ON A CHEBYSHEV-TYPE INEQUALITY FOR SUMS OF INDEPENDENT RANDOM VARIABLES[1]

By S. M. Samuels

Purdue University

1. Summary. Let $\mathcal{S}(\nu_1, \cdots, \nu_n)$ be the "class of all random variables", S_n, which are sums of n independent, non-negative random variables, $X_1, \cdots, X_n$, with $EX_i = \nu_i$, $i = 1, \cdots, n$. We consider the problem of finding

$$\inf_{S_n \varepsilon \mathcal{S}(\nu_1, \cdots, \nu_n)} P\{S_n < \lambda\} \tag{1.1}$$

where λ is a positive constant.

For $n = 1$, the infimum is $1 - \nu_1/\lambda$ from the well-known Markov inequality. The solution for $n = 2$ was given in [2]. We derive the solution for $n = 3$. From these results we conjecture what the solution is for arbitrary n. To lend support to the conjecture, we examine a sub-class of $\mathcal{S}(\nu_1, \cdots, \nu_n)$, namely those S_n's for which the problem reduces to one of considering the number of successes in independent trials. We show that, within this subclass, the conjectured value does minimize $P\{S_n < \lambda\}$.

2. Preliminaries. We shall henceforth assume $\lambda > \sum_{i=1}^n \nu_i$ since otherwise (1.1) is zero, which is attained by choosing $X_i \equiv \nu_i$ for all i.

We can restrict our attention to those S_n's in which each of the X_i's concentrates all mass at no more than two points.

This is because the class, $\mathfrak{F}_\nu$, of left-continuous probability distribution functions, with support in $[0, \infty)$ and mean ν, is convex, its extreme points comprising the subset, $\mathcal{E}_\nu$, of step functions with at most two jumps, and, for any bounded Borel function, g,

$$\inf_{F \varepsilon \mathfrak{F}_\nu} \int_0^\infty g \, dF = \inf_{F \varepsilon \mathcal{E}_\nu} \int_0^\infty g \, dF.$$

For a discussion of this aspect of the problem, see [6].

In our problem we take $g(x) = F^{(n-1)}(\lambda - x)$ where $F^{(n-1)}$ is the convolution of the distribution functions of $X_1, \cdots, X_{n-1}$. Then,

$$P\{S_n < \lambda\} = \int_0^\infty F^{(n-1)}(\lambda - x) \, dF_n(x), \tag{2.1}$$

where F_n is the distribution function of X_n, is minimized for some $F_n \varepsilon \mathcal{E}_{\nu_n}$. Obviously the same is true for the distribution function of any of the other X_i's.

From (2.1) it is evident that we can further restrict X_n to be bounded above by λ. Again, by the symmetry of the problem, the same is true for the other X_i's.

Received 9 October 1964; revised 9 April 1965.

[1] This paper is based on the author's Ph.D. thesis at Stanford University. Research supported by the Office of Naval Research under Contract No. Nonr(28) at Stanford University and by the Aerospace Research Laboratories under Contract No. AF 33(657-11737) at Purdue University.

We remark that the subset of $\mathfrak{F}_\nu$ with support in $[0, \lambda]$ is compact; hence (1.1) is actually attained.

Letting $S_0(\nu_1, \cdots, \nu_n)$ be the sub-class of $S(\nu_1, \cdots, \nu_n)$ to which we restrict our attention, we shall denote by a_i and b_i the lower and upper mass points, respectively, of X_i. Then

$$0 \leqq a_i \leqq \nu_i \leqq b_i \leqq \lambda,$$

$$P\{X_i = b_i\} = (\nu_i - a_i)/(b_i - a_i) = 1 - P\{X_i = a_i\}.$$

(If $P\{X_i = \nu_i\} = 1$, we shall, by convention, set $a_i = 0$, $b_i = \nu_i$.)

It is apparent that the minimizing S_n must have positive mass at λ. The following lemma makes a stronger assertion, namely that S_n must satisfy,

$$P\{S_n = \lambda \mid X_i = b_i\} > 0 \qquad \text{for all } i. \tag{2.2}$$

LEMMA 2.1. *If, for some* i,

$$P\{S_n = \lambda \mid X_i = b_i\} = 0, \tag{2.3}$$

then there is an S_n', *also in* $\mathcal{S}_0(\nu_1, \cdots, \nu_n)$, *such that*

$$P\{S_n' < \lambda\} < P\{S_n < \lambda\}.$$

PROOF. We can write

$$\begin{aligned} P\{S_n < \lambda\} = {} & [(b_i - \nu_i)/(b_i - a_i)]P\{S_n < \lambda \mid X_i = a_i\} \\ & + [(\nu_1 - a_i)/(b_i - a_i)]P\{S_n < \lambda \mid X_i = b_i\}. \end{aligned} \tag{2.4}$$

If (2.3) holds, there is a $\delta > 0$ such that, if we replace X_i by X_i' with $a_i' = a_i$, $b_i' = b_i - \delta$, then

$$P\{S_n' < \lambda \mid X_i' = b_i - \delta\} = P\{S_n < \lambda \mid X_i = b_i\}.$$

Hence, from (2.4), b_i can be replaced by $b_i - \delta$ without increasing $P\{S_n < \lambda\}$. In fact, the probability will be strictly decreased if $P\{S_n < \lambda \mid X_i = b_i\} < P\{S_n < \lambda \mid X_i = a_i\}$. An extended argument establishes the result even when these two conditional probabilities are equal.

The next three lemmas will enable us, in deriving (1.1) for $n \leqq 3$, to further restrict our attention to those S_n's for which $a_i = 0$ for all i.

LEMMA 2.2. *Suppose* (1.1) *is attained for* $S_n = \sum X_i$ *with lower mass points* a_i. *Let* $Y_i = X_i - a_i$ *and* $T_n = \sum Y_i$. *Then* T_n *attains* (1.1) *with* λ *replaced by* $\lambda - \sum a_i$ *and* $(\nu_1, \cdots, \nu_n)$ *replaced by* $(\nu_1 - a_1, \cdots, \nu_n - a_n)$.

PROOF. Obviously $T_n \,\varepsilon\, \mathcal{S}_0(\nu_1 - a_1, \cdots, \nu_n - a_n)$ and

$$P\{S_n < \lambda\} = P\{T_n < \lambda - \textstyle\sum a_i\}. \tag{2.5}$$

On the other hand, given *any* $T_n \,\varepsilon\, \mathcal{S}_0(\nu_1 - a_1, \cdots, \nu_n - a_n)$, we can evidently construct from it an $S_n \,\varepsilon\, \mathcal{S}_0(\nu_1, \cdots, \nu_n)$ such that (2.5) holds.

The preceding lemma shows that there are indeed choices of λ and the means for which (1.1) is attained by an S_n with $a_i = 0$ for all i. One can also easily show that, for all choices, the minimizing S_n must have at least one of the a_i's equal to zero.

LEMMA 2.3. *Among all* $S_n \varepsilon \mathcal{S}_0(\nu_1, \cdots, \nu_n)$ *which have only one mass point in the interval* $[0, \lambda)$, $P\{S_n < \lambda\}$ *is minimized only if* S_n *is of the form*:

$$P\{X_{i_j} = \nu_{i_j}\} = 1 \qquad \text{if } j \leqq k,$$

$$P\{X_{i_j} = \lambda - \textstyle\sum_{r=1}^{k} \nu_{i_r}\} = 1 - P\{X_{i_j} = 0\} \qquad \text{if } j > k,$$

for some $k = 0, 1, \cdots$, *or* $n - 1$.

PROOF. By Lemma 2.1, S_n must satisfy (2.2). (Note that if S_n has one mass point in $(0, \lambda)$, then so has the S_n' constructed in the Lemma.) Letting $A = \sum a_i$, this condition plus the hypothesis implies

$$b_i - a_i = \lambda - A \qquad \text{for all } i. \tag{2.6}$$

(If, for some i, $P\{X_i = \nu_i\} = 1$, we shall, in this proof, adopt the convention of setting $a_i = \nu_i$ and defining b_i by (2.6). Elsewhere, we set $a_i = 0$, $b_i = \nu_i$). Hence

$$P\{S_n < \lambda\} = \textstyle\prod_{i=1}^{n} (\lambda - A + a_i - \nu_i)/(\lambda - A). \tag{2.7}$$

Conversely, whenever

$$0 \leqq a_i \leqq \nu_i, \quad \text{for all } i; \qquad \textstyle\sum a_i = A \tag{2.8}$$

holds, there is an S_n satisfying the hypothesis for which (2.7) holds. If we fix A and minimize (2.7) over the set of $a_1, \cdots, a_n$ which satisfy (2.8), the minimum is attained only if at most one of the a_i's (say a_1) is different from 0 or ν_i. Finally, as a_1 ranges over the interval $[0, \nu_1]$, (2.7) is minimized only at an end-point. Thus S_n must be of the stated form.

LEMMA 2.4. *If, among those* $S_n \varepsilon \mathcal{S}_0(\nu_1, \cdots, \nu_n)$ *which have* $a_i = 0$, *for all* i, $P\{S_n < \lambda\}$ *is minimized by an* S_n^* *which has only one mass point in* $[0, \lambda)$, *and if this result holds for all choices of* λ *and the means, then* $P\{S_n^* < \lambda\}$ *is minimum among* all $S_n \varepsilon \mathcal{S}(\nu_1, \cdots, \nu_n)$.

PROOF. Let S_n^0 be any random variable which minimizes $P\{S_n < \lambda\}$ among all $S_n \varepsilon \mathcal{S}_0(\nu_1, \cdots, \nu_n)$. By applying, successively, Lemma 2.2, the hypothesis, the fact that, if $T_n = \sum Y_i \varepsilon \mathcal{S}(\nu_1 - a_1, \cdots, \nu_n - a_n)$ has only one mass piont in the interval $[0, \lambda - \sum a_i)$, then $S_n = \sum (Y_i + a_i) \varepsilon \mathcal{S}(\nu_1, \cdots, \nu_n)$ has only one mass point in $[0, \lambda)$, and, finally, Lemma 2.3, we can construct an $S_n^* \varepsilon \mathcal{S}(\nu_1, \cdots, \nu_n)$ with $a_i = 0$, for all i, and $P\{S_n^* < \lambda\} = P\{S_n^0 < \lambda\}$. In fact, from Lemma 2.3, we can show that S_n^0 itself must be of this form.

3. The solution for $n \leqq 3$. We have already mentioned the Markov inequality which states that, for $n = 1$, (1.1) is equal to $1 - \nu_1/\lambda$. We present here a modified derivation of the solution for $n = 2$, originally given in [2]:

THEOREM 3.1. *For* $n = 2$, (1.1) *equals*

$$\min\,[(1 - \nu_1/\lambda)(1 - \nu_2/\lambda),\ 1 - \nu_2/(\lambda - \nu_1)], \tag{3.1}$$

if $\nu_1 \leqq \nu_2$. *The first value is attained for* $a_1 = a_2 = 0$, $b_1 = b_2 = \lambda$, *while the second is attained for* $b_1 = \nu_1$, $a_2 = 0$, $b_2 = \lambda - \nu_1$.

Proof. We consider only those S_2's with $a_1 = a_2 = 0$. Then the four (not necessarily distinct) mass points of S_2 are $0, b_1, b_2, b_1 + b_2$. We consider separately the cases defined by specifying which of these four mass points are $\geqq \lambda$. The four possible cases are:

(1) $b_1 + b_2 \geqq \lambda > \max(b_1, b_2)$;
(2) $b_1 \geqq \lambda > b_2$;
(3) $b_2 \geqq \lambda > b_1$;
(4) $\min(b_1, b_2) \geqq \lambda$.

Since each X_i must have support in $[0, \lambda]$, the only S_n in case (4) is the one with $b_1 = b_2 = \lambda$, which attains the first value in (3.1). From Lemma 2.1 we can ignore cases (2) and (3) since no S_n in these cases can satisfy (2.2). An S_2 in case (1) satisfies (2.2) only if $b_1 + b_2 = \lambda$. Among these S_2's, it is easy to verify that $P\{S_2 < \lambda\}$ is minimized only if $b_1 = \nu_1$, $b_2 = \lambda - \nu_1$, if $\nu_1 \leqq \nu_2$, which gives the other value in (3.1).

To complete the proof, we apply Lemma 2.4.

A simple calculation shows that $1 - \nu_2/(\lambda - \nu_1) \leqq (1 - \nu_1/\lambda)(1 - \nu_2/\lambda)$ if and only if

$$\lambda \leqq \tfrac{1}{2}[\nu_1 + 2\nu_2 + (\nu_1^2 + 4\nu_2)^{\frac{1}{2}}].$$

Hence, if $\nu_1 = \nu_2$ and λ is large, (1.1) is attained when X_1 and X_2 are identically distributed. In Section six, we shall discuss the "identically distributed problem."

Theorem 3.2. *For* $n = 3$, (1.1) *equals*

$$\text{(3.2)} \quad \min [(1 - \nu_1/\lambda)(1 - \nu_2/\lambda)(1 - \nu_3/\lambda),$$
$$(1 - \nu_2/(\lambda - \nu_1))(1 - \nu_3/(\lambda - \nu_1)), 1 - \nu_3/(\lambda - \nu_1 - \nu_2)],$$

if $\nu_1 \leqq \nu_2 \leqq \nu_3$. *The three values are attained for*

$$b_i = \nu_i \qquad \text{if } i \leqq k,$$
$$a_i = 0, \qquad b_i = \lambda - \textstyle\sum_{j=1}^k \nu_j \qquad \text{if } i > k,\ k = 0, 1, \text{ or } 2.$$

Proof. We again consider only those S_3's with $a_1 = a_2 = a_3 = 0$. Then the eight mass points of S_3 are: $0, b_1, b_2, b_3, b_1 + b_2, b_1 + b_3, b_2 + b_3, b_1 + b_2 + b_3$. As before, we define cases by specifying which of these points are $\geqq \lambda$. However, to avoid redundancy, we condense them as follows:

(1) $b_1 + b_2 + b_3 \geqq \lambda > \max(b_i + b_j)$,
(2) $b_i + b_j \geqq \lambda > b_i + b_k, b_j + b_k$,
(3) $b_i + b_j, b_i + b_k \geqq \lambda > b_j + b_k, b_i$,
(4) $\min(b_i + b_j) \geqq \lambda > \max(b_i)$,
(5) $b_i \geqq \lambda > b_j + b_k$,
(6) $b_i, b_j + b_k \geqq \lambda > b_j, b_k$,
(7) $b_i, b_j \geqq \lambda > b_k$,
(8) $\min(b_i) \geqq \lambda$,

where $i \neq j \neq k \neq i$ and $i, j, k = 1, 2, 3$.

The only S_3 in case (8) for which each X_i has support in $[0, \lambda]$ is the one with $b_1 = b_2 = b_3 = \lambda$ which attains the first value in (3.2). By Lemma 2.1 we can ignore cases (2), (5), and (7) since no S_3 in these cases can satisfy (2.2). In the remaining cases, (2.2) implies:

(1) $b_1 + b_2 + b_3 = \lambda$,
(3) $b_j = b_k = \lambda - b_i < \lambda/2$,
(4) $b_j = b_k = \lambda - b_i \geqq \lambda/2$,
(6) $b_j + b_k = \lambda = b_i$.

Cases (1) and (4) are dealt with by Theorem 5.1 below. Cases (1), (3), and (6) are covered by Lemmas 3.1 and 3.2 below. These two lemmas establish that, for any S_3 belonging to one of the three cases, one of the following holds:

(a) There is an S_3' with $P\{S_3' < \lambda\} < P\{S_3 < \lambda\}$.
(b) There is an S_3' with $X_i' \equiv \nu_i$ for some i and $P\{S_3' < \lambda\} \leqq P\{S_3 < \lambda\}$.

Condition (b) is sufficient to establish the theorem since, by Theorem 3.1 and (4.3) below, (3.2) is $\leqq$

$$\min_{i,j,k=1,2,3,\, i\neq j\neq k\neq i} [\min_{S_2 \varepsilon \mathcal{S}(\nu_j,\nu_k)} P\{S_2 < \lambda - \nu_i\}].$$

To complete the proof, we apply Lemma 2.4.

For a comparison of the three values in (3.2), see Section 4.

The next two lemmas are needed in the proof of Theorem 3.2. We state them for arbitrary n and for S_n's with $a_i = 0$ for all i. In Section 6 we shall mention their applicability to the problem, $n = 4$.

Lemma 3.1. *If* $S_n \geqq \lambda$ *implies* $X_n = b_n > \nu_n$, *with* $n \geqq 3$, *then*

$$P\{S_n < \lambda\} \geqq 1 - (\nu_n/b_n)[1 - \min_{S_{n-1} \varepsilon \mathcal{S}(\nu_1,\cdots,\nu_{n-1})} P\{S_{n-1} < \lambda - b_n\}].$$

If, in addition, $b_i = \lambda - b_n$, $i = 1, \cdots, n-1$, *(hence, necessarily,* $b_n > \lambda/2$, *since* $n \geqq 3$*), then*

$$P\{S_n < \lambda\} > 1 - \nu_n/(\lambda - \textstyle\sum_{i=1}^{n-1} \nu_i).$$

Proof. The first part is immediate since, by hypothesis,

$$P\{S_n < \lambda\} = 1 - (\nu_n/b_n)[1 - P\{S_n < \lambda \mid X_n = b_n\}].$$

The second part follows by direct computation:

$$\begin{aligned} P\{S_n < \lambda\} &- [1 - \nu_n/(\lambda - \textstyle\sum_{i=1}^{n-1} \nu_i)] \\ &= [\nu_n/b_n(\lambda - \textstyle\sum_{i=1}^{n-1} \nu_i)] \\ &\quad \cdot\{-(\lambda - b_n - \textstyle\sum_{i=1}^{n-1} \nu_i) + (\lambda - \sum_{i=1}^{n-1} \nu_i)\prod_{i=1}^{n-1} [1 - \nu_i/(\lambda - b_n)]\} \\ &> [\nu_n/b_n(\lambda - b_n)(\lambda - \textstyle\sum_{i=1}^{n-1} \nu_i)](b_n - \sum_{i=1}^{n-1} \nu_i)(\lambda - b_n - \sum_{i=1}^{n-1}\nu_i), \end{aligned}$$

which is clearly positive whether or not $\lambda - b_n > \sum_{i=1}^{n-1} \nu_i$.

The proof of the next lemma is immediate.

Lemma 3.2. *If* $b_n = \lambda$, *then*

$$P\{S_n < \lambda\} \geqq (1 - \nu_n/\lambda) \min_{S_{n-1} \varepsilon \mathcal{S}(\nu_1,\cdots,\nu_{n-1})} P\{S_{n-1} < \lambda\}.$$

4. A conjecture. From Theorems 3.1 and 3.2, and the Markov inequality, we have the following result for $n \leqq 3$:

$$\min_{S_n \varepsilon \mathcal{S}(\nu_1, \cdots, \nu_n)} P\{S_n < \lambda\} = \min_{k=0,1,\cdots,n-1} P_{k,n}(\lambda), \tag{4.1}$$

where,

$$P_{0,n}(\lambda) = \prod_{i=1}^{n} (1 - \nu_i/\lambda), \tag{4.2}$$
$$P_{k,n}(\lambda) = \prod_{i=k+1}^{n} [1 - \nu_i/(\lambda - \textstyle\sum_{j=1}^{k} \nu_j)], \quad \text{for } k = 1, \cdots, n-1,$$

if $\nu_1 \leqq \cdots \leqq \nu_n$. We conjecture that (4.1) holds for all n. Theorem 5.1 below lends additional support to the conjecture.

The value $P_{0,n}(\lambda)$ is attained when $a_i = 0$, $b_i = \lambda$ for all i. For $k \geqq 1$, $P_{k,n}(\lambda)$ is attained when $b_i = \nu_i$, if $i \leqq k$; $a_i = 0$, $b_i = \lambda - \sum_{j=1}^{k} \nu_j$, if $i > k$. The justification for setting the k X_i's with the *smallest* means identically equal to them follows from the fact that, if $\nu_k \leqq \nu_{k+1}$, then

$$[1 - \nu_k/(\lambda - \nu_{k+1} - \textstyle\sum_{j=1}^{k-1} \nu_j)] \prod_{i=k+2}^{n} [1 - \nu_i/(\lambda - \nu_{k+1} - \sum_{j=1}^{k-1} \nu_j)] \tag{4.3}$$
$$\geqq \min(P_{k,n}, P_{k+1,n}).$$

It is easy to see that

$$\min_{k=0,1,\cdots,n-1} P_{k,n}(\lambda) = P_{n-1,n}(\lambda) \qquad \text{if } \lambda \text{ is close to } \textstyle\sum_{i=1}^{n} \nu_i$$
$$= P_{0,n}(\lambda) \qquad \text{for sufficiently large } \lambda$$

(we can show that $\lambda > 2n\nu_n$ is sufficiently large). Moreover, $P_{k,n}(\lambda) - P_{k+m,n}(\lambda)$ has a single root, say $\alpha_{k,k+m,n}$, in the interval $\sum_{i=1}^{k+m} \nu_j + \nu_n < \lambda < \infty$ and is negative if and only if $\lambda > \alpha_{k,k+m,n}$. Hence the set on which $\min_{k=0,1,\cdots,n-1} P_{k,n}(\lambda) = P_{k_0,n}(\lambda)$ is either an interval or the empty set. That both possibilities may occur is easily verified for $n = 3$. If $\nu_1 = \nu_2 = \nu_3$, $P_{1,3}(\lambda)$ is never minimum, while, if $\nu_2 = \nu_3$ and ν_1 is sufficiently small, there is an interval on which $P_{1,3}(\lambda)$ is minimum. Thus, for fixed n, the roots $\alpha_{k,k+1,n}$ do not, in general, form a monotone sequence.

5. Support for the conjecture. In this section we exhibit a reasonably large subset of $\mathcal{S}(\nu_1, \cdots, \nu_n)$ in which $P\{S_n < \lambda\}$ is minimized by the conjectured value, (4.1).

We define $\mathcal{B}_k(\nu_1, \cdots, \nu_n; \lambda)$, for $k = 0, 1, \cdots, n-1$, to be the set of those S_n's, with $a_i = 0$ for all i, for which the sum of any k of the b_i's is less than λ while the sum of any $k+1$ of the b_i's is greater than or equal to λ. (In Theorem 3.2, $\mathcal{B}_0(\nu_1, \nu_2, \nu_3; \lambda)$ is Case 8, $\mathcal{B}_1$ is Case 4, and $\mathcal{B}_2$ is Case 1.) Let

$$\mathcal{B}(\nu_1, \cdots, \nu_n; \lambda) = \bigcup_{k=0}^{n-1} \mathcal{B}_k(\nu_1, \cdots, \nu_n; \lambda).$$

THEOREM 5.1. *If* $\nu_1 \leqq \cdots \leqq \nu_n$, *then*

$$\min_{S_n \varepsilon \mathcal{B}(\nu_1, \cdots, \nu_n; \lambda)} P\{S_n < \lambda\} = \min_{k=0,1,\cdots,n-1} P_{k,n}(\lambda) \tag{5.1}$$

where $P_{k,n}(\lambda)$ *is defined by* (4.2).

PROOF. The only $S_n \varepsilon \mathcal{B}_0$ for which each X_i has support in $[0, \lambda]$ is the one with $b_i = \lambda$ for all i which yields the value $P_{0,n}(\lambda)$.

We now proceed by induction on n. Suppose $S_n \varepsilon \mathcal{B}_k$, $k \geqq 1$, has $b_i = \nu_i$ for some i. Then $S_n - X_i \varepsilon \mathcal{B}_{k-1}(\nu_1, \cdots, \nu_{i-1}, \nu_{i+1}, \cdots, \nu_n; \lambda - \nu_i)$. By the induction hypothesis,

$$P\{S_n < \lambda\} = P\{S_n - X_i < \lambda - \nu_i\} \geqq \min_{k=0,1,\cdots,n-2} P_{k,n-1}(\lambda - \nu_i).$$

But, by (4.3) and the definition, (4.2),

$$\min_{k=0,1,\cdots,n-2} P_{k,n-1}(\lambda - \nu_i) \geqq \min_{k=0,1,\cdots,n-1} P_{k,n}(\lambda).$$

Hence, to complete the proof, it will be sufficient to show that, if $S_n \varepsilon \mathcal{B} - \mathcal{B}_0$, then there is an $S_n' \varepsilon \mathcal{B} - \mathcal{B}_0$, with $X_i' \equiv \nu_i$ for some i and $P\{S_n' < \lambda\} \leqq P\{S_n < \lambda\}$.

For simplicity we shall assume that the X_i's are always ordered so that $b_1 \leqq \cdots \leqq b_n$. Hence, of course, $\nu_i = EX_i$ may not be less than or equal to $\nu_{i+1} = EX_{i+1}$.

We define the following subsets of $\mathcal{B}_k$, $k \geqq 1$:

(a) $\sum_{i=n-k+1}^{n} b_i < \lambda \leqq \sum_{i=1}^{k+1} b_i$,

(b) $\sum_{i=1}^{k} b_i + b_j = \lambda, j = k + 1, k + 2, \cdots, n$,

(c) $b_2 = \cdots = b_n = b = (\lambda - b_1)/k, \lambda/(k + 1) \leqq b \leqq (\lambda - \nu_1)/k$,

(d) $b = \lambda/(k + 1)$ or $b = (1/k)(\lambda - \nu_1)$,

(e) $b = (1/k)(\lambda - \nu_1)$.

In view of our assumption about the b_i's subset (a) is just $\mathcal{B}_k$ itself. We shall show that, if S_n belongs to one of these subsets, then there is an S_n', either in the succeeding subset or with $b_i = \nu_i$ for some i, satisfying $P\{S_n' < \lambda\} \leqq P\{S_n < \lambda\}$. Since subset (e) has $b_1 = \nu_1$, this will complete the proof.

That (b) dominates (a), in the above sense, follows immediately from Lemma 2.1. To simplify the remainder of the proof, we introduce the following notation:

$p_i = \nu_i/b_i$,

$f(k)$ = probability of k successes in n independent trials with probabilities p_i of success on the ith trial,

$F(k) = \sum_{j=0}^{k} f(j)$,

$f_{i_1,\cdots,i_m}(k)$ = probability of k successes in the $n - m$ trials obtained from the original n by excluding trials $i_1, \cdots, i_m$,

$F_{i_1,\cdots,i_m}(k) = \sum_{j=0}^{k} f_{i_1,\cdots,i_m}(j)$.

Then, if $S_n \varepsilon \mathcal{B}_k$, $P\{S_n < \lambda\} = F(k)$.

If $S_n \varepsilon$ subset (b), choose any two of the first k indices, say i and j, and fix $b_i + b_j = 2B$. Then $S_n \varepsilon \mathcal{B}_k$ if and only if

$$\max(\nu_i, 2B - b_{k+1}) \leqq b_i \leqq \min(2B - \nu_j, b_{k+1}).$$

In this interval we differentiate $F(k)$ twice with respect to b_i :

$$\begin{aligned} F(k) &= p_i p_j F_{ij}(k-2) + [p_i(1-p_j) + p_j(1-p_i)]F_{ij}(k-1) \\ &\quad + (1-p_i)(1-p_j)F_{ij}(k) \\ &= -p_i p_j f_{ij}(k-1) - (p_i + p_j - p_i p_j) f_{ij}(k) + F_{ij}(k), \\ F''(k) &= -(p_i'' p_j + 2p_i' p_j' + p_i p_j'') f_{ij}(k-1) \\ &\quad - [p_i''(1-p_j) - 2p_i' p_j' + p_j''(1-p_i)] f_{ij}(k). \end{aligned}$$

Since $p_i = \nu_i/b_i$ and $p_j = \nu_j/(2B - b_i)$, $F''(k)$ is negative, so $P\{S_n < \lambda\}$ is minimized either by setting b_i or b_j equal to the corresponding mean or by setting b_i or b_j equal to b_{k+1}. We can continue the procedure until either $b_i = \nu_i$ for some i or all but one of the b_i's are equal. Thus subset (c) dominates subset (b).

If $S_n \varepsilon$ subset (c) we differentiate $F(k)$ with respect to b. From the chain rule and the fact that

$$F(k) = p_i F_i(k-1) + (1-p_i)F_i(k) = F_i(k) - p_i f_i(k),$$

we have

$$\begin{aligned} F'(k) &= -\textstyle\sum_{i=1}^n p_i' f_i(k) \\ &= -p_1' f_1(k) + (1/b)[p_1 \textstyle\sum_{i=2}^n p_i f_{i1}(k-1) + (1-p_1)\textstyle\sum_{i=2}^n p_i f_{i1}(k)]. \end{aligned}$$

A well-known lemma, derived in [8], states that

$$kf(k) = \textstyle\sum_{i=1}^n p_i f_i(k-1). \tag{5.2}$$

Hence

$$\begin{aligned} F'(k) &= -p_1' f_1(k) + (1/b)[kp_1 f_1(k) + (k+1)(1-p_1) f_1(k+1)] \\ &= \{k\nu_1[(k+1)b - \lambda] f_1(k+1)/b(\lambda - kb)^2\} \\ &\quad \cdot(\{(k+1)(\lambda - kb)(\lambda - \nu_1 \text{[illegible]} \nu_1[(k+1)b - \lambda]\} - f_1(k)/f_1(k+1)) \\ &= A(b)\{B(b) - C(b)\}, \end{aligned}$$

where $A(b)$ is positive in the open interval

$$\lambda/(k+1) < b < (\lambda - \nu_1)/k, \tag{5.3}$$

and $B(b)$ is decreasing. Since $p_i(b)$ is decreasing, for $i = 2, \cdots, n$, we can apply a result derived in [8] (a special case of a result in [3]), which states that $C(b)$ is increasing. Hence $F'(k)$ has exactly one sign change in the interval (5.3), from positive to negative, so $P\{S_n < \lambda\}$ is minimized at an end-point. (If $\nu_i \geqq \lambda/(k+1)$ for some i, then $b = \nu_i$ is one of the end-points.) Thus subset (d) dominates subset (c).

If $S_n \varepsilon$ subset (d), with $b = \lambda/(k+1)$, then b_1 also equals $\lambda/(k+1)$ so we can assume, without loss of generality, $\nu_1 = \max(\nu_1, \cdots, \nu_n)$. We shall define functions $p_i(x)$, $0 \leqq x \leqq 1$, $i = 2, \cdots, n$, satisfying

$$0 \leqq p_i(x) \leqq 1,$$
$$p_i(0) = (k+1)\nu_i/\lambda,$$
$$p_i(1) = k\nu_i/(\lambda - \nu_1),$$

and will derive a function $p_1(x)$, satisfying

$$0 \leqq p_1(x) \leqq 1,$$
$$(5.4) \qquad p_1(0) = (k+1)\nu_1/\lambda,$$
$$p_1(1) = 1,$$

for which $(d/dx)F(k) \leqq 0$ for $0 < x < 1$.

Letting $p_1(0) = q$, we define

$$(5.5) \quad p_i(x) = [(k+1)\nu_i/\lambda][k/(k+1-q)]^x \qquad \text{for } 0 \leqq x \leqq 1, i = 2, \cdots, n,$$
$$c = \ln[1 + (1-q)/k].$$

For simplicity, we suppress the x:

$$F'(k) = -\sum_{i=1}^n p_i' f_i(k)$$
$$= -p_1' f_1(k) + c\sum_{i=2}^n p_i[p_1 f_{i1}(k-1) + (1-p_1)f_{i1}(k)].$$
$$= -p_1' f_1(k) + c[kp_1 f_1(k) + (k+1)(1-p_1)f_1(k+1)]$$

by (5.2). Now $f_1(k+1)/f_1(k)$ is decreasing in x, since $p_i(x) \leqq p_i(0)$ for $i = 2, \cdots, n$, $0 \leqq x \leqq 1$, and the condition $\sum_{i=1}^n \nu_i < \lambda$ is equivalent to $q + \sum_{i=2}^n p_i(0) < k+1$. We show in [8] that this implies $f_1(k) > f_1(k+1)$ for $x = 0$ and, hence, for all x, provided $q = \max(q, p_2(0), \cdots, p_n(0))$, or equivalently, $\nu_1 = \max(\nu_1, \cdots, \nu_n)$. Hence

$$F'(k) < f_1(k)\{-p_1' + c(k+1-p_1)\}$$

provided $p_1(x) \leqq 1$. If we set the expression in the bracket equal to zero and solve for $p_1(x)$, we have, with the boundary condition $p_1(0) = q$,

$$p_1(x) = (k+1) - (k+1-q)[k/(k+1-q)]^x$$

which satisfies (5.4). Thus subset (e) dominates subset (d) and the theorem is proved.

A curious feature of the preceding proof is that, in order to prove that subset (e) dominates subset (d), we were obliged to set the X_i with the *largest* mean equal to it—despite the fact that the values $P_{k,n}(\lambda)$ are obtained by setting the X_i's , with the *smallest* means equal to them.

We also remark that we have *not* proved: $\min_{S_n \in \mathcal{B}_k(\nu_1, \cdots, \nu_n)} P\{S_n < \lambda\} = P_{k,n}(\lambda)$. While equality does hold if all the means are equal, or if $\sum \nu_i \leqq k\lambda/(k+1)$, it fails to hold in certain cases where λ is close to the sum of the means. From the case of equal means we can derive the inequality

$\sum_{i=0}^{k} \binom{n}{i} p^i(1-p)^{n-i} \geqq [1 - p/(k+1-kp)]^{n-k}$ if $0 \leqq p \leqq (k+1)/n$.

For a derivation, see [8].

6. Related problems. (a) *The case* $n = 4$. If we proceed analogously to the proof of Theorem 3.2, there are 25 cases to consider. However, eight of them cannot satisfy (2.2), five each are covered by Lemmas 3.1 and 3.2, and two more by Theorem 5.1. Three other cases can be handled by another simple lemma, leaving only two cases as yet unresolved.

(b) *Lower bound for* $n > 3$ *from Theorem* 3.2. From the Markov inequality, we have, for any n,

$$P\{S_n < \lambda\} \geqq 1 - \sum \nu_i/\lambda. \tag{6.1}$$

Similarly, it follows from (3.2) that

$$P\{S_n < \lambda\} \geqq \min [(1 - \beta_1/\lambda)(1 - \beta_2/\lambda)(1 - \beta_3/\lambda), (1 - \beta_2/(\lambda - \beta_1))(1 - \beta_3/(\lambda - \beta_1)), 1 - \beta_3/(\lambda - \beta_1 - \beta_2)] \tag{6.2}$$

where, for any integers k_1, k_2, with $k_1 \leqq k_2 \leqq n$:

$$\beta_1 = \sum_{j=1}^{k_1} \nu_{i_j}, \qquad \beta_2 = \sum_{j=k_1+1}^{k_2} \nu_{i_j}, \qquad \beta_3 = \sum_{j=k_2+1}^{n} \nu_{i_j}.$$

No matter how the β's are chosen, the inequality in (6.2) is, in general, strict. It is easy to show that the choice of the β's which maximizes the right side of (6.2) is the one which makes them "as nearly equal as possible." More precisely, if the means are all equal (to ν), the optimal choice of the β's is

$$\begin{aligned} &\beta_1 = \beta_2 = \beta_3 = m\nu && \text{if } n = 3m, \\ &\beta_1 = (m-1)\nu, \quad \beta_2 = \beta_3 = m\nu && \text{if } n = 3m - 1, \\ &\beta_1 = \beta_2 = (m-1)\nu, \quad \beta_3 = m\nu && \text{if } n = 3m - 2. \end{aligned}$$

It can be shown that, if $n = 3m$ or $3m - 2$, or if $n = 3m - 1 > 20$, then the value $[1 - \beta_2/(\lambda - \beta_1)][1 - \beta_3/(\lambda - \beta_1)]$ is never (i.e. not for any λ) minimal.

(c) *Sums of independent, identically distributed random variables.* The argument we used to restrict our attention to X_i's with no more than two mass points is no longer valid here. It is shown in [5] that, for $n = 2$, the X_i's can be assumed to have no more than four mass points. On the other hand, if the means are all equal to ν, then $\min_{k=0,1,\cdots,n-1} P_{k,n}(\lambda) = P_{0,n}(\lambda)$ if $\lambda > 2n\nu$. Hence, if the conjecture is correct, the "identically distributed problem" is also solved except for small λ.

(d) *X_i's with common mean.* If the conjecture is correct, then, by Lemma 2.4, this is not really a "special case" of the problem with arbitrary means.

(e) *Constraints on the X_i's.* If the X_i's are constrained to have support in specified finite intervals, lower bounds for $P\{S_n < \lambda\}$ are obtained in [4] by minimizing $E \exp h(S_n - \lambda)$ where h is an arbitrary positive constant. We have remarked that (1.1) is always attained by S_n's for which each X_i has support in $[0, \lambda]$. With this constraint, the lower bound in [4] is even smaller than (6.1).

(f) *Equivalent form of the problem.* If we let $\mathfrak{I}(\mu_1, \cdots, \mu_n, \nu_1, \cdots, \nu_n)$ be the class of random variables, T_n, which are sums of n independent random variables $Y_1, \cdots, Y_n$, with $EY_i = \mu_i$, var $Y_i = \nu_i$, then it follows immediately that

$$\min_{T_n \varepsilon \mathfrak{I}(\mu_1, \cdots, \mu_n, \nu_1, \cdots, \nu_n)} P\{\textstyle\sum_{i=1}^n (Y_i - \mu_i)^2 < \lambda\}$$

is the same as (1.1).

(g) *The second moment problem.* On the other hand, the problem of finding

$$(6.3) \qquad \varphi_n(\lambda) = \inf_{T_n \varepsilon \mathfrak{I}(\mu_1, \cdots, \mu_n, \nu_1, \cdots, \nu_n)} P\{|\textstyle\sum_{i=1}^n (Y_i - \mu_i)|^2 < \lambda\}$$

is distinct from the one we have considered. For the special case where T_n is a sum of identically distributed random variables with common variance ν, it is shown in [7] that

$$(6.4) \qquad \varphi_n(\lambda) > 1 - n\nu/\lambda \quad \text{for} \quad n > 1, \quad \lambda \text{ sufficiently large,}$$

$$(6.5) \qquad \lim_{\lambda \to \infty} \lambda[1 - \varphi_n(\lambda)] = n\nu \quad \text{for} \quad n \geqq 1.$$

(The emphasis in (6.4) is on the strictness of the inequality. Equality, of course, holds for $n = 1$.) If we define $\varphi_n(\lambda)$ by (1.1) rather than by (6.3), then it is easy to show that (6.4) and (6.5) remain valid.

(h) *Redundant structures in reliability theory.* The eight cases considered in Theorem 3.2, and, in general, the set of cases which arise for any n, correspond to "coherent redundant structures" in reliability theory (see [1]). A redundant structure is a system of two-state (function or fail) components which functions as a whole even if certain components fail. The structure is defined by specifying which subsystems have the property that, if all components within the subsystem function, then the whole system functions. Such subsystems are called "paths." A structure is said to be "coherent" if every subsystem containing a path is also a path.

The sets of cases in our problem correspond to the sets of all coherent redundant structures with n components having the further property that the components can be ranked so that if a path contains component i but not component j $(j > i)$, then the subsystem obtained by replacing i with j is also a path.

7. Acknowledgment. I wish to thank Professor S. Karlin, my Ph.D. dissertation advisor, for his help and encouragement.

REFERENCES

[1] Birnbaum, Z. W., Esary, J. D. and Saunders, S. C. (1961). Multi-component systems and structures and their reliability. *Technometrics* **3** 55–77.

[2] Birnbaum, Z. W., Raymond, J. and Zuckerman, H. S. (1947). A generalization of Tshebyshev's inequality to two dimensions. *Ann. Math. Statist.* **18** 70–79.

[3] Ghurye, S. G. and Wallace, D. L. (1959). A convolution class of monotone likelihood ratio families. *Ann. Math. Statist.* **30** 1158–1164.

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

[5] HOEFFDING, W. and SHRIKHANDE, S. S. (1955). Bounds for the distribution function of a sum of independent, identically distributed random variables. *Ann. Math. Statist.* **26** 439– 449.

[6] MULHOLLAND, H. P. and ROGERS, C. A. (1958). Representation theorems for distribution functions. *Proc. London Math. Soc. Ser. 3* **8** 177–223.

[7] ROBBINS, H. (1948). Some remarks on the inequality of Tchebycheff. *Studies and Essays.* Interscience, New York.

[8] SAMUELS, S. M. (1965). On the number of successes in independent trials. *Ann. Math. Statist.* **36** 1272–1278.