The Annals of Probability
1988, Vol. 16, No. 4, 1644–1664

BOUNDS ON THE COARSENESS OF RANDOM SUMS[1]

BY JAMES ALLEN FILL

Stanford University

Let $X_1^0, \dots, X_N^0$ be integer-valued random variables and let $a_1, \dots, a_N$ be (fixed) nonzero vectors. We introduce the notion of coarseness of a discrete distribution and obtain upper bounds on the coarseness of the distribution of $S = \sum X_i^0 a_i$ by comparison with the case $a_i \equiv a$. The bounds derived are seen to be tight and to apply for example when S is formed (a) from independent summands or (b) by using any of a large family of sampling schemes. We show how such bounds can easily and efficiently substitute for use of Berry–Esseen theorems and other analytical methods in applications.

1. Introduction. Throughout this paper H is a Hilbert space over $\mathbb{R}$ with inner product $\langle \cdot, \cdot \rangle$ and norm $\|\cdot\|$, and $N \geq 0$ and $K \geq 1$ are deterministic integers. All random variables are defined on a common probability space. The value y is said to be a possible value of the random variable Y if $P\{Y = y\} > 0$. The symbols $\lfloor x \rfloor$ and $\lceil x \rceil$ denote, respectively, the largest integer no larger than x and the smallest integer no smaller than x.

The major results of this paper are of the following form.

GENERIC RESULT. *Let $X_1^0, X_2^0, \dots, X_N^0$ be random variables with values in the integers $\mathbb{Z}$. Let $a_1, \dots, a_N$ be vectors in H, each with length at least unity. Set $S^0 = \sum_{n=1}^N X_n^0$ and $S = \sum_{n=1}^N X_n^0 a_n$. If $R_1, \dots, R_K$ are sets in H each of diameter strictly less than unity, then, under suitable conditions,*

$$P\left\{S \in \bigcup_{k=1}^K R_k\right\} \leq P\{S^0 \in L_{S^0}(K)\} \equiv \pi_{S^0}(K), \tag{1.1}$$

where $L_{S^0}(K)$ is a set of K most likely values of S^0.

Toward understanding bounds of the form (1.1), which bound probabilities for a random sum S in terms of the distribution of S in the "null" case $a_i \equiv a$, we introduce some terminology. Given a discrete random variable Y with values in an arbitrary measurable space, let $L_Y(K)$ be a set of K most likely values of Y. We shall call any such set a *top K set* of values and its probability $\pi_Y(K)$ the *top K probability*. Due to ties there may be more than one top K set, but the top K probability is uniquely determined. Note that $\pi_Y(K)$ is invariant under one-to-one transformation of the random variable Y.

Received April 1987; revised November 1987.

[1]Written in part while on leave at The University of Chicago. This manuscript was prepared using computer facilities supported in part by National Science Foundation Grants DMS-86-01732 and DMS-84-04941 to the Department of Statistics at The University of Chicago.

AMS 1980 *subject classifications.* Primary 60E15; secondary 60G50.

Key words and phrases. Random sums, coarseness, Berry–Esseen theorems, top K set, top K probability, sampling, stochastic monotonicity, log concavity.

Let Y and Z be discrete random variables taking values in possibly different spaces. If for each integer J no bigger than a fixed $K \le \infty$ the top J probabilities for Y and Z satisfy $\pi_Y(J) \le \pi_Z(J)$, we shall say that (the distribution of) Z is *K-coarser than* (that of) Y (though not necessarily strictly so). When $K = \infty$ we shall simply say that Z *is coarser than* Y and write $Y \le^c Z$. When, as in (1.1), we have the stronger result that for any $K < \infty$ and any sets $R_1, \dots, R_K$ in the Hilbert space of values of Y with $\operatorname{diam}(R_k) < 1$ for every k,

$$P\left\{Y \in \bigcup_{k=1}^{K} R_K\right\} \le \pi_Z(K),$$

we shall say that Z *is strongly coarser than* Y and write $Y \le^{sc} Z$.

We note in passing that the ideas of coarseness can be put into the language of majorization. For example, $Y \le^c Z$ means precisely that the probability mass function of Y is majorized by that of Z in the sense of Marshall and Olkin (1979), page 16.

When $K = 1$ a top K set $L_Y(K)$ is any singleton with element a (primary) mode for Y. Suppose next that Y has a distribution $\mathscr{L}(Y)$ that is unimodal over $\mathbb{Z}$, which for this paper we shall understand to mean that for some (not necessarily unique) $y_0 \in \mathbb{Z}$ the values $P\{Y = y\}$ are nondecreasing for $y \le y_0$ and nonincreasing for $y \ge y_0$. Then each top K set comprises K consecutive integers including a mode. If $\mathscr{L}(Y)$ is unimodal over $\mathbb{Z}$ and symmetric about μ with $2\mu \in \mathbb{Z}$, then $L_Y(K) = \{\lfloor \mu - (K-1)/2 \rfloor, \dots, \lfloor \mu + (K-1)/2 \rfloor\}$.

The first result of the form (1.1) for general Hilbert space H was stated in combinatorial language and proved by Kleitman (1970), who generalized a result of Littlewood and Offord (1943) and established a conjecture of Erdös (1945). Kleitman's specific condition was that $X_1^0, X_2^0, \dots, X_N^0$ be i.i.d. with $P\{X_n^0 = 0\} = 1/2 = P\{X_n^0 = 1\}$. Then, since S^0 has the binomial$(N, 1/2)$ distribution, $L_{S^0}(K)$ is uniquely determined as

$$\{\lfloor (N - (K-1))/2 \rfloor, \dots, \lfloor (N + (K-1))/2 \rfloor\}$$

if $N + K$ is odd. If $N + K$ is even, either the same recipe or

$$\{\lceil (N - (K-1))/2 \rceil, \dots, \lceil (N + (K-1))/2 \rceil\}$$

can serve as $L_{S^0}(K)$.

Kleitman (1970) also established (1.1) under the more general condition that $X_1^0, X_2^0, \dots, X_N^0$ are independent random variables with X_n^0 uniformly distributed on the integers $\{m_n, \dots, M_n\}$.

The results of Section 2, which all follow from the powerful Theorem 2.1, extend Kleitman's result in several directions. Theorem 2.3 relaxes the assumption of uniform distributions by requiring only that each of the independently distributed summands X_n^0 have a (possibly unbounded) symmetric unimodal distribution on $\mathbb{Z}$. In Theorem 2.7 the assumption of independence of $X_1^0, \dots, X_N^0$ is dropped altogether and replaced by a sampling setup explained following the theorem's proof. The class of sampling schemes covered by Theorem 2.7 is very large and includes sampling both with (Corollary 2.13) and without (Corollary 2.9) replacement.

When symmetry conditions are dropped, inequality (1.1) can fail. The problem, as illustrated by Example 3.1, lies with the possibly large probability of cancellation in forming the sum S. When $H = \mathbb{R}$ this difficulty can be overcome by requiring that $a_n \geq 1$ for all n. In Section 3 we obtain theorems analogous to those of Section 2 when $H = \mathbb{R}$ and, roughly speaking, symmetric unimodality of $\mathscr{L}(X_n^0)$ is traded for stochastic monotonicity of both $\mathscr{L}(X_n^0 | \Sigma_{i=n}^N X_i^0 = \ell)$ and $\mathscr{L}(\Sigma_{i=n+1}^N X_i^0 | \Sigma_{i=n}^N X_i^0 = \ell)$ in ℓ. In most applications of Theorem 3.8, the asymmetric analogue of the symmetric sampling theorem (Theorem 2.7), the appropriate stochastic monotonicity condition is easy to verify. The monotonicity condition for Theorem 3.5, which like Theorem 2.3 deals with independent summands X_n^0, holds whenever each X_n^0 has a log-concave probability mass function (see Lemma 3.6).

Section 4 consists of the proofs of the general theorems 2.1 and 3.2.

One possible use for our bounds on coarseness is explained in Example 2.6. A similar, but more substantial, application of our new results inspired the present work. In Fill (1987b) we use Corollary 2.9 and its analogue for odd population size to complement a theorem of Freedman and Lane (1980) by showing that if a subset C of size t is drawn at random from the integers $0, 1, \ldots, n-1$ and $\min(t, n-t) \to \infty$, then the probability that no Fourier coefficient of the sequence $(I_C(n))$ vanishes approaches unity. As shown by this application and by Example 2.6, the inequality (1.1) can serve as a simple efficient alternative to existing deep analytical results (such as Berry–Esseen theorems), and is even easy to apply in certain settings where traditional analysis fails to provide informative bounds.

A more complete version of this paper is Fill (1987a), which also contains a discussion of results of Kleitman (1976) related to those presented in this paper.

2. Results for general Hilbert spaces. When compared with the generic result stated in Section 1, the results of this section exhibit two new features. First, the class of random sums $S = \Sigma_{n=1}^N X_n$ under investigation is broadened. In the generic result of Section 1 the values ja_n of each X_n form a set of equally spaced collinear points in H, the common distance between successive points being at least unity. In Theorem 2.3 dealing with independent summands both the equispacing and collinearity requirements are relaxed; it is merely assumed [see (2.5)] that the distances measured in some direction a_n between successive values of X_n are all at least unity. Except for a parity consideration discussed momentarily, the foregoing remark about (2.5) serves also to explain condition (2.1) in the abstract Theorem 2.1, which generates the two main theorems 2.3 (independent summands) and 2.7 (sampling) of this section.

The second new feature apparent in Theorem 2.1 is the auxiliary random vector $(Y_1, \ldots, Y_N)$ introduced for the following purposes. In proving Theorem 2.3, it is convenient that the points $\mu_n \in \frac{1}{2}\mathbf{Z}$ of symmetry for the summands X_n^0 [see (2.6)] all vanish. This can be arranged by the transformation $X_n^0 \to 2(X_n^0 - \mu_n)$, but of course parity is thereby fixed. In this setting the Y_n's deterministically indicate the parity of $2(X_n^0 - \mu_n)$, i.e., of $2\mu_n$. In the proof of Theorem 2.7 the Y_n's again play a role in keeping track of parity, but are genuinely random and have another more important use; see the proof of Theorem 2.7 for details.

All the results of the present section are derived from the following theorem.

THEOREM 2.1. *Let* $(X_1^0, \dots, X_N^0)$ *and* $(Y_1, \dots, Y_N)$ *be* $\mathbb{Z}^N$*-valued random vectors. For* $n = 1, \dots, N$, $j \in \mathbb{Z}$ *and* $(y_n, \dots, y_N)$ *such that* $P\{(Y_n, \dots, Y_N) = (y_n, \dots, y_N)\} > 0$, *let* $x_n(j; y_n, \dots, y_N) \in H$. *Let* X_n *be the random variable* $x_n(X_n^0; Y_n, \dots, Y_N)$ *and set*

$$S^0 = \sum_{n=1}^{N} X_n^0, \qquad S = \sum_{n=1}^{N} X_n, \qquad T = \sum_{n=1}^{N} Y_n.$$

For integers $k \geq 1$ *and* $t \in \mathbb{Z}$, *define*

$$V(t, k) = \{\lambda, \lambda + 2, \dots, \lambda + 2(k - 1)\},$$

where $\lambda = -(k-1)$ *or* $\lambda = -k$ *according as* $t + k$ *is odd or even. If the three conditions* (2.1)–(2.3) *hold, then the probability that* S *lies in the union of any* K *sets in* H *each of diameter* < 1 *is no more than* $P\{S^0 \in V(T, K)\}$.

Here are the three conditions to which Theorem 2.1 refers. For (2.1) we suppress the allowed dependence of a_n, $x_n(j)$ and $\xi_n(j)$ on $(y_n, \dots, y_N)$.

(2.1) There exist unit vectors $a_n \in H$ such that if $\xi_n(j)a_n$ is the projection of $x_n(j)$ onto a_n [i.e., if $\xi_n(j)$ is the inner product $\langle x_n(j), a_n\rangle$], then $\xi_n(j) - \xi_n(j-2) \geq 1$ for all j.

(2.2) $(X_1^0, \dots, X_{n-1}^0; Y_1, \dots, Y_{n-1})$ and X_n^0 are conditionally independent given $(Y_n, \dots, Y_N)$.

(2.3) If
$$P\{(Y_n, \dots, Y_N) = (y_n, \dots, y_N)\} > 0,$$
then $\mathscr{L}(X_n^0 | (Y_n, \dots, Y_N) = (y_n, \dots, y_N))$ is symmetric about 0 and unimodal over the set of integers with the same parity as y_n.

For the proof of Theorem 2.1, see Section 4.1.

REMARK 2.2. (a) $V(t, k)$ consists of the k integers of smallest absolute value that are of the same parity as t. If $t + k$ is even, we have for definiteness chosen the set that is skewed to the left. But note that the theorem applies indifferently to S and $-S$, so that $P\{S^0 \in V(T, K)\} = P\{S^0 \in -V(T, K)\}$.

(b) From condition (2.3) we see that the parities of S^0 and T agree. Therefore the bound $P\{S^0 \in V(T, K)\}$ can also be written as

$$P\{S^0 \in V(0, K) \cup V(1, K)\} = \sum_{k=-K}^{K-1} P\{S^0 = k\}. \tag{2.4}$$

Note, however, that in applications for which (the parity of) T is nonrandom, half of the $2K$ terms in (2.4) will automatically vanish. This is the situation prevailing in Theorem 2.3 and in Theorem 2.7 when the sample size T is fixed.

2.1. *Independent summands.*

THEOREM 2.3. *Let $X_1^0, \dots, X_N^0$ be independent $\mathbb{Z}$-valued random variables. For $n = 1, \dots, N$ and $j \in \mathbb{Z}$, let $x_n(j) \in H$. Let $X_n = x_n(X_n^0)$ and set $S^0 = \sum_{n=1}^N X_n^0$ and $S = \sum_{n=1}^N X_n$. If for $n = 1, \dots, N$,*

(2.5) *there exist unit vectors $a_n \in H$ such that $\langle x_n(j) - x_n(j-1), a_n\rangle \ge 1$ for all j*

and

(2.6) *$\mathscr{L}(X_n^0)$ is symmetric about $\mu_n \in \frac{1}{2}\mathbb{Z}$ and unimodal over $\mathbb{Z}$,*

then $S \le^{sc} S^0$, and the top K sets $L_{S^0}(K)$ are those consisting of K distinct integers $v_1, \dots, v_K$ that minimize $\sum_{k=1}^K |v_k - \sum_{n=1}^N \mu_n|$.

PROOF. Apply Theorem 2.1, replacing each instance of $(X_n^0, Y_n, x_n(j), X_n)$ in the statement of that theorem by the quantities $(2(X_n^0 - \mu_n), Y_n, x_n(\mu_n + \frac{1}{2}j), X_n)$ of the present theorem, where here

$$Y_n = 1 - I_{\mathbb{Z}}(\mu_n) = \begin{cases} 1, & \text{if } 2\mu_n \text{ is odd,} \\ 0, & \text{if } 2\mu_n \text{ is even.} \end{cases}$$

The conclusion is that when $\operatorname{diam}(R_k) < 1$ for every k, the probability $P\{S \in \bigcup_{k=1}^K R_k\}$ to be bounded does not exceed

$$P\left\{2\left(S^0 - \sum_{n=1}^N \mu_n\right) \in V(\#\{n: \mu_n \notin \mathbb{Z}\}, K)\right\} = \pi_S^0(K). \qquad \square$$

REMARK 2.4. It follows from the theorem's conclusion that S^0 has a unimodal distribution, symmetric about $\sum_{n=1}^N \mu_n$. This can also be shown directly à la Wintner (1938), Theorem 11.4.

Theorem 2.3 is stronger than Kleitman's (1970) Theorem II in two important ways. First, the assumption that each X_n^0 has a uniform distribution is greatly relaxed by (2.6); one need not even assume that the X_n^0's are bounded. The second improvement, the extension beyond sums of the form $S = \sum_{n=1}^N X_n^0 a_n$, is highlighted in the following example.

EXAMPLE 2.5. Suppose that $X_1^0, \dots, X_N^0$ are independent, with X_n^0 uniformly distributed on $\{1, \dots, M_n\}$. Let $X_1, \dots, X_N$ be independent H-valued random variables, and suppose that each X_n has a uniform distribution on M_n distinct points in H; the M_n points need *not* be as required by Kleitman's theorem, namely, equally spaced collinear points ja_n with $\|a_n\| \ge 1$ and j in a set of M_n consecutive integers. Put $S^0 = \sum_{n=1}^N X_n^0$ and $S = \sum_{n=1}^N X_n$; then $S \le^c S^0$.

To see this, one need only find a direction a_n separating the given M_n points from each other and apply Theorem 2.3 to a scalar multiple of S.

When N is large it may be difficult to compute exactly the bounds $\pi_{S^0}(K)$ resulting from Theorem 2.3. Nevertheless, Theorem 2.3 may still be useful in asymptotic settings, even when the Hilbert space is $\mathbb{R}$.

EXAMPLE 2.6. Consider the simple case of Theorem 2.3 with $K = 1$ when $X_1^0, \dots, X_N^0$ are i.i.d. symmetric ± 1 random variables. The result is

$$\begin{aligned} P\{S \in R\} &\le P\{\text{binomial}(N, 1/2) = \lfloor N/2 \rfloor\} \\ &= 2^{-N}\binom{N}{\lfloor N/2 \rfloor} \le \sqrt{2/\pi}\, N^{-1/2}, \end{aligned} \tag{2.7}$$

the last inequality being also an asymptotic equivalence. Suppose one has at hand real numbers $a_{ij} \ge 1$; $i = 1, \dots, N$; $j = 1, \dots, J(N) = o(N^{1/2})$ with $J(N) \to \infty$; and one is interested in showing

$$P\left\{ \sum_{i=1}^{N} X_i^0 a_{ij} = 0 \text{ for some } j \right\} \to 0, \quad \text{as } N \to \infty. \tag{2.8}$$

Indeed, such a problem, in a somewhat more complicated setting, motivated the present paper; see Fill (1987b). Without studying the joint distribution of the $J(N)$ sums of interest, one may resort to subadditivity to obtain the bound

$$\sum_{j=1}^{J(N)} P\left\{ \sum_{i=1}^{N} X_i^0 a_{ij} = 0 \right\} \tag{2.9}$$

for (2.8). Use of (2.7) yields (2.8) immediately.

Alternatively, one might attempt to apply a central limit theorem directly to each of the $J(N)$ terms in (2.9). In general, $\sum_{i=1}^{N} X_i^0 a_{ij}$ will not be of lattice type, so no local limit theorem will be available. Nonetheless, with j fixed if $a_{Nj}^2 / \sum_{i=1}^{N} a_{ij}^2 \to 0$, then the global central limit theorem [Feller (1971), Theorem XV.6.1] implies $P\{\sum_{i=1}^{N} X_i^0 a_{ij} = 0\} \to 0$ as $N \to \infty$; but this estimate falls far short of (2.8). On the other hand, application of the Berry–Esseen theorem [Feller (1971), Theorem XVI.5.2] yields

$$P\left\{ \sum_{i=1}^{N} X_i^0 a_{ij} = 0 \right\} \le 12 N^{-1/2} \left[\left(N^{-1} \sum_{i=1}^{N} a_{ij}^3 \right)^{1/3} \Big/ \left(N^{-1} \sum_{i=1}^{N} a_{ij}^2 \right)^{1/2} \right]^3. \tag{2.10}$$

This bound is always at least $12 N^{-1/2}$, which is itself bigger than the bound in (2.7) by a factor of $12\sqrt{\pi/2}$. It is even possible to find a_{ij}'s satisfying the central limit theorem condition but for which the right side in (2.10) is *not* $O(N^{-1/2})$.

If, for example, it is known that $|a_{ij}| \le B$ for some $B < \infty$ and all i and j, then the Lévy concentration function can also be used in the standard fashion to show that $P\{\sum_{i=1}^{N} X_i^0 a_{ij} = 0\} = O(N^{-1/2})$ uniformly in j. However, the general case $a_{ij} \ge 1$ seems difficult to treat using concentration functions.

At any rate, we remind the reader that the bound $2^{-N}\binom{N}{\lfloor N/2 \rfloor}$ on $P\{\sum_{i=1}^{N} X_i^0 a_{ij} = 0\}$ is best possible assuming only $a_{ij} \geq 1$ for all i and j.

In summary, results like Theorem 2.3 can serve as simple efficient substitutes for existing deep analytical methods, and can apply in certain more general settings, such as in the setup just described when $a_{Nj}^2/\sum_{i=1}^{N} a_{ij}^2$ does *not* converge to 0.

2.2. *Sampling from a paired population.* We turn our attention next to theorems of the general form discussed in Section 1 when the summands $(X_1^0, X_2^0, \dots, X_N^0)$ form a vector of counts for a sample from the indices $1, \dots, N$. Of course when the sample size $S^0 = \sum_{n=1}^{N} X_n^0$ is fixed, bounds of the type $S \leq^c S^0$ for $S = \sum_{n=1}^{N} X_n^0 a_n$ are useless, for then $\pi_{S^0}(1) = 1(!)$. However, when the vectors a_n can be paired, as in the next result, informative bounds on the coarseness of the sample sum S can be derived. Theorem 2.7 can yield useful results (see Remark 2.10) even when no "natural" pairing presents itself.

THEOREM 2.7. *Let $a_1, \dots, a_N$ and $b_1, \dots, b_N$ be vectors in H such that*

$$\|a_n - b_n\| \geq 1, \quad \textit{for all } n = 1, \dots, N. \tag{2.11}$$

Let $(A_1, \dots, A_N)$ and $(B_1, \dots, B_N)$ be $\mathbb{Z}^N$-valued random vectors. Set

$$S^0 = \sum_{n=1}^{N} (A_n - B_n), \qquad S = \sum_{n=1}^{N} (A_n a_n + B_n b_n), \qquad T = \sum_{n=1}^{N} (A_n + B_n).$$

If for $n = 1, \dots, N$,

(2.12) *$(A_1, \dots, A_{n-1}; B_1, \dots, B_{n-1})$ and A_n are conditionally independent given $(A_n + B_n, \dots, A_N + B_N)$*

and

(2.13) *$P\{(A_n + B_n, \dots, A_N + B_N) = (c_n, \dots, c_N)\} > 0$ implies $\mathscr{L}(A_n | (A_n + B_n, \dots, A_N + B_N) = (c_n, \dots, c_N))$ is symmetric about $c_n/2$ and unimodal over $\mathbb{Z}$,*

then the conclusion of Theorem 2.1 obtains.

PROOF. Apply Theorem 2.1 to $X_n^0 = A_n - B_n$ and $Y_n = A_n + B_n$, with $x_n(j; y_n, \dots, y_N) = \frac{1}{2}(y_n + j)a_n + \frac{1}{2}(y_n - j)b_n$, so that $X_n = A_n a_n + B_n b_n$. □

From now on we interpret Theorem 2.7 as a theorem about sampling, as follows. A sample of (possibly random) size T is drawn from the population $a_1, \dots, a_N; b_1, \dots, b_N$ according to some sampling scheme possibly allowing repetitions. Let $A_n \geq 0$ and $B_n \geq 0$ count, respectively, the number of times that a_n and b_n are included in the sample. Let S be the sample sum. Subject to the matching condition (2.11) on the population and the restrictions (2.12) and (2.13) governing the choice of sampling scheme, the conclusion of Theorem 2.7 obtains;

in particular, if the parity of T is nonrandom, then the coarsest distribution for S arises from the choice $a_n = -b_n = a$ with $\|a\| \geq \frac{1}{2}$.

The conditions (2.12)–(2.13) can be understood with the aid of the following easily verified lemma. Let

$$C_n = A_n + B_n$$

count the number of times that vectors with index n (a_n or b_n) are included in the sample. (We recall in passing that in proving Theorem 2.7 the auxiliary variable Y_n of Theorem 2.1 was taken to be precisely C_n.)

LEMMA 2.8. *The two conditions*

(2.14) $(A_1, B_1; \cdots; A_{n-1}, B_{n-1}; A_{n+1}, B_{n+1}; \cdots; A_N, B_N)$ *and* A_n *are conditionally independent given* C_n

and

(2.15) $P\{C_n = c\} > 0$ *implies* $\mathscr{L}(A_n | C_n = c)$ *is symmetric about* $c/2$ *and unimodal over* $\mathbb{Z}$

are together sufficient for (2.12)–(2.13).

We interpret Lemma 2.8 as follows. Suppose that a particular sampling scheme of interest can be performed in the following way. First, a sample is drawn from the indices $1, \dots, N$ in a completely arbitrary manner. Let $C_1, \dots, C_N$ be the respective counts and set $T = \sum_{n=1}^N C_n$. The remaining steps are carried out conditionally given $(C_1, \dots, C_n) = (c_1, \dots, c_n)$. Let A_n denote the number of times the vector a_n is to be included in the sample. $A_1, \dots, A_N$ are chosen independently according to (conditional) distributions satisfying (2.15). After the value of A_n is determined, the remaining $B_n = C_n - A_n$ sampled instances of index n are allocated to the vector b_n. Then (2.14) holds, so if (2.11) also holds, S satisfies the probability bounds of Theorem 2.7.

Suppose for example that, conditionally given $C_n = c$, A_n is obtained by drawing a sample of size c without replacement from a population consisting of $\gamma_n(c)$ copies each of a_n and b_n, where $c/2 \leq \gamma_n(c) \leq \infty$. [If $\gamma_n(c) = \infty$, we mean that A_n is gotten by ordinary sampling *with* replacement from $\{a_n, b_n\}$.] Then $\mathscr{L}(A_n | C_n = c)$ is the hypergeometric$(\gamma_n(c), \gamma_n(c); c)$ distribution [if $\gamma_n(c) = \infty$, $\mathscr{L}(A_n | C_n = c) = \text{binomial}(c, 1/2)$], so Theorem 2.7 applies.

The setup of the preceding two paragraphs is in force with $\gamma_n(c) = \gamma_n$ when the overall sampling scheme for Theorem 2.7 draws a random sample of fixed size t without replacement from a population consisting of $\gamma_n < \infty$ copies each of a_n and b_n, $n = 1, \dots, N$. Then $S^0 = 2\sum_{n=1}^N A_n - t$ is a linear transformation of $\sum_{n=1}^N A_n \sim \text{hypergeometric}(\sum_{n=1}^N \gamma_n, \sum_{n=1}^N \gamma_n; t)$. We have proved the following result.

COROLLARY 2.9 (Sampling without replacement). *Let* $a_1, \dots, a_N$ *and* $b_1, \dots, b_N$ *satisfy* (2.11). *Let* S *be the sum of a random sample of fixed size* t *without replacement from a population consisting of* $\gamma_n < \infty$ *copies each of* a_n *and* b_n, $n = 1, \dots, N$. *Then* $S \leq^{sc} Z$, *where* $Z \sim$

hypergeometric($\Sigma_{n=1}^{N}\gamma_n, \Sigma_{n=1}^{N}\gamma_n; t$), *and the top* K *sets* $L_Z(K)$ *are those consisting of* K *distinct integers* $v_1, \ldots, v_K$ *that minimize* $\Sigma_{k=1}^{K}|v_k - t/2|$.

REMARK 2.10. Suppose that a sample of size t is drawn without replacement from a population consisting of $\gamma < \infty$ copies each of vectors $w_1, \ldots, w_{2N}$ in H. If no $N+1$ of the $2N$ vectors agree, then Corollary 2.9 provides information about the coarseness of the sample sum S. Indeed, it is then possible to divide the vectors into two groups of size N each, say, $a_1, \ldots, a_N$ and $b_1, \ldots, b_N$, so that $a_n \neq b_n$ for all n. Let $\varepsilon = \min\{\|a_n - b_n\|\colon n \in \{1, \ldots, N\}\} > 0$. Then $\varepsilon^{-1}S \leq^{sc} Z$ with Z as in Corollary 2.9. In particular, $S \leq^{c} Z$.

Of course, the larger is the value of ε, the stronger is the result $\varepsilon^{-1}S \leq^{sc} Z$. One can show by induction on N that when $w_1 \leq \cdots \leq w_{2N}$ in $H = \mathbb{R}$, the pairing giving the largest value of ε is achieved by taking $a_n = w_{N+n}$ and $b_n = w_n$ for $n = 1, \ldots, N$.

Our next application of Theorem 2.7 includes sampling with replacement. Suppose that t independent draws are made from a population $a_1, \ldots, a_N$; $b_1, \ldots, b_N$ satisfying (2.11). At the jth of the t successive draws we suppose that the probability of selecting either given one of the vectors a_n, b_n is $\frac{1}{2}p_j(n)$, where each p_j is a probability mass function on $\{1, \ldots, N\}$. Then Theorem 2.7 applies, and $S^0 = 2\Sigma_{n=1}^{N}A_n - t$ is a linear transformation of $\Sigma_{n=1}^{N}A_n \sim$ binomial$(t, 1/2)$. We have proved the following result.

COROLLARY 2.11. *Let* $a_1, \ldots, a_N$ *and* $b_1, \ldots, b_N$ *satisfy* (2.11). *For* $j = 1, \ldots, t$ *let* p_j *be a probability mass function on* $\{1, \ldots, N\}$ *and let* W_j *be an* H-*valued random variable satisfying*

$$P\{W_j = a_n\} = \tfrac{1}{2}p_j(n) = P\{W_j = b_n\}, \qquad n = 1, \ldots, N.$$

Suppose $W_1, W_2, \ldots, W_t$ *are independent and set* $S = \Sigma_{j=1}^{t}W_j$. *Then* $S \leq^{sc} Z$, *where* $Z \sim$ *binomial*$(t, 1/2)$.

A simple truncation argument extends Corollary 2.11 to the case $N = \infty$.

EXAMPLE 2.12. Under certain circumstances Corollary 2.11 yields a bound on the coarseness of the "null" sum S^0 itself in Theorem 2.3. For simplicity suppose that each summand X_n^0 in Theorem 2.3 is uniformly distributed on a subset $\{m_n, \ldots, M_n\}$ of $\mathbb{Z}$ having even cardinality. Corollary 2.11 then implies that $S^0 \leq^{c}$ binomial$(t, 1/2)$.

When $p_j(n) \equiv N^{-1}$ in Corollary 2.11 we have the following result.

COROLLARY 2.13 (Sampling with replacement). *Let* $a_1, \ldots, a_N$ *and* $b_1, \ldots, b_N$ *satisfy* (2.11). *Let* S *be the sum of a random sample of fixed size* t *with replacement from the population* $a_1, \ldots, a_N$; $b_1, \ldots, b_N$. *Then* $S \leq^{sc} Z$, *where* $Z \sim$ *binomial*$(t, 1/2)$.

Corollary 2.14 in Fill (1987a) is a complementary result to Theorem 2.7 for fixed size sampling from a population with an *odd* number of elements.

3. Additional results for the real line. Let $X_1^0, \dots, X_N^0$ be independent $\mathbb{Z}$-valued random variables and for vectors $a_1, \dots, a_N$ in a Hilbert space H with length at least unity, set $S = \sum_{n=1}^N X_n^0 a_n$. Let $a \in H$ have unit length, let $\varepsilon_1, \dots, \varepsilon_N$ be any deterministic sequence of $(+1)$'s and (-1)'s and set $S^1 = \sum_{n=1}^N X_n^0 \varepsilon_n a$. If for each n the distribution of X_n^0 is symmetric about $\mu_n \in \frac{1}{2}\mathbb{Z}$ and unimodal over $\mathbb{Z}$, then Theorem 2.3 implies that $S \le^{sc} S^1$.

That is, in the symmetric case, any choice of (a_n) of the form $(\varepsilon_n a)$ yields the strongly coarsest distribution of S. What if the assumptions of symmetry are dropped? The answer is that, in contrast with the symmetric case, there may exist no choice of $a_1, \dots, a_N$ yielding simultaneously the K-coarsest distribution of S for every K.

EXAMPLE 3.1. Let $H = \mathbb{R}$ and $N = 2$. Let X_1^0 and X_2^0 be independent Bernoulli(2/3) random variables. Choices for a_1, a_2 of the form $a_1 = a = -a_2$, and only of this form, yield the largest modal probability for S_2 (namely, 5/9 at the mode 0). On the other hand, only choices of the form $a_1 = a = a_2$ yield the largest top 2 probability (namely, $\pi_{S_2}(2) = P\{S_2 = a\} + P\{S_2 = 2a\} = 4/9 + 4/9 = 8/9$).

In this section we obtain for asymmetrically distributed summands results bounding the coarseness of $S = \sum_{n=1}^N X_n^0 a_n$ and other somewhat more general sums in terms of the coarseness of $S^0 = \sum_{n=1}^N X_n^0$. We do this by narrowing our focus to the Hilbert space $H = \mathbb{R}$ and requiring $a_n \ge 1$ rather than just $|a_n| \ge 1$.

The main theorems 3.5 (independent summands) and 3.8 (sampling) of this section are the asymmetric analogues of Theorems 2.3 and 2.7, respectively. The following all-purpose result is analogous to Theorem 2.1 and is proved in Section 4.2.

THEOREM 3.2. *Let $(X_1^0, \dots, X_N^0)$ be a $\mathbb{Z}^N$-valued random vector. For $n = 1, \dots, N$ and $j \in \mathbb{Z}$, let $x_n(j) \in \mathbb{R}$. Let X_n be the random variable $x_n(X_n^0)$ and set*

$$S^0 = \sum_{n=1}^N X_n^0, \qquad S = \sum_{n=1}^N X_n.$$

If the four conditions (3.1)–(3.4) *hold, then* $S \le^{sc} S^0$.

To state the four conditions we need some notation. Let N_1 and N_2 be nonnegative integers sch that $N_1 + N_2 = N$. We divide the summands X_i^0 into two groups as follows. For $i = 1, \dots, N_1$, set $X_i^0(1) = X_i^0$; for $i = 1, \dots, N_2$, set $X_i^0(2) = X_{N_1+i}^0$. For $m = 1, 2$ define $S_n^0(m) = \sum_{i=1}^n X_i^0(m)$ for $n = 0, \dots, N_m$ and let $S^0(m) = S_{N_m}^0(m)$ be the sum of all X_i^0 in the mth group, so that $S^0 = S^0(1) + S^0(2)$.

Here are the four conditions to which Theorem 3.2 refers. Conditions (3.2)–(3.3) are assumed to hold for $m = 1, 2$ and $n = 1, \dots, N_m$. The transformation $m \to (3 - m)$ is used to switch between indices 1 and 2.

(3.1) The inequality $x_n(j) - x_n(j-1) \geq 1$ holds for all $n = 1, \dots, N$ and all j.

(3.2) $(X_1^0(m), \dots, X_{n-1}^0(m);\ X_1^0(3-m), \dots, X_{N_{3-m}}^0(3-m))$ and $X_n^0(m)$ are conditionally independent given $S^0(m) - S_{n-1}^0(m)$.

(3.3) Both $\mathscr{L}(X_n^0(m) | S^0(m) - S_{n-1}^0(m) = \ell)$ and $\mathscr{L}(S^0(m) - S_n^0(m) | S^0(m) - S_{n-1}^0(m) = \ell)$ are stochastically nondecreasing as ℓ increases through the possible values of $S^0(m) - S_{n-1}^0(m) = X_n^0(m) + (S^0(m) - S_n^0(m))$.

(3.4) If (ℓ_1, ℓ_2) and (ℓ_1', ℓ_2') are possible values of $(S^0(1), S^0(2))$ and $\ell_1 < \ell_1'$, then $\ell_2 \leq \ell_2'$.

Remark 3.3. (a) The division of $X_1^0, \dots, X_N^0$ into two groups allows for the establishment of sampling results. Condition (3.1) generalizes the definition $S = \sum_{n=1}^N X_n^0 a_n$ and condition (3.2) generalizes the requirement that the X_n^0's be independent. Condition (3.3) is the appropriate replacement for the symmetric unimodality condition (2.3) used in Theorem 2.1. Condition (3.4) is technical and is discussed at length in Remark 3.4.

(b) Neither the theorem nor its proof provides information as to which K values for S^0 form the top K set $L_{S^0}(K)$. This deficiency is addressed in Example 3.7.

Remark 3.4. (a) To understand condition (3.4), consider the following equivalent version. Let $V_1 = \{\ell_1: P\{S^0(1) = \ell_1\} > 0\}$ be the set of possible values for $S^0(1)$. For $\ell_1 \in V_1$ let $V_2(\ell_1) = \{\ell_2: P\{S^0(1) = \ell_1, S^0(2) = \ell_2\} > 0\} \neq \varnothing$ be the set of possible values for $S^0(2)$ when $S^0(1) = \ell_1$. Then (3.4) is equivalent to the assertion that $V_2(\ell)$ is nondecreasing elementwise in $\ell \in V_1$, in the precise sense that if $\ell_1 \in V_1$ and $\ell_1' \in V_1$ with $\ell_1 < \ell_1'$ and $\ell_2 \in V_2(\ell_1)$ and $\ell_2' \in V_2(\ell_1')$, then $\ell_2 \leq \ell_2'$.

(b) Suppose in particular that $S^0(2) = g(S^0(1))$ for some nondecreasing function g on V_1. Then for $\ell \in V_1$ we have $V_2(\ell) = \{g(\ell)\}$, and so condition (3.4) holds. This observation will be used in the proofs of Theorems 3.5 and 3.8.

(c) The following consequence of (3.4) is used twice in the proof of Theorem 3.2: If $(\ell_1, \ell_2) \neq (\ell_1', \ell_2')$ are possible values of $(S^0(1), S^0(2))$ and $\ell_1 + \ell_2 \leq \ell_1' + \ell_2'$, then $\ell_1 \leq \ell_1'$ and $\ell_2 \leq \ell_2'$ and at least one of these two inequalities is strict. Indeed, $\ell_1 > \ell_1'$ is easily ruled out, $\ell_1 = \ell_1'$ leads immediately to the desired conclusion, and $\ell_1 < \ell_1'$ implies $\ell_2 \leq \ell_2'$.

3.1. *Independent summands.*

Theorem 3.5. *Let $X_1^0, \dots, X_N^0$ be independent $\mathbb{Z}$-valued random variables. For $n = 1, \dots, N$ and $j \in \mathbb{Z}$, let $x_n(j) \in \mathbb{R}$. Let $X_n = x_n(X_n^0)$, $S^0 = \sum_{n=1}^N X_n^0$,*

$S = \sum_{n=1}^{N} X_n$. *If* (3.1) *holds and for* $n = 1, \ldots, N$,

(3.5) *both* $\mathscr{L}\left(X_n^0 \,\middle|\, \sum_{i=n}^{N} X_i^0 = \ell\right)$ *and* $\mathscr{L}\left(\sum_{i=n+1}^{N} \,\middle|\, X_i^0 \sum_{i=n}^{N} X_i^0 = \ell\right)$ *are stochastically nondecreasing as* ℓ *increases through the possible values of* $\sum_{i=n} R^N X_i^0$,

then $S \leq^{\mathrm{sc}} S^0$.

PROOF. Apply Theorem 3.2 with $N_2 = 0$ and note that one can take $g \equiv 0$ in Remark 3.4(b). □

The following simple sufficient condition for (3.5) follows from the main theorem of and Remark 3.1 in Efron (1965).

LEMMA 3.6. *In the setting of Theorem* 3.5 *it is sufficient for condition* (3.5) *that the logarithm of the probability mass function of every summand* X_n^0 *be a concave function on* $\mathbb{Z}$. *In that case the probability mass function of* S^0 *is also log concave and hence unimodal over* $\mathbb{Z}$.

Exact calculation of the bound $\pi_{S^0}(K)$ in Theorem 3.5 is typically difficult. However, if, for example by use of Lemma 3.6, one knows that $\mathscr{L}(S^0)$ is unimodal over $\mathbb{Z}$, then a top K set $L_{S^0}(K)$ consists of K consecutive integers and includes a mode $\mu \in \mathbb{Z}$ for S^0.

EXAMPLE 3.7. Suppose in Theorem 3.5 that $X_n^0 \sim$ Bernoulli(p). Then the conditions of Lemma 3.6 are trivially met, and $S^0 \sim$ binomial(N, p) with mode $\mu = \lfloor Np \rfloor$. So we have an asymmetric extension of Theorem I in Kleitman (1970). Practically speaking, however, there remains the problem of determining $\pi_{S^0}(K)$ when $K \geq 2$.

Typical values K of interest will be small (cf. Example 2.6). Then one can turn to uniform local limit theorems, such as Theorem XV.5.3 in Feller (1971) or its analogue for nonidentical components, to estimate $\pi_{S^0}(K)$ when N is large. In the present example we arrive at

$$\pi_{S^0}(K) \sim \frac{K}{\sqrt{2\pi p(1-p)}} N^{-1/2},$$

as $N \to \infty$ with K fixed.

3.2. *Sampling.* The following general theorem deals with the coarseness of the sum of a sample of fixed size from a (mixed) population of real numbers $a_1, \ldots, a_{N_1} \geq 1$ and $b_1, \ldots, b_{N_2} \leq -1$.

THEOREM 3.8. *Let* N_1 *and* N_2 *be nonnegative integers and set* $N = N_1 + N_2$. *Let* $a_1, \ldots, a_{N_1}$ *and* $b_1, \ldots, b_{N_2}$ *be real numbers such that* $a_n \geq 1$ *for all* $n = 1, \ldots, N_1$ *and* $b_n \leq -1$ *for all* $n = 1, \ldots, N_2$. *Let* $A_1, \ldots, A_{N_1}$ *and* $B_1, \ldots, B_{N_2}$ *be*

$\mathbb{Z}$-valued random variables. Set

$$S^0 = \sum_{n=1}^{N_1} A_n - \sum_{n=1}^{N_2} B_n, \qquad S = \sum_{n=1}^{N_1} A_n a_n + \sum_{n=1}^{N_2} B_n b_n.$$

If for some $t \in \mathbb{Z}$,

$$P\left\{ \sum_{n=1}^{N_1} A_n + \sum_{n=1}^{N_2} B_n = t \right\} = 1$$

and the conditions (3.6)–(3.7) *hold, then $S \le^{sc} S^0$.*

Here are the conditions to which Theorem 3.8 refers. For $n = 1, \dots, N_1$ we require:

(3.6a) $(A_1, \dots, A_{n-1}; B_1, \dots, B_{N_2})$ and A_n are conditionally independent given $\Sigma_{i=n}^{N_1} A_i$.

(3.7a) Both $\mathscr{L}(A_n | \Sigma_{i=n}^N A_i = \ell)$ and $\mathscr{L}\left(\Sigma_{i=n+1}^N \middle| A_i \sum_{i=n}^N A_i = \ell \right)$ are stochastically nondecreasing as ℓ increases through the possible values of $\Sigma_{i=n}^N A_i$.

For $n = 1, \dots, N_2$ we require conditions (3.6b) and (3.7b) which are obtained from (3.6a) and (3.7a) by interchanging the symbols A and B and the symbols N_1 and N_2.

PROOF OF THEOREM 3.8. Apply Theorem 3.2 to $X_n^0(1) = A_n$ and $X_n^0(2) = -B_n$, with $x_n(j)$ defined as ja_n for $n = 1, \dots, N_1$ and as $-jb_{n-N_1}$ for $n = N_1 + 1, \dots, N$. Observe that one can take $g(\ell) = \ell - t$ in Remark 3.4(b). □

The interpretation of Theorem 3.8 as a theorem about fixed size sampling is straightforward; compare the discussion following Theorem 2.7. The result here says that the coarsest distribution for S results from the choice $a_n \equiv +1$ and $b_n \equiv -1$.

For brevity, it is left to the reader to apply Theorem 3.8 to formulate an appropriate analogue of Corollary 2.11 in the present setting and to prove the following analogues of Corollaries 2.9 and 2.13. See Fill (1987a), Section 3, for details and further discussion.

COROLLARY 3.9 (Sampling without replacement). *Let $a_1, \dots, a_{N_1}$ and $b_1, \dots, b_{N_2}$ be as in Theorem 3.8. Let S be the sum of a random sample of fixed size t without replacement from a population consisting of $\alpha_n < \infty$ copies of a_n, $n = 1, \dots, N_1$, and $\beta_n < \infty$ copies of b_n, $n = 1, \dots, N_2$. Then $S \le^{sc} Z$, where $Z \sim$ hypergeometric$(\Sigma_{n=1}^{N_1} \alpha_n, \Sigma_{n=1}^{N_2} \beta_n; t)$.*

COROLLARY 3.10 (Sampling with replacement). *Let $a_1, \dots, a_{N_1}$ and $b_1, \dots, b_{N_2}$ be as in Theorem 3.8. Let S be the sum of a random sample of fixed* *size t with replacement from the population $a_1, \dots, a_{N_1}$; $b_1, \dots, b_{N_2}$. Then $S \leq^{sc} Z$, where $Z \sim$ binomial$(t, N_1/N)$.*

4. Proofs. In this section we present proofs of the general Theorems 2.1 and 3.2.

4.1. *Proof of Theorem* 2.1. Theorem 2.1 follows immediately upon setting $n = N$ in the next lemma.

LEMMA 4.1. *Adopt the notation and hypotheses of Theorem* 2.1—*in particular conditions* (2.1)–(2.3). *For $n = 0, \dots, N$, define*

$$S_n^0 = \sum_{i=1}^{n} X_i^0, \qquad S_n = \sum_{i=1}^{n} X_i, \qquad T_n = \sum_{i=1}^{n} Y_i, \qquad Y_{n+1}^* = (Y_{n+1}, \dots, Y_N),$$

and, for given $y_{n+1}, \dots, y_N$,

$$y_{n+1}^* = (y_{n+1}, \dots, y_N).$$

Let $R_1, \dots, R_K$ be sets in H each of diameter < 1. Then for each $n = 0, \dots, N$,

$$P\left\{S_n \in \bigcup_{k=1}^{K} R_k | Y_{n+1}^* = y_{n+1}^*\right\} \leq P\{S_n^0 \in V(T_n, K) | Y_{n+1}^* = y_{n+1}^*\}, \tag{4.1}$$

provided $P\{Y_{n+1}^* = y_{n+1}^*\} > 0$.

The proof of Lemma 4.1 is reminiscent of the region rearrangement proofs in Kleitman (1970). Although the following treatment is self-contained, the reader is advised to consult those proofs first in order to gain a basic understanding of the ideas involved.

PROOF. We use induction on $n \in \{0, \dots, N\}$. Since $S_0^0 \equiv 0 \equiv T_0$ and $0 \in V(0, K)$, the basis is trivial.

Induction step $n \in \{1, \dots, N\}$. We may suppose with no loss of generality that $R_1, \dots, R_K$ are disjoint. Using (2.2), we find

$$\begin{aligned} \text{LHS}(4.1) &= \sum_{y_n} \sum_{j} P\left\{S_{n-1} \in \bigcup_{k=1}^{K} R_k - x_n(j; y_n^*),\right. \\ &\qquad\qquad \left. X_n^0 = j, Y_n = y_n | Y_{n+1}^* = y_{n+1}^*\right\} \\ &= \sum_{y_n} \sum_{j} P\left\{S_{n-1} \in \bigcup_{k=1}^{K} R_k - x_n(j; y_n^*) | Y_n^* = y_n^*\right\} \\ &\qquad\qquad \times P\{X_n^0 = j | Y_n^* = y_n^*\} P\{Y_n = y_n | Y_{n+1}^* = y_{n+1}^*\}. \end{aligned} \tag{4.2}$$

In (4.2) the sums are over y_n such that $P\{Y_n^* = y_n^*\} > 0$. We shall show that for each possible value y_n^* of Y_n^*,

$$\sum_j P\left\{S_{n-1} \in \bigcup_{k=1}^{K} R_k - x_n(j;\, y_n^*) | Y_n^* = y_n^*\right\} P\{X_n^0 = j | Y_n^* = y_n^*\} \tag{4.3}$$
$$\leq P\{S_n^0 \in V(T_n, K) | Y_n^* = y_n^*\}.$$

We can then conclude

$$\text{LHS}(4.1) \leq \sum_{y_n} P\{S_n^0 \in V(T_n, K) | Y_n^* = y_n^*\} P\{Y_n = y_n | Y_{n+1}^* = y_{n+1}^*\}$$
$$= P\{S_n^0 \in V(T_n, K) | Y_{n+1}^* = y_{n+1}^*\} = \text{RHS}(4.1),$$

as desired.

We suppose until further notice that M, the essential supremum of $|X_n^0|$ conditionally given $Y_n^* = y_n^*$, is finite and proceed by induction on M. By (2.3), y_n has the same parity as M. If $M = 0$, then by the induction hypothesis for n,

$$\text{LHS}(4.3) = P\left\{S_{n-1} \in \bigcup_{k=1}^{K} (R_k - x_n(0;\, y_n^*)) | Y_n^* = y_n^*\right\}$$
$$\leq P\{S_{n-1}^0 \in V(T_{n-1}, K) | Y_n^* = y_n^*\} = \text{RHS}(4.3).$$

The case $M = 1$ is left to the reader; use an appropriate simplification of the argument to follow for $M \geq 2$.

For the remainder of the proof of Lemma 4.1 the following conventions shall be in force. *All* probabilities are to be computed conditionally given $Y_n^* = y_n^*$. We write $x(j)$ for $x_n(j;\, y_n^*)$ and $p(j)$ for $P\{X_n^0 = j | Y_n^* = y_n^*\}$. In keeping with the parity condition of (2.3), all sums over j [including the one in LHS(4.3)] are restricted to j having the same parity as the given value y_n.

Let $2 \leq M < \infty$ be given. The cases $K = 1$ and $K \geq 2$ are treated in Lemmas 4.2 and 4.3. Our proof of (4.3) is complete when $M < \infty$. A simple truncation argument handles the general case: For any $\mu < \infty$ we can immediately state

$$\sum_{|j| \leq \mu} P\left\{S_{n-1} \in \bigcup_{k=1}^{K} R_k - x(j)\right\} P\{X_n^0 = j\}$$
$$\leq \sum_{|j| \leq \mu} P\{S_{n-1}^0 + j \in V(T_n, K)\} P\{X_n^0 = j\}.$$

Now simply let $\mu \uparrow \infty$. □

Lemma 4.2. *Suppose* $2 \leq M < \infty$. *Then* (4.3) *holds when* $K = 1$.

Proof. We observe, writing R for R_1 and noting by the symmetry condition of (2.3) the equality $p(-M) = p(M)$,

$$\text{LHS}(4.3) = \sum_j p(j) P\{S_{n-1} \in R - x(j)\}$$
$$= p(M) \sum_{|j| \leq M} P\{S_{n-1} \in R - x(j)\} \tag{4.4}$$
$$+ \sum_{|j| < M} [p(j) - p(M)] P\{S_{n-1} \in R - x(j)\}.$$

As j varies over the $M + 1$ integers having the same parity as M and satisfying $|j| \le M$, the sets $R - x(j)$ are by (2.1) and the inequality $\mathrm{diam}(R) < 1$ disjoint. The induction hypothesis for n thus yields

$$\begin{aligned}\sum_{|j|\le M} P\{S_{n-1} \in R - x(j)\} &\le P\{S^0_{n-1} \in V(T_{n-1}, M+1)\} \\ &= P\{S^0_{n-1} \in V(T_n - M, M+1)\}.\end{aligned} \tag{4.5}$$

According to the unimodality condition of (2.3), the numbers $p^0(j) \equiv p(j) - p(M)$, $|j| < M$ with j of the same parity as M, either vanish identically [in which case (4.6) follows immediately] or can be rescaled to form a unimodal probability mass function $\tilde{p}$. By the symmetry part of (2.3), $\tilde{p}$ is also symmetric; moreover, the essential supremum of a random variable $\tilde{X}^0_n$ having probability mass function $\tilde{p}$ (conditionally given $Y^*_n = y^*_n$) is strictly smaller than M. Substituting $\tilde{p}$ for p and using induction on M—with due regard to (2.2), this can be justified by making the induction hypothesis suitably broad—we conclude

$$\begin{aligned}&\sum_{|j|<M} [p(j) - p(M)] P\{S_{n-1} \in R - x(j)\} \\ &\qquad \le \sum_{|j|\le M} [p(j) - p(M)] P\{S^0_{n-1} + j \in V(T_n, 1)\}.\end{aligned} \tag{4.6}$$

This is the $\tilde{p}$ version of (4.3) when $K = 1$; note in general

$$\mathrm{RHS}(4.3) = \sum_j p(j) P\{S^0_{n-1} + j \in V(T_n, K)\}.$$

Combining (4.4)–(4.6), we find

$$\begin{aligned}\mathrm{LHS}(4.3) \le\ & p(M) P\{S^0_{n-1} \in V(T_n - M, M+1)\} \\ &+ \sum_{|j|\le M} [p(j) - p(M)] P\{S^0_{n-1} + j \in V(T_n, 1)\}.\end{aligned} \tag{4.7}$$

Decomposition according to the parity of T_n establishes the identity

$$P\{S^0_{n-1} \in V(T_n - M, M+1)\} = \sum_{|j|\le M} P\{S^0_{n-1} + j \in V(T_n, 1)\},$$

which upon substitution in (4.7) yields

$$\mathrm{LHS}(4.3) \le \sum_{|j|\le M} p(j) P\{S^0_{n-1} + j \in V(T_n, 1)\} = \mathrm{RHS}(4.3). \qquad \square$$

LEMMA 4.3. *Suppose $2 \le M < \infty$. Then* (4.3) *holds when $K \ge 2$.*

PROOF. To prove (4.3) when $K \ge 2$, we will rearrange the probabilities of the $K(M + 1)$ regions $R_k - x(j)$ appearing in LHS(4.3) to get the greatest mileage out of our induction hypothesis. To begin, for $i = 1, 2$ let H_i be the hyperplane normal to $a_n = a_n(y^*_n)$ at $\eta_i a_n$, where η_1 and η_2 are the inf and sup, respectively, of the set $\{\langle z, a_n\rangle : z \in \bigcup_{k=1}^K R_k\}$. Note that H_1 and H_2 are "supporting" hyperplanes for the set $\bigcup_{k=1}^K R_k$, which is sandwiched between the two. We may suppose that the R_k's are ordered in such a way that H_1 intersects the closure of R_1 and H_2 intersects that of R_K.

The rearrangement analogous to (4.4) is in this case

$$\begin{aligned}
\mathrm{LHS}(4.3) = p(M)\Bigg[& P\left\{S_{n-1} \in \bigcup_{k=1}^{K} R_k - x(M)\right\} \\
& + \sum_{-M \le j < M} P\{S_{n-1} \in R_K - x(j)\}\Bigg] \\
& + p(M)\Bigg[\sum_{|j|<M} P\{S_{n-1} \in R_1 - x(j)\} \\
& + P\left\{S_{n-1} \in \bigcup_{k=1}^{K-1} R_k - x(-M)\right\}\Bigg] \\
& + p(M) \sum_{|j|<M} P\left\{S_{n-1} \in \bigcup_{k=2}^{K-1} R_k - x(j)\right\} \\
& + \sum_{|j|<M} [p(j) - p(M)] P\left\{S_{n-1} \in \bigcup_{k=1}^{K} R_k - x(j)\right\}.
\end{aligned}$$

If $-M \le j < M$, the set $R_K + x(M) - x(j)$ lies strictly to the $(+a_n)$ side of H_2 since $\mathrm{diam}(R_K) < 1$ and $\langle x(M) - x(j), a_n\rangle \ge 1$ by condition (2.1). One sees that the $(K + M)$ sets $R_k - x(M)$, $1 \le k \le K$, and $R_K - x(j)$, $-M \le j < M$, are mutually disjoint. Using this and similar observations, one follows along the lines of the proof of (4.7) to arrive in the present case at

$$\begin{aligned}
\mathrm{LHS}(4.3) \le p(M)\Bigg[& P\{S_{n-1}^0 \in V(T_n - M, K + M)\} \\
& + P\{S_{n-1}^0 \in V(T_n - M, K + M - 2)\} \\
& + \sum_{|j|<M} P\{S_{n-1}^0 + j \in V(T_n, K - 2)\}\Bigg] \\
& + \sum_{|j|<M} [p(j) - P(M)] P\{S_{n-1}^0 + j \in V(T_n, K)\}.
\end{aligned}$$

In particular, in applying the induction hypothesis to get the third of the four terms in this bound one replaces p by a uniform pmf.

Let v_k be the kth smallest member of $V(0, K)$, $k = 1, \dots, K$. By analysis of

$$\begin{aligned}
P\{S_n^0 \in V(T_n, K), T_n \text{ even}\} &= \sum_j p(j) P\{S_{n-1}^0 \in V(0, K) - j, T_n \text{ even}\} \\
&= \sum_j p(j) P\left\{S_{n-1}^0 \in \bigcup_{k=1}^{K} \{v_k\} - j, T_n \text{ even}\right\}
\end{aligned}$$

and likewise of $P\{S_n^0 \in V(T_n, K),\ T_n \text{ odd}\}$ parallel to the foregoing bounding of

$$\text{LHS}(4.3) = \sum_j p(j) P\left\{S_{n-1} \in \bigcup_{k=1}^{K} R_k - x(j)\right\},$$

one discovers that our bound on LHS(4.3) in fact equals

$$P\{S_n^0 \in V(T_n, K),\ T_n \text{ even}\} + P\{S_n^0 \in V(T_n, K),\ T_n \text{ odd}\} = \text{RHS}(4.3). \quad \square$$

4.2. *Proof of Theorem* 3.2. Theorem 3.2 is derived from the following lemma, proved later. Throughout this section we write $R_1 > R_2$ to mean that every member of the set R_1 is strictly larger than every member of the set R_2.

LEMMA 4.4. *Adopt the notation and hypotheses of Theorem* 3.2 *and the conditions* (3.1)–(3.4) *given previously. Define notation for the summands X and the sums S completely analogous to those for X^0 and S^0* [*so that, for example,* $S = S(1) + S(2)$]. *For each* $k = 1, \dots, K$ *let* $R_k(\ell_1, \ell_2)$, $\ell_1 \in \mathbb{Z}$, $\ell_2 \in \mathbb{Z}$, *be a collection of intervals, each of length* < 1, *such that*

$$R_k(\ell_1 - 1, \ell_2) > R_k(\ell_1, \ell_2) \quad \text{and} \quad R_k(\ell_1, \ell_2 - 1) > R_k(\ell_1, \ell_2), \tag{4.8}$$

for $\ell_1 \in \mathbb{Z}$ *and* $\ell_2 \in \mathbb{Z}$. *Then for each* $n_1 = 0, \dots, N_1$ *and* $n_2 = 0, \dots, N_2$,

$$\begin{aligned}\sum_{\ell_1 \in \mathbb{Z}} \sum_{\ell_2 \in \mathbb{Z}} P\Bigg\{ & S_{n_1}(1) + S_{n_2}(2) \in \bigcup_{k=1}^{K} R_k(\ell_1, \ell_2), \\ & S^0(1) - S_{n_1}^0(1) = \ell_1,\ S^0(2) - S_{n_2}^0(2) = \ell_2\Bigg\} \le \pi_{(S^0(1), S^0(2))}(K).\end{aligned} \tag{4.9}$$

PROOF OF THEOREM 3.2. Let $R_1, \dots, R_K$ be any K intervals each of length < 1. For each $k \in \{1, \dots, K\}$ and $\ell_1 \in \mathbb{Z}$ and $\ell_2 \in \mathbb{Z}$, define $R_k(\ell_1, \ell_2) = R_k - (\ell_1 + \ell_2)$ to satisfy the assumptions of Lemma 4.4. Now apply the lemma with $n_1 = N_1$ and $n_2 = N_2$ to deduce

$$P\left\{S \in \bigcup_{k=1}^{K} R_k\right\} \le \pi_{(S^0(1), S^0(2))}(K).$$

But we show

$$\pi_{(S^0(1), S^0(2))}(K) \le \pi_{S^0}(K). \tag{4.10}$$

Indeed, given distinct $t_1 = (t_1(1), t_2(2)), \dots, t_K = (t_K(1), t_K(2))$ in $\mathbb{Z}^2$, arrange them so that the values $s_k \equiv t_k(1) + t_k(2)$ are in nondecreasing order. Assume that $P\{(S^0(1), S^0(2)) = t_k\} > 0$ for every k; it should be clear how to proceed otherwise. Then by (3.4) [see Remark 3.4(c)], the values s_k are in *strictly* ascending order. Therefore

$$\sum_{k=1}^{K} P\{(S^0(1), S^0(2)) = t_k\} \le \sum_{k=1}^{K} P\{S^0 = s_k\} \le \pi_{S^0}(K),$$

which establishes (4.10) and completes the proof of the theorem. $\square$

PROOF OF LEMMA 4.4. We use induction on $n \equiv n_1 + n_2 \in \{0, \ldots, N\}$. For $\ell_1 \in \mathbb{Z}$ and $\ell_2 \in \mathbb{Z}$, we define $R(\ell_1, \ell_2) = \bigcup_{k=1}^{K} R_k(\ell_1, \ell_2)$.

Basis $n = 0$. If $n = 0$, then

$$\mathrm{LHS}(4.9) = \sum_{\ell_1} \sum_{\ell_2} I_{R(\ell_1, \ell_2)}(0) P\{S^0(1) = \ell_1, S^0(2) = \ell_2\},$$

so we need only show that 0 belongs to at most K of the sets $R(\ell_1, \ell_2)$ for which $P\{S^0(1) = l_1, S^0(2) = \ell_2\} > 0$. Indeed, if 0 were to belong to more than K of the sets, then for some integer k satisfying $1 \le k \le K$ we would have $0 \in R_k(\ell_1, \ell_2) \cap R_k(\ell_1', \ell_2')$ for two possible values $(\ell_1, \ell_2) \ne (\ell_1', \ell_2')$ of $(S^0(1), S^0(2))$ such that $\ell_1 + \ell_2 \le \ell_1' + \ell_2'$. Using (3.4) [see Remark 3.4(c)], we find that $\ell_1 \le \ell_1'$ and $\ell_2 \le \ell_2'$, with at least one of these inequalities strict. But then we have by (4.8) the contradiction that $R_k(\ell_1, \ell_2)$ and $R_k(\ell_1', \ell_2')$ are disjoint.

Induction step $n \in \{1, \ldots, N\}$. Suppose without loss of generality that $n_1 \ge 1$. We have

$$\begin{aligned} \mathrm{LHS}(4.9) &= \sum_{\ell_2} \sum_{\ell_1} \sum_{j} P\bigl\{S_{n_1-1}(1) + S_{n_2}(2) \in R(\ell_1, \ell_2) - x_{n_1}(j),\ X^0_{n_1} = j, \\ &\qquad\qquad S^0(1) - S^0_{n_1-1}(1) = l_1 + j,\ S^0(2) - S^0_{n_2}(2) = \ell_2\bigr\} \\ &= \sum_{\ell_2} \sum_{\ell_1} \sum_{j} P\bigl\{S_{n_1-1}(1) + S_{n_2}(2) \in R(\ell_1 - j, \ell_2) - x_{n_1}(j),\ X^0_{n_1} = j, \\ &\qquad\qquad S^0(1) - S^0_{n_1-1}(1) = l_1,\ S^0(2) - S^0_{n_2}(2) = \ell_2\bigr\}. \end{aligned} \tag{4.11}$$

We shall show how to write (4.11) as an average of expressions of the same form as LHS(4.9) with n_1 changed to $n_1 - 1$. The induction hypothesis then yields

$$\mathrm{LHS}(4.9) \le \pi_{(S^0(1), S^0(2))}(K),$$

as desired.

For the remainder of the proof, let $\mathscr{P}$ denote the set of possible values of $S^0(1) - S^0_{n_1-1}(1)$. Let the stochastic process A be as described in Lemma 4.5. Given $\ell_1 \in \mathscr{P}$, define the random variable $J(\ell_1)$ via the inequality

$$A_{J(\ell_1)-1} \le \ell_1 < A_{J(\ell_1)}; \tag{4.12}$$

the property (4.13) for A insures that $J(\ell_1)$ is well defined. For given $\ell_1 \in \mathscr{P}$ and each $\ell_2 \in \mathbb{Z}$, consider the (random) intervals $\tilde{R}_k(\ell_1, \ell_2) \equiv R_k(\ell_1 - J(\ell_1), \ell_2) - x_{n_1}(J(\ell_1))$. Observe that for $\ell_1 < \ell_1'$ both in $\mathscr{P}$, $\tilde{R}_k(\ell_1, \ell_2) > \tilde{R}_k(\ell_1', \ell_2)$. In light of (4.8) and (3.1) and the strictly subunit length of each interval, this requires checking only that $\ell_1 - J(\ell_1) \le \ell_1' - J(\ell_1')$, i.e., that $A_{J(\ell_1) + \ell_1' - \ell_1} > \ell_1'$. But this follows from (4.12) and (4.15).

Now extend the definition of $\tilde{R}_k(\ell_1, \ell_2)$ to all $\ell_1 \in \mathbb{Z}$ in any fashion such that (4.8) holds for the resulting family. If we write

$$f((r_k(\ell_1, \ell_2))) = \sum_{\ell_1 \in \mathbb{Z}} \sum_{\ell_2 \in \mathbb{Z}} P\left\{S_{n_1-1}(1) + S_{n_2}(2) \in \bigcup_{k=1}^{K} r_k(\ell_1, \ell_2), \right.$$
$$\left. S^0(1) - S^0_{n_1}(1) = \ell_1,\ S^0(2) - S^0_{n_2}(2) = \ell_2 \right\}$$

for any family of intervals $(r_k(\ell_1, \ell_2))$ satisfying (4.8), then

$$f\big((\tilde{R}_k(\ell_1,\ell_2))\big) = \sum_{\ell_1 \in \mathscr{P}} \sum_{\ell_2 \in \mathbb{Z}} \sum_{j \in \mathbb{Z}} I_{\{A_{j-1} \le \ell_1 < A_j\}}$$

$$\times P\big\{S_{n_1-1}(1) + S_{n_2}(2) \in R(\ell_1 - j, \ell_2) - x_{n_1}(j),$$

$$S^0(1) - S^0_{n_1-1}(1) = \ell_1,\ S^0(2) - S^0_{n_2}(2) = \ell_2\big\}$$

is a nonnegative Q-random variable, and

$$\int f\big((\tilde{R}_k(\ell_1,\ell_2))\big)\,dQ$$

$$= \sum_{\ell_1 \in \mathscr{P}} \sum_{\ell_2} \sum_{j} P\big\{X^0_{n_1} = j | S^0(1) - S^0_{n_1-1}(1) = \ell_1\big\}$$

$$\times P\big\{S_{n_1-1}(1) + S_{n_2}(2) \in R(\ell_1 - j, \ell_2) - x_{n_1}(j),$$

$$S^0(1) - S^0_{n_1-1}(1) = \ell_1,\ S^0(2) - S^0_{n_2}(2) = \ell_2\big\}$$

$$= \text{expression (4.11)}.$$

In the first equality here we have used (4.14); in the second, condition (3.2). Thus (4.11) is expressed as the appropriate average, and the proof of Lemma 4.4 is complete. □

LEMMA 4.5. *Let I be the set of all nondecreasing sequences $\alpha = (\alpha_j)_{j \in \mathbb{Z}}$ with values in the set $\mathbb{Z} \cup \{-\infty, +\infty\}$ of extended integers that satisfy*

$$\lim_{j \to -\infty} \alpha_j = \inf(\mathscr{P}) \quad \text{and} \quad \lim_{j \to +\infty} \alpha_j = +\infty. \tag{4.13}$$

There exists a stochastic process $A = (A_j)_{j \in \mathbb{Z}}$ defined on a probability space $(\Omega, \mathscr{A}, Q)$ with A taking values in I and having for every $j \in \mathbb{Z}$ and $\ell < \ell'$, both in $\mathscr{P}$, the following two properties:

$$Q\{A_j \le \ell\} = P\big\{X^0_{n_1} > j | S^0(1) - S^0_{n_1-1}(1) = \ell\big\}; \tag{4.14}$$

$$\text{if } A_j > \ell, \text{ then } A_{j+\ell'-\ell} > \ell'. \tag{4.15}$$

PROOF. Denote RHS(4.14) by $F_j(\ell)$ and observe from (3.3) that $F_j(\ell)$ is nondecreasing in $\ell \in \mathscr{P}$. Define an inverse probability transform $F_j^{\sim}: (0,1) \to \mathscr{P} \cup \{-\infty, +\infty\}$ by the recipe

$$F_j^{\sim}(t) = \inf\{\ell \in \mathscr{P}: F_j(\ell) \ge t\}, \qquad 0 < t < 1. \tag{4.16}$$

Let U be a uniform $(0,1)$ random variable on a probability space $(\Omega, \mathscr{A}, Q)$, and set

$$A_j = F_j^{\sim}(U), \qquad j \in \mathbb{Z}. \tag{4.17}$$

It is a routine exercise to use (4.16), (4.17) and (3.3) to verify that A has the required properties. □

Acknowledgments. I would like to thank Professors Persi Diaconis and Michael Wichura for their many helpful comments and suggestions. I am grateful also to the referee for pointing out an error in the original version of this paper.

REFERENCES

EFRON, B. (1965). Increasing properties of Pólya frequency functions. *Ann. Math. Statist.* **36** 272–279.

ERDÖS, P. (1945). On a lemma of Littlewood and Offord. *Bull. Amer. Math. Soc.* **51** 898–902.

FELLER, W. (1971). *An Introduction to Probability Theory and Its Applications* **2**, 2nd ed. Wiley, New York.

FILL, J. A. (1987a). Bounds on the coarseness of random sums. Technical Report 208, Dept. Statistics, Univ. Chicago.

FILL, J. A. (1987b). The Radon transform on $\mathbb{Z}_n$ and on the circle. Technical Report 210, Dept. Statistics, Univ. Chicago.

FREEDMAN, D. and LANE, D. (1980). The empirical distribution of Fourier coefficients. *Ann. Statist.* **8** 1244–1251.

KLEITMAN, D. (1970). On a lemma of Littlewood and Offord on the distributions of linear combinations of vectors. *Adv. in Math.* **5** 155–157.

KLEITMAN, D. (1976). Some new results on the Littlewood–Offord problem. *J. Combin. Theory Ser. A* **20** 89–113.

LITTLEWOOD, J. E. and OFFORD, A. C. (1943). On the number of real roots of a random algebraic equation. III. *Math. USSR Sb.* **12** 277–285. [In *The Collected Papers of J. E. Littlewood* **2** (1982) 1333–1344. Clarendon Press, Oxford.]

MARSHALL, A. and OLKIN, I. (1979). *Inequalities: Theory of Majorization and Its Applications.* Academic, New York.

WINTNER, A. (1938). *Asymptotic Distributions and Infinite Convolutions.* Edwards Brothers, Ann Arbor, Mich.

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA 94305