UNIVERSITY OF NORTH CAROLINA
Department of Statistics
Chapel Hill, N. C.

PROBABILITY INEQUALITIES FOR SUMS OF BOUNDED RANDOM VARIABLES

by

Wassily Hoeffding

May 1962

Contract No. AF(638)-261

Upper bounds are derived for the probability that the sum S of n independent random variables exceeds its mean ES by a positive number nt. It is assumed that the range of each summand of S is bounded or bounded above. The bounds for $\Pr\{S-ES \geq nt\}$ depend only on the endpoints of the ranges of the summands and the mean, or the mean and the variance of S. These results are then used to obtain analogous inequalities for certain sums of dependent random variables such as U statistics and the sum of a random sample without replacement from a finite population.

This research was supported by the Mathematics Division of the Air Force Office of Scientific Research.

Institute of Statistics
Mimeo Series No. 326

PROBABILITY INEQUALITIES FOR SUMS OF BOUNDED RANDOM VARIABLES[1]

Wassily Hoeffding
University of North Carolina

Summary. Upper bounds are derived for the probability that the sum S of n independent random variables exceeds its mean ES by a positive number nt. It is assumed that the range of each summand of S is bounded or bounded above. The bounds for $\Pr\{S - ES \geq nt\}$ depend only on the endpoints of the ranges of the summands and the mean, or the mean and the variance of S. These results are then used to obtain analogous inequalities for certain sums of dependent random variables such as U statistics and the sum of a random sample without replacement from a finite population.

1. Introduction. Let $X_1, X_2, \ldots, X_n$ be independent random variables with finite first and second moments,

$$S = X_1 + \ldots + X_n \quad , \qquad \overline{X} = S/n \quad , \tag{1.1}$$

$$\mu = E\overline{X} = ES/n \quad , \qquad \sigma^2 = n \operatorname{var}(\overline{X}) = (\operatorname{var} S)/n \quad . \tag{1.2}$$

(Thus if the X_i have a common mean then its value is μ and if they have a common variance then its value is σ^2.) In section 2 upper bounds are given for the probability,

$$\Pr\{\overline{X} - \mu \geq t\} = \Pr\{S - ES \geq nt\} \quad , \tag{1.3}$$

where $t > 0$, under the additional assumption that the range of each random variable X_i is bounded (or at least bounded from above). These upper bounds depend only on t, n, the endpoints of the ranges of the X_i, and on μ, or on μ and σ. We assume $t > 0$ since for $t \leq 0$ no nontrivial upper bound exists under our assumptions. The proofs are given in section 3. Note that an upper bound for $\Pr\{\overline{X} - \mu \geq t\}$ implies in an obvious way an upper bound for $\Pr\{-\overline{X} + \mu \geq t\}$ and hence also for

[1] This research was supported by the Air Force Office of Scientific Research.

$$\Pr\left\{ |\bar{X} - \mu| \geq t \right\} = \Pr\left\{ \bar{X} - \mu \geq t \right\} + \Pr\left\{ -\bar{X} + \mu \geq t \right\} \quad . \tag{1.4}$$

Known upper bounds for these probabilities include the Bienaymé-Chebyshev inequality

$$\Pr\left\{ |\bar{X} - \mu| \geq t \right\} \leq \frac{\sigma^2}{nt^2} \quad , \tag{1.5}$$

Chebyshev's[2] inequality

$$\Pr\left\{ \bar{X} - \mu \geq t \right\} \leq \frac{1}{1 + \frac{nt^2}{\sigma^2}} \tag{1.6}$$

(which do not require the assumption of bounded summands) and the inequalities of Bernstein and Prohorov (see formulas (2.22) and (2.23) below). Surveys of inequalities of this type have been given by Godwin [4], Savage [11] and Bennett [1]. Bennett also derived new inequalities, in particular inequality (2.21) below, and made instructive comparisons between different bounds.

The method employed to derive the inequalities, which has often been used (apparently first by S. N. Bernstein), is based on the following simple observation. The probability $\Pr\left\{ S - ES \geq nt \right\}$ is the expected value of the function which takes the values 0 and 1 according as $S - ES - nt$ is < 0 or ≥ 0. This function does not exceed $\exp\left\{ h(S - ES - nt) \right\}$, where h is an arbitrary positive constant. Hence

$$\Pr\left\{ \bar{X} - \mu \geq t \right\} = \Pr\left\{ S - ES \geq nt \right\} \leq Ee^{h(S-ES-nt)} \quad . \tag{1.7}$$

If, as we here assume, the summands of S are independent, then

---

[2] Inequality (1.6) has been attributed to various authors. Chebyshev [12] seems to be the first to have announced an inequality which implies (1.6) as an illustration of a general class of inequalities.

$$E\, e^{h(S-ES-nt)} = e^{-hnt} \prod_{i=1}^{n} E\, e^{h(X_i - EX_i)} \qquad (1.8)$$

It remains to obtain an upper bound for the expected value in (1.8) and to minimize this bound with respect to h. The bounds (2.1) and (2.12) of Theorems 1 and 3 are the best that can be obtained by this method under the assumptions of the theorems. They are not the best possible bounds for the probability in (1.7). The bounds derived in this paper are better than the Chebyshev bounds (1.5) and (1.6) except for small values of t. Typically, if t is held fixed, they tend to zero at an exponential rate as n increases.

In section 4 the results of the preceding sections are used to obtain probability bounds for certain sums of dependent random variables such as U statistics and sums of m-dependent random variables. In section 5 a relation between samples with and without replacement from a finite population is established which implies probability bounds for the sum of a sample without replacement.

The following facts about convex functions will be used; for proofs see [5]. A continuous function f(x) is convex in the interval I if and only if $f(px + (1 - p)y) \leq pf(x) + (1 - p)f(y)$ for $0 < p < 1$ and all x and y in I. If this is true for all real x and y, the function is simply called convex. A continuous function is convex in I if it has a nonnegative second derivative in I. If f(x) is continuous and convex in I then for any positive numbers $p_1, \ldots, p_N$ such that $p_1 + \ldots + p_N = 1$ and any numbers $x_1, \ldots, x_N$ in I

$$f\left(\sum_{i=1}^{N} p_i x_i\right) \leq \sum_{i=1}^{N} p_i f(x_i) \qquad (1.9)$$

This is known as Jensen's inequality.

2. Sums of independent random variables. In this section probability bounds for sums of independent random variables are stated and discussed. The proofs are given in section 3.

Let $X_1, X_2, \ldots, X_n$ be independent random variables and let S, $\overline{X}$, μ and σ^2 be defined by (1.1) and (1.2). First we consider bounds which do not depend on σ^2.

Theorem 1. If $X_1, X_2, \ldots, X_n$ are independent and $0 \leq X_i \leq 1$ for $i = 1, \ldots, n$, then for $0 < t < 1 - \mu$

$$\Pr\left\{ \overline{X} - \mu \geq t \right\} \leq \left\{ \left(\frac{\mu}{\mu+t}\right)^{\mu+t} \left(\frac{1-\mu}{1-\mu-t}\right)^{1-\mu-t} \right\}^n \tag{2.1}$$

$$\leq e^{-nt^2 g(\mu)} \tag{2.2}$$

$$\leq e^{-2nt^2} , \tag{2.3}$$

where

$$g(\mu) = \frac{1}{1-2\mu} \ln \frac{1-\mu}{\mu} \quad \text{for } 0 < \mu < \frac{1}{2} , \quad g(\mu) = \frac{1}{2\mu(1-\mu)} \quad \text{for } \frac{1}{2} \leq \mu < 1 , \tag{2.4}$$

The assumption $0 \leq X_i \leq 1$ has been made to give the bounds a simple form. If instead we assume $a \leq X_i \leq b$, the values μ and t in the three upper bounds of the theorem are to be replaced by $(\mu - a)/(b - a)$ and $t/(b - a)$, respectively.

If $t > 1 - \mu$, then under the assumptions of Theorem 1 the probability in (2.1) is zero. Inequality (2.1) remains true for $t = 1 - \mu$ if the right hand side is replaced by its limit as t tends to $1 - \mu$, which is μ^n. In this special case the sign of equality in (2.1) can be attained. Indeed, if $t = 1 - \mu$, then $\Pr\left\{ X - \mu \geq t \right\} = \Pr\left\{ \overline{X} = 1 \right\} = \Pr\left\{ S = n \right\}$, and $\Pr\left\{ S = n \right\} = \mu^n$ if

$$\Pr\left\{ X_i = 0 \right\} = 1 - \mu , \quad \Pr\left\{ X_i = 1 \right\} = \mu , \quad i = 1, \ldots, n , \tag{2.5}$$

that is, if S has the binomial distribution with parameters n and μ.

The bound in (2.1) is the best that can be obtained from inequality (1.7) under the assumptions of the theorem. Indeed, it is the minimum with respect to h of the right hand side of (1.7) when the X_i have the distribution (2.5).

Denote the bounds in (2.1) and (2.2) by A_1 and A_2, respectively. Then $A_1 \leq A_2$ and the first bound is appreciably better than the second if the ratio A_1/A_2 is not close to 1. We can write

$$A_1 = e^{-nt^2 G(t,\mu)} \quad , \tag{2.6}$$

where

$$t^2 G(t,\mu) = (\mu + t)\ln(1 + \frac{t}{\mu}) + (1 - \mu - t)\ln(1 - \frac{t}{1 - \mu}) \quad . \tag{2.7}$$

If t is small, we can approximate $G(t,\mu)$ by the first terms of its expansion in powers of t,

$$G(t,\mu) = \frac{1}{2\mu(1 - \mu)} + \frac{2\mu - 1}{6\mu^2(1 - \mu)^2} t + \frac{1 - 3\mu(1 - \mu)}{12\mu^3(1 - \mu)^3} t^2 + \ldots \quad . \tag{2.8}$$

If $\mu > \frac{1}{2}$, then all coefficients in the expansion (2.8) are positive and we have

$$\frac{A_1}{A_2} < \exp\left\{ - \frac{2\mu - 1}{6\mu^2(1 - \mu)^2} \; t^3 n \right\} \quad . \tag{2.9}$$

If $\mu = \frac{1}{2}$, then $A_1/A_2 < \exp(-4t^4 n/3)$. In this case the bounds in (2.2) and (2.3) are equal.

If $\mu < \frac{1}{2}$, then $g(\mu) < \frac{1}{2\mu(1 - \mu)}$ and for t small A_1 is appreciably better than A_2 unless $\left[\frac{1}{2\mu(1-\mu)} - g(\mu) \right] t^2 n$ is small. We have $A_1 = A_2$ if $\mu < \frac{1}{2}$ and $t = 1 - 2\mu$.

For the special (binomial) case (2.5) the inequalities of Theorem 1 (except for (2.2) with $\mu < \frac{1}{2}$) have been derived by Okamoto [9].

The following theorem gives an extension of bound (2.3) to the case where the ranges of the summands need not be the same.

<u>Theorem 2.</u> If $X_1, X_2, \ldots, X_n$ are independent and $a_i \leq X_i \leq b_i$ $(i = 1, 2, \ldots, n)$, then for $t > 0$

$$\Pr\left\{\bar{X} - \mu \geq t\right\} \leq e^{-\frac{2n^2t^2}{\Sigma_{i=1}^n (b_i - a_i)^2}} . \tag{2.10}$$

As an application of Theorem 2 we obtain the following bound for the distribution function of the difference of two sample means.

<u>Corollary.</u> If $Y_1, \ldots, Y_m, Z_1, \ldots, Z_n$ are independent random variables with values in the interval $[a, b]$, and if $\bar{Y} = (Y_1 + \ldots + Y_m)/m$, $\bar{Z} = (Z_1 + \ldots + Z_n)/n$, then for $t > 0$

$$\Pr\left\{\bar{Y} - \bar{Z} - (E\bar{Y} - E\bar{Z}) \geq t\right\} \leq e^{-\frac{2t^2}{(m^{-1}+n^{-1})(b-a)^2}} . \tag{2.11}$$

The inequalities of the next theorem depend also on the variance σ^2/n of $\bar{X}$. We now assume that the X_i have a common mean. For simplicity the mean is taken to be zero.

<u>Theorem 3.</u> If $X_1, X_2, \ldots, X_n$ are independent, $EX_i = 0$, $X_i \leq b$ $(i = 1, 2, \ldots, n)$, then for $0 < t < b$

$$\Pr\left\{\bar{X} \geq t\right\} \leq \left\{ \left(1 + \frac{bt}{\sigma^2}\right)^{-\left(1 + \frac{bt}{\sigma^2}\right)\frac{\sigma^2}{b^2+\sigma^2}} \left(1 - \frac{t}{b}\right)^{-\left(1 - \frac{t}{b}\right)\frac{b^2}{b^2+\sigma^2}} \right\}^n \tag{2.12}$$

$$\leq e^{-\frac{nt}{b}\left[\left(1 + \frac{\sigma^2}{bt}\right)\ln\left(1 + \frac{bt}{\sigma^2}\right) - 1\right]} . \tag{2.13}$$

Here the summands are assumed to be bounded only from above. However, to obtain from this theorem an upper bound for $\Pr\left\{|\bar{X}| \geq t\right\}$, we must assume that the summands are bounded on both sides.

Inequality (2.12) is the best that can be obtained from (1.7) under the present assumptions. It is the minimum with respect to h of the right hand side of (1.7) when the X_i have the distribution

$$(2.14) \qquad \Pr\left\{ X_i = -\frac{\sigma^2}{b} \right\} = \frac{b^2}{b^2 + \sigma^2} , \quad \Pr\left\{ X_i = b \right\} = \frac{\sigma^2}{b^2 + \sigma^2} ,$$

$$i = 1, \ldots, n .$$

Inequality (2.12) is true also for $t = b$ if the right hand side is replaced by its limit as t tends to b, which is $[\sigma^2/(b^2 + \sigma^2)]^n$. In this case the sign of equality in (2.12) is attained when the distribution is (2.14).

The bound (2.13) is due to Bennett ([1], inequality (8b)). (Bennett's notation is different from mine. His first proof assumes $|X_i| \leq b$ (= his M), a second proof (pp. 42-43) uses only $X_i \leq b$.)

Let the upper bounds in (2.12) and (2.13) be denoted by B_1 and B_2, respectively. The ratio B_1/B_2 can be written in the form

$$(2.15) \qquad \frac{B_1}{B_2} = e^{-n\phi(v,w)} ,$$

where

$$(2.16) \qquad v = \frac{1}{1 + \frac{\sigma^2}{bt}} , \qquad w = \frac{t}{b} ,$$

$$(2.17) \qquad \phi(v,w) = w \frac{\rho(v) + \rho(w)}{v^{-1} + w^{-1} - 1} ,$$

$$(2.18) \qquad \begin{aligned} \rho(x) &= x^{-2}\left[(1 - x)\ln(1 - x) + x - \tfrac{1}{2}x^2\right] \\ &= \frac{1}{2\cdot 3}x + \frac{1}{3\cdot 4}x^2 + \frac{1}{4\cdot 5}x^3 + \ldots \end{aligned}$$

Since $0 < t < b$, both v and w are between 0 and 1. The function $\phi(v,w)$ is an increasing function of both v and w. Hence $0 < \phi(v,w) < 1$. If b and σ^2 are both fixed and t approaches 0, then

$$\phi(v,w) \sim \frac{t^3}{6b\sigma^2} \qquad (2.19)$$

Hence B_1 is appreciably smaller than B_2 if $n\, t^3/(6b\sigma^2)$ is not small.

If we let

$$\lambda = \frac{bt}{\sigma^2} \quad , \qquad \tau = \frac{nt}{b} \quad , \qquad (2.20)$$

Bennett's inequality (bound (2.13)) can be written

$$\Pr\left\{ \overline{X} \geq t \right\} \leq e^{-\tau h_1(\lambda)} \quad , \quad h_1(\lambda) = (1 + \tfrac{1}{\lambda})\ln(1 + \lambda) - 1 \quad . \qquad (2.21)$$

Bennett has shown that (2.21) is better than Bernstein's

$$\Pr\left\{ \overline{X} \geq t \right\} \leq e^{-\tau h_2(\lambda)} \quad , \quad h_2(\lambda) = \frac{\lambda}{2(1 + \frac{1}{3}\lambda)} \quad . \qquad (2.22)$$

Inequality (2.21) is also better than Prohorov's [10]

$$\Pr\left\{ \overline{X} \geq t \right\} \leq e^{-\tau h_3(\lambda)} \quad ,$$

$$h_3(\lambda) = \frac{1}{2} \operatorname{arcsinh} \frac{\lambda}{2} = \frac{1}{2} \ln\left(\frac{\lambda}{2} + \left[1 + \left(\frac{\lambda}{2}\right)^2\right]^{1/2}\right) \quad . \qquad (2.23)$$

Indeed, it can be shown that the bound in (2.21) is the best bound of the form $\exp\{-\tau\, h(\lambda)\}$ that can be obtained from (2.12) and hence from (1.7). If λ is small, Bernstein's bound (2.22) does not differ much from Bennett's (2.21).

Under certain conditions $\overline{X}$ is approximately normally distributed when n is large, so that, for $y = \sqrt{n}\, t/\sigma$ fixed,

$$\Pr\left\{ \overline{X} - \mu \geq t \right\} = \Pr\left\{ \overline{X} - \mu \geq \frac{\sigma y}{\sqrt{n}} \right\} \longrightarrow \frac{1}{\sqrt{2\pi}} \int_y^{\infty} e^{-x^2/2}\, dx = \Phi(-y) \qquad (2.24)$$

as $n \longrightarrow \infty$. (Sufficient conditions are $n\sigma^2 \longrightarrow \infty$ and $\Sigma E|X_i - EX_i|^3/(\sigma\sqrt{n})^3 \longrightarrow 0$.) It is instructive to compare the present bounds with the upper bound for

$\bar{\Phi}(-y)$ which results from inequality (1.7) when $\bar{X}$ is normally distributed. In this case the right hand side of (1.7) is $\exp(-hnt + h^2 n\sigma^2/2)$. If we minimize with respect to h we obtain

$$\Pr\left\{ \bar{X} - \mu \geq t \right\} = \bar{\Phi}\left(-\frac{\sqrt{n}\, t}{\sigma}\right) \leq e^{-\frac{nt^2}{2\sigma^2}} \tag{2.25}$$

or $\bar{\Phi}(-y) \leq \exp(-y^2/2)$, where $y > 0$. This bound for $\bar{\Phi}(-y)$ is rather crude, especially when y is large, in which case $\bar{\Phi}(-y)$ is approximated by $\frac{1}{y\sqrt{2\pi}}\exp(-y^2/2)$. In contrast, the bounds (2.1) and (2.12) are attainable at the largest nontrivial values of t. It is interesting to note that the bound (2.2) with $\mu \geq 1/2$ is equal to the right hand side of (2.25) in the binomial case (2.5). The bound (2.10) of Theorem 2 is equal to the right hand side of (2.25) in the case where $\Pr\left\{ X_i = a_i \right\} = \Pr\left\{ X_i = b_i \right\} = \frac{1}{2}$ for all i. Bernstein's bound (2.22) is close to the right hand side of (2.25) when $\lambda = bt/\sigma^2$ is small. The same is true of the bounds of Theorem 3.

The inequalities of this section can be strengthened in the following way. Let $S_m = X_1 + \ldots + X_m$ for $m = 1, 2, \ldots, n$. It follows from a theorem of Doob ([2], p. 314) that

$$\Pr\left\{ \max_{1 \leq m \leq n} (S_m - ES_m) \geq nt \right\} \leq E\, e^{h(S_n - ES_n - nt)} \tag{2.26}$$

for $h > 0$. The right hand side is the same as that of inequality (1.7) (where $S = S_n$). Since the inequalities of Theorems 1, 2 and 3 have been obtained from (1.7), the right hand sides of those inequalities are upper bounds for the probability in (2.26) under the stated assumptions. This stronger result is analogous to an inequality of Kolmogorov (see, e.g., Feller [3], p. 220).

Furthermore, the inequalities of Theorems 1 and 2 remain true if the assumption that $X_1, X_2, \ldots, X_n$ are independent is replaced by the weaker assumption that the sequence $S'_m = S_m - ES_m$, $m = 1, 2, \ldots, n$, is a martingale, that is,

$$(2.27) \qquad E(S'_m \mid S'_1, \ldots, S'_j) = S'_j \quad , \quad 1 \leq j \leq m \leq n \quad ,$$

with probability one. Indeed, Doob's inequality (2.26) is true under this assumption. On the other hand, (2.27) implies that the conditional mean of X_m for S'_{m-1} fixed is equal to its unconditional mean. A slight modification of the proofs of Theorems 1 and 2 yields the stated result.

3. Proofs of the theorems of section 2. Let X be a random variable such that $a \leq X \leq b$. Since the exponential function exp(hX) is convex, its graph is bounded above on the interval $a \leq X \leq b$ by the straight line which connects its ordinates at X = a and X = b. Thus

$$(3.1) \qquad e^{hX} \leq \frac{b - X}{b - a} e^{ha} + \frac{X - a}{b - a} e^{hb} \quad , \quad a \leq X \leq b \quad .$$

Hence we obtain

Lemma 1. If X is a random variable such that $a \leq X \leq b$, then for any real number h

$$(3.2) \qquad E\, e^{hX} \leq \frac{b - EX}{b - a} e^{ha} + \frac{EX - a}{b - a} e^{hb} \quad .$$

We now prove Theorem 1. By (1.7) and (1.8) we have for $h > 0$

$$(3.3) \qquad Pr\left\{ \bar{X} - \mu \geq t \right\} \leq e^{-hnt-hn\mu} \prod_{i=1}^{n} E\, e^{hX_i} \quad .$$

By assumption $0 \leq X_i \leq 1$. Let $\mu_i = EX_i$. Then $n\mu = \mu_1 + \mu_2 + \ldots + \mu_n$. By Lemma 1 with $X = X_i$, a = 0, b = 1 we have

$$(3.4) \qquad \prod_{i=1}^{n} E\, e^{hX_i} \leq \prod_{i=1}^{n} (1 - \mu_i + \mu_i e^h) \quad .$$

Since the geometric mean does not exceed the arithmetic mean,

$$(3.5)\qquad \left\{ \prod_{i=1}^{n} (1 - \mu_i + \mu_i e^h) \right\}^{\frac{1}{n}} \le \frac{1}{n} \sum_{i=1}^{n} (1 - \mu_i + \mu_i e^h) = 1 - \mu + \mu e^h \quad .$$

It follows from (3.3), (3.4) and (3.5) that

$$(3.6)\qquad \Pr\left\{ \bar{X} - \mu \ge t \right\} \le \left\{ e^{-ht-h\mu}(1 - \mu + \mu e^h) \right\}^n \quad .$$

The right hand side of (3.6) attains its minimum at $h = h_o$, where

$$(3.7)\qquad h_o = \ln \frac{(1 - \mu)(\mu + t)}{(1 - \mu - t)\mu} \quad .$$

Since $0 < t < 1 - \mu$, h_o is positive. Inserting $h = h_o$ in (3.6) we obtain inequality (2.1) of Theorem 1.

To prove inequality (2.2) we write the right hand side of (2.1) in the form $\exp(-nt^2 G(t,\mu))$ (as in (2.6)), where

$$(3.8)\qquad G(t,\mu) = \frac{\mu + t}{t^2} \ln \frac{\mu + t}{\mu} + \frac{1 - \mu - t}{t^2} \ln \frac{1 - \mu - t}{1 - \mu} \quad .$$

Inequality (2.2) will be proved by showing that $g(\mu)$ as defined in (2.4) is the minimum of $G(t,\mu)$ with respect to t, where $0 \le t < 1 - \mu$. The derivative $\partial G(t,\mu)/\partial t$ can be written in the form

$$(3.9)\qquad \begin{aligned} t^2 \frac{\partial}{\partial t} G(t,\mu) &= \left(1 - 2\,\frac{1 - \mu}{t}\right)\ln\left(1 - \frac{t}{1-\mu}\right) - \left(1 - 2\,\frac{\mu+t}{t}\right)\ln\left(1 - \frac{t}{\mu+t}\right) \\ &= H\left(\frac{t}{1-\mu}\right) - H\left(\frac{t}{\mu+t}\right) \quad , \end{aligned}$$

where $H(x) = (1 - 2x^{-1})\ln(1 - x)$. By assumption $0 \le t/(\mu + t) < 1$ and $0 \le t/(1 - \mu) < 1$. For $|x| < 1$ we have the expansion

$$(3.10)\qquad H(x) = 2 + \left(\tfrac{2}{3} - \tfrac{1}{2}\right)x^2 + \left(\tfrac{2}{4} - \tfrac{1}{3}\right)x^3 + \left(\tfrac{2}{5} - \tfrac{1}{4}\right)x^4 + \ldots \quad ,$$

where the coefficients are positive. Thus $H(x)$ increases for $0 < x < 1$. It follows from (3.9) that $\partial G/\partial t > 0$ if and only if $t/(1 - \mu) > t/(\mu + 1)$, that is, $t > 1 - 2\mu$. Hence if $1 - 2\mu > 0$, $G(t,\mu)$ has its minimum at $t = 1 - 2\mu$ and the value of the minimum is $(\ln \frac{1-\mu}{\mu})/(1 - 2\mu) = g(\mu)$. If $1 - 2\mu \leq 0$, then $G(t,\mu)$ has its minimum at $t = 0$ and the value of the minimum is $1/[2\mu(1 - \mu)] = g(\mu)$ (see (2.8)). This proves inequality (2.2).

It is easily seen that $g(\mu) \geq g(\frac{1}{2}) = 2$. This implies inequality (2.3). The proof of Theorem 1 is complete.

We next prove Theorem 2. The proof will also indicate a short direct derivation of the simple bound (2.3).

In Theorem 2 we assume $a_i \leq X_i \leq b_i$. Let again $\mu_i = EX_i$. By (1.7) and (1.8),

$$\Pr\{\bar{X} - \mu \geq t\} \leq e^{-hnt} \prod_{i=1}^{n} E\, e^{h(X_i-\mu_i)} \quad . \tag{3.11}$$

By Lemma 1,

$$E\, e^{h(X_i-\mu_i)} \leq e^{-h\mu_i}\left(\frac{b_i-\mu_i}{b_i-a_i} e^{ha_i} + \frac{\mu_i-a_i}{b_i-a_i} e^{hb_i}\right) = e^{L(h_i)} \quad , \tag{3.12}$$

where

$$L(h_i) = -h_i p_i + \ln(1 - p_i + p_i e^{h_i}) \quad , \tag{3.13}$$

$$h_i = h(b_i - a_i) \quad , \quad p_i = \frac{\mu_i - a_i}{b_i - a_i} \quad . \tag{3.14}$$

The first two derivatives of $L(h_i)$ are

$$L'(h_i) = -p_i + \frac{p_i}{(1 - p_i)e^{-h_i} + p_i} \quad ,$$

$$L''(h_i) = \frac{p_i(1 - p_i)\, e^{-h_i}}{[(1 - p_i)e^{-h_i} + p_i]^2} \quad .$$

The last ratio is of the form $u(1 - u)$ where $0 < u < 1$. Hence $L''(h_i) \leq 1/4$. Therefore by Taylor's formula

$$(3.15) \qquad L(h_i) \leq L(0) + L'(0)h_i + \frac{1}{8} h_i^2 = \frac{1}{8} h_i^2 = \frac{1}{8} h^2(b_i - a_i)^2 \quad .$$

Hence by (3.12)

$$(3.16) \qquad E\, e^{h(x_i - \mu_i)} \leq e^{\frac{1}{8} h^2(b_i - a_i)^2}$$

and by (3.11)

$$(3.17) \qquad \Pr\left\{ \bar{X} - \mu \geq t \right\} \leq e^{-hnt + \frac{1}{8} h^2 \sum_{i=1}^{n} (b_i - a_i)^2} \quad .$$

The right hand side of (3.17) has its minimum at $h = 4nt/\Sigma(b_i - a_i)^2$. Inserting this value in (3.17) we obtain inequality (2.10) of Theorem 2.

To prove Theorem 3 we need two lemmas.

Lemma 2. If X is a random variable such that $EX = 0$, $EX^2 = \sigma^2$ and $X \leq b$, then for any positive number h

$$(3.18) \qquad E\, e^{hX} \leq \frac{b^2}{b^2 + \sigma^2}\, e^{-\frac{\sigma^2}{b} h} + \frac{\sigma^2}{b^2 + \sigma^2}\, e^{bh} \quad .$$

A proof of this inequality can be found in Bennett [1] .

Lemma 3. If $c > 0$, the function

$$(3.19) \qquad f(u) = \ln\left(\frac{1}{1 + u}\, e^{-cu} + \frac{u}{1 + u}\, e^{c}\right)$$

has a negative second derivative for $u \geq 0$.

To prove this we write $f(u) = c + \ln f_1(y)$, where $y = 1 + u$ and

$$f_1(y) = y^{-1} e^{-cy} - y^{-1} + 1 \quad .$$

For the second derivative $f''(u)$ we have $f_1^2(y)f''(u) = f_1(y)f_1''(y) - (f_1'(y))^2$. Now

$$f_1'(y) = (-y^{-2} - cy^{-1})e^{-cy} + y^{-2} \quad ,$$

$$f_1''(y) = (2y^{-3} + 2cy^{-2} + c^2y^{-1})e^{-cy} - 2y^{-3}$$

$$= -2y^{-3}e^{-cy}(e^{cy} - 1 - cy - \tfrac{1}{2}c^2y^2) \quad ,$$

which is negative for $cy > 0$. Since $f_1(y) > 0$ for $y > 1$, it follows that $f''(u) < 0$ for $u > 0$.

We now can prove Theorem 3. By assumption, $EX_i = 0$ and $X_i \leq b$. Let $\sigma_i^2 = EX_i^2$, so that $n\sigma^2 = \sigma_1^2 + \sigma_2^2 + \ldots + \sigma_n^2$. By (1.7), (1.8) and Lemma 2,

$$\Pr\left\{ \bar{X} \geq t \right\} \leq e^{-hnt} \prod_{i=1}^{n} \left(\frac{b^2}{b^2 + \sigma_i^2} e^{-\frac{\sigma_i^2}{b}h} + \frac{\sigma_i^2}{b^2 + \sigma_i^2} e^{bh}\right)$$

(3.20)

$$= e^{-hnt + \sum_{i=1}^{n} f(\sigma_i^2/b^2)} \quad ,$$

where f is the function defined by (3.19), with $c = bh$. Since, by Lemma 3, $f(u)$ has a negative second derivative, $-f(u)$ is convex for $u \geq 0$. Therefore by Jensen's inequality (1.9)

$$\frac{1}{n} \sum_{i=1}^{n} f(\sigma_i^2/b^2) \leq f\left(\frac{1}{n} \sum_{i=1}^{n} \sigma_i^2/b^2\right) = f(\sigma^2/b^2)$$

(3.21)

$$= \ln\left(\frac{b^2}{b^2 + \sigma^2} e^{-\frac{\sigma^2}{b}h} + \frac{\sigma^2}{b^2 + \sigma^2} e^{bh}\right) \quad .$$

It follows from (3.20) and (3.21) that

$$(3.22) \qquad \Pr\left\{ \bar{X} \geq t \right\} \leq \left(\frac{b^2}{b^2+\sigma^2} e^{-(t+\frac{\sigma^2}{b})h} + \frac{\sigma^2}{b^2+\sigma^2} e^{(b-t)h} \right)^n .$$

The right hand side of (3.22) attains its minimum at $h = h_1$, where

$$h_1 = \frac{b}{b^2+\sigma^2} \ln \frac{1 + \frac{tb}{\sigma^2}}{1 - \frac{t}{b}} .$$

Inserting this value in (3.22), we obtain inequality (2.12) of Theorem 3.

Inequality (2.13) follows from equations (2.15) to (2.18). The proof of Theorem 3 is complete.

As noted above, the upper bound (2.13) for $\Pr\left\{ \bar{X} \geq t \right\}$ has been derived by Bennett [1]. An alternative direct proof goes as follows. By Lemma 3, if $u > 0$, then $f''(u) < 0$ and hence $f(u) \leq f(0) + f'(0)u = (e^c - 1 - c)u$. Applying this inequality to the right side of (3.20) (where $c = bh$) and minimizing with respect to h we obtain the bound (2.13).

4. Sums of dependent random variables. The inequalities of sections 2 and 3 can be used to obtain probability bounds for certain sums of dependent random variables. Suppose that T is a random variable which can be written in the form

$$(4.1) \qquad T = p_1 T_1 + p_2 T_2 + \dots + p_N T_N ,$$

where each of $T_1, T_2, \dots, T_N$ is a sum of independent random variables and $p_1, p_2, \dots, p_N$ are nonnegative numbers, $p_1+p_2+ \dots + p_N = 1$. The random variables $T_1, T_2, \dots, T_N$ need not be mutually independent. For $h > 0$

$$\Pr\left\{ T \geq t \right\} \leq e^{-ht} E e^{hT} .$$

Since the exponential function is convex, we have by Jensen's inequality (1.9)

$$\exp(hT) = \exp(h \sum_{i=1}^{N} p_i T_i) \leq \sum_{i=1}^{N} p_i \exp(h\, T_i) \quad .$$

Therefore

$$\Pr \{ T \geq t \} \leq \sum_{i=1}^{N} p_i E e^{h(T_i - t)} \quad . \tag{4.2}$$

Since each T_i is a sum of independent random variables, the expectations on the right can be bounded as in section 3. If the random variables T_i are identically distributed or if the upper bound for $E \exp(h(T_i - t))$ is independent of i, then the upper bound we obtain for $\Pr \{ T \geq t \}$ is also an upper bound for $\Pr \{ T_i \geq t \}$. The bounds obtained in this way will be rather crude but may be useful.

We now consider several types of random variables T which can be represented in the form (4.1).

4a. <u>One-sample U statistics.</u> Let $X_1, X_2, \ldots, X_n$ be independent random variables (real or vector valued). For $n \geq r$ consider a random variable of the form

$$U = \frac{1}{n^{(r)}} \Sigma_{n,r}\, g(X_{i_1}, \ldots, X_{i_r}) \quad , \tag{4.3}$$

where $n^{(r)} = n(n-1) \ldots (n - r + 1)$ and the sum $\Sigma_{n,r}$ is taken over all r-tuples $i_1, \ldots, i_r$ of distinct positive integers not exceeding n. Random variables of the form (4.3) have been called (one-sample) U statistics. For example, if $X_i = (Y_i, Z_i)$, $i = 1, \ldots, n$, are independent random vectors with two components which have continuous distributions, then Kendall's rank correlation coefficient is of the form (4.3) with $r = 2$ and $g(X_i, X_j)$ equal to the sign of $(Y_i - Y_j)(Z_i - Z_j)$. Other examples of U statistics can be found in [6].

Let

(4.4) $$V(X_1, \ldots, X_n) = \frac{1}{k} \left\{ g(X_1, \ldots, X_r) + g(X_{r+1}, \ldots, X_{2r}) + \ldots + g(X_{kr-r+1}, \ldots, X_{kr}) \right\} ,$$

where $k = [n/r]$, the largest integer contained in n/r. Then

(4.5) $$U = \frac{1}{n!} \Sigma_{n,n} V(X_{i_1}, \ldots, X_{i_n}) ,$$

where (in accordance with the notation in (4.3)) the sum $\Sigma_{n,n}$ is taken over all permutations $i_1, i_2, \ldots, i_n$ of the integers 1, 2, ..., n. Each term in the sum on the right is a sum of k independent random variables. Thus (4.5) gives a representation of U in the form (4.1) with $N = n!$ and $p_i = 1/n!$.

If the function g is bounded,

(4.6) $$a \leq g(x_1, \ldots, x_r) \leq b ,$$

it follows from (4.2) and the proof of Theorem 2 that

(4.7) $$\Pr \left\{ U - EU \geq t \right\} \leq e^{-2kt^2/(b-a)^2} ,$$

where $k = [n/r]$. This is an extension of the bound (2.3). To obtain simple extensions of the other inequalities of theorems 1 and 3 we assume that the random variables $X_1, X_2, \ldots, X_n$ are identically distributed. In this case, if $0 \leq g(X_1, \ldots, X_r) \leq 1$, then the bounds of Theorem 1 with n replaced by $[n/r]$ and $\mu = Eg(X_1, \ldots, X_r)$ are upper bounds for $\Pr \left\{ U - EU \geq t \right\}$, where $EU = \mu$. If $g(X_1, \ldots, X_r) \leq EU + b$, then the right hand sides of (2.12) and (2.13) with n replaced by $[n/r]$ and $\sigma^2 = \text{var}\, g(X_1, \ldots, X_r)$ aee upper bounds for $\Pr \left\{ U - EU \geq t \right\}$.

4b. Two-sample U statistics. Let $X_1, X_2, \ldots, X_m, Y_1, Y_2, \ldots, Y_n$ be $m + n$ independent random variables. For $m \geq r$ and $n \geq s$ consider a random variable of the form

$$U = \frac{1}{m^{(r)} n^{(s)}} \Sigma_{m,r;n,s}\, g(X_{i_1}, \ldots, X_{i_r}, Y_{j_1}, \ldots, Y_{j_s}) \qquad (4.8)$$

where the sum $\Sigma_{m,r;n,s}$ is taken over all r-tuples $i_1, \ldots, i_r$ of distinct positive integers $\leq m$ and all s-tuples $(j_1, \ldots, j_s)$ of distinct positive integers $\leq n$. A random variable of the form (4.8) has been called a two-sample U statistic. For example, let X_i and Y_j be real and let U' denote the number of pairs (X_i, Y_j) such that $Y_j < X_i$. (This is one form of the Wilcoxon-Mann-Whitney statistic [13], [8].) Then U'/mn is of the form (4.8) with $r = s = 1$ and $g(x,y) = 1$ or 0 according as $y < x$ or $y \geq x$. Other examples of two-sample U statistics can be found in [7].

Let

$$\begin{aligned} V(X_1, \ldots, X_m, Y_1, \ldots, Y_n) = \frac{1}{k} \{ & g(X_1, \ldots, X_r, Y_1, \ldots, Y_s) \\ & + g(X_{r+1}, \ldots, X_{2r}, Y_{s+1}, \ldots, Y_{2s}) + \ldots \\ & + g(X_{kr-r+1}, \ldots, X_{kr}, Y_{ks-s+1}, \ldots, Y_{ks}) \} \end{aligned} \qquad (4.9)$$

where

$$k = \min([m/r], [n/s]) \qquad (4.10)$$

Then U as defined in (4.8) can be written as

$$U = \frac{1}{m!n!} \Sigma_{m,m;n,n}\, V(X_{i_1}, \ldots, X_{i_m}, Y_{j_1}, \ldots, Y_{j_n}) \qquad (4.11)$$

Each term on the right is a sum of k independent random variables. Thus (4.11) represents U in the form (4.1).

If $a \leq g \leq b$, then for U as defined by (4.8) we have inequality (4.7) where k is now given by (4.10). If we assume that $X_1, \ldots, X_m$ have a common distribution and $Y_1, \ldots, Y_n$ have a common distribution (not necessarily the same as that of X_i), then the terms in (4.11) are identically distributed and we obtain extensions of the inequalities of Theorems 1 and 3 analogous to those discussed at the end of section 4a, where now n is replaced by k as defined by (4.10).

4c. Sums related to U statistics. Let again $X_1, X_2, \ldots, X_n$ be independent and consider the random variable

$$W = \frac{1}{n^r} \sum_{i_1=1}^{n} \cdots \sum_{i_r=1}^{n} g(X_{i_1}, \ldots, X_{i_r}) \qquad . \tag{4.12}$$

For example, the Cramér-von Mises goodness of fit statistic ω^2 is defined by

$$\omega^2 = \int_{-\infty}^{\infty} [\, F_n(x) - G(x) \,]^2 \, dG(x) \qquad , \tag{4.13}$$

where $G(x)$ is a given cumulative distribution function and $n\, F_n(x)$ denotes the number of those $X_1, \ldots, X_n$ which are $\leq x$. If $G(x)$ is continuous, we can write ω^2 in the form (4.12) with $r = 2$ and

$$g(x_1,x_2) = \frac{1}{3} + \frac{1}{2} G^2(x_1) + \frac{1}{2} G^2(x_2) - \max \{ G(x_1), G(x_2) \} \qquad . \tag{4.14}$$

A random variable W of the form (4.12) can be written as a U statistic,

$$W = \frac{1}{n^{(r)}} \Sigma_{n,r}\, g^*(X_{i_1}, \ldots, X_{i_r}) \qquad , \tag{4.15}$$

where $g^*(x_1, \ldots, x_r)$ is a weighted arithmetic mean of certain values of g. For example, for $r = 2$ and $r = 3$ we have, respectively,

$$g^*(x_1,x_2) = \frac{n-1}{n} g(x_1,x_2) + \frac{1}{n} g(x_1,x_1) \qquad , \tag{4.16}$$

$$\begin{aligned} g^*(x_1,x_2,x_3) = {} & \frac{(n-1)(n-2)}{n^2} g(x_1,x_2,x_3) + \frac{n-1}{n^2} \{ g(x_1,x_1,x_2) + g(x_1,x_2,x_1) \\ & + g(x_2,x_1,x_1) \} + \frac{1}{n^2} g(x_1,x_1,x_1) \qquad . \end{aligned} \tag{4.17}$$

(The function g^* for which (4.15) is satisfied is not uniquely determined. For example, in (4.16) the value $g(x_1,x_1)$ may be replaced by $\frac{1}{2}\, g(x_1,x_1) + \frac{1}{2}\, g(x_2,x_2)$.)

Thus the results of section 4a can be directly applied to obtain upper bounds for $\Pr\{W - EW \geq t\}$. Note also that since g^* is an arithmetic mean of values of g, $a \leq g \leq b$ implies $a \leq g^* \leq b$. Hence the right hand side of (4.7) with $k = [n/r]$ is also an upper bound for $\Pr\{W - EW \geq t\}$ if (4.6) is satisfied. (In some cases, as in example (4.14), the range of g^* is smaller than the range of g, but the difference is negligible when n is large.)

4d. Sums of m-dependent random variables. Let

$$S = Y_1 + Y_2 + \dots + Y_n \qquad , \tag{4.18}$$

where the sequence of random variables $Y_1, Y_2, \dots, Y_n$ is (r-1)-dependent; that is, the random vectors $(Y_1, \dots, Y_i)$ and $(Y_j, \dots, Y_n)$ are independent if $j - i \geq r$, where r is a positive integer. (Example: $S = X_1X_r + X_2X_{r+1} + \dots + X_nX_{r+n-1}$, where $X_1, X_2, \dots$ are independent.) Then the random variables Y_i, Y_{r+i}, Y_{2r+i}, ... are independent. For $i = 1, \dots, r$ let

$$S_i = Y_i + Y_{r+i} + Y_{2r+i} + \dots + Y_{n_i r-r+i} \;, \quad n_i = \left[\frac{n-i+r}{r}\right] \quad . \tag{4.19}$$

Then $S = S_1 + S_2 + \dots + S_r$ and S_i is a sum of n_i independent random variables. If we put $p_i = n_i/n$ then the equation

$$\frac{1}{n}(S - ES) = \sum_{i=1}^{r} p_i \frac{1}{n_i} (S_i - ES_i) \tag{4.20}$$

represents $(S - ES)/n$ in the form (4.1). Hence by (4.2)

$$\Pr\left\{\frac{1}{n}(S - ES) \geq t\right\} \leq \sum_{i=1}^{r} p_i\, e^{-ht}\, E\, e^{\frac{h}{n_i}(S_i - ES_i)} \quad . \tag{4.21}$$

If n is a multiple of r, $n = kr$, then $n_i = k$ for all i and we can obtain in a straightforward way explicit upper bounds similar to those of section 4a. In general $n_i \geq [n/r]$ and it is easy to see that the bounds for the expected values in (4.21) remain valid if n_i is replaced by $[n/r]$. Explicitly, if $a \leq Y_j \leq b$, then $\Pr\{S - ES \geq nt\} \leq \exp(-2[n/r]t^2/(b-a)^2)$. If $Y_1, Y_2, \ldots, Y_n$ are identically distributed and $0 \leq Y_j \leq 1$, then the bounds of Theorem 1 with n replaced by $[n/r]$ and $\mu = EY_1$ are upper bounds for $\Pr\{S - ES \geq nt\}$, where $ES = n\mu$. If the Y_j are identically distributed and $Y_j - EY_j \leq b$, then the right hand sides of (2.12) and (2.13) with n replaced by $[n/r]$ and $\sigma^2 = \operatorname{Var} Y_1$ are upper bounds for $\Pr\{S - ES \geq nt\}$.

5. Sampling from a finite population. In this section it will be shown that the inequalities of section 2 yield probability bounds for the sum of a random sample without replacement from a finite population. Let the population C consist of N values $c_1, c_2, \ldots, c_N$. Let $X_1, X_2, \ldots, X_n$ denote a random sample without replacement from C and let $Y_1, Y_2, \ldots, Y_n$ denote a random sample with replacement from C. The random variables $Y_1, \ldots, Y_n$ are independent and identically distributed with mean μ and variance σ^2, where

$$\mu = \frac{1}{N}\sum_{i=1}^{N} c_i \quad , \quad \sigma^2 = \frac{1}{N}\sum_{i=1}^{N}(c_i - \mu)^2 \quad . \tag{5.1}$$

If the c_i are bounded, Theorems 1, 2 and 3 give upper bounds for $\Pr\{\bar{Y} - \mu \geq t\}$, where $\bar{Y} = (Y_1 + \ldots + Y_n)/n$. It will now be shown that the same bounds, with μ and σ^2 defined by (5.1), are upper bounds for $\Pr\{\bar{X} - \mu \geq t\}$, where $\bar{X} = (X_1 + \ldots + X_n)/n$. (Note that $E\bar{X} = E\bar{Y} = \mu$ but $\operatorname{Var}\bar{X} = \frac{N-n}{N-1}\frac{\sigma^2}{n} < \frac{\sigma^2}{n} = \operatorname{Var}\bar{Y}$.) This will be an immediate consequence of

Theorem 4. If the function f(x) is continuous and convex then

$$Ef\left(\sum_{i=1}^{n} X_i\right) \leq Ef\left(\sum_{i=1}^{n} Y_i\right) \quad . \tag{5.2}$$

Applied to $f(x) = \exp(hx)$ the theorem yields the claimed result if we recall that the bounds of Theorems 1 to 3 have been obtained from inequality (1.7). (Note that the inequality $\text{Var}\, \overline{X} \leq \text{Var}\, \overline{Y}$ is a special case of (5.2).)

To prove Theorem 4 we first observe that for an arbitrary function g of n variables we have, in the notation of (4.3),

$$\text{(5.3)} \qquad Eg(X_1, \ldots, X_n) = \frac{1}{N^{(n)}} \Sigma_{N,n}\, g(c_{i_1}, \ldots, c_{i_n}) \quad ,$$

$$\text{(5.4)} \qquad Eg(Y_1, \ldots, Y_n) = \frac{1}{N^n} \sum_{i_1=1}^{N} \cdots \sum_{i_n=1}^{N} g(c_{i_1}, \ldots, c_{i_n}) \quad .$$

The right hand sides are of the same form as U in (4.3) and W in (4.12), respectively. It has been observed in section 4c that W can be written as U with g replaced by an arithmetic mean g^* of values of g. It follows that

$$\text{(5.5)} \qquad Eg(Y_1, \ldots, Y_n) = Eg^*(X_1, \ldots, X_n) \quad .$$

As mentioned after (4.17), the function g^* is not uniquely determined. The version of $g^*(x_1, \ldots, x_n)$ which is symmetric in $x_1, \ldots, x_n$ will be denoted by $\overline{g}(x_1, \ldots, x_n)$. Here we are concerned with the special case $g(x_1, \ldots, x_n) = f(x_1 + \ldots + x_n)$. In this case, if $n = 2$,

$$\overline{g}(x_1,x_2) = \frac{N-1}{N} f(x_1 + x_2) + \frac{1}{2N} f(2x_1) + \frac{1}{2N} f(2x_2) \quad .$$

In general $\overline{g}$ can be written as

$$\text{(5.6)} \qquad \overline{g}(x_1, \ldots, x_n) = \Sigma' p(k, r_1, \ldots, r_k, i_1, \ldots, i_k) f(r_1 x_{i_1} + \ldots + r_k x_{i_k}) \quad ,$$

where the sum Σ' is taken over the positive integers $k, r_1, \ldots, r_k, i_1, \ldots, i_k$ such that $k = 1, \ldots, n$; $r_1 + \ldots + r_k = n$; and $i_1, \ldots, i_k$ are all different and do not exceed n. The coefficients p are positive and do not depend on the function f. In accordance with (5.5) we have

$$(5.7) \qquad Ef(Y_1 + \dots + Y_n) = E\,\bar{g}(X_1, \dots, X_n) \quad .$$

If we let $f(x) = 1$, we see from (5.6) and (5.7) that

$$(5.8) \qquad \Sigma' p(k, r_1, \dots, r_k, i_1, \dots, i_k) = 1 \quad .$$

If we put $f(x) = x$, then $\bar{g}(x_1, \dots, x_n)$ is a linear symmetric function of $x_1, \dots, x_n$ and hence equal to $K\cdot(x_1 + \dots + x_n)$, where K is a constant factor. Since $E(Y_1 + \dots + Y_n) = E(X_1 + \dots + X_n)$, it follows from (5.7) that $K = 1$. Thus

$$(5.9) \qquad \Sigma' p(k, r_1, \dots, r_k, i_1, \dots, i_k)(r_1 x_{i_1} + \dots + r_k x_{i_k}) = x_1 + \dots + x_n .$$

If $f(x)$ is continuous and convex, it follows from (5.6), (5.8), (5.9) and Jensen's inequality (1.9) that

$$(5.10) \qquad \bar{g}(x_1, \dots, x_n) \geq f(x_1 + \dots + x_n) \quad .$$

Hence $E\,\bar{g}(X_1, \dots, X_n) \geq E\, f(X_1 + \dots + X_n)$. With (5.7) this implies Theorem 4.

REFERENCES

[1] Bennett, George, "Probability inequalities for the sum of independent random variables," Journal of the American Statistical Association, 57 (1962), 33-45.

[2] Doob, J. L., Stochastic Processes. New York: John Wiley and Sons, 1953.

[3] Feller, William, An Introduction to Probability Theory and Its Applications, Volume I, Second edition. New York: John Wiley and Sons, 1957.

[4] Godwin, H. J., "On generalizations of Tchebychef's inequality," Journal of the American Statistical Association, 50 (1955), 923-945.

[5] Hardy, G. H., Littlewood, J. E., and Polya, G., Inequalities. Cambridge: University Press, 1952.

[6] Hoeffding, Wassily, "A class of statistics with asymptotically normal distribution," Annals of Mathematical Statistics, 19 (1948), 293-325.

[7] Lehmann, E. L., "Consistency and unbiasedness of certain nonparametric tests," Annals of Mathematical Statistics, 22 (1951), 165-181.

[8] Mann, H. B., and Whitney, D. R., "On a test of whether one of two random variables is stochastically larger than the other," Annals of Mathematical Statistics, 18 (1947), 50-60.

[9] Okamoto, Masashi, "Some inequalities relating to the partial sum of binomial probabilities," Annals of the Institute of Statistical Mathematics, 10 (1958), 29-35.

[10] Prohorov, Yu. V., "An extremal problem in probability theory" (Russian; English summary), Teoriya Veroyatnostei i ee Primeneniya, 4 (1959), 211-214. English translation in Theory of Probability and Its Applications, 4 (1959), 201-203.

[11] Savage, I. R., "Probability inequalities of the Tchebycheff type," Washington: National Bureau of Standards Report No. 1744, 1952.

[12] Tchebichef, P., "Sur les valeurs limites des intégrales," Journal de Mathématiques Pures et Appliquées, Ser. 2, 19 (1874), 157-160.

[13] Wilcoxon, Frank, "Individual comparisons by ranking methods," Biometrics 1 (1943), 80-83.