k*th-*order jackknife variance estimator defined in* (1.8). *$V^{(k)}(\varphi)$ is unbiased for a linear projection statistic of the form $\varphi(\mathbf{x}) = \sum_{i=1}^{p} w(x_i)$. Let*

$$b^{(k)}(\varphi) = E\{V^{(k)}(\varphi)\} - \operatorname{Var}\{\varphi(X_1, \cdots, X_p)\} \tag{1.10}$$

*be the bias of the k*th-*order jackknife variance estimator of φ. There exists a positive continuous decreasing function $g(\xi)$ defined for $\xi \ge 0$ such that*

$$b^{(k)}(\varphi) = \int_0^\infty e^{-k\xi} g(\xi)\, d\xi, \qquad k = 1, 2, 3, \cdots. \tag{1.11}$$

*Thus, the bias of the k*th-*order jackknife variance estimate of φ is a completely monotone sequence in k. In particular, $E\{V^{(k)}(\varphi)\}$ is decreasing convexly at an algebraic rate to* Var *$\varphi(X_1, \cdots, X_p)$ as k increases to ∞.*

It is worth emphasizing the fact that the bias sequence $b^{(k)}$ is decreasing, convex, with negative third order divided differences, etc.; in fact, they form a moment sequence such that

$$b^{(k)}(\varphi) = \int_0^1 u^k g(-\log u)\frac{du}{u} = \int_0^1 u^k h(u)\, du,$$

where $h(u)$ is positive and integrable. Moreover, $b^{(k)}(\varphi)$ can be embedded as part of a continuous completely monotone function $b^{(t)}(\varphi) = \int_0^\infty e^{-t\xi} g(\xi)\, d\xi$. On the other hand, the conditional variance sequence $\{c_k\}$ of (1.3) is increasing convex, with positive third order difference; in fact, generating an absolutely monotone sequence which cannot, in general, be embedded as part of an absolutely monotone function. The intrinsic regularity properties of $\{b^{(k)}(\varphi)\}$ contrast with those of $\{c_k\}$ reflecting the differences between completely monotone sequences and absolutely monotone sequences; consult Widder (1946) for more discussion of this topic.

Theorem A applies when X_i are i.i.d. vector valued, such that $\varphi(\mathbf{x}_1, \cdots, \mathbf{x}_p)$ is a symmetric function of p vector-variables. For example, $\mathbf{X}_i$, $i = 1, \cdots, p$ can be independent bivariate observations and $\varphi(\mathbf{X}_1, \cdots, \mathbf{X}_p)$ can be taken to be the correlation coefficient based on the p pairs.

The orderings on $b^{(k)}(\varphi)$ of Theorem A involves a single function of p arguments, while the sample size $\{X_1, \cdots, X_{p+k}\}$ increases as k increases.

We consider next successive jackknife variance estimates for a sequence of statistics based on the symmetric functions $\varphi_1(x_1)$, $\varphi_2(x_1, x_2)$, $\cdots$, $\varphi_p(x_1, x_2, \cdots, x_p)$ defined as follows:

$$\varphi_r(x_1, \cdots, x_r) = E\{\varphi_p(x_1, \cdots, x_r, X_{r+1}, \cdots, X_p)\}, \tag{1.12}$$
$$r = 1, 2, \cdots, p-1, \qquad \varphi_p = \varphi,$$

where the expectation is taken with respect to $(X_{r+1}, \cdots, X_p)$. Clearly $\varphi_r(x_1, \cdots, x_r) = E\{\varphi_{r+1}(x_1, \cdots, x_r, X_{r+1})\}$. In this setting, φ_r performs like a marginal function induced by φ_{r+1}.

In the following variance comparisons, the total sample size $n = k + r$ is held constant so that k varies along with r. The expected kth-order jackknife variance estimator for the rth statistic φ_r is (cf. (1.8))

$$V^{(k)}(\varphi_r) = \frac{1}{\binom{n-1}{r}} \sum_{\mathbf{i}\in\mathscr{D}_r} \{\varphi_r(\mathbf{X_i}) - \bar{\varphi}_r\}^2, \qquad r = 1, \cdots, p, \qquad k = n - r, \tag{1.13}$$

where

$$\bar{\varphi}_r = \frac{1}{\binom{n}{r}} \sum_{\mathbf{i}\in\mathscr{D}_r} \varphi_r(\mathbf{X_i}),$$

and $\mathscr{D}_r$ consists of all r-tuples of indices of the form $\mathbf{i} = (i_1, \cdots, i_r)$, $1 \le i_1 < \cdots < i_r \le n$.

The elementary ℓth degree symmetric function of p variables

$$\varphi_p(x_1, \cdots, x_p) = S_\ell(x_1, \cdots, x_p) = \sum_{1\le i_1<\cdots<i_\ell\le p} \left(\prod_{\nu=1}^{\ell} x_{i_\nu}\right)$$

under the operation (1.12) passes into $\varphi_r(x_1, \cdots, x_r) = S_\ell(x_1, \cdots, x_r)$ provided $EX_i = 0$. When $EX_i = a$, $\varphi_p(x_1, \cdots, x_p) = S_\ell(x_1, \cdots, x_p)$ entails $\varphi_{p-1}(x_1, \cdots, x_{p-1}) = S_\ell(x_1, \cdots, x_{p-1}) + aS_{\ell-1}(x_1, \cdots, x_{p-1})$.

The following theorem provides comparisons between the kth-order expected jackknife variance estimators for the succession of statistics (1.12) with fixed sample size.

THEOREM B. *Let* $X_1, \cdots, X_n$ *be a sample of* i.i.d. *random variables and consider the sequence of symmetric functions* $\varphi_r(x_1, \cdots, x_r)$, *defined in* (1.12). *Define the bias of the* kth-*order jackknife variance estimate based on* φ_r *to be*

$$B_r^{(k)} = E\{V^{(k)}(\varphi_r)\} - \operatorname{Var}\{\varphi_r(X_1, \cdots, X_r)\}, \quad r = 1, \cdots, p, k = n - r. \tag{1.14}$$

Then

$$B_r^{(k)} \le B_{r+1}^{(k-1)}, \quad r = 1, \cdots, p - 1. \tag{1.15}$$

In heuristic terms, the above inequality asserts that the jackknife bias grows as more of the variables of the proposed statistics "interact."

It is important to emphasize that the variance inequalities of (1.15) refer to r and $k = n - r$ varying involving a constant sample size $n = r + k$, whereas in the comparisons of Theorem A, p is held fixed and the order of jackknifing k increases together with the sample size $n = p + k$.

Consider next a two-sample statistic $\varphi(x_1, \cdots, x_p; y_1, \cdots, y_q)$ invariant with respect to permutations of the x's and y's separately. Let $X_1, \cdots, X_n$ and $Y_1, \cdots, Y_m$, $n > p$, $m > q$ be two independent samples, each consisting of i.i.d. random variables. We construct the jackknife estimate based on $\varphi(\mathbf{x}; \mathbf{y})$ of order $n - p$ in the x's and $m - q$ in the y's analogous to (1.8) and inquire as to the validity of Theorem A in this two-sample context. Perhaps surprising is that the universal inequality parallel to (1.9) does *not* hold. It will be shown that the analog of inequality (1.9) carries over to the two-sample case *only* under the restriction $p/n = q/m$, i.e., when the proportion of the number of the x and y variables in the function φ to the corresponding sample sizes coincide. For the case $p/n = q/m$, we prescribe

$$K = K(p, q, n, m) = \left\{\left(1 - \frac{p}{n}\right)\binom{n}{p}\binom{m}{q}\right\}^{-1}$$

and determine

$$V(\varphi) = K \textstyle\sum_{i_1<\cdots<i_p, j_1<\cdots<j_q}\{\varphi(X_{i_1}, \cdots, X_{i_p}; Y_{j_1}, \cdots, Y_{j_q}) - \bar{\varphi}\}^2, \tag{1.16}$$

where

$$\bar{\varphi} = \frac{1}{\binom{n}{p}\binom{m}{q}} \textstyle\sum_{i_1<\cdots<i_p, j_1<\cdots<j_q} \varphi(X_{i_1}, \cdots, X_{i_p}; Y_{j_1}, \cdots, Y_{j_q}).$$

Consider the class of linear symmetric functions $L_{\alpha,\beta}(\mathbf{x}, \mathbf{y}) = \alpha \sum_{i=1}^{p} u(x_i) + \beta \sum_{j=1}^{q} v(y_j)$. It is elementary to check that $E\{V(L_{\alpha,\beta})\} = \operatorname{Var}(L_{\alpha,\beta})$ independently of α and β if and only if $p/n = q/m$.

THEOREM C. *If* $p/n = q/m$, *then for any symmetric function* $\varphi(x_1, \cdots, x_p; y_1, \cdots, y_q)$

$$E\{V(\varphi)\} \ge \operatorname{Var}\{\varphi(\mathbf{X}, \mathbf{Y})\}. \tag{1.17}$$

For $p/n \ne q/m$ *with any prescribed normalizing constant there exists no inequality of the form* (1.17) *valid for all two-sample statistics.*

Another class of two-sample extensions of (1.1), where the sequences of random variables $\{X_i\}_1^{p+1}$ and $\{Y_j\}_1^{q+1}$ are each i.i.d, and the X's and Y's are mutually independent, is based on the additive jackknife estimator for the variance of $\varphi(X_1, \cdots, X_p; Y_1, \cdots, Y_q)$ of the form

$$J_{\mathbf{x},\mathbf{y}}^{(1)} = \frac{1}{q+1} \textstyle\sum_{j=1}^{q+1} \sum_{i=1}^{p+1} (\varphi_{\hat{x}_i,\hat{y}_j} - \bar{\varphi}_{\mathbf{x},\hat{y}_j})^2 + \frac{1}{p+1} \sum_{i=1}^{p+1} \sum_{j=1}^{q+1} (\varphi_{\hat{x}_i,\hat{y}_j} - \bar{\varphi}_{\hat{x}_i,\mathbf{y}})^2, \tag{1.18}$$

where

$$\varphi_{\hat{x}_i, \hat{y}_j} = \varphi(X_1, \cdots, X_{i-1}, X_{i+1}, \cdots, X_{p+1}; Y_1, \cdots, Y_{j-1} Y_{j+1}, \cdots, Y_{q+1}),$$

(1.19)

$$\bar{\varphi}_{\mathbf{x}, \hat{y}_j} = \frac{1}{p+1} \sum_{i=1}^{p+1} \varphi_{\hat{x}_i, \hat{y}_j}, \quad \bar{\varphi}_{\hat{x}_i, \mathbf{y}} = \frac{1}{q+1} \sum_{j=1}^{q+1} \varphi_{\hat{x}_i, \hat{y}_j}.$$

A two-sample additive kth-order jackknife estimator $J^{(k)}_{\mathbf{x},\mathbf{y}}$ involving the factors (1.8) separately with respect to the $\mathbf{X}$'s and $\mathbf{Y}$'s averaged over the variables of the other set can be formed paraphrasing (1.18).

THEOREM C′. *Let $\varphi(\mathbf{x}, \mathbf{y})$, $\{X_i\}_1^{k+p}$ and $\{Y_j\}_1^{k+q}$ be as in Theorem* C, *but with no restrictions on k, p, and q. Then the two-sample jackknife estimator $J^{(k)}_{\mathbf{x},\mathbf{y}}$ satisfies*

$$\operatorname{Var}\{\varphi(X_1, \cdots, X_p; Y_1, \cdots, Y_p)\} \le E\{J^{(k)}_{\mathbf{x},\mathbf{y}}\}. \tag{1.20}$$

Theorems C and C′ can be extended to functions of multiple samples $\varphi(\mathbf{x}^{(1)}, \cdots, \mathbf{x}^{(\lambda)})$, where $\mathbf{x}^{(j)}$ has p_j components with corresponding sample size n_j, $j = 1, \cdots, \lambda$.

The representation (1.11) of Theorem A can be extended to the multiple sample case when $p_1 = \cdots = p_\lambda$ and $n_1 = \cdots = n_\lambda$. In this case, a representation of the bias analogous to (1.11) holds with $k = n_j - p_j$ and an appropriate function g.

In Section 4 we describe the ANOVA decomposition in the two-sample case, and some corresponding results for conditional variance terms. For example, consider

$$c_{r,s} = E\{[E\{\varphi(X_1, \cdots, X_p, Y_1, \cdots, Y_q) \mid X_1, \cdots, X_r, Y_1, \cdots, Y_s\}]^2\}.$$

The approach taken in the one-sample case leads here to a variety of inequalities. We feature the following multidimensional convexity-type inequality $c_{r+1,s+1} + c_{r,s} \ge c_{r,s+1} + c_{r+1,s}$.

The U-statistics of Hoeffding are constructed as symmetrized expressions of the form

$$\varphi(x_1, \cdots, x_n) = \binom{n}{q}^{-1} \sum_{\mathbf{i}} \psi(x_{i_1}, \cdots, x_{i_q}), \tag{1.21}$$

where ψ is a symmetric real-valued function of q variables, and the sum extends over all q-tuples, $\mathbf{i} = (i_1, \cdots, i_q)$, $1 \le i_1 < \cdots < i_q \le n$. In the course of establishing asymptotic limit theorems for the statistics

$$Z_n = \varphi(X_1, \cdots, X_n), \qquad n = q, q+1, \cdots,$$

Hoeffding (1948) achieved the ordering

$$n \operatorname{Var} Z_n \ge (n+1) \operatorname{Var} Z_{n+1}. \tag{1.22}$$

His proof uses the property (assuming for convenience $c_0 = 0$) that c_r/r (c_r defined in (1.3)) is decreasing as r increases which is in turn a direct consequence of the property that c_r describes a convex sequence. By exploiting more fully the higher order convexity endowments in (1.4), we derive a hierarchy of variance inequalities.

THEOREM D. *For φ as defined in* (1.21) *we have*

$$(-1)^r \Delta^r \left\{ \binom{n}{r} \operatorname{Var} Z_n \right\} \ge 0, \qquad n > 0, \qquad r > 0. \tag{1.23}$$

where Δ^r is the rth difference operator defined following (1.4) *applied with respect to n.*

Inequality (1.22) is the special case of (1.23) with $r = 1$.

It should be noted that the bulk of studies on U-statistics pertains to asymptotics (e.g., central limit theorems, law of the iterated logarithm) as sample size grows to infinity. The present paper concentrates on exact variance comparisons for varying sample sizes.

We conclude this introduction by indicating the organization of the remainder of the paper. Section 2 presents the details of the ANOVA-orthogonal decomposition in a general setting for $\varphi(X_1, \cdots, X_p)$ where X_i are independent, not necessarily identically distributed, and φ is a general function of p variables. The proof of (1.4) and ramifications are presented in Section 2. Several examples are examined and interpreted in this setting (Section 3). In Section 4, we single out the corresponding ANOVA-orthogonal decomposition for the case of $\varphi(X_1, \cdots, X_p; Y_1, \cdots, Y_q)$ where $\{X_i\}$ and $\{Y_j\}$ are separately i.i.d. Section 5 is devoted to the proofs of Theorems A through D.

2. Generalized ANOVA-type orthogonal decomposition. For purposes of the analysis of Theorems A through D, stated in Section 1, we rederive (1.2) in a more convenient format. In our approach the construction of the H-component functions appear as Fourier transforms with respect to a suitable direct-product discrete group.

Let $\boldsymbol{\varepsilon} = (\varepsilon_1, \cdots, \varepsilon_p)$ be a p-vector with $\varepsilon_i = 0$ or 1. Similarly, $\boldsymbol{\delta} = (\delta_1, \cdots, \delta_p)$ is a p-vector having $\delta_i = 0$ or 1. We use the special symbols $\mathbf{1} = (1, \cdots, 1)$ and $\mathbf{0} = (0, \cdots, 0)$. The ordering $\boldsymbol{\delta} \le \boldsymbol{\varepsilon}$ signifies that $\delta_i \le \varepsilon_i$ for all i. The number of nonzero components of the vector $\boldsymbol{\varepsilon}$ is denoted by $|\boldsymbol{\varepsilon}| = \sum_{i=1}^{p} \varepsilon_i$.

We will use the traditional inner product notation $(\boldsymbol{\delta}, \boldsymbol{\varepsilon}) = \sum_{i=1}^{p} \delta_i \varepsilon_i$. For each $\boldsymbol{\varepsilon}$ we define

$$\lambda(\boldsymbol{\varepsilon}) = (\lambda_1(\boldsymbol{\varepsilon}), \cdots, \lambda_s(\boldsymbol{\varepsilon})), \quad s = |\boldsymbol{\varepsilon}|, \tag{2.1}$$

to be the set of indices of unit components of $\boldsymbol{\varepsilon}$; e.g., $\lambda(0, 1, 1, 0) = (2, 3)$. We designate $\mathbf{X}_{\lambda(\varepsilon)} = (X_{\lambda_1(\varepsilon)}, \cdots, X_{\lambda_s(\varepsilon)})$.

We adopt the convention and notation

$$\tilde{\mathbf{X}}_{\delta_i} = (\tilde{X}_{\delta_1}, \cdots, \tilde{X}_{\delta_p}); \qquad \tilde{X}_{\delta_i} = \begin{cases} X_i \text{ when } \delta_i = 0 \\ x_i \text{ when } \delta_i = 1 \end{cases} \tag{2.2}$$

that is, $\tilde{X}_{\delta_i}$ is the random variable X_i when $\delta_i = 0$, while $\tilde{X}_{\delta_i}$ is the deterministic variable x_i for $\delta_i = 1$. Then

$$E\{\varphi(\tilde{\mathbf{X}}_\delta)\} = E\{\varphi(\tilde{X}_{\delta_1}, \tilde{X}_{\delta_2}, \cdots, \tilde{X}_{\delta_p})\} \tag{2.3}$$

is the expectation taken with respect to those components having index values 0, whereas the remaining components take prescribed values. Specifically, if $|\boldsymbol{\delta}| = s$, $E\{\varphi(\tilde{\mathbf{X}}_\delta)\}$ is a function of $\mathbf{x}_{\lambda(\delta)} = (x_{\lambda_1(\delta)}, \cdots, x_{\lambda_s(\delta)})$.

For any given $\boldsymbol{\varepsilon}$ with $|\boldsymbol{\varepsilon}| = r$, we define

$$H_\varepsilon(\mathbf{x}_{\lambda(\varepsilon)}) = H_\varepsilon(x_{\lambda_1(\varepsilon)}, \cdots, x_{\lambda_r(\varepsilon)}) = (-1)^{|\varepsilon|} \textstyle\sum_{\delta:\delta \le \varepsilon} (-1)^{(\delta,\varepsilon)} E\{\varphi(\tilde{\mathbf{X}}_\delta)\}, \tag{2.4}$$

where the sum extends over all $\boldsymbol{\delta}$ satisfying $\boldsymbol{\delta} \le \boldsymbol{\varepsilon}$. Obviously, $H_0 = E\{\varphi(X_1, \cdots, X_p)\}$.

It is illuminating to exhibit some concrete cases of (2.4). Indeed, let $\boldsymbol{\varepsilon}_{(i)} = (0, \cdots, 0, 1, 0, \cdots, 0)$ with a single unit value in coordinate i and zeros otherwise. Analogously we define $\boldsymbol{\varepsilon}_{(i,j)} = (0, \cdots, 0, 1, 0, \cdots, 0, 1, 0, \cdots, 0)$, the $\boldsymbol{\varepsilon}$ vector with two unit components in places i and j. From the definition (2.4), we have

$$H_{\varepsilon_{(i)}}(x_i) = E\{\varphi(X_1, \cdots, X_{i-1}, x_i, X_{i+1}, \cdots, X_p)\} - E\{\varphi(X_1, \cdots, X_p)\} \tag{2.5a}$$

and

$$\begin{aligned} H_{\varepsilon_{(i,j)}}(x_i, x_j) &= E\{\varphi(X_1, \cdots, X_{i-1}, x_i, X_{i+1}, \cdots, X_{j-1}, x_j, X_{j+1}, \cdots, X_p)\} \\ &\quad - E\{\varphi(X_1, \cdots, X_{i-1}, x_i, X_{i+1}, \cdots, X_p)\} \\ &\quad - E\{\varphi(X_1, \cdots, X_{j-1}, x_j, X_{j+1}, \cdots, X_p)\} \\ &\quad + E\{\varphi(X_1, \cdots, X_p)\}. \end{aligned} \tag{2.5b}$$

We claim the following orthogonality relations.

PROPOSITION 2.1.

$$E\{H_\varepsilon(\mathbf{X}_{\lambda(\varepsilon)})H_{\tilde{\varepsilon}}(\mathbf{X}_{\lambda(\tilde{\varepsilon})})\} = 0 \quad \textit{provided} \quad \boldsymbol{\varepsilon} \neq \tilde{\boldsymbol{\varepsilon}}. \tag{2.6}$$

PROOF. Consider first $\boldsymbol{\varepsilon} = (1, \varepsilon_2, \varepsilon_3, \cdots, \varepsilon_p) = (1, \boldsymbol{\varepsilon}^{(2)})$ where $\boldsymbol{\varepsilon}^{(2)} = (\varepsilon_2, \cdots, \varepsilon_p)$. From the definition (2.4), we have ($r = |\boldsymbol{\varepsilon}|$)

$$\begin{aligned} E\{H_\varepsilon(X_1, x_{\lambda_2(\varepsilon)}, \cdots, x_{\lambda_r(\varepsilon)})\} &= (-1)^{|\varepsilon|} \textstyle\sum_{\delta \le \varepsilon, \delta = (0, \delta^{(2)})} (-1)^{(\delta^{(2)}, \varepsilon^{(2)})} E\{\varphi(X_1, \tilde{\mathbf{X}}_{\delta^{(2)}})\} \\ &\quad + (-1)^{|\varepsilon|} \textstyle\sum_{\delta \le \varepsilon, \delta = (1, \delta^{(2)})} (-1)^{\delta_1}(-1)^{(\delta^{(2)}, \varepsilon^{(2)})} E\{\varphi(X_1, \tilde{\mathbf{X}}_{\delta^{(2)}})\} = 0, \end{aligned} \tag{2.7}$$

where the equality to zero results since the second sum terms are, respectively, negatives of the terms of the first sum. In (2.7) we assumed $\lambda_1(\boldsymbol{\varepsilon}) = 1$ and computed the expectation of H_ε with respect to X_1. Clearly, the expectation of H_ε with respect to any of the variables $X_{\lambda_1(\varepsilon)}, X_{\lambda_2(\varepsilon)}, \cdots, X_{\lambda_r(\varepsilon)}$ vanishes by the same argument.

In order to prove the orthogonality property, we may assume without loss of generality, since $\boldsymbol{\varepsilon} \neq \tilde{\boldsymbol{\varepsilon}}$, that $\boldsymbol{\varepsilon} = (1, \boldsymbol{\varepsilon}^{(2)})$ while $\tilde{\boldsymbol{\varepsilon}} = (0, \tilde{\boldsymbol{\varepsilon}}^{(2)})$. Taking expectations in (2.6) with respect to X_1 first, the expectation vanishes since $H_{\tilde{\varepsilon}}$ is constant with respect to X_1 and (2.7) shows that the expectation of H_ε with respect to X_1 vanishes.

Now completing the expectations with respect to the remaining variables yields

$$E\{H_\varepsilon(\mathbf{X}_{\lambda(\varepsilon)})H_{\tilde{\varepsilon}}(\mathbf{X}_{\lambda(\tilde{\varepsilon})})\} = 0.$$

We next validate the representation formula (1.2) in its general setting, that is, when φ is not necessarily symmetric and the X_i are independent but not necessarily identically distributed.

PROPOSITION 2.2. *Let $X_1, \cdots, X_p$ be independent random variables and let $\varphi(\mathbf{x}) = \varphi(x_1, \cdots, x_p)$ be a function satisfying $E\{\varphi^2(\mathbf{X})\} < \infty$. Then the following representation holds*

$$\varphi(x_1, \cdots, x_p) = \textstyle\sum_\varepsilon H_\varepsilon(x_{\lambda(\varepsilon)}) = \sum_\varepsilon (-1)^{|\varepsilon|} \sum_{\delta \le \varepsilon} (-1)^{(\delta, \varepsilon)} E\{\varphi(\tilde{\mathbf{X}}_\delta)\}. \tag{2.8}$$

PROOF. Take $\boldsymbol{\delta}$, for definiteness, to be $\boldsymbol{\delta} = (1, \cdots, 1, 0, \cdots, 0)$ with k initial unit components, $k < p$ and all zero components subsequently. We collect all the terms from the double sum (2.8) involving $E\{\varphi(\tilde{\mathbf{X}}_\delta)\}$ for this $\boldsymbol{\delta}$. These include all $\boldsymbol{\varepsilon}$ of the form $(1, 1, \cdots, 1, \varepsilon_{k+1}, \cdots, \varepsilon_p)$, $\varepsilon_{k+1}, \cdots, \varepsilon_p$ independently 0 and 1; this $\boldsymbol{\varepsilon}$ is associated with the sign whose exponent is $|\boldsymbol{\varepsilon}| + (\boldsymbol{\delta}, \boldsymbol{\varepsilon}) = 2k + \sum_{i=k+1}^p \varepsilon_i$. The coefficient of $E\{\varphi(\tilde{\mathbf{X}}_\delta)\}$ is then

$$\textstyle\sum_{(\varepsilon_{k+1}, \cdots, \varepsilon_p)} (-1)^{\sum_{i=k+1}^p \varepsilon_i} = 0.$$

Accordingly, there remains a single term in the double sum of (2.8)corresponding to $\delta = (1, \cdots, 1)$, namely $E\{\varphi(\tilde{\mathbf{X}}_\delta)\} = \varphi(x_1, \cdots, x_p)$ and (2.8) is proved.

An important formula which plays a vital role in our later analyses involves the following array of conditional variance coefficients.

DEFINITION. For any prescribed $\boldsymbol{\eta} = (\eta_1, \cdots, \eta_p)$, $\eta_i = 0$ or 1,

$$C_\eta = E\{[E\{\varphi(\mathbf{X}) \mid \mathbf{X}_{\lambda(\eta)}\}]^2\}, \tag{2.9}$$

where the inner expectation applies to the variables of indices where $\eta_i = 0$ and the second expectation extends over the conditional variables corresponding to the indices where $\eta_i = 1$.

For example, in the case $\boldsymbol{\eta} = (1, \cdots, 1, 0, \cdots, 0)$ with k initial unit values followed by zeros, we have

$$C_\eta = \int \left\{ \int \varphi(x_1, \cdots, x_k, x_{k+1}, \cdots, x_p)\, dF_{k+1}(x_{k+1}) \cdots dF_p(x_p) \right\}^2 \quad (2.10)$$
$$dF(x_1) \cdots dF(x_k).$$

THEOREM 2.1. *For any η, we have under the conditions of Proposition 2.2 that*

$$C_\eta = \textstyle\sum_{\varepsilon \le \eta} a_\varepsilon \quad \textit{where} \quad a_\varepsilon = E[\{H_\varepsilon(\mathbf{X}_{\lambda(\varepsilon)})\}^2]. \quad (2.11)$$

The above sum runs over all ε satisfying $\varepsilon \le \eta$.

PROOF. To ease the exposition of the proof we take $\eta = (1, \cdots, 1, 0, \cdots, 0)$ with an initial segment of k ones. Now substitute for $\varphi(\mathbf{X})$ its orthogonal representation (2.8) $\varphi(x_1, \cdots, x_p) = \sum_\varepsilon H_\varepsilon(x_{\lambda(\varepsilon)})$, and consider the expectation with respect to $X_{k+1}, \cdots, X_p$:

$$E\{\varphi(\tilde{\mathbf{X}}_\eta)\} = E\{\varphi(x_1, \cdots, x_k, X_{k+1}, \cdots, X_p)\}. \quad (2.12)$$

If $\varepsilon_i = 1$ for some $i \ge k + 1$, then the variable X_i will appear in H_ε and in this case, by (2.7), $E\{H_\varepsilon\}$ vanishes. Accordingly, the expectation in (2.12) reduces to

$$E\{\varphi(\tilde{\mathbf{X}}_\eta)\} = \hat{\varphi}(x_1, \cdots, x_k) = E\{\varphi(x_1, \cdots, x_k, X_{k+1}, \cdots, X_p)\} = \textstyle\sum_{\varepsilon \le \eta} H_\varepsilon(\mathbf{x}_{\lambda(\varepsilon)}). \quad (2.13)$$

Consider now $C_\eta = E[\{\hat{\varphi}(X_1, X_2, \cdots, X_k)\}^2]$. Using the orthogonality relationship confirmed in Proposition 2.1 leads to the formula (2.11) for the specified η. An analogous argument applies to any η. The proof of Theorem 2.1 is complete.

The equation (2.13) can be construed as the inverse Fourier transform of (2.4).

The following corollary specializes (2.11) to the symmetric case, i.e., for $\varphi(x_1, \cdots, x_p)$ symmetric and $\{X_i\}$ i.i.d. In this case, each $H_\varepsilon(\mathbf{x}_{\lambda(\varepsilon)})$, $|\varepsilon| = r$, reduces to a common symmetric function of r variables which we write as $H_r(\mathbf{x}_{\lambda(\varepsilon)})$. Moreover, because the X_i are i.i.d.,

$$E[\{H_r(\mathbf{X}_{\lambda(\varepsilon)})\}^2] = a_r \quad (2.14)$$

independent of ε, provided $|\varepsilon| = r$. If $|\eta| = t$, the number of those ε obeying $\varepsilon \le \eta$, with $|\varepsilon| = \nu$ is $\binom{t}{\nu}$.

COROLLARY 2.1. *For $\varphi(x_1, \cdots, x_p)$ symmetric and $\{X_i\}$* i.i.d., *we have for $|\eta| = t$,*

$$C_\eta = c_t = \textstyle\sum_{\nu=0}^{p} \binom{t}{\nu} a_\nu, \quad t = 0, 1, \cdots, p, \quad (2.15)$$

where c_t is defined in (1.3).

With the aid of this formula, there exists a natural extension of c_t, $t = 0, 1, \cdots, p$ defined for all integers, viz.,

$$c_t = \textstyle\sum_{\nu=0}^{p} \binom{t}{\nu} a_\nu, \quad t = 0, 1, 2, \cdots \quad (2.16)$$

where $\binom{t}{\nu} = 0$ for $\nu > t$.

The relation in (2.15) can be inverted to the expression $a_\nu = \sum_{t=0}^{p} (-1)^{t+\nu} \binom{\nu}{t} c_t$; see Hoeffding (1948, Lemma 5.1).

By (2.11) we can view C_η as the distribution function $P(\mathbf{Y} \le \eta)$ where the random vector $\mathbf{Y}$ is defined by $P(\mathbf{Y} = \varepsilon) = a_\varepsilon$. It is convenient to introduce the lattice notation $\eta^1 \vee \eta^2$ and $\eta^1 \wedge \eta^2$ for the p-tuples of components with values $\max(\eta_i^1, \eta_i^2)$ and $\min(\eta_i^1, \eta_i^2)$, $i =$

1, $\cdots$, p, respectively. For intersections of events we have the equality

$$\{\mathbf{Y} \le \boldsymbol{\eta}^1\} \cap \{\mathbf{Y} \le \boldsymbol{\eta}^2\} \cap \cdots \cap \{\mathbf{Y} \le \boldsymbol{\eta}^k\} = \{\mathbf{Y} \le \boldsymbol{\eta}^1 \wedge \boldsymbol{\eta}^2 \wedge \cdots \wedge \boldsymbol{\eta}^k\}. \tag{2.17a}$$

However, with respect to unions of events, the containing relation (which usually is strict) prevails:

$$\{\mathbf{Y} \le \boldsymbol{\eta}^1\} \cup \{\mathbf{Y} \le \boldsymbol{\eta}^2\} \cup \cdots \cup \{\mathbf{Y} \le \boldsymbol{\eta}^k\} \subseteq \{\mathbf{Y} \le \boldsymbol{\eta}^1 \vee \boldsymbol{\eta}^2 \vee \cdots \vee \boldsymbol{\eta}^k\}. \tag{2.17b}$$

The inclusion-exclusion identity (Feller, 1950, Chapter 4.1) and (2.17) imply that

$$C_{\eta^1\vee\cdots\vee\eta^k} - \textstyle\sum_i C_{\eta^i} + \sum_{i<j} C_{\eta^i\wedge\eta^j} - \cdots + (-1)^k C_{\eta^1\wedge\cdots\wedge\eta^k} \ge 0. \tag{2.18}$$

Other set theoretic relationships lead to further inequalities satisfied by C_η. For example, by using the extension of the inclusion-exclusion identity to the probability that exactly m among the k events $\{\mathbf{Y} \le \boldsymbol{\eta}^i\}$ occur (Feller, 1950, Chapter 4.3) we obtain for

$$\boldsymbol{\eta} = \textstyle\bigvee_{1\le i_1<\cdots<i_m\le k} (\boldsymbol{\eta}^{i_1} \wedge \cdots \wedge \boldsymbol{\eta}^{i_m}),$$

using the notation $(\wedge_{\nu=1}^r \boldsymbol{\eta}^{i_\nu}) = \boldsymbol{\eta}^{i_1} \wedge \cdots \wedge \boldsymbol{\eta}^{i_r}$, that

$$C_\eta - \textstyle\sum_{i_1<\cdots<i_m} C_{(\wedge_{\nu=1}^m \eta^{i_\nu})} + \binom{m+1}{m} \sum_{i_1<\cdots<i_{m+1}} C_{(\wedge_{\nu=1}^{m+1} \eta^{i_\nu})}$$
$$- \binom{m+2}{m} \textstyle\sum_{i_1<\cdots<i_{m+2}} C_{(\wedge_{\nu=1}^{m+2} \eta^{i_\nu})} + \cdots \pm \binom{k}{m} C_{\eta^1\wedge\cdots\wedge\eta^k} \ge 0.$$

In the symmetric case, we apply (2.18) with the choice of $\boldsymbol{\eta}^i$ having unit values in its first $r + k$ components except for a zero in the $r + i$th place ($i = 1, \cdots, k$), and zeros elsewhere to obtain the inequality (1.4). This proves that c_t, $t = 0, 1, 2, \cdots$ constitutes an *absolutely monotone sequence* as asserted in the introduction.

The inequalities (1.4) can also be deduced directly from formula (2.16) using elementary manipulations with generating functions to get

$$\Delta^r c_t = \textstyle\sum_\nu \binom{t}{\nu - r} a_\nu. \tag{2.19}$$

The quantity is nonnegative since $a_\nu = E[\{H_\nu(\mathbf{X})\}^2]$ is plainly nonnegative.

For later purposes, it is worth recording that for a U-statistic of order m,

$$U(X_1, X_2, \cdots, X_n) = \textstyle\sum_{i_1<\cdots<i_m} \psi(X_{i_1}, X_{i_2}, \cdots, X_{i_m}) \tag{2.20}$$

based on n random variables $X_1, X_2, \cdots, X_n$, we have $H_\varepsilon(\mathbf{x}_{\lambda(\varepsilon)}) \equiv 0$ for all $\boldsymbol{\varepsilon}$ having $|\boldsymbol{\varepsilon}| > m$. In this case

$$c_t = \textstyle\sum_{\nu=0}^m \binom{t}{\nu} a_\nu, \quad t = 0, 1, 2, \cdots, n. \tag{2.21}$$

The terms of the ANOVA expansion can also be characterized as projections in the context of an L^2-norm best approximation procedure. To this end, let $\varphi(x_1, \cdots, x_p)$ be such that $\operatorname{Var}\{\varphi(X_1, \cdots, X_p)\} < \infty$. We pose the following problem. Determine functions $u_{\mathbf{i}}^*(\mathbf{x}_{\mathbf{i}})$ attaining

$$\min_{u_{\mathbf{i}}} E[\{\varphi(X_1, \cdots, X_p) - \textstyle\sum_{\mathbf{i}} u_{\mathbf{i}}(\mathbf{X}_{\mathbf{i}})\}^2] \tag{2.22}$$

where the sum extends over $\mathbf{i} = (i_1, \cdots, i_r)$, $1 \le i_1 < \cdots < i_r \le p$. It can be shown that the minimum is attained when $\sum_{\mathbf{i}} u_{\mathbf{i}}^*(\mathbf{X}_{\mathbf{i}}) = \sum_{\varepsilon:|\varepsilon|\le r} H_\varepsilon(\mathbf{x}_{\lambda(\varepsilon)})$.

Note that the right hand side consists of the terms of the decomposition (2.8) depending on at most r variables.

3. Some concrete examples of the ANOVA decomposition. In this section we determine explicitly the components of the ANOVA orthogonal decomposition for several classes of functions.

EXAMPLE 1. *Elementary symmetric polynomials.* The elementary symmetric polynomial of degree r has the form

$$\varphi(\mathbf{x}) = P_r(x_1, \cdots, x_n) = \textstyle\sum_{1\le i_1<\cdots<i_r\le n} (\prod_{\nu=1}^{r} x_{i_\nu}). \tag{3.1a}$$

We deal first with

$$\varphi(\mathbf{x}) = P_n(\mathbf{x}) = \textstyle\prod_{i=1}^{n} x_i. \tag{3.1b}$$

Let $X_1, X_2, \cdots, X_n$ be independent random variables with $E\{X_i\} = \mu_i$, and $\operatorname{Var} X_i = \sigma_i^2$. Take $\boldsymbol{\varepsilon} = (1, \cdots, 1, 0, \cdots, 0)$ with k initial unit components. Then the span of $\boldsymbol{\delta} \le \boldsymbol{\varepsilon}$ consists of all vectors such as $(\delta_1, \delta_2, \cdots, \delta_k, 0, \cdots, 0)$ having $\delta_i = 0$ or 1.

Abiding by the convention (2.2), we obtain for an admissible $\boldsymbol{\delta}$,

$$E\{\varphi(\tilde{\mathbf{X}}_\delta)\} = \textstyle\prod_{i=1}^{k} (x_i^{\delta_i}\mu_i^{1-\delta_i}) \prod_{\nu=k+1}^{n} \mu_\nu = \prod_{\nu=1}^{n} \mu_\nu \prod_{i=1}^{k} \left(\dfrac{x_i}{\mu_i}\right)^{\delta_i}.$$

Therefore, in the case at hand, by the definition (2.4), we have for $\boldsymbol{\varepsilon} = (1, \cdots, 1, 0, \cdots, 0)$ with $|\boldsymbol{\varepsilon}| = k$,

$$\begin{aligned} H_\varepsilon(x_{\lambda(\varepsilon)}) &= (-1)^k \textstyle\prod_{\nu=1}^{n} \mu_\nu \left\{\sum_{(\delta_1,\cdots,\delta_k)} (-1)^{|\delta|} \prod_{i=1}^{k} \left(\dfrac{x_i}{\mu_i}\right)^{\delta_i}\right\} \\ &= (-1)^k \textstyle\prod_{\nu=1}^{n} \mu_\nu \left\{\prod_{i=1}^{k} \left(1 - \dfrac{x_i}{\mu_i}\right)\right\} = \prod_{\nu=k+1}^{n} \mu_\nu \prod_{i=1}^{k} (x_i - \mu_i). \end{aligned}$$

Generally,

$$H_\varepsilon(\mathbf{x}_{\lambda(\varepsilon)}) = \textstyle\prod_{\nu=1}^{n-|\varepsilon|} \mu_{\lambda_\nu(\mathbf{1}-\varepsilon)} \prod_{i=1}^{|\varepsilon|} (x_{\lambda_i(\varepsilon)} - \mu_{\lambda_i(\varepsilon)}); \tag{3.2}$$

for the notation, see (2.1), and note that $\lambda_\nu(\mathbf{1} - \boldsymbol{\varepsilon})$ traverses the set of zero indices in $\boldsymbol{\varepsilon}$.

In particular, we record

$$E[\{H_\varepsilon(\mathbf{X}_{\lambda(\varepsilon)})\}^2] = \textstyle\prod_{\nu=1}^{n-|\varepsilon|} \mu^2_{\lambda_\nu(1-\varepsilon)} \prod_{i=1}^{|\varepsilon|} \sigma^2_{\lambda_i(\varepsilon)}, \tag{3.3a}$$

where σ_i^2 is the variance of X_i. When the X_i are i.i.d., then for $|\boldsymbol{\varepsilon}| = r$,

$$E[\{H_r(X_1, \cdots, X_r)\}^2] = \mu^{2(n-r)}\sigma^{2r}, \quad r = 0, 1, \cdots, n. \tag{3.3b}$$

We deal next with the rth ($r < n$) elementary symmetric function $P_r(\mathbf{x})$, see (3.1). To use the result of the calculation (3.2), we introduce the extra variables $Z_{r+1}, \cdots, Z_n$ such that

$$X_1 \cdot X_2 \cdots X_r = X_1 \cdot X_2 \cdots X_r Z_{r+1} \cdots Z_n \tag{3.4}$$

as the special case of the nth degree polynomial $P_n(\mathbf{x})$ where X_i, $i = 1, 2, \cdots r$ are distributed following $F_i(x)$ while Z_j, $j = r + 1, \cdots, n$ are degenerate random variables concentrating at 1.

Consider any $\boldsymbol{\varepsilon} = (\varepsilon_1, \cdots, \varepsilon_n)$. When $\varepsilon_j = 1$ for some $j > r$, we obtain $\sum_{\delta\le\varepsilon} (-1)^{(\delta,\varepsilon)} E\{P_n(\tilde{\mathbf{X}}_\delta)\} = 0$ because Z_j is degenerate and then the two terms corresponding to $(\delta_1, \cdots, \delta_{j-1}, 0, \delta_{j+1}, \cdots, \delta_n)$ and $(\delta_1, \cdots, \delta_{j-1}, 1, \delta_{j+1}, \cdots, \delta_n)$ add to zero for any specification of $(\delta_1, \cdots, \delta_{j-1}, \delta_{j+1}, \cdots, \delta_n)$. Thus for the function (3.4), the only possible nonzero components in its ANOVA orthogonal representation consist of the H_ε for $\boldsymbol{\varepsilon}$ of the form $\boldsymbol{\varepsilon} = (\varepsilon_1, \cdots, \varepsilon_r, 0, 0, \cdots, 0)$. The calculation hereafter reduces to that of the Example (3.1b) for the product function $X_1X_2 \cdots X_r$ of r factors rather than n. It follows for $\boldsymbol{\varepsilon} \le (1, 1, \cdots, 1, 0, \cdots, 0) = \boldsymbol{\xi}$, the latter vector with r ones, that

$$H_\varepsilon(\mathbf{x}_{\lambda(\varepsilon)}) = \textstyle\prod_{j=1}^{r-|\varepsilon|} \mu_{\lambda_j(\xi-\varepsilon)} \prod_{i=1}^{|\varepsilon|} (x_{\lambda_i(\varepsilon)} - \mu_{\lambda_i(\varepsilon)}).$$

The construction of H_ε is additive for sums of functions so that for $P_r(\mathbf{x})$, we arrive at the formula

$$H_\varepsilon^{(P_r)}(\mathbf{x}_{\lambda(\varepsilon)}) = \textstyle\sum_{\eta:\eta\geq\varepsilon,|\eta|=r} \prod_{j=1}^{r-|\varepsilon|} \mu_{\lambda_j(\eta-\varepsilon)} \prod_{i=1}^{|\varepsilon|} (x_{\lambda_i(\varepsilon)} - \mu_{\lambda_i(\varepsilon)}), \quad |\varepsilon| \leq r. \tag{3.5}$$

Moreover, we find that

$$H_\varepsilon^{(P_r)}(\mathbf{x}_{\lambda(\varepsilon)}) \equiv 0 \quad \text{for} \quad |\varepsilon| > r. \tag{3.6}$$

In the i.i.d. case with $k \leq r$, then

$$H_k^{(P_r)}(x_1, x_2, \cdots, x_k) = \binom{n-k}{r-k}\mu^{r-k} \textstyle\prod_{i=1}^{k} (x_i - \mu)$$

and a straightforward calculation yields

$$a_k = E[\{H_k^{(P_r)}(X_1, \cdots, X_k)\}^2] = \binom{n-k}{r-k}^2 \mu^{2(r-k)}\sigma^{2k}. \tag{3.7}$$

EXAMPLE 2. *Sample variance.* We record the familiar second order U-statistic example of the sample variance for ready reference; e.g., see Efron and Stein (1981) and Serfling (1980, page 173). Consider $S(X_1, X_2, \cdots, X_n) = n^{-1} \sum_{i=1}^{n} (X_i - \bar{X})^2$, $\bar{X} = n^{-1} \sum_{i=1}^{n} X_i$. Here

$$H_0 = \frac{n-1}{n}\sigma^2, \quad H_1(X_i) = \frac{n-1}{n^2}\{(X_i - \xi)^2 - \sigma^2\}, \tag{3.8}$$

$$\mathrm{H}_2(\mathrm{X}_i, X_j) = -\frac{2}{n^2}(X_1 - \xi)(X_2 - \xi),$$

with $\xi = E\{X_i\}$, $\sigma^2 = \operatorname{Var} X_i$. The coefficients $a_i = E\{H_i^2\}$ are

$$a_0 = \left(\frac{n-1}{n}\right)^2 \sigma^4, \quad a_1 = \frac{(n-1)^2}{n^4}(\gamma^4 - \sigma^4), \quad a_2 = \frac{4}{n^4}\sigma^4, \quad \gamma^4 = E\{(X - \xi)^4\}. \tag{3.9}$$

4. The ANOVA-type decomposition for multi-sets of sample statistics. In dealing with mixed functionals like the two-sample Wilcoxon statistic and others, we will need to ascertain the associated ANOVA orthogonal representation. To this end, consider a function of the form $\varphi(x_1, \cdots, x_p; y_1, \cdots, y_q)$ distinguishing two groups of variables. Let $\boldsymbol{\varepsilon} = (\varepsilon_1, \cdots, \varepsilon_p)$ be a p-tuple as before with each $\varepsilon_i = 0$ or 1 and $\boldsymbol{\beta} = (\beta_1, \cdots, \beta_q)$ an independent q-tuple with $\beta_i = 0$ or 1. Adapting the constructions of (2.4), we define

$$H_{\varepsilon,\beta}(\mathbf{x}_{\lambda(\varepsilon)}, \mathbf{y}_{\lambda(\beta)}) = (-1)^{|\varepsilon|+|\beta|} \textstyle\sum_{\delta\leq\varepsilon,\alpha\leq\beta} (-1)^{(\delta,\varepsilon)+(\alpha,\beta)} E\{\varphi(\tilde{\mathbf{X}}_\delta, \tilde{\mathbf{Y}}_\alpha)\},$$

where the expectations involve conditioning on the components of $\mathbf{X}$ and $\mathbf{Y}$, corresponding to the unit values of $\boldsymbol{\delta}$ and $\boldsymbol{\alpha}$, respectively. Paraphrasing the analysis of Section 2, we achieve the representation formula

$$\varphi(\mathbf{x}, \mathbf{y}) = \textstyle\sum_{\varepsilon,\beta} H_{\varepsilon,\beta}(\mathbf{x}_{\lambda(\varepsilon)}, \mathbf{y}_{\lambda(\beta)})$$

such that $H_{\varepsilon,\beta}$ and $H_{\tilde{\varepsilon},\tilde{\beta}}$ are mutually orthogonal if either $\boldsymbol{\varepsilon} \neq \tilde{\boldsymbol{\varepsilon}}$ and/or $\boldsymbol{\beta} \neq \tilde{\boldsymbol{\beta}}$. For any $\boldsymbol{\eta} = (\eta_1, \cdots, \eta_p)$ and $\boldsymbol{\xi} = (\xi_1, \xi_2, \cdots, \xi_q)$ paralleling (2.9), we consider the generalized conditional variance coefficients,

$$C_{\eta,\xi} = E[\{E(\varphi(\mathbf{X}, \mathbf{Y}) \mid \mathbf{X}_{\lambda(\eta)}, Y_{\lambda(\xi)})\}^2].$$

The following result obtains (cf. Theorem 2.1):

THEOREM 4.1. *If $X_1, \cdots, X_p$ and $Y_1, \cdots, Y_q$ are all independent and $E\{\varphi^2(\mathbf{X}, \mathbf{Y})\} < 0$, then the quantities $C_{\eta,\xi}$ provide an additive set function defined on the direct product* *of the Boolean algebras spanned by* η *and* ξ. *More explicitly, we have*

$$C_{\eta,\xi} = \sum_{\varepsilon\le\eta,\beta\le\xi} E[\{H_{\varepsilon,\beta}(X_{\lambda(\varepsilon)}, Y_{\lambda(\beta)})\}^2], \tag{4.1}$$

the summation encompassing all ε *and* β *obeying the indicated constraints.*

In the case that X_i are i.i.d., and Y_j separately i.i.d., X's and Y's independent, and φ symmetric under permutations of the p x's and the q y's separately, then $C_{\eta,\xi}$ depends only on $|\eta|, |\xi|$. With $|\eta| = r$, $|\xi| = s$, reduces to the expression

$$c_{r,s} = \sum_{\nu,\mu} \binom{r}{\nu}\binom{s}{\mu} a_{\nu\mu}, \tag{4.2}$$

with $a_{\nu,\mu} = E[\{H_{\nu,\mu}(X_1, \cdots, X_\nu, Y_1, \cdots, Y_\mu)\}^2] \ge 0$.

In particular, $c_{r,s}$ is absolutely monotone in each variable r and s separately. Moreover, $c_{r,s}$ determines an additive set function with respect to the pair of variables, in the manner

$$c_{r+1,s+1} + c_{r,s} \ge c_{r,s+1} + c_{r+1,s} \quad \text{for all integer } r \text{ and } s. \tag{4.3}$$

We display these formulas for the case of the two-sample Wilcoxon statistic

$$\varphi(X_1 \cdots X_p; Y_1 \cdots Y_q) = \#\{X_i < Y_j\}.$$

Let X_i follow the density $f(x)$ and Y_j that of $g(y)$. A direct computation produces

$$H_{0,0} = pq \int f(\xi)\{1 - G(\xi)\}\, d\xi,$$

$$H_{1,0}(x_i) = q[1 - G(x_i) - \int f(\xi)\{1 - G(\xi)\}\, d\xi],$$

$$H_{0,1}(y_j) = p[F(y_j) - \int f(\xi)\{1 - G(\xi)\}\, d\xi],$$

$$H_{1,1}(x_i, y_j) = I(x_i < y_j) - \{1 - G(x_i)\} - F(y_j) + \int f(\xi)\{1 - G(\xi)\, d\xi\},$$

where I is the indicator function taking the value 1 if $x_i < y_j$. Obviously, $H_{r,s} \equiv 0$ for $r \ge 2$ or $s \ge 2$.

5. Proofs of Theorems A Through D. Statements of Theorems A through D are given in Section 1.

PROOF OF THEOREM A. For easy reference, we recall the notation. Let $\varphi(x_1, \cdots, x_p)$ be a symmetric function of p variables and $\{X_1, \cdots, X_n\}$ be i.i.d. random variables with $n = p + k$, $k > 0$. The associated kth-order jackknife variance is defined in (1.8) and its expected value is

$$E\{V^{(k)}(\varphi)\} = \frac{1}{\binom{p+k-1}{p}}\left[\sum_{\mathbf{i}\in\mathscr{D}} E\{\varphi(\mathbf{X_i})^2\} - \binom{p+k}{p}^{-1} \sum_{\mathbf{i},\mathbf{j}\in\mathscr{D}} E\{\varphi(\mathbf{X_i})\varphi(\mathbf{X_j})\}\right]. \tag{5.1}$$

Clearly, for any $\mathbf{i} = (i_1, i_2, \cdots, i_p)$,

$$E\{[\varphi(X_{\mathbf{i}})]^2\} = c_p, \tag{5.2}$$

with c_p as defined in (1.3). Using elementary combinatorics, we obtain for each fixed $\mathbf{i}$,

$$E\{\varphi(X_{\mathbf{i}}) \sum_{\mathbf{j}\in\mathscr{D}} \varphi(X_{\mathbf{j}})\} = \sum_i \binom{p}{i}\binom{k}{p-i} c_i$$

independent of $\mathbf{i}$. Therefore,

$$E\{V^{(k)}(\varphi)\} = \frac{\binom{p+k}{p}}{\binom{p+k-1}{p}} c_p - \frac{\sum_{i=0}^{p} \binom{p}{i}\binom{k}{p-i} c_i}{\binom{p+k-1}{p}}. \tag{5.3}$$

Substituting for c_r from (2.15), relying on the orthogonal ANOVA decomposition of $\varphi(x_1, \cdots, x_p)$, we obtain for the bias

$$\begin{aligned} b^{(k)}(\varphi) &= E\{V^{(k)}(\varphi)\} - \operatorname{Var}\varphi \\ &= \frac{p+k}{k} c_p - \frac{\sum_i \binom{p}{i}\binom{k}{p-i} c_i}{\binom{p+k-1}{p}} - c_p + c_0 \\ &= \frac{p}{k} \sum_\nu \binom{p}{\nu} a_\nu + a_0 - \frac{\sum_\nu \sum_i \binom{p}{i}\binom{k}{p-i}\binom{i}{\nu} a_\nu}{\binom{p+k-1}{p}} \\ &= \frac{p}{k} \sum_\nu \binom{p}{\nu} a_\nu + a_0 - \frac{\sum_\nu a_\nu \binom{p}{\nu} \sum_i \binom{p-\nu}{i-\nu}\binom{k}{p-i}}{\binom{p+k-1}{p}}. \end{aligned}$$

The inner final sum is a convolution which can be summed (generating functions here facilitate the calculation) to give

$$\begin{aligned} b^{(k)}(\varphi) &= \sum_\nu \binom{p}{\nu} a_\nu \left\{ \frac{p}{k} - \frac{\binom{p-\nu+k}{p-\nu}}{\binom{p+k-1}{p}} \right\} + a_0 \\ &= \sum_\nu \binom{p}{\nu} a_\nu \left[\frac{p}{k} \left\{ 1 - \frac{(p-1)(p-2)\cdots(p-\nu+1)}{(p+k-1)(p+k-2)\cdots(p+k-\nu+1)} \right\} \right] + a_0. \end{aligned} \tag{5.4}$$

Direct verification yields that the coefficients of a_0 and a_1 vanish; indeed, the normalizations were set such that $V^{(k)}(\varphi)$ is unbiased for φ linear.

We observe that for $2 \le \nu \le p$

$$\prod_{\ell=1}^{\nu-1} \left(\frac{p-\ell}{s+p-\ell} \right) = \int_0^\infty e^{-s\xi} f_\nu(\xi)\, d\xi, \tag{5.5}$$

where $f_\nu(\xi)$ is a convolution of $\nu - 1$ exponential densities with scale parameters $p-1$, $p-2, \cdots, p-\nu+1$. Integration by parts reveals that

$$\frac{1}{k}\left\{ 1 - \frac{(p-1)(p-2)\cdots(p-\nu+1)}{(p+k-1)(p+k-2)\cdots(p+k-\nu+1)} \right\} = \int_0^\infty e^{-k\xi}\{1 - F_\nu(\xi)\}\, d\xi,$$

where $F_\nu(\xi)$ is the cumulative of $f_\nu(\xi)$. Since the a_ν are nonnegative (see (2.11)), it follows that

$$b^{(k)}(\varphi) = \int_0^\infty e^{-k\xi} g(\xi)\, d\xi, \tag{5.6}$$

where $g(\xi) = \sum_{\nu=2}^{p} \binom{p}{\nu} a_\nu\{1 - F_\nu(\xi)\}$ for $\xi > 0$ is plainly a decreasing positive continuous function. The claim associated with (1.11) of Theorem A ensues immediately from the representation (5.6). The proof of Theorem A is complete.

PROOF OF THEOREM B. Recall the expression (1.14)

$$B_r^{(k)} = E\{V_r^{(k)}\} - \operatorname{Var} \varphi_r,$$

where $k + r = n$ is held fixed. Here we examine $B_r^{(k)}$ as a function of r. Following (5.4), inserting $k = n - r$, we have that

$$B_r^{(k)} = \sum_{\nu=1}^{r} a_\nu \binom{r}{\nu} \frac{n}{n-r}\left\{1 - \binom{r}{\nu}\Big/\binom{n}{\nu}\right\} - \sum_{\nu=1}^{r} a_\nu \binom{r}{\nu}.$$

Consider the coefficient of a_ν in the difference of the biases, that is

$$\begin{aligned} &\text{Coeff. } a_\nu \text{ of } (B_{r+1}^{(k-1)} - B_r^{(k)}) \\ &= \binom{r+1}{\nu}\left\{\frac{r+1}{n-r-1} - \frac{(r+1)r(r-1)\cdots(r-\nu+2)}{(n-r-1)(n-1)(n-2)\cdots(n-\nu+1)}\right\} \\ &\quad - \binom{r}{\nu}\left\{\frac{r}{n-r} - \frac{r(r-1)(r-2)\cdots(r-\nu+1)}{(n-r)(n-1)(n-2)\cdots(n-\nu+1)}\right\} \end{aligned} \tag{5.7}$$

which we now prove is nonnegative for $1 \le \nu \le r < n$. The nonnegativity of (5.7) reduces to

$$\frac{(r+1)(r+1)}{(r+1-\nu)(n-r-1)} - \frac{r}{n-r} > \frac{r(r-1)\cdots(r-\nu+2)}{(n-1)(n-2)\cdots(n-\nu+1)}\left\{\frac{(r+1)^2}{(n-r-1)(r+1-\nu)} - \frac{(n-\nu+1)}{n-r}\right\},$$

or

$$\begin{aligned} &\{n(r+1) + \nu r(n-r-1)\}\{(n-1)(n-2)\cdots(n-\nu+1)\} \\ &\quad - \{r(r-1)(r-2)\cdots(r-\nu+2)\}\{(r+1)^2(n-r) \\ &\quad - (n-\nu+1)(n-r-1)(r+1-\nu)\} \ge 0. \end{aligned} \tag{5.8}$$

We view the left hand expression in (5.8) as a polynomial inequality in the variable n over the range $n \ge r + 1$. Substituting $n - r - 1 = x$ or $n = x + r + 1$, it is manifest since $\nu \le r$ that the resulting polynomial $P(x)$ has all positive coefficients of x^i, $i \ge 2$ and $P(0) = 0$. Evaluating $P(1)$, apart from a positive factor, we obtain $2(r + 1)^2 + (r + 1)(\nu - 2)(3\nu - 1) - \nu(\nu - 1)(\nu - 2)$ which is positive for integer ν, $1 \le \nu \le r$. It follows that $P(x)$ is positive for all integer $x > 0$, and the proof of Theorem B is complete.

We next discuss briefly an example of a class of statistics for which the inequality (1.15) of Theorem B is reversed, in order to emphasize the dependence of these variance comparisons on the structure of the class.

Consider the class of statistics

$$\psi_p(X_1, \cdots, X_p) = \frac{1}{p} \sum_{i=1}^{p} (X_i - \bar{X})^2, \quad p = 2, 3 \cdots$$

The sequence $\{\psi_p\}$ is obviously not constructed as a sequence of marginal functions of the type (1.12). Parallel to (1.14), for a sample $X_1, \cdots, X_n$ we define

$$B_p^{(k)} = E\{V^{(k)}(\psi_p)\} - \operatorname{Var} \psi_p, \quad 2 \le p < n, p + k = n.$$

The variance orderings are now the reverse of those in Theorem B.

PROPOSITION 5.1. *For the sequence ψ_p as displayed above*

$$B_p^{(k)} \geq B_{p+1}^{(k-1)}, \quad 2 \leq p \leq n-1. \tag{5.9}$$

PROOF. First recall from (3.9) the calculation $a_2 = (4/p^4)\,\sigma^4$ where $\sigma^2 = \operatorname{Var} X_i$, while $a_\nu = 0$ for $\nu > 2$. From the calculation preceding (5.5), and its subsequent discussion, we see that only a_2 appears in the expression of $B_p^{(k)}$. Specifically,

$$B_p^{(k)} = \binom{p}{2}\left(\frac{4\sigma^4}{p^4}\right)\left\{\frac{p}{k}\left(1 - \frac{p-1}{p+k-1}\right)\right\},$$

$$B_{p+1}^{(k-1)} = \binom{p+1}{2}\frac{4\sigma^4}{(p+1)^4}\left\{\frac{p+1}{k-1}\left(1 - \frac{p}{p+k-1}\right)\right\}.$$

Direct calculation reveals that (5.9) is equivalent to $(p-1)/p^2 \geq p/(p+1)^2$ which is valid for $p \geq 2$.

PROOF OF THEOREM C. Consider the function $\varphi(\mathbf{x}, \mathbf{y}) = \varphi(x_1, \cdots, x_p; y_1, \cdots, y_q)$. We assume φ to be symmetric with respect to permutations of the x's and y's separately. Let $X_1, \cdots, X_n$ and $Y_1, \cdots, Y_m$ $(n > p,\ m > q)$ be two independent samples, each consisting of i.i.d. random variables. We define the jackknife variance estimate as displayed in (1.16) where $K = \left\{\left(1 - \frac{p}{n}\right)\binom{n}{p}\binom{m}{q}\right\}^{-1}$ is a normalizing constant that guarantees unbiasedness for all symmetric linear functions of the form $\alpha \sum_{i=1}^{p} u(x_i) + \beta \sum_{j=1}^{q} v(y_j)$ provided $p/n = q/m$. If $p/n \neq q/m$ it can be shown by direct computation that a normalization which guarantees unbiasedness for all α, β does not exist. Henceforth, we restrict attention to the case $p/n = q/m$.

We now proceed to study the expected jackknife variance of $V(\varphi)$. Calculations analogous to (5.3) produce

$$E[V(\varphi)] = K\binom{n}{p}\binom{m}{q}\left\{c_{pq} - \binom{n}{p}^{-1}\binom{m}{q}^{-1}\sum_{i,j}\binom{p}{i}\binom{n-p}{p-i}\binom{q}{j}\binom{m-q}{q-j}c_{ij}\right\}$$

and $\operatorname{Var} \varphi(\mathbf{X}, \mathbf{Y}) = c_{pq} - c_{00}$ where c_{ij} is defined in (4.2).

We substitute $c_{ij} = \sum \binom{i}{\nu}\binom{j}{\mu} a_{\nu\mu}$ from (4.2) into the expression for $E\{V(\varphi)\} - \operatorname{Var} \varphi(\mathbf{X}, \mathbf{Y})$ and examine the resulting coefficients of $a_{\nu,\mu}$. With some circumspect manipulations of the binomial factors, we find them to be always nonnegative provided $p/n = q/m$. The details are a bit arduous but straightforward.

PROOF OF THEOREM C′. We will discuss only the proof for the jackknife estimator of order one. A direct calculation establishes the equation

$$E\{J_{\mathbf{x},\mathbf{y}}^{(1)}\} - \operatorname{Var} \varphi(\mathbf{X}, \mathbf{Y}) = (p+q-1)c_{p,q} - pc_{p-1,q} - qc_{p,q-1} + c_{00}.$$

Substituting $c_{p,q} = \sum_{\mu,\nu}\binom{p}{\mu}\binom{q}{\nu} a_{\mu,\nu}$ the right hand side reduces to $\sum_{\mu,\nu}(\mu + \nu - 1)\binom{p}{\mu}\binom{q}{\nu} a_{\mu\nu} + a_{00}$, which is nonnegative because $a_{\mu,\nu} \geq 0$. The proof for the case $k > 1$ is analogous, but more tedious.

PROOF OF THEOREM D. We deal here with a U-statistic of order q for ψ a symmetric function of q variables of the form $\varphi(X_1, X_2, \cdots, X_n) = \binom{n}{q}^{-1} \sum_{\mathbf{i}} \psi(X_{i_1}, X_{i_2}, \cdots, X_{i_q})$; $\mathbf{i}$ traverses the set of all q-tuples $1 \leq i_1 < \cdots < i_q \leq n$. Since $X_1, \cdots, X_n$ are i.i.d., we obtain on the basis of the ANOVA decomposition

$V_n = \mathrm{Var}\{\varphi(X_1, X_2, \cdots, X_n)\}$

$$= \frac{1}{\binom{n}{q}} \sum_{3i=0}^{q} \binom{q}{i}\binom{n-q}{q-i} c_i = \sum_{\nu=0}^{q} \frac{\binom{q}{\nu}\binom{n-\nu}{q-\nu}}{\binom{n}{q}} a_\nu = \sum_{\nu=0}^{q} \frac{\binom{q}{\nu}^2}{\binom{n}{\nu}} a_\nu.$$

We consider next for r fixed $\binom{n}{r} V_n = \sum_{\nu=0}^{q} a_\nu \binom{q}{\nu}^2 \binom{n}{r} \Big/ \binom{n}{\nu}$. For each $\nu > r$ we find, compared to (5.5), that $\binom{n}{r} \Big/ \binom{n}{\nu} = \int_0^\infty e^{-n\xi} f_{r,\nu}(\xi)\, d\xi$, where $f_{r,\nu}(\xi)$ is a positive bounded function. Moreover, for $1 \le \nu \le r$, $\binom{n}{r} \Big/ \binom{n}{\nu}$ is a polynomial in n of degree $\le r - 1$. It follows that $\Delta^r \left\{ \binom{n}{r} \Big/ \binom{n}{\nu} \right\} = 0$ for $1 \le \nu \le r$, while for $\nu > r$, $(-1)^r \Delta^r \left\{ \binom{n}{r} \Big/ \binom{n}{\nu} \right\} > 0$, the difference operator taken with respect to $n \ge q$. Then, as $a_\nu \ge 0$, $1 \le \nu \le q$, we have $(-1)^r \Delta^r \left\{ \binom{n}{r} V_n \right\} \ge 0$ yielding equality $(-1)^r \Delta^r \left\{ \binom{n}{r} V_n \right\} = 0$ only when φ is a U-statistic of order $\le r$.

REFERENCES

BHARGAVA, R. P. (1980). A property of the jackknife estimation of the variance when more than one observation is omitted. Tech. Report. No. 152, Dept. of Statist., Stanford Univ.

EFRON, B. and STEIN, C. (1981). The jackknife estimate of the variance. *Ann. Statist.* **9** 586–596.

FELLER, W. (1950). *An Introduction to Probability Theory and Its Applications.* Wiley, New York.

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

HOEFFDING, W. (1961). The strong law of large numbers for U-statistics. Inst. of Statist., Univ. of North Carolina, Mimeograph Series No. 302.

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

KARLIN, S. (1968). *Total Positivity.* Stanford Univ. Press, Stanford, Calif.

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

SERFLING, R. J. (1980). *Approximation Theorems of Mathematical Statistics.* Wiley, New York.

WIDDER, D. V. (1946). *The Laplace Transform.* Princeton Univ. Press, Princeton, N.J.

DEPARTMENT OF MATHEMATICS
STANFORD UNIVERSITY
STANFORD, CA 94305

DEPARTMENT OF STATISTICS
THE HEBREW UNIVERSITY OF JERUSALEM
JERUSALEM, ISRAEL