ON AN INEQUALITY OF HOEFFDING

By Bengt Rosén

Uppsala University

1. Introduction. Let $X_1, X_2, \cdots, X_n$ be a sample drawn without replacement from the finite population π, and let $Y_1, Y_2, \cdots, Y_n$ be a sample drawn with replacement from the same population. We set

$$S_n = X_1 + X_2 + \cdots + X_n \quad \text{and} \quad Z_n = Y_1 + Y_2 + \cdots + Y_n .$$

The distribution of S_n is, to a greater extent concentrated at its mean than is the distribution of Z_n. A quantitative formulation of this fact is, for example, the well-known relation

$$\sigma^2(S_n) \leqq \sigma^2(Z_n) \tag{1.1}$$

where σ^2 denotes variance.

The following theorem, due to Hoeffding (Theorem 4 in [3]), is a considerably more informative result in this direction.

Theorem. *For any convex and continuous function $\varphi(x)$ we have*

$$E\varphi(S_n) \leqq E\varphi(Z_n). \tag{1.2}$$

(E denotes here, and in the sequel, mathematical expectation.)

If we, for example, choose $\varphi(x) = (x - n\mu_\pi)^2$, where μ_π is the mean in the population π, then (1.2) becomes just (1.1).

The purpose of this paper is to generalize Hoeffding's result in two directions, one of which consists in showing that (1.2) holds for certain types of sample functions other than convex continuous functions of the sample sum. The other direction of generalization is to prove that (1.2) holds not only when $Y_1, Y_2, \cdots, Y_n$ are sampled with replacement, but for a broader class of sampling procedures, here called symmetric sampling procedures. Loosely speaking, a sampling procedure is said to be symmetric if all elements in the population are treated symmetrically during the drawing procedure. So for example sampling with replacement every second time is symmetric while sampling with probability proportional to size is not. In Section 2 we formally define sampling procedures and in particular symmetric procedures. In recent years problems in survey sampling have inspired various formalizations of the concept of sampling procedure, see Godambe [1], of which our formalization is a special case.

2. About sampling in general. Let (m) stand for the set of the first m integers, i.e. $(m) = (1, 2, \cdots, m)$ and let $\Omega(n, N) = (N) \times (N) \times \cdots \times (N)$ (n factors, $\times$ denotes Cartesian product). Points in $\Omega(n, N)$ will be denoted by $(i_1, i_2, \cdots, i_n)$ or by ω.

Received 31 May 1966.

DEFINITION 1. By a *sampling procedure* yielding an (ordered) sample of size n from (N), we mean a random experiment $(\Omega(n, N), P)$, where P is a probability on $\Omega(n, N)$.

As a random quantity the sample will be denoted by $I_1, I_2, \cdots, I_n$, where $I_\nu(i_1, i_2, \cdots, i_n) = i_\nu$, $\nu = 1, 2, \cdots, n$.

When n and N are given, the sampling procedure is determined by P, and we shall sometimes refer to a sampling procedure only by P.

DEFINITION 2. A sampling procedure $(\Omega(n, N), P)$ is said to be *symmetric* if for an arbitrary permutation $u(\cdot)$ of the elements in (N) we have

$$P(u(i_1), u(i_2), \cdots, u(i_n)) = P(i_1, i_2, \cdots, i_n).$$

Usually sampling procedures P are not specified explicitly, but by a description of how to obtain the sample. Some examples are:

1. Sampling without replacement;
2. Sampling with replacement;
3. Sampling with replacement every kth time;
4. Sampling with replacement with a fixed replacement probability.

We shall be particularly interested in *sampling without replacement* and we shall throughout the paper denote this sampling procedure by Q. Its formal definition is: for $\omega = (i_1, i_2, \cdots, i_n) \in \Omega(n, N)$

$$Q(\omega) = [N(N-1) \cdots (N-n+1)]^{-1} \quad \text{if all } i_\nu\text{'s are different}$$
$$= 0 \quad \text{otherwise.}$$

The concept of a symmetric sampling procedure is very general, and it is easily seen that the sampling procedures listed above are all symmetric.

LEMMA 2.1. *Let $I_1, I_2, \cdots, I_n$ be drawn according to a symmetric sampling procedure $(\Omega(n, N), P)$. Then*

$$P(I_\nu = i) = N^{-1}, \qquad i = 1, 2, \cdots, N; \nu = 1, 2, \cdots, n.$$

By a (finite) *population* $\pi = (a_1, a_2, \cdots, a_N)$, we mean a finite collection of real numbers. The number of elements in π will be called its *size*, and it will be denoted by N_π. Furthermore, we define

$$\mu_\pi = N^{-1} \textstyle\sum_1^N a_\nu .$$

DEFINITION 3. By a *random sample* of size n from $\pi = (a_1, a_2, \cdots, a_N)$ drawn according to the sampling procedure $(\Omega(n, N), P)$, we mean the random vector $a_{I_1}, a_{I_2}, \cdots, a_{I_n}$, where $I_1, I_2, \cdots, I_n$ is a sample from (N), drawn according to $(\Omega(n, N), P)$.

We usually use the notation $(X_1, X_2, \cdots, X_n) = (a_{I_1}, a_{I_2}, \cdots, a_{I_n})$ and we say that $X_1, X_2, \cdots, X_n$ is a symmetric sample from π if $(\Omega(n, N), P)$ is a symmetric sampling procedure.

LEMMA 2.2. *If $X_1, X_2, \cdots, X_n$ is a symmetric sample from π, then*

$$E \textstyle\sum_1^k X_\nu = k\mu_\pi . \tag{2.1}$$

PROOF. According to Lemma 2.1 it holds that

$$EX_\nu = \sum_{i=1}^{N} a_i P(I_\nu = i) = N^{-1}\sum_{i=1}^{N} a_i = \mu_\pi$$

and (2.1) follows.

Let $X_1, X_2, \cdots, X_n$ be a sample from π, drawn according to $(\Omega(n, N_\pi), P)$, and let $f(x_1, x_2, \cdots, x_n)$ be a function of n variables. The distribution of the random variable $f(X_1, X_2, \cdots, X_n)$ will depend on π and P. For the mathematical expectation of $f(X_1, X_2, \cdots, X_n)$ we shall indicate this dependence by the notation

$$E(P, \pi)f(X_1, X_2, \cdots, X_n).$$

3. Inequalities of Hoeffding's type. We can now formulate the problem which we shall consider.

A function $f(x_1, x_2, \cdots, x_n)$ of n variables (which may take infinite values) is said to belong to $\mathcal{H}^{(n)}$ if, for all populations π for which $N_\pi \geqq n$ and for all symmetric sampling procedures P, we have

$$(3.1) \quad E(P, \pi)f(X_1, X_2, \cdots, X_n) \geqq E(Q, \pi)f(X_1, X_2, \cdots, X_n).$$

(As stated before, Q denotes sampling without replacement.)

We would like to give a complete characterization of the functions in $\mathcal{H}^{(n)}$. However, we are not able to do this, but we shall exhibit some general principles, which are useful in the construction of functions in $\mathcal{H}^{(n)}$.

The following lemma yields a reduction of the problem.

LEMMA 3.1. *For the function f to belong to $\mathcal{H}^{(n)}$ it is necessary and sufficient that* (3.1) *is fulfilled for all π such that $N_\pi = n$ and for all symmetric sampling procedures* $(\Omega(n, n), P)$.

REMARK. It is easily seen from the proof that Lemma 3.1 also holds if we consider equality, instead of inequality, in (3.1).

The necessity part in the lemma is immediate. To prove the sufficiency part we shall first construct a "refinement" of the sample space $\Omega(n, N)$.

Let $C_1, C_2, \cdots, C_{\binom{N}{n}}$ be the different combinations of n elements from (N) and let

$$\Lambda(C_k) = \{(\omega)_k : \omega = (i_1, i_2, \cdots, i_n) \in \Omega(n, N), i_\nu \in C_k, \nu = 1, 2, \cdots, n\}.$$

The set $\Lambda(C_k)$ can be described as the set of all possible samples of size n from C_k, each of which is labelled with the subscript k.

Let

$$\Omega^*(n, N) = \bigcup_{k=1}^{\binom{N}{n}} \Lambda(C_k).$$

Elements in $\Omega^*(n, N)$ will be denoted by ω^*. $\Omega^*(n, N)$ is the desired "refinement" of $\Omega(n, N)$.

EXAMPLE. For $N = 3$ and $n = 2$ we have

$$\Omega(n, N) = \{(1, 1), (1, 2), (1, 3), (2, 1), (2, 2), (2, 3), (3, 1), (3, 2), (3, 3)\},$$

$$C_1 = \{1, 2\}, \qquad C_2 = \{1, 3\}, \qquad C_3 = \{2, 3\},$$
$$\Lambda(C_1) = \{(1, 1)_1, (1, 2)_1, (2, 1)_1, (2, 2)_1\},$$
$$\Lambda(C_2) = \{(1, 1)_2, (1, 3)_2, (3, 1)_2, (3, 3)_2\},$$
$$\Lambda(C_3) = \{(2, 2)_3, (2, 3)_3, (3, 2)_3, (3, 3)_3\}.$$

We have a natural mapping from Ω^* to Ω by disregarding the label. This mapping will be denoted by a bar, i.e. $\bar{\omega}^* = (\bar{\omega})_k = \omega$.

Henceforth, $\#A$ denotes the number of elements in the (finite) set A.

For a probability P on $\Omega(n, N)$ we define a corresponding probability P^0 on $\Omega^*(n, N)$ as follows

$$P^0(\omega^*) = P(\bar{\omega}^*)/\#\{\nu^*: \bar{\nu}^* = \bar{\omega}^*, \nu^* \varepsilon \Omega^*\}.$$

In words: all elements ω^* in Ω^* which agree except with regard to the label, share equally the probability at $\bar{\omega}^*$.

The following result is easily verified.

LEMMA 3.2. *If $(\Omega(n, N), P)$ is a symmetric sampling procedure, then*

$$P^0(\Lambda(C_k)) = \binom{N}{n}^{-1}, \qquad k = 1, 2, \cdots, \binom{N}{n}.$$

For a probability P^* on Ω^*, $P^*(\cdot \mid \Lambda(C_k))$ denotes the conditional probability on $\Lambda(C_k)$. Such a conditional probability can be regarded as defining a sampling procedure for drawing a sample of size n from C_k.

The next result has a straight-forward proof, which we omit.

LEMMA 3.3. *If $(\Omega(n, N), P)$ is a symmetric sampling procedure, then $P^0(\cdot \mid \Lambda(C_k))$ defines a symmetric sample from C_k.*

Let $\pi = (a_1, a_2, \cdots, a_N)$ be a population. For given $(\Omega^*(n, N), P^*)$ we define a sample $X_1, X_2, \cdots, X_n$ from π as follows. Let $\omega^* = (i_1, i_2, \cdots, i_n)_k \varepsilon \Omega^*(n, N)$. Then $(X_1, X_2, \cdots, X_n)(\omega^*) = (a_{i_1}, a_{i_2}, \cdots, a_{i_n})$.

We give another lemma, the proof of which is straight-forward enough to be omitted.

LEMMA 3.4. *Let $X_1, X_2, \cdots, X_n$ and $Y_1, Y_2, \cdots, Y_n$ be random samples from π, according to $(\Omega(n, N_\pi), P)$ and $(\Omega^*(n, N_\pi), P^*)$ respectively. If $P^* = P^0$, then $X_1, X_2, \cdots, X_n$ and $Y_1, Y_2, \cdots, Y_n$ have identical distributions.*

We shall use the following notation.

$E(P^*, \pi)f(X_1, X_2, \cdots, X_n)$ is the mathematical expectation of the random variable $f(X_1, \cdots, X_n)$ when $X_1, X_2, \cdots, X_n$ is a sample from π according to $(\Omega^*(n, N_\pi), P^*)$.

$E(P^*(\cdot \mid \Lambda(C_k)), \pi)f(X_1, X_2, \cdots, X_n)$ is the conditional expectation of $f(X_1, X_2, \cdots, X_n)$ given $\Lambda(C_k)$, when $X_1, X_2, \cdots, X_n$ is a sample from π according to $(\Omega^*(n, N_\pi), P^*)$.

We are now prepared to prove Lemma 3.1.

PROOF OF LEMMA 3.1. We consider the sufficiency part, and we assume that $(\Omega(n, N_\pi), P)$ is symmetric and that

(3.2) $\quad f$ satisfies (3.1) when $N_\pi = n$ and P is symmetric.

According to Lemma 3.4, we have

$$E(P, \pi)f(X_1, X_2, \cdots, X_n) = E(P^0, \pi)f(X_1, X_2, \cdots, X_n)$$
$$= \sum_{k=1}^{\binom{N}{n}} E(P^0(\cdot \mid \Lambda(C_k)), \pi)f(X_1, \cdots, X_n) \cdot P^0(\Lambda(C_k)).$$

C_k contains n elements. Thus, by (3.2), Lemma 3.3 and Lemma 3.2, we can continue

$$\geqq \sum_{k=1}^{\binom{N}{n}} E(Q^0(\cdot \mid \Lambda(C_k)), \pi)f(X_1, \cdots, X_n)Q^0(\Lambda(C_k))$$
$$= E(Q^0, \pi)f(X_1, X_2, \cdots, X_n) = E(Q, \pi)f(X_1, X_2, \cdots, X_n).$$

Thereby Lemma 3.1 is proved.

4. Some general properties of $\mathcal{H}^{(n)}$. We introduce the following subclasses of $\mathcal{H}^{(n)}$:

$\mathcal{H}_1^{(n)}$: The symmetric functions in $\mathcal{H}^{(n)}$.

$\mathcal{H}_2^{(n)}$: The function in $\mathcal{H}_1^{(n)}$ for which (3.1) holds with equality for all π and all symmetric P.

We have $\mathcal{H}_2^{(n)} \subset \mathcal{H}_1^{(n)} \subset \mathcal{H}^{(n)}$.

LEMMA 4.1. (a) *The function classes* $\mathcal{H}^{(n)}$, $\mathcal{H}_1^{(n)}$ *and* $\mathcal{H}_2^{(n)}$ *are closed under*

1. *addition,*
2. *multiplication with non-negative constants* ($\mathcal{H}_2^{(n)}$ *is closed under multiplication with arbitrary constants*),
3. *pointwise convergence.*

(b) *If* $f(x_1, x_2, \cdots, x_n)$ *belongs to an* $\mathcal{H}$*-class, then the following functions belong to the same* $\mathcal{H}$*-class:*

1. $f(\lambda(x_1), \lambda(x_2), \cdots, \lambda(x_n))$, *where* $\lambda(x)$ *is an arbitrary function,*
2. $f(x_{\sigma(1)}, x_{\sigma(2)}, \cdots, x_{\sigma(n)})$, *where* $\sigma(\cdot)$ *is an arbitrary permutation of the elements in* (n).

PROOF. (a) follows easily from the linearity of the expectation operator.

We prove (b) only for $\mathcal{H}^{(n)}$. Let $\pi = (a_1, a_2, \cdots, a_N)$ be a population and $\lambda(x)$ a function of one variable. Then $\lambda(\pi)$ denotes the population $\lambda(\pi) = (\lambda(a_1), \lambda(a_2), \cdots, \lambda(a_N))$. With this notation we have

$$(4.1) \quad E(P, \pi)f(\lambda(X_1), \cdots, \lambda(X_n)) = E(P, \lambda(\pi))f(X_1, \cdots, X_n).$$

Thus, when $f \in \mathcal{H}^{(n)}$

$$E(P, \pi)f(\lambda(X_1), \lambda(X_2), \cdots, \lambda(X_n)) = E(P, \lambda(\pi))f(X_1, \cdots, X_n)$$
$$\geqq E(Q, \lambda(\pi))f(X_1, \cdots, X_n)$$
$$= E(Q, \pi)f(\lambda(X_1), \cdots, \lambda(X_n))$$

and (b)1 is verified. To prove (b)2, let $(X_1^*, \cdots, X_n^*) = (X_{\sigma(1)}, \cdots, X_{\sigma(n)})$. When $X_1, \cdots, X_n$ is a sample from π drawn according to P, $X_1^*, \cdots, X_n^*$

can also be regarded as a sample from π drawn according to $V(P)$, which is determined by P. It is easily seen that if P is symmetric, $V(P)$ is symmetric also. Thus, if $f \varepsilon \mathcal{H}^{(n)}$ we have

$$\begin{aligned} E(P, \pi)f(X_{\sigma(1)}, \cdots, X_{\sigma(n)}) &= E(V(P), \pi)f(X_1, \cdots, X_n) \\ &\geqq E(Q, \pi)f(X_1, \cdots, X_n) \\ &= E(Q, \pi)f(X_{\sigma(1)}, \cdots, X_{\sigma(n)}). \end{aligned}$$

In the last step we used the fact that a sample drawn without replacement is exchangeable, i.e. its distribution is invariant under permutation of the variables. Thus, (b)2 is proved.

5. On $\mathcal{H}_2^{(n)}$. We first consider $\mathcal{H}_2^{(n)}$, the smallest of the function classes introduced.

THEOREM 1. *Necessary and sufficient for f to belong to $\mathcal{H}_2^{(n)}$, is that f be of the form*

$$f(x_1, x_2, \cdots, x_n) = \textstyle\sum_{\nu=1}^{n} \lambda(x_\nu) \tag{5.1}$$

for some function $\lambda(x)$.

PROOF. The sufficiency part follows immediately from Lemma 2.2 and (4.1).

Let $\pi = (a_1, a_2, \cdots, a_n)$, and let P stand for the particular symmetric procedure $(\Omega(n, n), P)$, where

$$\begin{aligned} P(i_1, i_2, \cdots, i_n) &= n^{-1} \quad &&\text{if } i_1 = i_2 = \cdots = i_n \\ &= 0 &&\text{otherwise.} \end{aligned}$$

If $f \varepsilon \mathcal{H}_2^{(n)}$, we then have

$$\begin{aligned} n^{-1} \textstyle\sum_{\nu=1}^{n} f(a_\nu, a_\nu, \cdots, a_\nu) &= E(P, \pi)f(X_1, \cdots, X_n) \\ &= E(Q, \pi)f(X_1, \cdots, X_n) = f(a_1, a_2, \cdots, a_n). \end{aligned} \tag{5.2}$$

As $f \varepsilon \mathcal{H}_2^{(n)}$, (5.2) holds for arbitrary values of $a_1, a_2, \cdots, a_n$, and we obtain that f is of the form (5.1) for $\lambda(x) = n^{-1}f(x, x, \cdots, x)$. Thus, Theorem 1 is proved.

6. On $\mathcal{H}_1^{(n)}$. According to Lemma 3.1 it is sufficient to consider populations of size n when determining whether or not a function f belongs to $\mathcal{H}^{(n)}$. For such populations the corresponding sample space is $\Omega(n, n)$. By the *kernel* in $\Omega(n, n)$ we mean the elements $(i_1, i_2, \cdots, i_n)$ which are permutations of $(1, 2, \cdots, n)$.

The following facts are crucial:

(1) Q has its whole mass concentrated on the kernel.

(2) If $f(x_1, x_2, \cdots, x_n)$ is a symmetric function, then $f(X_1, X_2, \cdots, X_n)$ is constant on the kernel.

LEMMA 6.1. (*Maximum principle*). *If f_1 and f_2 both belongs to $\mathcal{H}_1^{(n)}$, then* $\max(f_1, f_2) \varepsilon \mathcal{H}_1^{(n)}$.

PROOF. We assume π to have size n and P to be symmetric. Then, we have if $f_1, f_2 \varepsilon \mathcal{H}_1^{(n)}$,

$$E(P, \pi) \max (f_1, f_2)(X_1, \cdots, X_n)$$
$$\geqq \max (E(P, \pi)f_1(X_1, \cdots, X_n), E(P, \pi)f_2(X_1, \cdots, X_n))$$
$$\geqq \max (E(Q, \pi)f_1(X_1, \cdots, X_n), E(Q, \pi)f_2(X_1, \cdots, X_n))$$
$$= E(Q, \pi) \max (f_1, f_2)(X_1, X_2, \cdots, X_n)$$

since Q is concentrated on the kernel and $f_1(X_1, \cdots, X_n)$ and $f_2(X_1, \cdots, X_n)$ are constant there.

Hence Lemma 6.1 is proved.

THEOREM 2. *If $\varphi(x)$ is continuous and convex, then for an arbitrary function $\lambda(x)$ the following function belongs to $\mathcal{K}_1^{(n)}$:*

$$\varphi(\lambda(x_1) + \lambda(x_2) + \cdots + \lambda(x_n)). \tag{6.1}$$

REMARKS. (1) It is immediate that it suffices to assume that $\varphi(x)$ is convex and continuous on the interval $[n \cdot \inf_x \lambda(x), n \sup_x \lambda(x)]$. ($\varphi$ may even take infinite values at the ends of the interval.)

(2) The theorem by Hoeffding is included in Theorem 2.

PROOF. When $\varphi(x)$ is convex and continuous it can be represented

$$\varphi(x) = \lim_{k \to \infty} \max (L_1(x), L_2(x), \cdots, L_k(x)), \tag{6.2}$$

where $L_s(x)$, $s = 1, 2, \cdots$, are linear functions. Thus

$$\varphi(\textstyle\sum_{\nu=1}^n \lambda(x_\nu)) = \lim_{k \to \infty} \max (L_1(\sum_{\nu=1}^n \lambda(x_\nu)), \cdots, L_k(\sum_{\nu=1}^n \lambda(x_\nu))). \tag{6.3}$$

According to Theorem 1, $L_s(\sum_{\nu=1}^n \lambda(x_\nu))$ belongs to $\mathcal{K}_2^{(n)}$ and thus to $\mathcal{K}_1^{(n)}$. The assertion in Theorem 2 now follows from (6.2) and Lemmas 6.1 and 4.1(a)3.

LEMMA 6.2. *If $f \in \mathcal{K}_1^{(n)}$ and if $\varphi(x)$ is convex and non-decreasing, then $\varphi(f) \in \mathcal{K}_1^{(n)}$.*

PROOF. When $\varphi(x)$ is convex and non-decreasing it has a representation (6.2) where the $L_s(x)$ have non-negative slopes.

$$\varphi(f) = \lim_{k \to \infty} \max (L_1(f), L_2(f), \cdots, L_k(f)). \tag{6.4}$$

The lemma now follows from (6.4), Lemma 6.1 and (a)2 and (a)3 in Lemma 4.1.

EXAMPLES. The following functions belong to $\mathcal{K}_1^{(n)}$:

(a) $(\prod_{\nu=1}^n |x_\nu|)^\alpha$, α real,
(b) $(\sum_{\nu=1}^n |x_\nu|^\alpha)^{1/r}$, α real, $r \leqq 1$,
(c) $\min (x_1, x_2, \cdots, x_n)$,
(d) $\exp \{\prod_{\nu=1}^n |x_\nu| + \min (x_1, x_2, \cdots, x_n)\}$.

VERIFICATION. (a) follows from the representation

$$(\textstyle\prod_{\nu=1}^n |x_\nu|)^\alpha = \exp (\alpha \sum_{\nu=1}^n \log |x_\nu|)$$

and from Theorem 2, as the exponential function is convex and continuous.

(b) is a consequence of Theorem 2 and the fact that $x^{1/r}$ is convex for $x \geqq 0$ when $r \leqq 1$.

(c) can be verified by a direct application of Lemma 3.1. Let $\pi =$

$(a_1, a_2, \cdots, a_n)$ be an arbitrary population of size n. Then

$$E(Q, \pi) \min (X_1, \cdots, X_n) = \min (a_1, a_2, \cdots, a_n).$$

when $X_1, X_2, \cdots, X_n$ is a sample from π, we have $\min (X_1, \cdots, X_n) \geqq \min (a_1, a_2, \cdots, a_n)$. Thus, for an arbitrary P, we have

$$E(P, \pi) \min (X_1, \cdots, X_n) \geqq \min (a_1, \cdots, a_n) = E(Q, \pi) \min (X_1, \cdots, X_n)$$

and (c) follows from Lemma 3.1.

Another proof of (c) can be obtained from (b), Lemma 4.1 and the formula

$$\min (x_1, \cdots, x_n) = \lim_{A\to\infty} (\lim_{r\to-\infty} \{ \textstyle\sum_{\nu=1}^{n} |x_\nu + A|^r\}^{1/r} - A).$$

(d) follows from (a) and (c) and Lemma 6.2, as e^x is convex and non-decreasing.

It would be a nice solution to the problem of characterizing $\mathcal{H}_1^{(n)}$, if $\mathcal{H}_1^{(n)}$ were exactly the functions described in (6.1). This is, however, not true because $\min (x_1, x_2, \cdots, x_n)$, which we know belongs to $\mathcal{H}_1^{(n)}$, cannot be written in the form (6.1). We prove this by contradiction. Assume that

$$\min (x_1, x_2, \cdots, x_n) = \varphi(\lambda(x_1) + \cdots + \lambda(x_n)), \tag{6.5}$$

where φ is convex and continuous. By setting $x_1 = x_2 = \cdots = x_n$ in (6.5) we obtain that

$$x = \varphi(n\lambda(x)). \tag{6.6}$$

From (6.6) it follows that if $x_1 \neq x_2$, then also $\lambda(x_1) \neq \lambda(x_2)$. By letting $x_1 \leqq x_2 = x_3 = \cdots = x_n = x$ in (6.5) we get $x_1 = \varphi(\lambda(x_1) + (n - 1)\lambda(x))$, $x_1 \leqq x$. Thus, φ must assume the value x_1 at an infinity of points. This contradicts the assumption that φ is a convex continuous function.

Actually, Theorem 2 is true even if we remove the assumption that φ is continuous. For a treatment of general convex functions, we refer, for example, to the book by Hardy, Littlewood and Pólya [2].

THEOREM 3. *If $\varphi(x)$ is convex, then for an arbitrary $\lambda(x)$, we have*

$$\varphi(\lambda(x_1) + \lambda(x_2) + \cdots + \lambda(x_n)) \in \mathcal{H}_1^{(n)}.$$

REMARKS. (1) This result includes Theorem 2.

(2) The proof, which follows, is mainly a formalization of the idea of proof used by Hoeffding in [3].

PROOF. According to Lemma 3.1, it is sufficient to verify (3.1) for populations of size n.

We partition $\Omega(n, n)$ into $\Omega(n, n) = T_1 \cup T_2 \cup \cdots \cup T_S$, so that two elements $(i_1, \cdots, i_n)$ and $(j_1, \cdots, j_n)$ in $\Omega(n, n)$ belong to the same partitioning set T, if there is a permutation σ of $(1, 2, \cdots, n)$ such that $i_\nu = \sigma(j_\nu)$, $\nu = 1, 2, \cdots, n$. It is easily seen that if P is symmetric, then $P(\omega)$ is constant on every T. Thus, for $\pi = (a_1, a_2, \cdots, a_n)$ and P symmetric we have

$$
\begin{aligned}
&E(P, \pi)\varphi(X_1 + \cdots + X_n)\\
&\quad = \textstyle\sum_{\nu=1}^{s} E(P(\cdot \mid T_\nu), \pi)\varphi(X_1 + \cdots + X_n)P(T_\nu)\\
&\quad = \textstyle\sum_{\nu=1}^{s} P(T_\nu)\cdot(\#T_\nu)^{-1} \sum_{(i_1, i_2, \cdots, i_n)\varepsilon T_\nu} \varphi(a_{i_1} + a_{i_2} + \cdots + a_{i_n})\\
&\quad \geqq \textstyle\sum_{\nu=1}^{s} P(T_\nu)\varphi((\#T_\nu)^{-1} \sum_{(i_1, \cdots, i_n)\varepsilon T_\nu} (a_{i_1} + a_{i_2} + \cdots + a_{i_n}))
\end{aligned}
$$

by Jensen's inequality,

$$
\begin{aligned}
&\quad = \textstyle\sum P(T_\nu)\varphi(a_1 + a_2 + \cdots + a_n) = \varphi(a_1 + a_2 + \cdots + a_n)\\
&\quad = E(Q, \pi)\varphi(X_1 + X_2 + \cdots + X_n).
\end{aligned}
$$

Thus, $\varphi(x_1 + x_2 + \cdots + x_n) \,\varepsilon\, \mathfrak{K}_1^{(n)}$. Theorem 3 now follows from Lemma 4.1(b)1.

7. On non-symmetric functions in $\mathfrak{K}^{(n)}$. The existence of non-symmetric functions in $\mathfrak{K}^{(n)}$ follows, for example, from the following observation, which will be used without mentioning in the sequel. Let $f(x_1, \cdots, x_m) \,\varepsilon\, \mathfrak{K}^{(m)}$, where $m < n$ If we regard f as a function of n variables, then also $f \,\varepsilon\, \mathfrak{K}^{(n)}$. If f is not constant, then $f(x_1, \cdots, x_m)$ is not symmetric as a function of n variables.

The following principle will be useful in constructing a class of non-symmetric functions in $\mathfrak{K}^{(n)}$.

LEMMA 7.1. (*Domination principle*). *$f(x_1, x_2, \cdots, x_n)$ is given. Let $\{\cup\Pi_n^t, t \,\varepsilon\, T\}$ be a partitioning of Π_n = the class of populations of size n. Suppose that there exists a class of functions $\{f_t^*, t \,\varepsilon\, T\}$, $f_t^* \,\varepsilon\, \mathfrak{K}^{(n)}$ for every $t \,\varepsilon\, T$, such that for every symmetric P we have*

$$(7.1)\quad E(P, \pi)f(X_1, \cdots, X_n) \geqq E(P, \pi)f_t^*(X_1, \cdots, X_n) \quad \text{when} \quad \pi \,\varepsilon\, \Pi_n^t$$

$$(7.2)\quad E(Q, \pi)f_t^*(X_1, \cdots, X_n) \geqq E(Q, \pi)f(X_1, \cdots, X_n) \quad \text{when} \quad \pi \,\varepsilon\, \Pi_n^t.$$

Then, $f(x_1, x_2, \cdots, x_n) \,\varepsilon\, \mathfrak{K}^{(n)}$.

REMARK. The following conditions are easily seen to be sufficient for (7.1) and (7.2): When $\pi \,\varepsilon\, \Pi_n^t$,

$$f(X_1, \cdots, X_n) \geqq f_t^*(X_1, \cdots, X_n) \quad \text{on} \quad \Omega(n, n)$$

$$f(X_1, \cdots, X_n) = f_t^*(X_1, \cdots, X_n) \quad \text{on the kernel in} \quad \Omega(n, n).$$

PROOF. As $f_t^* \,\varepsilon\, \mathfrak{K}^{(n)}$ and as P is symmetric, we get from (7.1) and (7.2), when $\pi \,\varepsilon\, \Pi_n^t$

$$
\begin{aligned}
(7.3)\quad E(P, \pi)f(X_1, \cdots, X_n) &\geqq E(P, \pi)f_t^*(X_1, \cdots, X_n)\\
&\geqq E(Q, \pi)f_t^*(X_1, \cdots, X_n) \geqq E(Q, \pi)f(X_1, \cdots, X_n).
\end{aligned}
$$

By varying t over T we obtain that (7.3) holds for all $\pi \,\varepsilon\, \Pi_n$. Lemma 7.1 now follows from Lemma 3.1.

THEOREM 4. *Let $1 \leqq n_1 < n_2 < \cdots < n_k$. If $\varphi(x)$ is convex and monotone, then*

$$\max\,(\varphi(\textstyle\sum_{\nu=1}^{n_1} x_\nu), \varphi(\sum_{\nu=1}^{n_2} x_\nu), \cdots, \varphi(\sum_{\nu=1}^{n_k} x_\nu)) \,\varepsilon\, \mathfrak{K}^{(n_k)}.$$

Proof. We introduce some notation. Let $s_n = x_1 + x_2 + \cdots + x_n$. For a population $\pi = (a_1, a_2, \cdots, a_N)$ let $S_\pi = a_1 + a_2 + \cdots + a_N$.

We first prove the theorem for $\varphi(x) = \max(a, x)$, a real, by induction on k. We know from Theorem 2 that the assertion is true for $k = 1$. Suppose that it is true (for every a) for $k - 1$. Let

$$\begin{aligned} f(x_1, \cdots, x_{n_k}) &= \max(a, s_{n_1}, s_{n_2}, \cdots, s_{n_k}), \\ f_1^*(x_1, \cdots, x_{n_k}) &= \max(a, s_{n_1}, s_{n_2}, \cdots, s_{n_{k-1}}), \\ f_2^*(x_1, \cdots, x_{n_k}) &= \max(s_{n_1}, s_{n_2}, \cdots, s_{n_k}) \\ &= s_{n_1} + \max(0, s^*_{n_2-n_1}, s^*_{n_3-n_1}, \cdots, s^*_{n_k-n_1}) \end{aligned}$$

where $s_\nu^* = x_{n_1} + x_{n_1+1} + \cdots + x_{n_1+\nu}$.

From the induction hypothesis it follows that $f_1^* \varepsilon \mathcal{H}^{(nk)}$, and also that $\max(0, s_{n_2-n_1}, \cdots, s_{n_k-n_1}) \varepsilon \mathcal{H}^{(nk)}$. From Lemma 4.1(b)2, it then follows that also $\max(0, s^*_{n_2-n_1}, s^*_{n_3-n_1}, \cdots, s^*_{n_k-n_1}) \varepsilon \mathcal{H}^{(nk)}$. Thus, $f_2^* \varepsilon \mathcal{H}^{(nk)}$. Let

$$\begin{aligned} \Pi_n^{\,1} &= \{\pi : \pi \, \varepsilon \, \Pi_n \quad \text{and} \quad S_\pi \leqq a\}, \\ \Pi_n^{\,2} &= \{\pi : \pi \, \varepsilon \, \Pi_n \quad \text{and} \quad S_\pi > a\}. \end{aligned}$$

It is easily checked that when $\pi \, \varepsilon \, \Pi_n^{\,i}$, $i = 1, 2$,

$$\begin{aligned} f(X_1, \cdots, X_{n_k}) &\geqq f_i^*(X_1, \cdots, X_{n_k}) \quad \text{on} \quad \Omega(n_k, n_k), \\ f(X_1, \cdots, X_{n_k}) &= f_i^*(X_1, \cdots, X_{n_k}) \quad \text{on the kernel in} \quad \Omega(n_k, n_k). \end{aligned}$$

By the domination principle it now follows that $f \, \varepsilon \, \mathcal{H}^{(nk)}$. Thus, Theorem 4 holds for $\varphi(x) = \max(a, x)$.

Next, let $\varphi(x)$ be convex and non-decreasing. Then $\varphi(x)$ can be represented

$$\varphi(x) = \lim_{k\to\infty} (a + \textstyle\sum_{\nu=1}^{k} b_\nu (x - c_\nu)^+)$$

($\alpha^+ = \max(0, \alpha)$) for suitable choices of a, b_ν and c_ν where $b_\nu \geqq 0$. Then

$$\begin{aligned} (7.4) \quad \max(\varphi(s_{n_1}), \cdots, \varphi(s_{n_k})) &= \varphi(\max(s_{n_1}, s_{n_2}, \cdots, s_{n_k})) \\ &= \lim_{k\to\infty} (a + \textstyle\sum_{\nu=1}^{k} b_\nu[\max(c_\nu, s_{n_1}, \cdots, s_{n_k}) - c_\nu])^{\cdot} \end{aligned}$$

We have proved that $\max(c_\nu, s_{n_1}, \cdots, s_{n_k}) \, \varepsilon \, \mathcal{H}^{(nk)}$. From (7.4) and Lemma 4.1 it follows that Theorem 4 is true when φ is convex and non-decreasing.

To treat the case when φ is convex and non-increasing, we first observe that if $f(x_1, x_2, \cdots, x_n) \, \varepsilon \, \mathcal{H}^{(n)}$ then also $f(-x_1, -x_2, \cdots, -x_n) \, \varepsilon \, \mathcal{H}^{(n)}$. This is a consequence of Lemma 4.1(b)1.

Let $\varphi(x)$ be convex and non-increasing. Then

$$(7.5) \quad \max(\varphi(s_{n_1}), \varphi(s_{n_2}), \cdots, \varphi(s_{n_k})) = \max(\varphi(-(-s_{n_1})), \cdots, \varphi(-(-s_{n_k}))).$$

As $\varphi(-x)$ is convex and non-decreasing, it follows from what is already proved that the right hand side, and thus, the left hand side in (7.5) belong to $\mathcal{H}^{(nk)}$. Thereby, Theorem 4 is completely proved.

The author conjectures that Theorem 4 is true if it is only assumed that $\varphi(x)$ is convex and continuous, but not necessarily monotone.

Acknowledgment. The author is indebted to the referee for his careful reading of the manuscript and for his pointing out some references.

REFERENCES

[1] GODAMBE, V. P. (1965). A review of the contributions towards a unified theory of sampling from finite populations. *Rev. Int. Statist. Inst.* **33** 242–258.

[2] HARDY, G. H., LITTLEWOOD, J. E. and PÓLYA, G. (1959). *Inequalities.* Cambridge Univ. Press.

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