The Annals of Statistics
1974, Vol. 2, No. 6, 1065–1094

INEQUALITIES FOR SYMMETRIC SAMPLING PLANS I[1]

BY SAMUEL KARLIN

The Weizmann Institute of Science and Stanford University

Introduction. In recent years there has been much interest in evaluating certain probabilistic quantities arising in sampling with and/or without replacement, from finite populations, e.g., Korwar and Serfling (1970), Lanke (1972), Sen (1970), Serfling (1973), Kemperman (1973), and others. For these special sampling procedures a variety of inequalities have been derived for the moments of the sums and maximum of partial sums based on the observed sample.

Rosén (1972) investigated the validity of a central limit theorem for weighted sums of observation arising from quite general sampling schemes. In an earlier paper Rosén (1967) established a convexity type inequality comparing the expectations of certain functionals evaluated with respect to a "symmetric" sampling plan and for the special plan of sampling without replacement. It was this earlier paper that kindled our interest. The present work develops a variety of inequalities for functionals associated with sampling plans which can be interpreted as results on multivariate ordering relationships among certain distributions.

The organization of the contents is as follows. In Section 1 relevant definitions and preliminaries are introduced. Several important classes of sampling procedures are delimited including the "symmetric" sampling schemes, "random replacement policies," sampling procedures based on special partitionings of the sample space, sampling plans involving a prescribed number of distinct observations with given multiplicities, conditional sampling procedures and other

Received March 1973; revised January 1974.

[1] Research supported in part by ONR Contract N00014-67-A-0112-0015 at Stanford.

important sampling plans of a more technical nature not readily identified by a single name.

A series of inequalities for expectations of certain functionals of the observations with respect to symmetric sampling schemes are set forth in Section 2. In the process we underscore the special nature of sampling without replacement as it fits the hierarchy of inequalities. In a certain sense sampling without replacement can be characterized by a minimum variational problem (Theorem 2.1). A by-product of our work allows us to settle some problems posed by Rosén (1967).

The class of functions $\mathscr{C}$ described in Definition 1.3 and prominent in all the present investigations can be viewed as a more general form of the class of doubly alternating capacity functions occurring in the theory of robust statistics.

Section 3 is devoted to the study of inequalities for random replacement schemes. In this context we highlight some of the significance and importance of sampling with replacement. Sampling with replacement provides the maximum for a large class of expectations when computed with respect to all random replacement schemes (see Theorem 3.1).

The fourth section elaborates a series of applications of the main theorems of Sections 2 and 3 to the theory of dilation of measures induced by symmetric sampling schemes. A number of majorization inequalities in the sense of Hardy, Littlewood and Pólya (1934), Chapter 2, are set forth in culmination of this work. A number of the results of this section were inspired by an unpublished manuscript of Kemperman (1973). We are greatly indebted to him for making his preprint available.

In a future paper we hope to take up some suggestive martingale versions of the inequalities of Sections 2 and 3. Future prospects for new kinds of multivariate majorization relationships are also indicated by the present work (see the close of Section 3). The ideas developed herein further bear on various multivariate formulations of total positivity, generalized convexity and the concepts of measures of associations among vector random variables.

1. Definitions and preliminaries. Consider a population space Ω consisting of N elements $\{a_1, a_2, \cdots, a_N\}$. Henceforth, unless stated otherwise, the space Ω is assumed to contain N real values, (repetitions allowed). A sampling plan $\mathscr{S}$ of size n is defined by prescribing a probability distribution $P_{\mathscr{S}}$ on $\Omega^n = \Omega \times \Omega \times \cdots \times \Omega$ (n direct copies of Ω). Thus the specification of the probability $P_{\mathscr{S}}(x_1, x_2, \cdots, x_n)$ of each n-tuple $\mathbf{x} = (x_1, x_2, \cdots, x_n)$, $x_i \in \Omega$, determines $\mathscr{S}$. A sample according to $\mathscr{S}$ consists of a choice of a point of Ω^n following the distribution law $P_{\mathscr{S}}$.

Sampling with replacement ($\mathscr{S}$) corresponds to the sampling procedure where

$$P_{\mathscr{S}}(\mathbf{x}) = P_{\mathscr{S}}(x_1, \cdots, x_n) = \frac{1}{N^n} \tag{1.1}$$

for each $\mathbf{x} \in \Omega^n$. For convenience we use the distinguished symbol $\mathscr{I}$ (suggesting independent observations) for sampling with replacement so $P_{\mathscr{I}}(\mathbf{x}) = 1/N^n$.

Sampling without replacement ($\mathscr{W}$) is characterized by the probability distribution

$$(1.2) \qquad P_{\mathscr{W}}(\mathbf{x}) = P_{\mathscr{W}}(x_1, \cdots, x_n) = \frac{1}{N(N-1)\cdots(N-n+1)}$$

provided all x_i comprise different elements of Ω and $P_{\mathscr{W}}(\mathbf{x}) = 0$ otherwise.

DEFINITION 1.1. A sampling plan $\mathscr{S}$ is called symmetric if

$$P_{\mathscr{S}}(a_{\sigma(1)}, a_{\sigma(2)}, \cdots, a_{\sigma(n)})$$

has the same value for all permutations σ of the elements of Ω into itself, i.e., $P_{\mathscr{S}}$ is invariant with respect to the full group of permutations acting on Ω.

A broad class of symmetric sampling plans includes the so-called *Random Replacement Schemes*. They are formally defined as follows:

DEFINITION 1.2. Let $\{\pi_1, \pi_2, \cdots, \pi_{n-1}\}$ be given in advance satisfying $0 \leqq \pi_i \leqq 1$. Choose X_1 from Ω with each element a_i equally likely. Remove X_1 from Ω with probability $1 - \pi_1$ and with probability π_1 replace it. Then sample X_2 from the new population space with each element occurring equally likely. Remove X_2 with probability $1 - \pi_2$ and replace it with probability π_2. Continue in this manner, yielding the sample $(X_1, X_2, \cdots, X_n)$. The sampling plan so constructed is designated as $\mathscr{S} = \mathscr{R}(\boldsymbol{\pi})$ for the random replacement scheme with associated parameters $\boldsymbol{\pi} = (\pi_1, \pi_2, \cdots, \pi_{n-1})$. Note for $\mathbf{1} = \{1, 1, \cdots, 1\}$ that $\mathscr{R}(\mathbf{1}) = \mathscr{I}$ (sampling with replacement) and for $\mathbf{0} = \{0, 0, \cdots, 0\}$ that $\mathscr{R}(\mathbf{0}) = \mathscr{W}$ (sampling without replacement). It is easy to check that each $\mathscr{R}(\boldsymbol{\pi})$ generates a symmetric sampling plan.

We will use the following notation. For any given sampling plan $\mathscr{S}$, $E_{\mathscr{S}}\,\phi(X_1, \cdots, X_n) = E_{\mathscr{S}}\,\phi(\mathbf{X})$ denotes the expected value of $\phi(\mathbf{X})$ evaluated with respect to the probability law $P_{\mathscr{S}}$ where $\phi(\xi_1, \cdots, \xi_n)$ is defined on Ω^n.

The stimulus of our work stems from the following result of Rosén (1967).

THEOREM. *Let the elements of Ω be real-valued. For any continuous convex function $\phi(x)$ and setting $Z_n = X_1 + X_2 + \cdots + X_n$, we have*

$$(1.3) \qquad E_{\mathscr{W}}\,\phi(Z_n) \leqq E_{\mathscr{S}}\,\phi(Z_n)$$

for any symmetric sampling plan $\mathscr{S}$ where $\mathscr{W}$ denotes sampling without replacement.

This theorem generalized a simpler inequality of Hoeffding (1963) of the form (1.3) for the special sampling plans, comparing sampling with and without replacement.

The value of the theorem is manifold. It allows us to obtain lower estimates on moments for certain functions (e.g., the variance of the sample) with respect to general symmetric sampling plans in terms of the specific sampling plan $\mathscr{W}$. Upper estimates of importance are implied later in Theorem 3.1. Such results are relevant in establishing central limit theorems for special functions of Z_n based on certain symmetric sampling plans, (e.g., see Rosén (1972) for details, and other references), and serve also in other contexts.

Rosén raised the problem of characterizing more general functions $\phi(\xi_1, \cdots, \xi_n)$ not necessarily dependent on $(\xi_1, \cdots, \xi_n)$ through the sum $(\xi_1 + \cdots + \xi_n)$ defined on Ω^n for which the inequality

$$E_{\mathscr{W}}\phi(X_1, X_2, \cdots, X_n) \leqq E_{\mathscr{S}}\phi(X_1, \cdots, X_n) \tag{1.4}$$

is fulfilled for every symmetric sampling scheme $\mathscr{S}$. He pointed out, specifically, that (1.4) applies for

$$\phi(\xi_1, \cdots, \xi_n) = \psi(\lambda(\xi_1) + \cdots + \lambda(\xi_n)) \tag{1.5}$$

when ψ is convex and λ is arbitrary. He further proved that if (1.4) is satisfied for ϕ_1 and ϕ_2, with both these functions symmetric, then (1.4) also holds for

$$\phi(\boldsymbol{\xi}) = \max(\phi_1(\boldsymbol{\xi}), \phi_2(\boldsymbol{\xi}))\,. \tag{1.6}$$

Of course, limits of functions satisfying (1.4) also fulfill (1.4).

In this paper we shall settle the problem raised by Rosén pertaining to the scope of validity for (1.4). In the process we do much more. Indeed, we establish a whole hierarchy of comparison inequalities among certain classes of symmetric sampling plans, subsuming (1.4) as an application. In order to formulate the results we need to introduce some apparatus giving perhaps a more natural perspective on the nature of symmetric sampling schemes.

We first delineate the appropriate class of functions $\mathscr{C}$ to be involved in our considerations, generalizing those of the form (1.5), which also satisfies the property that the operation (1.6) preserves the class.

DEFINITION 1.3. A function $\phi(\xi_1, \cdots, \xi_n)$ is said to be of class $\mathscr{C}$ if the following two conditions are fulfilled:

(i) ϕ defined on Ω^n is a symmetric function of its arguments;

(ii)

$$\begin{aligned}\phi(a, a, \xi_3, \cdots, \xi_n) &+ \phi(b, b, \xi_3, \cdots, \xi_n) \\ &\geqq \phi(a, b, \xi_3, \cdots, \xi_n) + \phi(b, a, \xi_3, \cdots, \xi_n) \\ &= 2\phi(a, b, \xi_3, \cdots, \xi_n) \qquad \text{for all } a, b, \xi_3, \cdots, \xi_n \in \Omega\,.\end{aligned} \tag{1.7}$$

Clearly, owing to (i), a corresponding inequality to (1.7) applies for any pair of coordinates.

When ϕ is twice continuously differentiable, a sufficient condition for (1.7) is

$$\frac{\partial^2\phi}{\partial\xi_1\partial\xi_2}(\xi_1, \xi_2, \xi_3, \cdots, \xi_n) \geqq 0 \qquad \text{for all } \boldsymbol{\xi}\,. \tag{1.8}$$

The implication leading from (1.8) to (1.7) is straightforward. Indeed, integrate (1.8) over the rectangle $a \leqq \xi_1 \leqq b$, $a \leqq \xi_2 \leqq b$ where $\xi_3, \xi_4, \cdots, \xi_n$ are held fixed and (1.7) quickly ensues.

It is readily checked that functions of the form (1.5) obey (1.7). Also, it is elementary to show that the property of (1.6) is satisfied for functions of class $\mathscr{C}$. Requirements of the sort (1.7) arise in diverse contexts of generalized convexity (see Section 4).

Sampling plans induced by sets of ordered integers. There is a natural partitioning of Ω^n induced by the following set of symmetric sampling schemes which will play a central role in all our developments.

DEFINITION 1.4. For each prescribed set of positive integers $k_1 \geqq k_2 \geqq \cdots \geqq k_p$ (p arbitrary but fixed) satisfying $\sum_{i=1}^p k_i = n$, we determine the sampling plan $\mathscr{S}[\mathbf{k}]$ associated with $\mathbf{k} = (k_1, k_2, \cdots, k_p)$ as follows. (Designate $\mathscr{K} = [\mathbf{k}]$ as the collection of all such sets of decreasing integers.) Pick p elements $Y_1, Y_2, \cdots, Y_p$ successively from Ω at random without replacement. Construct thereby the set of sample points $(X_1, \cdots, X_n)$ where Y_1 is first repeated k_1 times, Y_2 then repeated k_2 times, etc., and Y_p repeated k_p times. The totality of all possible points obtained by this sampling procedure is denoted by $\Omega(n, \mathbf{k}) = \Omega(\mathbf{k}) = \Omega([k_1, k_2, \cdots, k_p])$. Thus, a point $(b_1, b_2, \cdots, b_n)$ belongs to $\Omega(\mathbf{k})$ if its composition consists of p distinct elements of Ω, one occurring k_1 times, a second represented k_2 times, etc., and the pth appearing k_p times *where each group of identical elements appears in adjacent positions and the blocks occur in decreasing order of size.*

The set $\Omega(\mathbf{k})$ contains $(N)_p = N(N-1) \cdots (N-p+1) = N!/(N-p)!$ points corresponding to all possible choices of sets of ordered p distinct elements from Ω.

The probability $P_{\mathscr{S}(\mathbf{k})}$ attached to each point $\mathbf{a} \in \Omega(\mathbf{k})$ where $\mathbf{k} = [k_1, k_2, \cdots, k_p]$, $k_1 \geqq k_2 \geqq \cdots \geqq k_p$, is obviously $P_{\mathscr{S}(\mathbf{k})}(\mathbf{a}) = (N-p)!/N!$ and $P_{\mathscr{S}[\mathbf{k}]}(\mathbf{a}) = 0$ for $\mathbf{a} \notin \Omega(\mathbf{k})$.

More generally, for any set of ordered positive integers $\gamma_1, \gamma_2, \cdots, \gamma_r$ with $\sum_{i=1}^r \gamma_i = n$ we can define a sampling plan $\mathscr{S}^*[\gamma_1, \gamma_2, \cdots, \gamma_r]$ which assigns probability $(N-r)!/N!$ to all n-tuples $\mathbf{x}$ in Ω^n of the form

$$\mathbf{x} = (\overbrace{\xi_1, \cdots, \xi_1}^{\gamma_1}, \overbrace{\xi_2, \cdots, \xi_2}^{\gamma_2}, \cdots, \overbrace{\xi_r, \cdots, \xi_r}^{\gamma_r})$$

such that $\xi_1 \in \Omega$ appears first γ_1 times and then ξ_2 appears γ_2 times, etc. Whenever γ_i are not arranged in decreasing order we employ the asterisk designation $\mathscr{S}^*[\gamma_1, \cdots, \gamma_r]$ to indicate this fact.

If $\phi(\boldsymbol{\xi})$ is symmetric it is clear that

$$E_{\mathscr{S}^*[\gamma_1,\gamma_2,\cdots,\gamma_r]}\phi(\mathbf{X}) = E_{\mathscr{S}[\delta_1,\delta_2,\cdots,\delta_r]}\phi(\mathbf{X})$$

where the integers $\{\delta_i\}$ coincide with $\{\gamma_i\}$ but are arranged in decreasing order, i.e.

$$\delta_1 \geqq \delta_2 \geqq \cdots \geqq \delta_r > 0 .$$

Some illustrations of the sampling schemes $\mathscr{S}[\mathbf{k}]$ are worth highlighting:

(i) $\mathscr{S}([n, 0, 0, \cdots, 0])$ is the sampling plan that assigns probability $1/N$ to each of the points $(a_i, a_i, \cdots, a_i)$, $i = 1, 2, \cdots, N$ where a_i is a labeling of the elements of Ω.

(ii) $\mathscr{S}([n-1, 1, 0, \cdots, 0])$ is the sampling scheme which assigns probability $1/N(N-1)$ to each of the n tuples where one element of Ω appears first $N-1$ times and a different element then appears once.

(iii) $\mathscr{S}[1, 1, \cdots, 1] = \mathscr{W}$ = sampling without replacement. If $N = 3$, $n = 2$, we have

$$\Omega[2, 0] = \{(a_1, a_1), (a_2, a_2), (a_3, a_3)\}$$
$$\Omega[1, 1] = \{(a_1, a_2), (a_2, a_1), (a_1, a_3), (a_3, a_1), (a_2, a_3), (a_3, a_2)\} .$$

The plan $\mathscr{S}[\mathbf{k}]$ is not symmetric, because of our insistence that the groups of identical elements appear adjacently. It is easy to convert $\mathscr{S}[\mathbf{k}]$ into a symmetric sampling plan, denoted henceforth by $\tilde{\mathscr{S}}[\mathbf{k}]$ (with a tilde inserted), by taking an element $\xi = (\xi_1, \cdots, \xi_n)$ of $\Omega([\mathbf{k}])$ and constructing with it the collection of all possible permutations of the components of ξ (designating the result by $\Lambda(\xi)$), and then spreading the probability assigned to ξ, $P_{\mathscr{S}}(\{\xi\})$ equally among the elements of $\Lambda(\xi)$. The observation space for $\tilde{\mathscr{S}}[\mathbf{k}]$ is denoted by $\tilde{\Omega}([\mathbf{k}]) = \tilde{\Omega}(n, k) = \tilde{\Omega}([k_1, k_2, \cdots, k_p])$.

In terms of the subspaces $\Omega([\mathbf{k}])$ of Ω^n we can delineate the general structure of symmetric sampling plans in a transparent manner.

PROPOSITION 1.1. *A symmetric sampling plan has the structure that the total probability $P_{\mathscr{S}}$ assigned to $\Omega(\mathbf{k}) = \Omega(n; [k_1, k_2, \cdots, k_p])$ is distributed uniformly among its elements. Moreover, if $\psi(\xi_1, \xi_2, \cdots, \xi_n)$ is symmetric then*

$$(1.9) \qquad E_{\mathscr{S}}\psi(X_1, X_2, \cdots, X_n) = \sum_{[k_1,\cdots,k_p]} c(\mathbf{k}, n, \mathscr{S}) \cdot E_{\mathscr{S}([k_1,\cdots,k_p])}\psi(X_1, X_2, \cdots, X_n)$$

where the sum is extended over all choices of nonnegative integers $k_1 \geqq k_2 \geqq \cdots \geqq k_p$ satisfying $\sum_{i=1}^p k_i = n$, i.e., $\mathbf{k} \in \mathscr{K}$ and $c(\mathbf{k}, n; \mathscr{S})$ $(= c(\mathbf{k})$ for brevity of notation) are nonnegative numbers satisfying the normalization $\sum_{\mathbf{k}\in\mathscr{K}} c(\mathbf{k}) = 1$.

PROOF. From the definition we find that for $\mathbf{k} = [k_1, k_2, \cdots, k_p]$ and $\mathbf{l} = [l_1, l_2, \cdots, l_q]$ not identical the sets $\tilde{\Omega}(n; [k_1, k_2, \cdots, k_p])$ and $\tilde{\Omega}(n; [l_1, l_2, \cdots, l_q])$ are disjoint. (Also $\Omega([\mathbf{k}])$ and $\Omega([\mathbf{l}])$ are nonintersecting.) Because $\mathscr{S}$ is symmetric the probability $P_{\mathscr{S}}$ assigned to each point of $\tilde{\Omega}(n; [k_1, k_2, \cdots, k_p])$ where nonzero agrees, since each two members of this set are mapped by an appropriate permutation of Ω one into the second.

Consider any point $\mathbf{x} = (x_1, x_2, \cdots, x_n)$ from Ω^n which involves p distinct specific elements $\mathbf{y} = (y_1, y_2, \cdots, y_p)$, y_1 appearing k_1 times among the coordinates of $\mathbf{x}$, y_2 appearing k_2 times, etc.

For any two points $\mathbf{x}$ and $\mathbf{x}'$ leading to the same $\mathbf{y}$ we have $\psi(\mathbf{x}) = \psi(\mathbf{x}')$ since ψ is symmetric. Moreover, because $\mathscr{S}$ is a symmetric sampling scheme, necessarily $P_{\mathscr{S}}(\mathbf{x}) = P_{\mathscr{S}}(\mathbf{x}')$.

Let $P_{\mathscr{S}}(\tilde{\Omega}(n; [k_1, \cdots, k_p]) = c(\mathbf{k}, n, \mathscr{S})$ denote the probability assigned by $\mathscr{S}$ to the extended set $\tilde{\Omega}$ comprised of $\Omega(n; [k_1, k_2, \cdots, k_p])$ and the totality of elements obtained by performing all permutations on the elements in $\Omega[n; \mathbf{k}]$.

Of course

$$E_{\mathscr{S}[k_1,\cdots,k_p]}\phi(X_1, X_2, \cdots, X_n) = \frac{(N-p)!}{N!} \sum_{\boldsymbol{\xi} \in \Omega(n;[\mathbf{k}])} \phi(\boldsymbol{\xi}) .$$

The formula (1.9) is now evident by the law of total probabilities. □

2. The principal comparison inequalities for symmetric sampling plans. We now state the first principal theorem of this paper.

THEOREM 2.1. *Let N and n be fixed. Consider* $[\mathbf{k}]$ *and* $[\mathbf{l}] \in \mathscr{K}$ *i.e.*,

$$[\mathbf{k}] = [k_1, k_2, \cdots, k_p], \quad k_1 \geqq k_2 \geqq \cdots \geqq k_p, \qquad \sum k_i = n$$
$$[\mathbf{l}] = [l_1, l_2, \cdots, l_q], \quad l_1 \geqq l_2 \geqq \cdots \geqq l_q, \qquad \sum l_j = n .$$

Then

$$E_{\mathscr{S}[\mathbf{k}]}\phi(X_1, X_2, \cdots, X_n) \geqq E_{\mathscr{S}([\mathbf{l}])}\phi(X_1, X_2, \cdots, X_n) \tag{2.1}$$

for every function ϕ *of class* $\mathscr{C}$ *(see Definition* 1.3) *if and only if*

$$k_1 + k_2 + \cdots + k_r \geqq l_1 + l_2 + \cdots + l_r \qquad r = 1, 2, \cdots, p . \tag{2.2}$$

Notation. The ordering relation between $[\mathbf{k}]$ and $[\mathbf{l}]$ implied by the set of inequalities (2.2) will be compactly written as

$$[\mathbf{k}] \succ [\mathbf{l}] . \tag{2.3}$$

This ordering coincides with the notion of *majorization* of two vectors of positive components provided suitable zero terms are appended so that the lengths of the two sequences agree (see Hardy, Littlewood and Pólya (1934), Chapter 2).

The details of the proof of Theorem 2.1 are deferred momentarily. We presently harvest a series of important corollaries. Manifestly,

$$[1] = [\overbrace{1, 1, \cdots, 1}^{n}] \prec [k_1, k_2, \cdots, k_p] \prec [n, 0, 0, \cdots, 0] \tag{2.4}$$

for any $[\mathbf{k}] \in \mathscr{K}$. Therefore

COROLLARY 2.1. *If* $\phi \in \mathscr{C}$, *then*

$$E_{\mathscr{W}}\phi(\mathbf{X}) = E_{\mathscr{S}[1]}\phi(\mathbf{X}) \leqq E_{\mathscr{S}[\mathbf{k}]}\phi(\mathbf{X}) \leqq E_{\mathscr{S}[n,0,\cdots,0]}\phi(\mathbf{X}) . \tag{2.5}$$

This inequality in conjunction with (1.9) yields

COROLLARY 2.2. *For any symmetric sampling plan* $\mathscr{S}$ *and* $\phi \in \mathscr{C}$, *we have*

$$E_{\mathscr{W}}\phi(\mathbf{X}) \leqq E_{\mathscr{S}}\phi(\mathbf{X}) . \tag{2.6}$$

The result of (2.6) substantially extends that of Rosén cited in Section 1, as the class $\mathscr{C}$ well embraces all the functions occurring in his work.

The parallel reasoning to that of (2.6) produces the upper bound

COROLLARY 2.3. *For any symmetric sampling plan $\mathscr{S}$ and $\phi \in \mathscr{C}$ we have*

$$E_{\mathscr{S}}\phi(\mathbf{X}) \leqq E_{\mathscr{S}[n,0,\cdots,0]}\phi(\mathbf{X}) .$$

PROOF OF THEOREM 2.1. It is convenient for clarity of exposition to divide the proof into several steps.

LEMMA 2.1. *Let ϕ be of class $\mathscr{C}$. For $n = 2m$ ($m \geqq 1$) we have*

$$(2.7) \qquad E_{\mathscr{S}[2m]}\phi(\mathbf{X}) \geqq E_{\mathscr{S}([2m-1,1])}\phi(\mathbf{X}) \geqq E_{\mathscr{S}([2m-2,2])}\phi(\mathbf{X}) \geqq \cdots \geqq E_{\mathscr{S}([m+1,m-1])}\phi(\mathbf{X}) \geqq E_{\mathscr{S}([m,m])}\phi(\mathbf{X}) .$$

For $n = 2m + 1$ ($m \geqq 1$), we have

$$(2.8) \qquad E_{\mathscr{S}([2m+1])}\phi(\mathbf{X}) \geqq E_{\mathscr{S}([2m,1])}\phi(\mathbf{X}) \geqq \cdots \geqq E_{\mathscr{S}[m+2,m-1]}\phi(\mathbf{X}) \geqq E_{\mathscr{S}([m+1,m])}\phi(\mathbf{X}) .$$

PROOF. We begin with the proof of (2.8). From the definition we obtain

$$(2.9) \qquad E_{\mathscr{S}([m+1,m])}\phi(\mathbf{X}) = \frac{1}{N(N-1)} \sum_{i,j=1;i\neq j}^{N} \phi(\overbrace{a_i, \cdots, a_i}^{m+1}, \overbrace{a_j, \cdots, a_j}^{m})$$

and

$$(2.10) \qquad E_{\mathscr{S}[m+2,m-1]}\phi(\mathbf{X}) = \frac{1}{N(N-1)} \sum_{i,j=1;i\neq j}^{N} \phi(\overbrace{a_i, \cdots, a_i}^{m+2}, \overbrace{a_j, \cdots, a_j}^{m-1})$$

For each pair $\{a_i, a_j\}$ the terms

$$\phi(a_i, a_i, \overbrace{a_i, \cdots, a_i}^{m}, \overbrace{a_j, \cdots, a_j}^{m-1})$$

and

$$\phi(a_j, a_j, \overbrace{a_i, a_i, \cdots, a_i}^{m}, \overbrace{a_j, \cdots, a_j}^{m-1})$$

occur exactly once in the sums (2.10) and (2.9) respectively. Moreover, condition (1.7) assures

$$\phi(\overbrace{a_i, \cdots, a_i}^{m+2}, \overbrace{a_j, \cdots, a_j}^{m-1}) + \phi(a_j, a_j, \overbrace{a_i, \cdots, a_i}^{m}, \overbrace{a_j, \cdots, a_j}^{m-1}) \geqq 2\phi(a_j, a_i, \overbrace{a_i, \cdots, a_i}^{m}, \overbrace{a_j, \cdots, a_j}^{m-1}) .$$

Summing gives

$$E_{\mathscr{S}[m+2,m-1]}\phi(\mathbf{X}) + E_{\mathscr{S}[m,m+1]}\phi(\mathbf{X}) \geqq 2E_{\mathscr{S}[m+1,m]}\phi(\mathbf{X})$$

which manifestly reduces to

$$(2.11) \qquad E_{\mathscr{S}[m+2,m-1]}\phi(\mathbf{X}) \geqq E_{\mathscr{S}[m+1,m]}\phi(\mathbf{X}) .$$

By similar pairing based on

$$\phi(\overbrace{a_i, \cdots, a_i}^{m+3}, \overbrace{a_j, \cdots, a_j}^{m-2}) + \phi(\overbrace{a_i, \cdots, a_i}^{m+1}, \overbrace{a_j, \cdots, a_j}^{m}) \geqq 2\phi(\overbrace{a_i, \cdots, a_i}^{m+2}, \overbrace{a_j, \cdots, a_j}^{m-1})$$

we deduce

$$E_{\mathscr{S}[m+3,m-2]}\phi(\mathbf{X}) + E_{\mathscr{S}[m+1,m]}\phi(\mathbf{X}) \geqq 2E_{\mathscr{S}[m+2,m-1]}\phi(\mathbf{X}) .$$

This coupled with (2.11) proves

$$E_{\mathscr{S}[m+3,m-2]}\phi(\mathbf{X}) \geqq E_{\mathscr{S}[m+2,m-1]}\phi(\mathbf{X}) .$$

Continuing in this manner, we obtain the series of inequalities

$$E_{\mathscr{S}[2m,1]}\phi(\mathbf{X}) \geqq E_{\mathscr{S}[2m-1,2]}\phi(\mathbf{X}) \geqq \cdots \geqq E_{\mathscr{S}[m+1,m]}\phi(\mathbf{X}) . \tag{2.12}$$

Observe, next that

$$E_{\mathscr{S}[2m+1]}\phi(\mathbf{X}) = \frac{1}{N} \sum_{i=1}^{N} \phi(a_i, \cdots, a_i) = \frac{1}{N(N-1)} \sum_{i=1}^{N} (N-1)\phi(a_i, \cdots, a_i) .$$

Now

$$\phi(a_i, \cdots, a_i) + \phi(\overbrace{a_i, \cdots, a_i}^{2m-1}, a_j, a_j) \geqq 2\phi(\overbrace{a_i, \cdots, a_i}^{2m}, a_j) .$$

Hence

$$(N-1)\phi(a_i, \cdots, a_i) + \sum_{j=1;j\neq i}^{N} \phi(a_i, \cdots, a_i, a_j, a_j) \geqq 2 \sum_{j=1;j\neq i}^{N} \phi(a_i, \cdots, a_i, a_j) .$$

Another summation gives

$$E_{\mathscr{S}(2m+1]}\phi(\mathbf{X}) + E_{\mathscr{S}[2m-1,2]}\phi(\mathbf{X}) \geqq 2E_{\mathscr{S}[2m,1]}\phi(\mathbf{X}) ,$$

and this in conjunction with (2.12) leads to the result

$$E_{\mathscr{S}[2m+1]}\phi(\mathbf{X}) \geqq E_{\mathscr{S}[2m,1]}\phi(\mathbf{X}) .$$

The validation of (2.8) is done.

We next turn to (2.7) with $n = 2m$. Note for each pair $\{a_i, a_j\}$

$$\phi(\overbrace{a_i, \cdots, a_i}^{m+1}, \overbrace{a_j, \cdots, a_j}^{m-1}) + \phi(\overbrace{a_i, \cdots, a_i}^{m-1}, \overbrace{a_j, \cdots, a_j}^{m+1}) \geqq 2\phi(\overbrace{a_i, \cdots, a_i}^{m}, \overbrace{a_j, \cdots, a_j}^{m}) .$$

Summing, this readily yields

$$E_{\mathscr{S}[m+1,m-1]}\phi(\mathbf{X}) \geqq E_{\mathscr{S}[m,m]}\phi(\mathbf{X}) .$$

The remaining reasoning replicates the preceding case. The proof of Lemma 2.1 is complete.

LEMMA 2.2. *Let ϕ be as in Lemma 2.1. For any integer k_s and integers α, β satisfying $0 \leqq \beta < \alpha \leqq [k_s/2]$ we have*

$$E_{\mathscr{S}^*[k_1,\cdots,k_{s-1},k_s-\alpha,\alpha,k_{s+1},\cdots,k_r]}\phi(\mathbf{X}) \leqq E_{\mathscr{S}^*[k_1,\cdots,k_{s-1},k_s-\beta,\beta,k_{s+1},\cdots,k_r]}\phi(\mathbf{X}) . \tag{2.13}$$

REMARK. The sampling plan $\mathscr{S}^*[\gamma_1, \gamma_2, \cdots, \gamma_r]$, where the γ_i, $\sum \gamma_i = n$, constitute a set of integers but not necessarily appearing in decreasing order, is defined after Definition 1.4.

PROOF. The left side of (2.13) can be evaluated by conditioning on the variables X_i corresponding to the indices $k_1, k_2, \cdots, k_{s-1}, k_{s+1}, \cdots, k_r$ and reliance on the law of total probabilities. (We drop the asterisk notation for ease of writing.) We obtain

$$\begin{aligned} &E_{\mathscr{S}[k_1,\cdots,k_{s-1},k_s-\alpha,\alpha,k_{s+1},\cdots,k_r]}\phi(\mathbf{X}) \\ &\quad = E_{\mathscr{S}[k_1,\cdots,\hat{k}_s,\cdots,k_r]}E_{\mathscr{S}[k_s-\alpha,\alpha,\Omega-\{a_i^0\}]} \\ &\quad\quad \times \phi(\overbrace{a_{i_1}^0, \cdots, a_{i_1}^0}^{k_1}, \overbrace{a_{i_2}^0, \cdots, a_{i_2}^0}^{k_2}, \cdots, \overbrace{a_{i_r}^0, \cdots, a_{i_r}^0}^{k_r}, Y_1, \cdots, Y_{k_s}) \end{aligned} \tag{2.14}$$

where the symbol $\mathscr{S}[k_s-\alpha, \alpha, \Omega-\{a_i^0\}]$ refers to the sampling plan $\mathscr{S}[k_s-\alpha, \alpha]$ applied to the sample space Ω with $a_{i_1}^0, a_{i_2}^0, \cdots, a_{k_{s-1}}^0, a_{k_{s+1}}^0, \cdots, a_{i_r}^0$ removed. The $Y_1, \cdots, Y_{k_s}$ denote the random variables corresponding to the k_s distinguished variables, and the notation $\mathscr{S}[k_1, \cdots, \widehat{k_s}, \cdots, k_k]$ stands for the sampling plan induced by the parameters $[k_1, \cdots, k_{s-1}, k_{s+1}, \cdots, k_r]$ of sample size $n - k_s$.

By Lemma 2.1 we know that

$$\begin{aligned} &E_{\mathscr{S}[k_s-\alpha,\alpha,\Omega-\{a_i^0\}]}\phi(\overbrace{a_{i_1}^0, \cdots, a_{i_1}^0}^{k_1}, \cdots, \widehat{a_{i_s}^0}, \cdots, \overbrace{a_{i_r}^0, \cdots, a_{i_r}^0}^{k_r}, Y_1, \cdots, Y_{k_s}) \\ &\quad \leqq E_{\mathscr{S}[k_s-\beta,\beta,\Omega-\{a_i^0\}]}\phi(\overbrace{a_{i_1}^0, \cdots, a_{i_1}^0}^{k_1}, \cdots, \widehat{a_{i_s}^0}, \cdots, \\ &\quad\quad \overbrace{a_{i_r}^0, \cdots, a_{i_r}^0}^{k_r}, Y_1, \cdots, Y_{k_s}) . \end{aligned} \tag{2.15}$$

The result (2.13) now follows from (2.15) by virtue of the representation (2.14) and the corresponding one that applies for the right side of (2.13). □

A vital corollary emanating from the preceding theorem is the following.

LEMMA 2.3. *For $\phi \in \mathscr{C}$ and $k_{i_0} - k_{j_0} \geqq 2$, $i_0 < j_0$ we have*

$$E_{\mathscr{S}[k_1,k_2,\cdots,k_{i_0},\cdots,k_{j_0},\cdots,k_r]}\phi(\mathbf{X}) \geqq E_{\mathscr{S}[k_1,k_2,\cdots,k_{i_0}-1,\cdots,k_{j_0}+1,\cdots,k_r]}\phi(\mathbf{X}) . \tag{2.16}$$

PROOF. To prove (2.16) we coalesce the indices k_{i_0} and k_{j_0} to a single index $k_{i_0} + k_{j_0} = k^*$ and consider

$$E_{\mathscr{S}[k_1,k_2,\cdots,k^*,\cdots,k_r]}\phi(\mathbf{X}) .$$

We can clearly apply the inequality (2.13) to secure (2.16). □

We now have available the necessary ingredients to complete the proof of Theorem 2.1.

PROOF OF THEOREM 2.1. Consider $[\mathbf{k}] = [k_1, k_2, \cdots, k_p] \succ [l_1, l_2, \cdots, l_q] = [\mathbf{l}]$. By adding zero components if necessary we may without loss of generality assume $p = q$. Now define the differences

$$k_1 + k_2 + \cdots + k_i - (l_1 + l_2 + \cdots + l_i) = \gamma_i .$$

The stipulation $[\mathbf{k}] \succ [\mathbf{l}]$ entails

$$\gamma_i \geqq 0 \qquad i = 1, 2, \cdots, p . \tag{2.17}$$

Now define the index value

$$c = \sum_{i=1}^p \gamma_i = I([\mathbf{k}] - [\mathbf{l}]) . \tag{2.18}$$

If $c = 0$ so that all $\gamma_i = 0$ it follows that $[\mathbf{k}] \equiv [\mathbf{l}]$, and manifestly equality holds in (2.1).

Assume by induction that the inequality (2.1) is established whenever $c < r$. Suppose next that $I([\mathbf{k}] - [\mathbf{l}]) = r \geqq 1$. Let j_0 be the least index $2 \leqq j_0 \leqq p$ where $\gamma_j < \gamma_{j-1}$. Such a j_0 certainly exists since $r \geqq 1$ and $\gamma_p = 0$. Thus $0 \leqq \gamma_1 \leqq \gamma_2 \leqq \cdots \leqq \gamma_{j_0-1}$ but $\gamma_{j_0} < \gamma_{j_0-1}$. From the definition we manifestly have

$$k_i \geqq l_i \qquad i = 1, 2, \cdots, j_0 - 1 ,$$

but $k_{j_0} < l_{j_0}$ and because $\gamma_{j_0-1} > 0$ there certainly exists some $i \leqq j_0 - 1$ such that $k_i > l_i$. Take the largest such i, call it i_0 so that

$$k_{i_0} > l_{i_0} \geqq l_{j_0} > k_{j_0} \tag{2.19}$$

and

$$k_{i_0+1} = l_{i_0+1}, \quad k_{i_0+2} = l_{i_0+2}, \cdots, k_{j_0-1} = l_{j_0-1} . \tag{2.20}$$

Clearly $\gamma_{i_0} > 0$, and it follows from (2.19) that

$$k_{i_0} - k_{j_0} \geqq 2 .$$

Now form the new $[\mathbf{k}^*] = (k_1^*, \cdots, k_p^*) \in \mathscr{K}$ where

$$k_\nu^* = k_\nu \qquad \text{for all} \quad \nu \neq i_0, j_0$$

and

$$k_{i_0}^* = k_{i_0} - 1 , \qquad k_{j_0}^* = k_{j_0} + 1 . \tag{2.21}$$

Lemma 3.3 affirms that

$$E_{\mathscr{S}[\mathbf{k}]} \phi(\mathbf{X}) \geqq E_{\mathscr{S}[\mathbf{k}^*]} \phi(\mathbf{X}) . \tag{2.22}$$

Because of the hypotheses $[\mathbf{k}] \succ [\mathbf{l}]$ and the definition of i_0 and j_0 and reference to (2.19) and (2.20), we easily verify that

$$[\mathbf{k}^*] \succ [\mathbf{l}] . \tag{2.23}$$

Moreover

$$I([\mathbf{k}^*] - [\mathbf{l}]) < I([\mathbf{k}] - [\mathbf{l}]) = r$$

and the induction postulate applies, yielding

$$E_{\mathscr{S}[\mathbf{k}^*]}\phi(\mathbf{X}) \geqq E_{\mathscr{S}[\mathbf{l}]}\phi(\mathbf{X}) . \tag{2.24}$$

Combining (2.22) and (2.24) produces the desired conclusion and the sufficiency part of the proof of Theorem 2.1 is complete.[2]

It remains to establish that if the ordering relation (2.3) between [k] and [l] does not hold, then the inequality (2.1) cannot persist for all $\phi \in \mathscr{C}$.
Assume

$$E_{\mathscr{S}[\mathbf{k}]}\phi(\mathbf{X}) \geqq E_{\mathscr{S}[\mathbf{l}]}\phi(\mathbf{X}) \qquad \text{for all} \quad \phi \in \mathscr{C} \tag{2.25}$$

and suppose to the contrary that

$$k_1 < l_1 . \tag{2.26}$$

Take $\phi(\xi_1, \cdots, \xi_n) = \sum \xi_{i_1} \xi_{i_2} \cdots \xi_{i_{l_1}}$ where the sum extends over all choices of products of l_1 variables from the collection $\{\xi_1, \cdots, \xi_n\}$. It is easy to check that ϕ is of class $\mathscr{C}$ provided Ω contains only nonnegative numbers. The sample space Ω is chosen to contain one large element a and all others of moderate magnitude. With this specification we find that $E_{\mathscr{S}[\mathbf{k}]}\phi(\mathbf{X})$ is of order a^{k_1} while $E_{\mathscr{S}[\mathbf{l}]}\phi(\mathbf{X})$ is of the order a^{l_1} and (2.25) is violated. It follows that $k_1 \geqq l_1$. Suppose next that

$$k_1 \geqq l_1 \qquad \text{and introduce the notation} \quad \alpha = k_1 + k_2 , \quad \beta = l_1 + l_2 . \tag{2.27}$$

Now take the symmetric sum $\phi(\boldsymbol{\xi}) = \sum \xi_{i_1} \xi_{i_2} \cdots \xi_{i_\beta}$ involving all possible products of β variables from $\{\xi_1, \cdots, \xi_n\}$. Determine Ω to contain two large elements of equal value a and the remaining of small value. Then $E_{\mathscr{S}[\mathbf{k}]}\phi(\mathbf{X})$ is of the order at most a^α while

$$E_{\mathscr{S}[\mathbf{l}]}\phi(\mathbf{X}) \quad \text{is of the order} \quad a^\beta .$$

In order to guarantee (2.25), we must have $\beta \leqq \alpha$. Continuing in this vein we establish that (2.25) entails

$$[\mathbf{k}] \succ [\mathbf{l}] .$$

The proof of Theorem 2.1 is complete.

3. Comparison of random replacement schemes. Let $\mathscr{S}$ and $\tilde{\mathscr{S}}$ designate two alternate symmetric sampling plans. The problem of ascertaining criteria for comparing $E_{\mathscr{S}}\phi(\mathbf{X})$ and $E_{\tilde{\mathscr{S}}}\phi(\mathbf{X})$ for ϕ of class $\mathscr{C}$ (see Definition 1.3), where

[2] We have learned from the referee of this manuscript that the argument commencing with (2.17) and concluding with (2.24) can be replaced by appeal to the following result of Folkman and Fulkerson (1969). ("Edge colorings in bipartite graphs" in *Combinatorial Mathematics and its Applications*, Chapter 31, 561–577.)

"If $a = \{a_1, \cdots, a_p\}$, $b = \{b_1, \cdots, b_p\}$ are integers arranged in decreasing order adding to n, $a_1 + \cdots + a_k \leqq b_1 + \cdots + b_k$ holds for all $k = 1, 2, \cdots, p$ $(a < b)$, then a can be derived from b by successive applications of a finite number of transfers. The vector b' is a transfer of b means that for some i and j, $b_i' = b_i - 1$, $b_j' = b_j + 1$, $b_k' = b_k$, $k \neq i, j$."

We have retained our proof intact so that the argument is complete in itself.

$\mathbf{X} = (X_1, \cdots, X_n)$ is the observed sample of size n, is of intrinsic value in the theory of sampling from finite populations. By virtue of (1.9) the problem can be stated in the following terms. Consider sums of the form

$$\sum_{\mathbf{k}=[k_1,\cdots,k_p]} a_{[\mathbf{k}]} E_{\mathscr{S}[\mathbf{k}]} \phi(X_1, \cdots, X_n) = E_{\mathscr{S}} \phi(\mathbf{X}) \qquad (3.1)$$
$$\sum_{\mathbf{k}} b_{[\mathbf{k}]} E_{\tilde{\mathscr{S}}[\mathbf{k}]} \phi(X_1, \cdots, X_n) = E_{\tilde{\mathscr{S}}} \phi(\mathbf{X})$$

where $a_{[\mathbf{k}]}, b_{[\mathbf{k}]} \geqq 0$ and $\sum a_{[\mathbf{k}]} = \sum b_{[\mathbf{k}]} = 1$. We would like to determine necessary and/or sufficient conditions on the coefficient arrays $\{a_{[\mathbf{k}]}\}$ and $\{b_{[\mathbf{k}]}\}$, implying the equality

$$E_{\mathscr{S}} \phi(\mathbf{X}) \leqq E_{\tilde{\mathscr{S}}} \phi(\mathbf{X}) \qquad \text{for all} \quad \phi \in \mathscr{C}. \qquad (3.2)$$

Recall that the comparison inequalities relating the special symmetric sampling plans $E_{\mathscr{S}[\mathbf{k}]}\phi(\mathbf{X})$, as delineated in Theorem 2.1, obey only a partial ordering relationship. The results deduced in Corollaries 2.2 and 2.3 emanate, since $\mathscr{S}_{[1,1,\cdots,1]}$ and $\mathscr{S}_{[n,0,\cdots,0]}$ are extremal sampling schemes in the ordering prescription of (2.3). The relations (2.1) and (2.2) can be construed as a multivariate case of the *majorization* concept of Hardy, Littlewood and Pólya, Chapter 2.

The general problem of ascertaining conditions for (3.2) appears exceptionally formidable. We will concentrate attention on the circumstance of (3.2) where

$$\mathscr{S} = \mathscr{R}(\mathbf{\Pi}) \qquad \text{and} \qquad \mathscr{S}^* = \mathscr{R}(\mathbf{\Pi}^*) \qquad (3.3)$$

are two symmetric sampling plans induced by random replacement procedures (consult here Definition 1.2). The principal theorem established in this context is:

THEOREM 3.1. *Let $\mathscr{R}(\bar{\Pi}) = \mathscr{R}(\Pi_1, \Pi_2, \cdots, \Pi_{n-1})$ be a general random replacement procedure and, $\mathscr{R}(\mathbf{1}) = \mathscr{R}(1, 1, \cdots, 1) = \mathscr{I}$ of course corresponds to sampling with replacement. Let $\phi(\xi_1, \xi_2, \cdots, \xi_n)$ be of class $\mathscr{C}$. Then*

$$E_{\mathscr{R}(\Pi)} \phi(\mathbf{X}) \leqq E_{\mathscr{I}} \phi(\mathbf{X}) \qquad (3.4)$$

holds in the following cases:

(i) *N is appropriately large compared to n. For example, it suffices to have $(N/(N-1))^{n-1} \leqq n/(n-3)$;*

(ii) *for all $N \geqq n$ provided*

(a) *$\phi(\xi_1, \cdots, \xi_n) = (\xi_1 + \cdots + \xi_n)^2$ or where*

(b) *$\phi(\boldsymbol{\xi}) = \psi(\xi_1 + \cdots + \xi_n)$, $\psi(\xi)$ is convex and the sample points in Ω have only two distinct values, say r points of common value α and $N - r = s$ points of common value β.*

REMARK 1. It is surmised that (3.4) persists for all $\phi \in \mathscr{C}$ provided only that $N \geqq n$ (this constraint is essential). The validation of (3.4) subject only to the requirement $\phi \in \mathscr{C}$, for $N = n$ appears quite delicate. We have verified the inequality (3.4) whenever $N \geqq n$, $n \leqq 12$ by direct means (the method to be indicated later in this section). The calculation of $E_{\mathscr{I}} \phi(X_1, \cdots, X_n)$ is relatively easy, since for the sampling scheme $\mathscr{I}$, $X_1, \cdots, X_n$ are independently uniformly distributed on Ω. The inequalities (3.4) manifest $E_{\mathscr{I}}\,\phi_{(\mathbf{x})}$ as an upper bound for a large class of expectations with respect to "random replacement schemes." The universal lower bound is that prescribed in Theorem 2.1 corresponding to $\mathscr{R}_{\{0,0,\cdots,0\}} = \mathscr{W}$ (sampling without replacement).

The proof of Theorem 3.2 is divided into a series of separate lemmas.

For $E_{\mathscr{R}(\Pi)}\,\phi(\mathbf{X})$ it is convenient to employ sometimes the explicit notation $E_{\mathscr{R}(\Pi)}\,\phi(\mathbf{X}) = E_{\{\Pi_1,\Pi_2,\cdots,\Pi_{n-1}\}}\,\phi(X_1, X_2, \cdots, X_n)$.

The first lemma presents the reduction of the problem via an induction argument to a basic inequality underlying (3.4).

LEMMA 3.1. *If*

$$E_{\{0,1,1,\cdots,1\}}\,\phi(\mathbf{X}) \leqq E_{\{1,1,\cdots,1\}}\,\phi(\mathbf{X}) = E_{\mathscr{I}}\,\phi(\mathbf{X}) \tag{3.5}$$

holds for each n, and all $N \geqq n$ and all $\phi \in \mathscr{C}$, then

$$E_{\{\Pi_1,\Pi_2,\cdots,\Pi_{n-1}\}}\,\phi(\mathbf{X}) \leqq E_{\mathscr{I}}\,\phi(\mathbf{X}) \tag{3.6}$$

for any random replacement scheme $\mathscr{R}(\Pi)$.

PROOF. Assume inductively that (3.6) is established for all random replacement schemes of size at most $n-1$ ($N \geqq n-1$), and (3.5) applies for the random replacement schemes of parameters $\{0, 1, \cdots, 1\}$ and $\{1, \cdots, 1\}$ comprised of n components. Conditioning on the outcome of the first observation, we obtain

$$\begin{aligned} E_{\{\Pi_1,\Pi_2,\cdots,\Pi_{n-1}\}}\,&\phi(X_1, \cdots, X_n) \\ &= \frac{1}{N}\sum_{k=1}^{N} \pi_1 E_{\{\Pi_2,\cdots,\Pi_{n-1}\}}\,\phi(a_k, X_2, \cdots, X_n) \\ &\quad + \frac{1}{N}\sum_{k=1}^{N} (1-\pi_1) E'_{\{\Pi_2,\cdots,\Pi_{n-1}\}}\,\phi(a_k, X_2, \cdots, X_n) \end{aligned}$$

where the prime in the second sum signifies that the element a_k was removed from the sample space of the observations $X_2, \cdots, X_n$. Note that $\psi(\xi_2, \cdots, \xi_n) = \phi(a_k, \xi_2, \cdots, \xi_n)$ is of class $\mathscr{C}$ for each a_k. Invoking the induction assumption to the terms of each sum leads to

$$E_{\{\Pi_1,\cdots,\Pi_{n-1}\}}\,\phi(X_1, \cdots, X_n) \leqq \pi_1 E_{\mathscr{I}}\,\phi(\mathbf{X}) + (1-\pi_1)E_{\{0,1,\cdots,1\}}\,\phi(\mathbf{X})\,. \tag{3.7}$$

The stipulation of (3.5) applied to the last expectation of (3.7) yields (3.6). The proof of the lemma is complete.

We will next deal with the special case

$$\theta(\xi_1, \xi_2, \cdots, \xi_n) = (\xi_1 + \cdots + \xi_n)^2\,. \tag{3.8}$$

LEMMA 3.2. *The inequality* (3.5), *and therefore* (3.6), *holds for the univariate function $\theta(\xi) = \xi^2$ and extended as in* (3.8).

REMARK. The case of Lemma 3.2 was established earlier by Rosén involving a more complicated analysis. Rosén used the inequality (3.4) to have estimates for proving central limit theorems for certain finite sampling schemes.

PROOF. Define $Y_a = X_2 + \cdots + X_n$ as the random variable where $X_2, \cdots, X_n$ are i.i.d. following the uniform distribution on $\Omega - \{a\}$; i.e. all values of Ω with the element $\{a\}$ deleted sampled equally likely. Conditioning on the outcome of X_1, we obtain

$$(3.9) \qquad E_{\{0,1,\cdots,1\}}\theta(\mathbf{X}) = \frac{1}{N} \sum_{k=1}^{N} [a_k^2 + 2a_k E(Y_{a_k}) + EY_{a_k}^2]$$

and we give the sum the short name L.

Set

$$\mu_1 = \frac{1}{N} \sum_{k=1}^{N} a_k\,, \qquad \mu_2 = \frac{1}{N} \sum_{k=1}^{N} a_k^2\,.$$

The assertion of the lemma is confirmed once we check that L does not exceed

$$(3.10) \qquad E(Z_1 + \cdots + Z_n)^2 = n\mu_2 + n(n-1)\mu_1^2$$

where Z_i are independently uniformly distributed over Ω. Note that

$$(3.11) \qquad E(Y_{a_k}) = \frac{(N\mu_1 - a_k)(n-1)}{N-1}$$

$$EY_{a_k}^2 = (n-1)\frac{N\mu_2 - a_k^2}{N-1} + (n-1)(n-2)\left(\frac{N\mu_1 - a_k}{N-1}\right)^2.$$

Inserting (3.11) into (3.9) the required inequality, after some straightforward manipulations, reduces to

$$(3.12) \qquad (\mu_2 - \mu_1^2)\frac{(n-1)}{(N-1)^2}(2N-n) \geqq 0\,.$$

This relation is manifestly valid for $N \geqq n$ (even for $N \geqq n/2$) since the moment inequality $\mu_2 \geqq \mu_1^2$ constantly prevails. The proof of Lemma 3.2 is complete.

REMARK 2. The method of Lemma 3.2 can be pursued to validate (3.5) for the function $\phi(\xi_1, \cdots, \xi_n) = (\xi_1 + \cdots + \xi_n)^3$ provided all elements of Ω are nonnegative. However, this type of argument is already too crude to handle the example $\phi(\xi_1, \cdots, \xi_n) = (\xi_1 + \cdots + \xi_n)^4$.

We will next deal with case (i) of Theorem 3.1. First, we write out $E_{\{0,1,\cdots,1\}}\phi(X_1, X_2, \cdots, X_n)$ explicitly obtaining

$$(3.13) \qquad E_{\{0,1,\cdots,1\}}\phi(X_1, \cdots, X_n) = \frac{1}{N(N-1)^{n-1}} \sum_{a_{i_1},\cdots,a_{i_n} \in \Omega;\, a_{i_1} \neq a_{i_\nu},\, \nu=2,\cdots,n} \phi(a_{i_1}, a_{i_2}, \cdots, a_{i_n})$$

where the sum is extended over all n tuples $(a_{i_1}, \cdots, a_{i_n})$, $a_{i_\nu} \in \Omega$ for all ν subject to the restriction that a_{i_1} is distinct from the other a_{i_ν}.

In order to appropriately represent groupings for future comparisons of the terms in (3.13) the evaluation of the following combinatorial situation will be useful. Consider $n-1$ indistinguishable points arranged on a line. We inquire as to how many ways can $p-1$ groups be formed consisting of $(k_1, k_2, \cdots, k_{p-1})$

points, respectively for a prescribed sequence

$$(3.14)\qquad k_1 \geqq k_2 \geqq \cdots \geqq k_{p-1} \qquad \textstyle\sum_{i=1}^{p-1} k_i = n-1 \,.$$

This is an elementary known combinatorial query whose solution by careful counting is seen to be

$$(3.15)\qquad \binom{n-1}{k_1, k_2, \cdots, k_{p-1}} \frac{1}{\alpha_1!\,\alpha_2!\cdots\alpha_m!} = \frac{(n-1)!}{k_1!\,k_2!\cdots k_{p-1}!}\,\frac{1}{\alpha_1!\,\alpha_2!\cdots\alpha_m!}$$

where α_i are the integers defined by the relations

$$k_1 = k_2 = \cdots = k_{\alpha_1} > k_{\alpha_1+1} = \cdots = k_{\alpha_1+\alpha_2} > \cdots > k_{\alpha_1+\cdots+\alpha_{m-1}+1} = \cdots = k_{\alpha_1+\cdots+\alpha_m} \,.$$

Of course $\sum_{i=1}^{m} \alpha_i = p-1$.

Examine the sum in (3.13) and identify the 2nd to the nth variables of $\phi(a_{i_1}, a_{i_2}, \cdots, a_{i_n})$ with the points of the previous experimental setup.

Next let $y_1, y_2, \cdots, y_{p-1}, y_p$ be p-ordered distinct elements of Ω fixed for the moment. Consider all terms of (3.13) where y_i appears exactly k_i times ($i = 1, 2, \cdots, p-1$) among the $(a_{i_2}, \cdots, a_{i_n})$ and $y_p = a_{i_1}$. Owing to symmetry of $\phi(\xi_1, \cdots, \xi_n)$ we instantly see that each such term has the same value and the number of such summands is (3.15). With these facts, strongly exploiting the symmetry of ϕ, we can represent the sum of (3.13) in the form

$$(3.16)\qquad \begin{aligned} &\textstyle\sum_{a_{i_2},\cdots,a_{i_n}\in\Omega;\, a_{i_1}\neq a_{i_\nu}} \phi(a_{i_1}, a_{i_2}, \cdots, a_{i_n}) \\ &= \textstyle\sum_{[k]\in\mathscr{K}} \binom{n-1}{k_1, k_2, \cdots, k_{p-1}} \frac{1}{\alpha_1!\,\alpha_2!\cdots\alpha_m!} \sum_{(y_1,y_2,\cdots,y_p)} \\ &\quad\times \phi(y_p, \overbrace{y_1, \cdots, y_1}^{k_1}, \overbrace{y_2, \cdots, y_2}^{k_2}, \cdots, \overbrace{y_{p-1}, \cdots, y_{p-1}}^{k_{p-1}}) \end{aligned}$$

where the inner sum runs over all choices of p elements of Ω and the outer sum extends through all prescriptions of sets of positive integers obeying the stipulations of (3.14), p arbitrary. The number of terms in the inner sum is clearly the number of ordered arrangements of p selections out of N elements which is $N!/(N-p)!$. Note that

$$(3.17)\qquad \begin{aligned} &E_{\mathscr{S}[k_1,k_2,\cdots,k_{p-1},k_p]}\phi(\mathbf{X}) \\ &= \frac{(N-p)!}{N!} \textstyle\sum_{(y_1,y_2,\cdots,y_p)} \\ &\quad\times \phi(\overbrace{y_1, \cdots, y_1}^{k_1}, \overbrace{y_2, \cdots, y_2}^{k_2}, \cdots, \overbrace{y_{p-1}, \cdots, y_{p-1}}^{k_{p-1}}, y_p) \end{aligned}$$

for $\mathbf{k} = [k_1, k_2, \cdots, k_p]$ as in (3.14) with $k_p = 1$. In view of (3.16) and (3.17), (3.13) becomes

$$(3.18)\qquad \begin{aligned} E_{\mathscr{R}\{0,1,1,\cdots,1\}}\phi(\mathbf{X}) &= \frac{1}{N(N-1)^{n-1}} \textstyle\sum_{[k_1,\cdots,k_{p-1}]} \binom{n-1}{k_1, k_2, \cdots, k_{p-1}} \\ &\quad\times \frac{1}{\alpha_1!\,\alpha_2!\cdots\alpha_m!}\,\frac{N!}{(N-p)!}\, E_{\mathscr{S}[k_1,k_2,\cdots,k_p]}\phi(\mathbf{X}) \,. \end{aligned}$$

Executing a similar analysis and expansion, we obtain

$$(3.19)\qquad E_{\mathscr{R}\{1,\cdots,1\}}\phi(\mathbf{X}) = \frac{1}{N^n}\sum\nolimits_{[l_1,l_2,\cdots,l_q]}\frac{1}{\gamma_1!\,\gamma_2!\cdots\gamma_r!}\binom{n}{l_1,l_2,\cdots,l_q} \times \frac{N!}{(N-q)!}E_{\mathscr{S}[l]}\phi(\mathbf{X})$$

where $l_1 \geqq l_2 \geqq \cdots \geqq l_q$, $\sum l_i = n$ and $\{\gamma_i\}$ count the blocks of equal l values analogous to that in (3.15). Obviously, all the expectations

$$E_{\mathscr{S}[\mathbf{k}]} = E_{\mathscr{S}\{k_1,k_2,\cdots,k_p,1\}}\phi(\mathbf{X})$$

occurring in (3.18) also occur in (3.19) and the latter sum includes other terms not present in (3.18); e.g., $E_{\mathscr{S}[n]}\phi(\mathbf{X})$, $E_{\mathscr{S}[n-2,2]}\phi(\mathbf{X})$ and similarly. For the array of (3.14) enlarged with $k_p = 1$ define $\beta_1, \beta_2, \cdots, \beta_r$ in the manner of $\alpha_1, \alpha_2, \cdots, \alpha_m$ with respect to

$$\{k_1, k_2, \cdots, k_{p-1}, k_p\} \qquad k_p = 1\,.$$

Clearly

$$(3.20)\qquad \alpha_i = \beta_i\,, \quad i = 1, 2, m-1 \quad \text{and either} \quad r = m \quad \text{with} \quad \alpha_m + 1 = \beta_m \quad \text{or} \quad r = m+1 \quad \text{and} \quad \beta_{m+1} = 1\,.$$

We next compare the coefficients $a_{\mathbf{k}}$ and $b_{\mathbf{k}}$ accompanying $E_{\mathscr{S}[\mathbf{k}]}$ displayed in the sums (3.18) and (3.19) respectively. It follows directly that the inequality

$a_{\mathbf{k}} \leqq b_{\mathbf{k}}$ is equivalent to

$$(3.21)\qquad \frac{1}{N(N-1)^{n-1}}\frac{1}{\alpha_1!\,\alpha_2!\cdots\alpha_m!}\binom{n-1}{k_1,k_2,\cdots,k_{p-1}} \leqq \frac{1}{N^n}\frac{1}{\beta_1!\,\beta_2!\cdots\beta_r!}\binom{n}{k_1,k_2,\cdots,k_{p-1},1}$$

which on account of (3.20) reduces to

$$(3.22)\qquad \left(\frac{N}{N-1}\right)^{n-1} \leqq \frac{n}{\beta_r}\,.$$

For any $[k_1, k_2, \cdots, k_{p-1}, 1]$ with $\beta_r \leqq n-3$ we see that (3.22) holds whenever condition (i) of Theorem 3.1 is in force. Actually, the only circumstances when (3.22) does not apply are for the arrays

$$(3.23)\qquad [k_1, \cdots, k_p] = [2, 1, 1, 1, \cdots, 1] \qquad \text{and} \qquad [1, 1, 1, \cdots, 1]\,.$$

Thus $a_{\mathbf{k}} \leqq b_{\mathbf{k}}$ for all $[\mathbf{k}] = [k_1, k_2, \cdots, k_{p-1}, 1]$ other than the last two cited. However, the result of Theorem 2.1 tells us that

$$(3.24)\qquad E_{\mathscr{S}\{1,1,\cdots,1\}}\phi(\mathbf{X}) \leqq E_{\mathscr{S}\{2,1,1,\cdots,1\}}\phi(\mathbf{X}) \leqq E_{\mathscr{S}[l]}\phi(\mathbf{X})$$

for all other $[l] \in \mathscr{K}$. Now

$$(3.25)\qquad \begin{aligned} &E_{\mathscr{R}\{1,1,\cdots,1\}}\phi(\mathbf{X}) - E_{\mathscr{R}\{0,1,\cdots,1\}}\phi(\mathbf{X}) \\ &\quad = \sum\nolimits_l c_l E_{\mathscr{S}[l]}\phi(\mathbf{X}) + (b_{[2,1,\cdots,1]} - a_{[2,1,1,\cdots,1]})E_{\mathscr{S}\{2,1,\cdots,1\}}\phi(\mathbf{X}) \\ &\qquad + (b_{[1,1,\cdots,1]} - a_{[1,1,\cdots,1]})E_{\mathscr{S}\{1,1,\cdots,1\}}\phi(\mathbf{X}) \end{aligned}$$

where the sum extends over all $[l]$ other than the ones of (3.23).

The fact of (3.22) implies

$$c_{\mathbf{k}} \geqq 0 \qquad \text{for all} \quad \mathbf{k} = [k_1, k_2, \cdots, k_q] \tag{3.26}$$

except possibly for the $\mathbf{k}$ of (3.23). Direct inspection reveals that

$$b_{[1,1,\cdots,1]} < a_{[1,1,\cdots,1]} \,. \tag{3.27}$$

The sign of the coefficient of $E_{\mathscr{S}\{2,1,\cdots,1\}}\phi(\mathbf{X})$ in (3.25) is undetermined and unnecessary for our needs. Note when $\phi(\boldsymbol{\xi}) \equiv 1$ the quantity of (3.25) obviously vanishes. In view of the facts of (3.24)—(3.27), and since $\sum_{\mathbf{l}} c_{\mathbf{l}} + (b_{[2,1,\cdots,1]} - a_{[2,1,\cdots,1]}) + (b_{[1,\cdots,1]} - a_{[1,\cdots,1]}) = 0$, we deduce

$$\begin{aligned}
&\textstyle\sum_{\mathbf{l}} c_{\mathbf{l}} E_{\mathscr{S}[\mathbf{l}]}\phi(\mathbf{X}) + (b_{[2,1,\cdots,1]} - a_{[2,1,\cdots,1]})E_{\mathscr{S}[2,1,\cdots,1]}\phi(\mathbf{X}) \\
&\quad + (b_{[1,1,\cdots,1]} - a_{[1,1,\cdots,1]})E_{\mathscr{S}[1,1,\cdots,1]}\phi(\mathbf{X}) \\
&\qquad > \{\textstyle\sum_{[\mathbf{l}]} c_{\mathbf{l}} + b_{[2,1,1,\cdots,1]} - a_{[2,1,1,\cdots,1]} + b_{[1,1,\cdots,1]} - a_{[1,1,\cdots,1]}\}E_{\mathscr{S}[2,1,\cdots,1]}\phi(\mathbf{X}) \\
&\qquad = 0 \,.
\end{aligned}$$

Therefore the inequality is established:

$$E_{\mathscr{R}\{1,\cdots,1\}}\phi(\mathbf{X}) > E_{\mathscr{R}\{0,1,\cdots,1\}}\phi(\mathbf{X}) \,.$$

Summing up this discussion we have case (i) of Theorem 3.1. This is now stated formally.

LEMMA 3.3. *Provided* $(N/(N-1))^{n-1} \leqq n/(n-3)$, *then* (3.6) *holds.*

We continue the developments pertaining to Theorem 3.1 concerned now with case (ii b). In this circumstance Ω consists of r points of value α and s points of value β $(N = r + s \geqq n)$. Consider $\phi \in \mathscr{C}$ of the form $\phi(\boldsymbol{\xi}) = \phi(\xi_1 + \xi_2 + \cdots + \xi_n)$. We will compute and compare

$$E_{\mathscr{R}\{0,1,\cdots,1\}}\phi(\mathbf{X}) \,, \qquad E_{\mathscr{R}\{1,1,\cdots,1\}}\phi(\mathbf{X}) \,. \tag{3.28}$$

By a change of scale and location, without restricting generality, we may take henceforth $\alpha = 0$, $\beta = 1$. In order to guarantee $\phi \in \mathscr{C}$ we stipulate $\phi(\eta)$ *to be convex* for $\eta \geqq 0$.

Direct evaluation for the case at hand gives

$$\begin{aligned}
&E_{\mathscr{R}\{0,1,\cdots,1\}}\phi(\mathbf{X}) \\
&\quad = \textstyle\sum_{k=0}^{n} \phi(k) \dfrac{r}{r+s}\dbinom{n-1}{k}\left(\dfrac{s}{r+s-1}\right)^{k}\left(\dfrac{r-1}{r+s-1}\right)^{n-1-k} \\
&\qquad + \textstyle\sum_{k=0}^{n} \phi(k) \dfrac{s}{r+s}\dbinom{n-1}{k-1}\left(\dfrac{s-1}{r+s-1}\right)^{k-1}\left(\dfrac{r}{r+s-1}\right)^{n-k}
\end{aligned} \tag{3.29}$$

and

$$E_{\mathscr{R}\{1,\cdots,1\}}\phi(\mathbf{X}) = \textstyle\sum_{k=0}^{n} \phi(k)\dbinom{n}{k}\left(\dfrac{s}{r+s}\right)^{k}\left(\dfrac{r}{r+s}\right)^{n-k} . \tag{3.30}$$

The difference of (3.29) and (3.30) apart from the factor n is

$$U = \sum_{k=0}^{n} \phi(k)\delta_k \Big[(n-k)\Big(\frac{r-1}{r}\Big)^{n-1-k} + k\Big(\frac{s-1}{s}\Big)^{k-1} - n\Big(\frac{r+s-1}{r+s}\Big)^{n-1} \Big] \tag{3.31}$$

with

$$\delta_k = \frac{r^{n-k}s^k}{(r+s)(r+s-1)^{n-1}} \binom{n}{k} .$$

We will establish that the sequence

$$c_k = (n-k)\lambda^{n-1-k} + k\mu^{k-1} - \gamma , \qquad k = 0, 1, 2, \cdots, n , \tag{3.32}$$

with $\lambda = (r-1)/r$, $\mu = (s-1)/s$, $\gamma = n((r+s-1)/(r+s))^{n-1}$, changes sign precisely twice in the order $-$, $+$, $-$. Interpretation of the quantities involved reveals the equations

$$\sum_k \delta_k c_k = \sum_k k\delta_k c_k = 0 . \tag{3.33}$$

In view of this fact and appeal to the theory of generalized convex function (see [4], Chapter 11) it follows that the expression U of (3.31) is non-positive.

The two moment equations (3.33) imply that $\{c_k\}$ changes sign at least twice (see [4], page 409). A direct check produces the inequalities $c_0 < 0$ and $c_n < 0$.

To prove the assertion of (3.32), it is enough to demonstrate that the derivative of $c(x) = (n-x)\lambda^{n-1-x} + x\mu^{x-1} - \gamma$ vanishes once as x traverses the interval $[0, n]$. The relevant zeros of $c'(x)$ (i.e. those located in $[0, n]$) satisfy $-[(n-x)\log\lambda + 1]\lambda^{n-1-x} + [x\log\mu + 1]\mu^{x-1} = 0$ or, equivalently, are solutions of

$$K(\mu\lambda)^x = \frac{(n-x)\log\lambda + 1}{x\log\mu + 1} , \qquad K = \frac{1}{\lambda^{n-1}\mu} . \tag{3.34}$$

Analysis of (3.34) reveals that if the inequality

$$A = \log\lambda + \log\mu + n(\log\lambda)(\log\mu) < 0 \tag{3.35}$$

prevails, then the right side of (3.34) increases while $(\lambda\mu)^x$ is decreasing in x (since $\log\lambda < 0$ and $\log\mu < 0$). For $r \geqq n$ and $s \geqq n$ we claim the validity of (3.35). In fact, since A is bilinear in $\log\lambda$ and $\log\mu$ it is enough to check (3.35) when ($\mu = 1$, $\lambda = (n-1)/n$), ($\mu = (n-1)/n$, $\lambda = 1$) and ($\mu = \lambda = (n-1)/n$). The first two cases are immediate and in the last case we have

$$A = \Big(\log\frac{n-1}{n}\Big)\Big[2 + n\log\frac{n-1}{n}\Big] < 0$$

when $n \geqq 2$. Thus for $r \geqq n$, $s \geqq n$, the relation (3.34) admits a single solution for $x > 0$, and therefore because $c(x)$ changes sign at most twice it must change sign exactly twice. The theory of generalized convex functions, as referred to earlier, (see especially [4], Theorem 5.4, Chapter 11) implies the conclusion that the quantity of (3.31) is $\leqq 0$.

We extend the result to the general case of (ii b) where we only stipulate $r, s \geqq 1$, $r + s \geqq n$.

The elementary theory of exponential polynomials tells us that there can exist at most three solutions of (3.34) and indeed either one or three are possible. In order to pass from that of a single zero to that of three there must be values r and s where (3.34) exhibits a zero of multiplicity 3. Writing down the conditions for the existence of a third order zero, we get at such an x point, the additional equations

$$(3.36) \qquad K(\mu\lambda)^x(\log \mu\lambda) = \frac{-A}{(x \log \mu + 1)^2}$$

$$K(\mu\lambda)^x(\log \mu\lambda)^2 = \frac{+2A \log \mu}{(x \log \mu + 1)^3}$$

with A defined in (3.35).

Eliminating x and K from (3.36) and (3.34) leads to the equation

$$(3.37) \qquad \tilde{B} = (\log \mu\lambda)[n(\log \mu)(\log \lambda) + \log \mu\lambda] + 4(\log \mu)(\log \lambda) = 0 .$$

Clearly since $\log \mu < 0$ and $\log \lambda < 0$, $\tilde{B}$ is diminished by replacing n by $r + s$ (recall $n \leqq r + s$). Doing this we write the resulting expression as B. We will show that $B > 0$ for $\lambda \geqq \frac{2}{3}$ and $\mu \geqq \frac{2}{3}$ $(r, s \geqq 2)$. (The case of $r = 1$ or $s = 1$ can be treated separately, where we can then check directly that $c(k)$ changes sign exactly twice; we omit this verification.)

Writing $C = (\log \lambda\mu)[(r + s)(\log \mu)(\log \lambda) + \log \mu\lambda] + D(\log \lambda)(\log \mu)$ we will prove, provided $D > 4(2 \log 2 - 1) \approx 1.6$, that $C > 0$. To this end, define

$$\phi(r, s) = \left[\log\left(1 - \frac{1}{r}\right) + \log\left(1 - \frac{1}{s}\right)\right] \times \left[r + s + \frac{1}{\log(1 - 1/r)} + \frac{1}{\log(1 - 1/s)}\right].$$

The inequality $C > 0$ is equivalent to $D > -\phi(r, s)$. We prove next that $\partial\phi/\partial r = \partial\phi/\partial s > 0$ for all $r > 2$. Then

$$D \geqq -\phi(2, 2) = 2 \log 2 \left(4 - \frac{2}{\log 2}\right) = 4(2 \log 2 - 1) \geqq -\phi(r, s)$$

for $r \geqq 2$, $s \geqq 2$ obtains.

A calculation gives

$$\frac{\partial\phi}{\partial r} = \frac{1}{r(r-1)}\left[r + s + \frac{1}{\log(1 - 1/s)} + \frac{1}{\log(1 - 1/r)}\right] + \left\{\log\left(1 - \frac{1}{r}\right)\left(1 - \frac{1}{s}\right)\right\}\left[1 - \frac{1}{r(r-1)\log^2(1 - 1/r)}\right].$$

The factor $r + [\log(1 - 1/r)]^{-1} = [r \log(1 - 1/r) + 1][\log(1 - 1/r)]^{-1}$ is positive since $r \log(1 - 1/r) + 1 < 0$. Similarly, we have $s + [\log(1 - 1/s)]^{-1} > 0$. Now we also verify $1 - [r(r-1)\log^2(1 - 1/r)]^{-1} > 0$ by expansion of $\log(1 - 1/r)$

in a series. The inequality $\partial\phi/\partial r > 0$, and analogously $\partial\phi/\partial s > 0$, is established. Thus for $r \geqq 2$, $s \geqq 2$ the sharp inequality $(\log \mu\lambda)[(r+s)(\log \mu)(\log \lambda) + \log \mu\lambda] + 4(2\log 2 - 1)(\log \mu)(\log \lambda) > 0$ holds.

Since B never vanishes we infer that (3.34) in all cases admits a single zero. This fact entails the property stated in connection with (3.32), and consequently $U \leqq 0$, as earlier. The above discussion contains the proof of case (ii b) of Theorem 3.1. The proof of Theorem 3.1 is complete.

It is instructive to exemplify further the analysis used in Lemma 3.3. Specifically, we will verify

$$E_{\mathscr{R}(0,1,\cdots,1)}\phi(\mathbf{X}) \leqq E_{\mathscr{R}(1,1,\cdots,1)}\phi(\mathbf{X}) \tag{3.38}$$

for all cases of $n \leqq 6$ with $\phi \in \mathscr{C}$.

Case ($n = 3$). We have

$$\begin{aligned} E_{\mathscr{R}\{0,1\}}\phi(\mathbf{X}) &= \frac{1}{N(N-1)^2} \textstyle\sum_{j\neq i; k\neq i} \phi(a_i, a_j, a_k) \\ &= \frac{1}{N-1} E_{\mathscr{S}[2,1]}\phi(\mathbf{X}) + \frac{N-2}{N-1} E_{\mathscr{S}[1,1,1]}\phi(\mathbf{X}) . \end{aligned} \tag{3.39}$$

Analogously,

$$\begin{aligned} E_{\mathscr{R}\{1,1\}}\phi(\mathbf{X}) &= \frac{1}{N^2} E_{\mathscr{S}[3]}\phi(\mathbf{X}) + \frac{3(N-1)}{N^2} E_{\mathscr{S}[2,1]}\phi(\mathbf{X}) \\ &\quad + \frac{(N-1)(N-2)}{N^2} E_{\mathscr{S}[1,1,1]}(\mathbf{X}) . \end{aligned} \tag{3.40}$$

Since $\mathscr{S}_{[3]} > \mathscr{S}_{[2,1]} > \mathscr{S}_{[1,1,1]}$ (the ordering is meant in the sense of (2.3)) and combining coefficients of (3.39) and (3.40) the result $E_{\mathscr{R}[1,1]}\phi(\mathbf{X}) \geqq E_{\mathscr{R}[0,1]}\phi(\mathbf{X})$ ensues, provided $N \geqq 3$.

Case ($n = 4$). We have

$$\begin{aligned} E_{\mathscr{R}\{0,1,1\}}\phi(\mathbf{X}) &= \frac{1}{N(N-1)^3} [N(N-1)E_{\mathscr{S}[3,1]}\phi(\mathbf{X}) \\ &\quad + \tbinom{3}{1}N(N-1)(N-2)E_{\mathscr{S}[2,1,1]}\phi(\mathbf{X}) \\ &\quad + N(N-1)(N-2)(N-3)E_{\mathscr{S}[1,1,1,1]}\phi(\mathbf{X})] ; \\ E_{\mathscr{R}[1,1,1]}\phi(\mathbf{X}) &= \frac{1}{N^4} [NE_{\mathscr{S}[4]}\phi(\mathbf{X}) + 4N(N-1)E_{\mathscr{S}[3,1]}\phi(\mathbf{X}) \\ &\quad + \frac{1}{2!}\binom{4}{2} N(N-1)E_{\mathscr{S}[2,2]}\phi(\mathbf{X}) \\ &\quad + \frac{1}{2!}\binom{4}{2,\,1,\,1} N(N-1)(N-2)E_{\mathscr{S}[2,1,1]}\phi(\mathbf{X}) \\ &\quad + N(N-1)(N-2)(N-3)E_{\mathscr{S}[1,1,1,1]}\phi(\mathbf{X}) . \end{aligned}$$

Again using the ordering (see (2.1))

$$\mathscr{S}[4] > \mathscr{S}[3,1] > \mathscr{S}[2,2] > \mathscr{S}[2,1,1] > \mathscr{S}[1,1,1,1]$$

we deduce the desired inequality for $N \geqq 4$.

We illustrate finally the analysis for the case $n = 6$. (*We suppress the* $\phi(\mathbf{X})$ *in the expectation symbols after the equation sign.*)

$$E_{\mathscr{R}[0,1,1,1,1]}\phi(\mathbf{X})$$

$$\begin{aligned} &= \frac{1}{N(N-1)^5}\Big\{N(N-1)E_{\mathscr{S}[5,1]} + \tbinom{5}{4,1}N(N-1)(N-2)E_{\mathscr{S}[4,1,1]} \\ &\quad + \tbinom{5}{3,2}N(N-1)(N-2)E_{\mathscr{S}[3,2,1]} \\ &\quad + \frac{1}{2!}\binom{5}{3,1,1}N(N-1)(N-2)(N-3)E_{\mathscr{S}[3,1,1,1]} \\ &\quad + \binom{5}{2,2,1}\frac{1}{2!}N(N-1)(N-2)(N-3)E_{\mathscr{S}[2,2,1,1]} \\ &\quad + \frac{1}{3!}\binom{5}{2,1,1,1}N(N-1)(N-2)(N-3)(N-4)E_{\mathscr{S}[2,1,1,1,1]} \\ &\quad + N(N-1)(N-2)(N-3)(N-4)(N-5)E_{\mathscr{S}[1,1,1,1,1,1]}\Big\}. \end{aligned} \tag{3.41}$$

$$E_{\mathscr{R}[1,1,1,1,1]}$$

$$\begin{aligned} &= \frac{1}{N^6}\Big[NE_{\mathscr{S}[6]} + \tbinom{6}{1}N(N-1)E_{\mathscr{S}[5,1]} + \tbinom{6}{2}N(N-1)E_{\mathscr{S}[4,2]} \\ &\quad + \frac{1}{2!}\binom{6}{4,1,1}N(N-1)(N-2)E_{\mathscr{S}[4,1,1]} \\ &\quad + \frac{1}{2!}\binom{6}{3,3}N(N-1)E_{\mathscr{S}[3,3]} + \tbinom{6}{3,2,1}N(N-1)(N-2)E_{\mathscr{S}[3,2,1]} \\ &\quad + \frac{1}{3!}\binom{6}{3,1,1,1}N(N-1)(N-2)(N-3)E_{\mathscr{S}[3,1,1,1]} \\ &\quad + \frac{1}{3!}\binom{6}{2,2,2}N(N-1)(N-2)E_{\mathscr{S}[2,2,2]} \\ &\quad + \frac{1}{2!\,2!}\binom{6}{2,2,1,1}N(N-1)(N-2)(N-3)E_{\mathscr{S}[2,2,1,1]} \\ &\quad + \frac{1}{4!}\binom{6}{2,1,1,1,1}N(N-1)(N-2)(N-3)(N-4)E_{\mathscr{S}[2,1,1,1,1]} \\ &\quad + N(N-1)(N-2)(N-3)(N-4)(N-5)E_{\mathscr{S}[1,1,1,1,1,1]}\Big]. \end{aligned} \tag{3.42}$$

We have the ordering relations

$$\mathscr{S}_{[6]} > \mathscr{S}_{[5,1]} > \mathscr{S}_{[4,2]} \begin{matrix} > \mathscr{S}_{[4,1,1]} \\ > \mathscr{S}_{[3,3]} \end{matrix} > \mathscr{S}_{[3,2,1]} > \tag{3.43}$$

$$\begin{matrix} \mathscr{S}_{[3,1,1,1]} \\ \mathscr{S}_{[2,2,2]} \end{matrix} > \mathscr{S}_{[2,2,1,1]} > \mathscr{S}_{[2,1,1,1,1]} > \mathscr{S}_{[1,1,1,1,1,1]}. \tag{3.44}$$

The coefficient of $E_{\mathscr{S}[3,1,1,1]}$ on the right of (3.41) does not exceed that occurring in (3.42). It is necessary now to use the differences of the coefficients of those in (3.43) to dominate out this negative part. Using the fact that all the coefficients of (3.43) in (3.42) exceed the corresponding terms in (3.41), the result $E_{\mathscr{R}[1,1,1,1,1]} \geqq E_{\mathscr{R}[0,1,1,1,1]}$ then obtains.

In pursuing this method for larger values of n, the lack of a complete ordering among $\mathscr{S}[\mathbf{k}]$ as $\mathbf{k}$ varies makes the comparisons of $E_{\mathscr{R}[0,1,\cdots,1]}\phi(\mathbf{X})$ and $E_{\mathscr{R}[1,1,\cdots,1]}\phi(\mathbf{X})$ quite delicate. We have actually verified the inequality of (3.4) for n up to 12. We are convinced of its universal validity, provided, of course that $N \geqq n$. Theorem 3.1 shows that (3.4) certainly prevails if N is large enough compared to n, and this fact usually suffices for all practical applications.

REMARK. It is worth pointing out that examples can be constructed to show that already for $n = 4$ the quantities

$$E_{\mathscr{R}[0,1,1,1]}\phi(\mathbf{X}) \qquad \text{and} \qquad E_{\mathscr{R}[1,0,1,1]}\phi(\mathbf{X})$$

are not comparable for all ϕ of class $\mathscr{C}$. We omit the details on this matter.

We conclude this section by stating our general surmise pertaining to random replacement schemes.

Consider $\mathscr{R}(\Pi_1, \Pi_2, \cdots, \Pi_{n-1})$ and $\mathscr{R}(\Pi_1', \Pi_2', \cdots, \Pi_{n-1}')$. It is conjectured that

$$E_{\mathscr{R}[\Pi_1,\cdots,\Pi_{n-1}]}\phi(\mathbf{X}) \leqq E_{\mathscr{R}[\Pi_1',\Pi_2',\cdots,\Pi_{n-1}']}\phi(\mathbf{X}) \tag{3.45}$$

holds for all ϕ in $\mathscr{C}$ iff

$$\Pi_i \leqq \Pi_i', \qquad i = 1, 2, \cdots, n-1 \ \text{ and } \ N \geqq n\,. \tag{3.46}$$

Theorem 3.1 affirms (3.45) when $\mathbf{\Pi}' = (1, 1, \cdots, 1)$ while Theorem 2.1 subsumes the result where $\mathbf{\Pi} = (0, 0, \cdots, 0)$.

4. Dilations of measures associated with symmetric sampling plans. Consider a specified sample space $\Omega = \{a_1, a_2, \cdots, a_N\}$ with the a_i regarded as distinct even if they agree. For a symmetric sampling plan $\mathscr{S}$ on Ω with observation $\mathbf{X} = (X_1, \cdots, X_n)$ consisting of n-value we associate the integer-valued N-tuple random vector

$$R = (R_1, R_2, \cdots, R_N) \tag{4.1}$$

such that

$$R_i = \{\text{number of components among } (X_1, X_2, \cdots, X_n) \text{ equal to } a_i\}\,, \quad i = 1, 2, \cdots, N\,. \tag{4.2}$$

Obviously, each R_i is a nonnegative integer and these obey the constraints

$$0 \leqq R_i\,, \qquad \textstyle\sum_{i=1}^{N} R_i = n\,. \tag{4.3}$$

Designate the discrete *simplex of N-tuples of integers* determined by the relations (4.3) as Δ_N.

Let f be a *convex function* defined on the discrete simplex Δ_N. Convexity of f

can be interpreted in our present context in the manner that for any $\mathbf{R}$ and $\mathbf{R}^*$ of Δ_N such that $\frac{1}{2}(\mathbf{R} + \mathbf{R}^*)$ also belongs to Δ_N then $f(\frac{1}{2}(\mathbf{R} + \mathbf{R}^*)) \leqq \frac{1}{2}f(\mathbf{R}) + \frac{1}{2}f(\mathbf{R}^*)$ holds.

We would like to compare $E_{\mathscr{S}}f(\mathbf{R})$ and $E_{\mathscr{S}'}f(\mathbf{R})$ for two different symmetric sampling schemes $\mathscr{S}$ and $\mathscr{S}'$. To this end we introduce a function $\phi(\xi_1, \xi_2, \cdots, \xi_n)$ defined on $\Omega^n = \Omega \otimes \Omega \otimes \cdots \otimes \Omega$ (n copies of Ω) in terms of f as follows.

Let $\boldsymbol{\xi} = (\xi_1, \xi_2, \cdots, \xi_n)$ be in Ω^n composed of $k_\mu(\boldsymbol{\xi})$ components equal to a_μ, $\mu = 1, 2, \cdots, N$. Set

$$\phi_f(\xi_1, \cdots, \xi_n) = f(\mathbf{R}) \qquad \text{where} \quad \mathbf{R}(\xi) = (k_1(\boldsymbol{\xi}), \cdots, k_N(\boldsymbol{\xi})) , \tag{4.4}$$

and $\mathbf{R}(\xi)$ obviously is a point of Δ_n.

We claim as a consequence of the convexity of f that this ϕ is of class $\mathscr{C}$ (consult Definition 1.3). Note first that ϕ is manifestly a symmetric function of its argument as the arrangement of the components in $\boldsymbol{\xi}$ does not affect the determination of $\mathbf{R}(\xi)$. Next, to check condition (1.7), take

$$\boldsymbol{\xi} = (a_\alpha, a_\beta, \xi_3, \cdots, \xi_n)$$
$$\boldsymbol{\eta} = (a_\beta, a_\beta, \xi_3, \cdots, \xi_n)$$

sharing the same $\xi_3, \xi_4, \cdots, \xi_n$ entries, and suppose $a_\alpha(a_\beta)$ occurs $l_\alpha(l_\beta)$ times in the collection $\{\xi_3, \cdots, \xi_n\}$. Then

$$\mathbf{R}(\xi) = (k_1, \cdots, l_\alpha + 2, \cdots, l_\beta, \cdots, k_N)$$
$$\mathbf{R}(\eta) = (k_1, \cdots, l_\alpha, \cdots, l_\beta + 2, \cdots, k_N)$$

where $\mathbf{R}(\xi)$ and $\mathbf{R}(\eta)$ agree in all components except in the αth and βth as indicated. Observe that

$$\boldsymbol{\zeta} = (a_\alpha, a_\beta, \xi_3, \cdots, \xi_n)$$

corresponds to

$$\mathbf{R}(\zeta) = (k_1, \cdots, l_\alpha + 1, \cdots, l_\beta + 1, \cdots, k_N) = \frac{\mathbf{R}(\xi) + \mathbf{R}(\eta)}{2} .$$

The inequality (1.7) i.e.,

$$\phi_f(\boldsymbol{\xi}) + \phi_f(\boldsymbol{\eta}) \geqq 2\phi_f(\zeta) \tag{4.5}$$

reduces instantly to

$$f(\mathbf{R}(\xi)) + f(\mathbf{R}(\eta)) \geqq 2f(\mathbf{R}(\zeta)) \tag{4.6}$$

which is valid by virtue of the convexity stipulation on f. Thus ϕ as determined in (4.4) is of class $\mathscr{C}$.

From the definitions, we have

$$E_{\mathscr{S}}\phi_f(\mathbf{X}) = E_{\mathscr{S}}f(\mathbf{R}) . \tag{4.7}$$

Appealing to Lemma 2.1, Theorem 2.1 and its corollaries, we secure the conclusion of the next theorem.

THEOREM 4.1. *Let $f(R_1, \cdots, R_N)$ be a convex function defined on Δ_N (see (4.3)). Then, if the ordering relation* $[\mathbf{k}] < [\mathbf{l}]$ *in the sense of* (2.3) *holds, the inequality*

$$E_{\mathscr{S}[\mathbf{k}]}f(\mathbf{R}) \leqq E_{\mathscr{S}[\mathbf{l}]}f(\mathbf{R})$$

maintains. (For the definitions of the sampling plans $\mathscr{S}[\mathbf{k}]$ and $\mathscr{S}[\mathbf{l}]$, see Definition 1.4.)

In particular, for a general symmetric sampling plan $\mathscr{S}$,

$$E_{\mathscr{W}}f(\mathbf{R}) \leqq E_{\mathscr{S}}f(\mathbf{R}) \leqq E_{\mathscr{S}[n,0,\cdots,0]}f(\mathbf{R}) \tag{4.8}$$

where $\mathscr{W}$ = sampling without replacement.

Examination of $\mathbf{R}$ reveals that this random vector is defined independently of the specific $\{a_i\}$ values that make up Ω and depends only on the sampling scheme $\mathscr{S}$. Thus for sampling with replacement $\mathbf{R}$ consists of the ordered vector counting the number of different elements observed in the sample, and the values of the elements are irrelevant.

An application of Theorem 3.1 in the present context yields the result

THEOREM 4.2. *Let f be as in Theorem* 4.1. *Suppose N and n satisfy $(N/(N-1))^{n-1} \leqq n/(n-2)$ (probably $N \geqq n$ suffices), then for any random replacement scheme $\mathscr{R}(\boldsymbol{\pi})$ (see Definision* 1.2), *we have*

$$E_{\mathscr{R}(\pi)}f(R_1, R_2, \cdots, R_N) \leqq E_{\mathscr{I}}f(R_1, \cdots, R_N) \tag{4.9}$$

for $\mathscr{I}$ = sampling with replacement.

The assertions of (4.8) and (4.9) can be expressed in the language of dilations of measure.

DEFINITION 4.1. Let K be a convex and compact metrizable subset of a locally convex topological space, such as a Euclidean space. Let λ and μ be given probability measures on K.

The measure μ is said to be a dilation of λ (written symbolically $\lambda < \mu$) if there exists a Markov kernel $P(y, A)$ with $y \in K$, $A \subset K$ such that

$$\mu(A) = \int P(y, A)\lambda(dy) \qquad \text{for all measurable} \quad A \subset K \tag{4.10}$$

and

$$y = \int zP(y, dz) \qquad \text{for all} \quad y \in K\,. \tag{4.11}$$

This amounts to a transformation of the λ-mass distribution into the μ-mass distribution by spreading out a unit mass at $y \in K$ to a mass distribution $\nu_y(A) = P(y, A)$ having its center of gravity at y. From Jensen's inequality we find that $\lambda < \mu$ implies that

$$\int \phi(y)\lambda(dy) \leqq \int \phi(y)\mu(dy) \qquad \text{for all} \quad \phi \in \Gamma(K)\,. \tag{4.12}$$

Here, $\Gamma(K)$ denotes the collection of real-valued continuous and convex functions on K.

If Y and Z are random variables taking values in K we write $Y \prec Z$ if the relation $\lambda \prec \mu$ holds for the corresponding probability measures $\lambda(A) = \Pr(Y \in A)$ and $\mu(A) = \Pr(Z \in A)$. It follows that

$$E\phi(Y) \leqq E\phi(Z) \qquad \text{for each} \quad \phi \in \Gamma(K) .$$

Condition (4.12) is not only necessary, but also sufficient, in order that $\lambda \prec \mu$.

In view of the above formulation, the inequalities (4.8) applying to all convex f defined on Δ_N can be expressed such that the measures $P_{\mathscr{W}}$, $P_{\mathscr{S}}$ for the sampling schemes $\mathscr{W}$ and $\mathscr{S}$ restricted to $\mathbf{R}$ (the induced measure of $P_{\mathscr{S}}$ on the sample space of $\mathbf{R}$ is henceforth denoted by $\lambda_{\mathscr{S}}$), satisfy

$$\lambda_{\mathscr{W}} \prec \lambda_{\mathscr{S}} \qquad \text{(i.e., } \lambda_{\mathscr{S}} \text{ dilates } \lambda_{\mathscr{W}}\text{)} \tag{4.13}$$

for any symmetric sampling plan $\mathscr{S}$. In particular,

$$\lambda_{\mathscr{W}} \prec \lambda_{\mathscr{I}} \tag{4.14}$$

states that the procedure of sampling with replacement dilates the procedure of sampling without replacement. Thus, in the statistical theoretic sense, sampling with replacement wastes effort in gaining a quantity of information on the underlying population as compared to sampling without replacement. The special case (4.14) was established first by Kemperman (1973) who proved (4.14) by explicitly constructing the Markov kernel connecting $\lambda_{\mathscr{W}}$ and $\lambda_{\mathscr{I}}$. Such an approach appears to be impractical for comparing two general symmetric sampling schemes. We are much indebted to Kemperman, who by his findings, stimulated us to uncover the general facts of Theorems 4.1 and 4.2. They fall out as nice applications of Theorems 2.1 and 3.1. Some special cases of Theorem 4.5 below were also discovered earlier by Kemperman.

The conclusion of (4.9) can be phrased in the form

$$\lambda_{\mathscr{R}(\pi)} \prec \lambda_{\mathscr{I}}$$

so that *sampling with replacement dilates any random replacement scheme.*

The results of Theorems 4.1 and 4.2 can be extended to the following framework. Suppose Ω is composed of q groups of distinct elements $a_1, a_2, \cdots, a_q$ with a_i replicated N_i times ($\sum_{i=1}^{q} N_i = N$), N_i prescribed and fixed. To any symmetric sampling scheme $(\mathscr{S}, P_{\mathscr{S}})$, associate the vector random variable $\mathbf{Y} = (Y_1, Y_2, \cdots, Y_q)$ where Y_i is the number of appearances of a_i in the n-sample $X = (X_1, \cdots, X_n)$.

Let $\sum_q$ denote the simplex of nonnegative integers

$$y_i \geqq 0 , \qquad \textstyle\sum_{i=1}^{q} y_i = n . \tag{4.15}$$

Let $g(y_1, y_2, \cdots, y_q)$ be a convex function defined on $\sum_q$. The construction of a function $\phi(\xi_1, \cdots, \xi_n)$ of class $\mathscr{C}$ in terms of $g(y_1, \cdots, y_q)$ is done completely analogously to (4.4). We then deduce, paraphrasing the proof of Theorem 4.1, the following assertions.

THEOREM 4.3. *Let $g(y_1, \cdots, y_q)$ be convex defined on $\sum_q$ (see (4.15)). Then for* *the random vectors* $\mathbf{Y} = (Y_1, \cdots, Y_q)$ *as defined above, we have*

$$E_{\mathscr{W}}\, g(\mathbf{Y}) \leqq E_{\mathscr{S}}\, g(\mathbf{Y}) \qquad \text{for any symmetric sampling plan } \mathscr{S} \tag{4.16}$$

and

$$E_{\mathscr{R}(\pi)}\, g(\mathbf{Y}) \leqq E_{\mathscr{I}}\, g(\mathbf{Y}) \tag{4.17}$$

where $\mathscr{R}(\pi)$ *is any random replacement scheme, and in the case of* (4.17) *we require* $N \geqq n$ *as in Theorem* 3.1.

As previously, the relations (4.16) and (4.17) can be expressed in the language of dilations of measures. The special case of (4.16) with $\mathscr{S} = \mathscr{I}$ was discovered earlier by Kemperman (1973). Thus, for this example (4.16), we have that the multinomial probability measure

$$P\{Y_i = \nu_i, i = 1, 2, \cdots, q\} = \frac{n!}{\nu_1!\, \nu_2! \cdots \nu_q!} \left(\frac{N_1}{N}\right)^{\nu_1} \left(\frac{N_2}{N}\right)^{\nu_2} \cdots \left(\frac{N_q}{N}\right)^{\nu_q}$$
$$\sum \nu_i = n$$

dilates the hypergeometric probability measure

$$\Pr\{Y_i = \nu_i, i = 1, 2, \cdots, q\} = \binom{N}{n}^{-1} \prod_{p=1}^{q} \binom{N_p}{\nu_p}.$$

The relations of Theorems 4.1–4.3 are independent of the specific $\{a_v\}$ values comprising Ω and actually express expectation inequalities for certain convex functions of the numbers of the distinct values observed. We now develop some inequalities where the elements $a_v \in \Omega$ are variables. To this end, let $A(\xi)$ be a convex function of a single real variable defined over the interval $[n(\min_v a_v), n(\max_v a_v)]$.

Consider a fixed set of integers

$$k_1 \geqq k_2 \geqq \cdots \geqq k_p > 0\,, \qquad \sum k_i = n \tag{4.18}$$

and fixed choice of p indices from Ω, and designate the variables at these indices as $\{b_1, b_2, \cdots, b_p\}$ to distinguish them from the collection of all $\{a_i\}_1^N$. Form the function

$$g(\mathbf{a}) = g(a_1, \cdots, a_N) = A(k_1 b_1 + k_2 b_2 + \cdots + k_p b_p) \tag{4.19}$$

defined over the collections of all $\{a_1, \cdots, a_N\}$ fulfilling, perhaps, some convex constraints. Denote the domain of definition of g by $\mathscr{A}$. We assume also that $\mathscr{A}$ is *a* symmetric set. Only the variables $\{b_1, \cdots, b_p\}$ obviously occur in the evaluation of $g(\mathbf{a})$. It is trivial to check (since $A(\xi)$ is convex) that g is a convex function of $(a_1, \cdots, a_N)$. With this fact we easily prove

LEMMA 4.1. *Let* $A(\xi)$ *be convex as prescribed above. Let* $(X_1, X_2, \cdots, X_n)$ *be a random sample based on the sampling plan* $\mathscr{S}[\mathbf{k}]$ *with* $\mathbf{k}$ *specified as in* (4.18). *Then*

$$E_{\mathscr{S}[k]} A(X_1 + \cdots + X_n) = \psi(a_1, \cdots, a_N) \tag{4.20}$$

is a symmetric convex function of $\mathbf{a} = \{a_i\}_{i=1}^N$, $\mathbf{a} \in \mathscr{A}$.

PROOF. Note that

$$(4.21)\qquad E_{\mathscr{S}[\mathbf{k}]}A(X_1 + \cdots + X_n) = \frac{(N-p)!}{N!} \sum_{i_\nu \neq i_\mu} A(k_1 a_{i_1} + k_2 a_{i_2} + \cdots + k_p a_{i_p})$$

where the sum extends over all distinct choices of p indices $\{i_1, \cdots, i_p\}$ from $\{1, 2, \cdots, N\}$. As established following (4.19), each term of the sum is a convex function of $(a_1, \cdots, a_N)$. Consequently, the total sum persists as a convex function and is clearly symmetric in the variables $(a_1, \cdots, a_N)$. The proof of the lemma is complete.

The result of the lemma in conjunction with the representation (1.9) readily implies

THEOREM 4.4. *If $A(\xi)$ is convex and $\mathscr{S}$ is a symmetric sampling plan then*

$$(4.22)\qquad E_{\mathscr{S}} A(X_1 + \cdots + X_n) = \Gamma(a_1, a_2, \cdots, a_N)$$

is a convex symmetric function of $\mathbf{a} = (a_1, a_2, \cdots, a_N)$ in $\mathscr{A}$, where $\mathscr{A}$ is any convex set in Euclidean N-space.

The main applications of the fact of Theorem 4.4 concern *majorization inequalities*, as depicted in Hardy, Littlewood and Pólya, (1934), Chapter 2.

DEFINITION 4.2. A real sequence

$$(4.23)\qquad \mathbf{a} = (a_1, \cdots, a_N) \text{ is said to majorize } \mathbf{c} = (c_1, \cdots, c_N)$$

iff there exists a doubly stochastic matrix $T = ||t_{ij}||_1^N$ such that

$$(4.24)\qquad \mathbf{c} = T\mathbf{a}$$

the relation of majorization is symbolized by $\mathbf{c} \prec \mathbf{a}$.

This concept is readily identifiable as the discrete version of dilation of measures put forth in (4.10) and (4.11).

An equivalent and more practical setup of the majorization conditions is as follows.

Arrange the components of the vector $\mathbf{a}$ in decreasing order where $\{a_1^*, \cdots, a_N^*\} = \{a_1, \cdots, a_N\}$ and $a_1^* \geqq a_2^* \geqq \cdots, \geqq a_N^*$, and do the same for the vector $\mathbf{c}$. Then (4.24) prevails if and only if

$$(4.25)\qquad \sum_{i=1}^r a_i^* \geqq \sum_{i=1}^r c_i^*\,, \qquad r = 1, 2, \cdots, N$$

with equality for $r = N$. (The equivalence of (4.24) and (4.25) is established in the reference cited prior to Definition 4.2.)

The characterization of the functions $\phi(\mathbf{a}) = \phi(a_1, \cdots, a_N)$ which are monotone with respect to the ordering of (4.25) is classical and of immense utility (e.g., Hardy, Littlewood and Pólya, Chapter 2). They carry the special name—Schur–Ostrowski functions. Included in this class are all symmetric convex functions. Thus, if $\theta(a_1, a_2, \cdots, a_N)$ is convex and symmetric for $\mathbf{a} \in \mathscr{A}$, then $\mathbf{a} \succ \mathbf{c}$ in the sense of (4.25) entails $\theta(\mathbf{a}) \geqq \theta(\mathbf{c})$.

The convex functions of (4.22) are of special interest. We obtain therefore

THEOREM 4.5. *Let $A(\xi)$ satisfy the conditions of* (4.22); *then if* $\mathbf{a} \succ \mathbf{c}$, $\mathbf{a}, \mathbf{c} \in \mathscr{A}$ *we have*

$$\Gamma(\mathbf{a}) \geqq \Gamma(\mathbf{c}) \tag{4.26}$$

where $\Gamma(\mathbf{a}) = E_{\mathscr{S}} A(X)$, *and* $\mathscr{S}$ *is a fixed symmetric sampling plan defined on the population space* $\Omega = \{a_i\}_1^N$ *and* $\mathscr{A}$ *is a symmetric convex set over which* $\mathbf{a}$ *ranges.*

We close this section with a concrete application of Theorem 4.4.

Let $\check{\mathscr{A}}$ consist of the collection of all N-tuples $\{a_1, a_2, \cdots, a_N\}$ of real values satisfying the constraints $-1 \leqq a_i \leqq 1$ and $\sum_{i=1}^N a_i = N\alpha$ (α given). It is elementary to show that with respect to the majorization ordering relation (4.25) the maximal element in this specific $\check{\mathscr{A}}$ has the form

$$\tilde{a} = (\tilde{a}_1, \tilde{a}_2, \cdots, \tilde{a}_N) \tag{4.27}$$

where all $\tilde{a}_i$ are either $+1$ or -1 except for at most one component. For the special case $\alpha = 0$, $N = 2M$, then explicitly $\tilde{a}_1 = \cdots = \tilde{a}_M = +1$, $\tilde{a}_{M+1} = \cdots = \tilde{a}_{2M} = -1$.

The upper bound for $\Gamma(\mathbf{a})$ of (4.26) with $\mathbf{a}$ traversing $\check{\mathscr{A}}$ is attained for (4.27) for each symmetric sampling plan.

Acknowledgments. It is a pleasure for me to express my thanks to Allan Pinkus and Yosi Rinott for some suggestions on improving the manuscript.

REFERENCES

[1] BASU, D. (1958). On sampling with and without replacement. *Sankhyā* **20** 2132–2134.

[2] HARDY., G. H., LITTLEWOOD, J. R. and PÓLYA, G. (1934). *Inequalities*. Cambridge Univ. Press.

[3] HOEFFDING, W. (1963). Probability inequalities for sums of bounded random variables. *J. Amer. Statist. Assoc.* **58** 13–30.

[4] KARLIN, S. and STUDDEN, W. J. (1966). *Tchebycheff Systems: Applications in Analysis and Statistics*. Interscience, New York.

[5] KEMPERMAN, J. H. B. (1973). Moment problems for sampling without replacement. Preprint.

[6] KORWAR, R. M. and Serfling, R. J. (1970). On averaging over distinct units in sampling with replacement. *Ann. Math. Statist.* **41** 2132–2134.

[7] LANKE, J. (1972). Sampling distinguishable elements with replacement. *Ann. Math. Statist.* **43** 1329–1332.

[8] PATHAK, P. K. (1961). On the evaluation of moments of distinct units in a sample. *Sankhyā Ser. A* **23** 415–420.

[9] RAJ, D. and KHAMIS, S. H. (1958). Some remarks on sampling with replacement. *Ann. Math. Statist.* **29** 555–557.

[10] ROSÉN, B. (1967). On an inequality of Hoeffding. *Ann. Math. Statist.* **38** 382–392.

[11] ROSÉN, B. (1972). Asymptotic theory for successive sampling with varying probabilities without replacement I. *Ann. Math. Statist.* **43** 373–397.

[12] SEN, P. KUMAR (1970). The Hájek-Rényi inequality for sampling from a finite population. *Sankhyā* **32** 181–188.

[13] SERFLING, R. J. (1973). Refinements of some probability inequalities of Hoeffding (abstract). *Bull. Inst. Math. Statist.* **2** 104.

DEPARTMENT OF MATHEMATICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA 94305